



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



**BARRERAS PARA EL USO DE INNOVACIONES TECNOLÓGICAS
EN PROFESORES DE EDUCACIÓN SUPERIOR DESDE LA
MINERÍA DE DATOS**

Trabajo recepcional bajo la modalidad de Tesis
que para obtener el grado de:

**Maestro en Administración
de Tecnologías de la Información**

Presenta:

Marcela de los Ángeles Yanes Pérez

Directores de Trabajo Recepcional:

Dra. Martha Patricia Silva Payró

Dr. Pablo Payró Campos

Cuerpo Académico: Gestión de Tecnologías de la Información

Línea de Generación y Aplicación del Conocimiento de la Maestría:

**Administración, diseño e implementación de
soluciones de TI**

Cunduacán, Tabasco

Octubre, 2023



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO
DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



**BARRERAS PARA EL USO DE INNOVACIONES TECNOLÓGICAS
EN PROFESORES DE EDUCACIÓN SUPERIOR DESDE LA
MINERÍA DE DATOS**

Trabajo recepcional bajo la modalidad de Tesis
que para obtener el grado de:

**Maestro en Administración
de Tecnologías de la Información**

Presenta:

Marcela de los Ángeles Yanes Pérez

Directores de Trabajo Recepcional:

Dra. Martha Patricia Silva Payró
Dr. Pablo Payró Campos

Cuerpo Académico: Gestión de Tecnologías de la Información

Jurado Revisor:

Dr. Gerardo Arceo Moheno
Dr. Guillermo de los Santos Torres
Dra. Verónica García Martínez

Línea de Generación y Aplicación del Conocimiento de la Maestría:

**Administración, diseño e implementación de
Soluciones de TI**

Cunduacán, Tabasco

Octubre, 2023



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



F8: Cesión de derechos

Cunduacán, Tabasco, a 22 de agosto de 2023.

Asunto: Cesión de Derechos.

A quien corresponda:

Los que suscriben la presente, declaramos que el trabajo de tesis titulado, "***Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos***" es de nuestra autoría intelectual y por lo tanto cedemos todos los **derechos** sobre este proyecto a la Universidad Juárez Autónoma de Tabasco, a la cual relevamos de cualquier sanción y asumimos responder a cualquier reclamo de derechos de autor ante las autoridades competentes.

Atentamente

Autores:

Nombre	Domicilio	Firma autógrafa
Marcela de los Ángeles Yanes Pérez	Cárdenas, Tab.	
Dra. Martha Patricia Silva Payró	Villahermosa, Tab.	
Dr. Pablo Payró Campos	Villahermosa Tab.	

c.c.p. MTE. Óscar Alberto González González – Director de la División Académica de Ciencias y Tecnologías de la Información.
Dr. Eddy Arquimedes Ancona Alcocer - Encargado del despacho de la Coordinación de Posgrado.
Estudiante.





UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



Carta de Autorización

A quien corresponda:

La que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que utilice tanto física como digitalmente la tesis de grado denominada, **"Barreras para el uso de innovaciones tecnológicas en profesores de Educación Superior desde la Minería de Datos"**, de la cual soy autora y titular de los Derechos de Autor.

La finalidad del uso por parte de la Universidad Juárez Autónoma de Tabasco de la tesis antes mencionada, será única y exclusivamente para difusión, educación y sin fines de lucro; autorización que se hace de manera enunciativa más no limitativa para subirla a la Red Abierta de Bibliotecas Digitales (RABID) y a cualquier otra red académica con las que la Universidad tenga relación institucional.

Por lo antes manifestado, libero a la Universidad Juárez Autónoma de Tabasco de cualquier reclamación legal que pudiera ejercer respecto al uso y manipulación de la tesis mencionada y para los fines estipulados en este documento.

Se firma la presente autorización en la ciudad de Villahermosa, Tabasco a los 22 días del mes de Septiembre del año 2023.

Autorizo


Marcela de los Angeles Yanes Pérez



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN



"2022, Año de Ricardo Flores Magón"

Cunduacán, Tabasco a 06 de junio de 2022
Oficio No. 0673/DACYTI/CP/2022

Asunto: Asignación de Director de Tesis

Dra. Martha Patricia Silva Payró
Profesora Investigadora

De conformidad con lo establecido en el Reglamento de Estudios de Posgrado Vigente, de la Universidad Juárez Autónoma de Tabasco, me permito informarle, que ha sido designada como Director de la Tesis titulada "**Barreras para el uso de Innovaciones Tecnológicas en profesores de Educación Superior desde la Minería de Datos.**", a realizar por el **C. Marcela De Los Ángeles Yanes Pérez**, para obtener el grado de Maestro en Administración de Tecnologías de la Información.

Sin otro particular, aprovecho la ocasión para enviarle un afectuoso saludo.

Atentamente

MTE. Oscar Alberto González González
Director

UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN

C.c.p. Dr. Eddy Arquimedes García Aleocer. Encargado del Despacho de la Coordinación de Posgrado

Alumno
Archivo

MTE OAGG/EAGA

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.
Cunduacán, Tabasco, México.
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870
E-mail: direccion.dacyti@ujat.mx

www.ujat.mx



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO
"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN



"2022, Año de Ricardo Flores Magón"

Cunduacán, Tabasco a 06 de junio de 2022

Asunto: Dirección de Tesis

Dr. Pablo Payró Campos
Profesor Investigador

De conformidad con lo establecido en el Reglamento de Estudios de Posgrado Vigente, de la Universidad Juárez Autónoma de Tabasco, me permito informarle, que ha sido designado como Director de la tesis titulada "**Barreras para el uso de innovaciones tecnológicas en profesores de Educación Superior desde la Minería de Datos**", a realizar por la **C. Marcela de los Ángeles Yanes Pérez** con matrícula **212H19002**, para obtener el grado de Maestra en Administración de Tecnologías de la Información.

Sin otro particular, aprovecho la ocasión para enviarle un afectuoso saludo.

UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN

Atentamente


MTE. Óscar Alberto González González
Director

C.c.p. Dr. Eddy Arquímedes García Alcocer. - Encargado del despacho de la Coordinación de Posgrado
Archivo.
Consecutivo

M.T.E. OAGG/EAGA X

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.
Cunduacán, Tabasco, México.
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870
E-mail: direccion.dacyti@ujat.mx

www.ujat.mx



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



F7: Respuesta de jurado

Cunduacán, Tabasco, a 09 de agosto de 2023.

MTE. Óscar Alberto González González
Director de la División Académica de Ciencias y Tecnologías de la Información
Presente

En atención a los oficios girados por usted, en los que se nos designa como parte del jurado para efectuar la revisión de la tesis titulada **"Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos"**, realizada por la **C. Marcela de los Ángeles Yanes Pérez**, estudiante de la Maestría en Administración de Tecnologías de la Información, nos permitimos informarle que, en virtud de que ha atendido las observaciones realizadas, otorgamos nuestra aprobación para que continúe los trámites para la obtención del grado.

Sin otro particular, aprovechamos la ocasión para enviarle un cordial saludo.

Atentamente integrantes del jurado

Dra. Martha Patricia Silva Payró

Dr. Pablo Payró Campos

Dr. Gerardo Arceo Moheno

Dr. Guillermo de los Santos Torres

Dra. Verónica García Martínez



c.c.p. **Dr. Eddy Arquímedes Ancona Alcocer**. Encargado del despacho de la Coordinación de Posgrado.
Estudiante.



**UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO**

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



**DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN**



Cunduacán, Tabasco a 08 de septiembre de 2023
Oficio No. 1323/DACYTI/CP/2023

Asunto: Autorización de impresión de Tesis

C. Marcela de los Ángeles Yanes Pérez
Matricula: 212H19002

En virtud de que cumple satisfactoriamente los requisitos establecidos en el Reglamento General de Estudios de Posgrado vigente en la Universidad, informo a Usted que se autoriza la impresión del trabajo recepcional "**Barreras para el uso de innovaciones tecnológicas en profesores de Educación Superior desde la Minería de Datos**", para presentar examen y obtener el Grado de Maestra en Administración de Tecnologías de la Información.

Sin otro particular, aprovecho la ocasión para enviarle un afectuoso saludo.

Atentamente

MTE. Óscar Alberto González González
Director

C.c.p. Dr. Eddy Arquímedes García Alcocer. - Encargado del Despacho de la Coordinación de Posgrado DACYTI
Archivo.
Consecutivo.
MTE/OAGG/EAGA X

Carretera Cunduacán-Jalpa Km. 1, Colonia Esmeralda, C.P. 86690.
Cunduacán, Tabasco, México.
Tel: (993) 358 1500 ext. 6727; (914) 336 0616; Fax: (914) 336 0870
E-mail: direccion.dacyti@ujat.mx

www.ujat.mx

Agradecimientos

Mi más profundo agradecimiento a la Doctora Martha Patricia Silva Payró y al Doctor Pablo Payró Campos, mis asesores y mentores, por su tiempo, esfuerzo y comprensión para ayudarme a tener éxito en mis estudios. Su vasta sabiduría y riqueza de experiencia me han inspirado a lo largo de mi formación profesional.

Además, me gustaría agradecer a la Doctora Verónica García Martínez, al Doctor Gerardo Arceo Moheno y al Doctor Guillermo de los Santos Torres por su asistencia y diversas revisiones a lo largo de nuestra investigación. Muchas gracias por dedicarme parte de su valioso tiempo.

Me gustaría expresar también mi gratitud a todos en la División Académica de Ciencias y Tecnologías de la Información. Gracias a su generosidad y aliento, el tiempo que pasé estudiando y viviendo en nuestra amada Universidad Juárez Autónoma de Tabasco ha sido verdaderamente gratificante.

Particularmente hago extensos mis agradecimientos al Consejo Nacional de Ciencia y Tecnología (CONACYT) por brindarme su apoyo a lo largo de este nuevo escalón profesional.

Dedicatorias

Dedico este proyecto a Dios Todopoderoso mi creador, mi pilar fuerte, mi fuente de inspiración, sabiduría, conocimiento y entendimiento. Él ha sido mi fortaleza a lo largo de este programa y solo en sus alas he volado.

Dedico esta tesis a una persona amada que ha significado y sigue significando mucho para mí. Aunque ya no estés en este mundo, tus recuerdos siguen regulando mi vida. A mi abuelita paterna Sebastiana Rodríguez, cuyo amor por mí no conoció límites y quien me enseñó el valor del trabajo duro. Muchas gracias Mamá Tana, nunca te olvidaré.

Dedico mi trabajo de tesis a mi familia. Un sentimiento especial de gratitud a mis amados padres, Marcelino Yanes Rodríguez y Noemi Pérez Campos, cuyas palabras de aliento y empuje por la tenacidad resuenan en mis oídos, gracias por ser los mejores papás del mundo, por siempre su Mau.

A mis queridos hermanos Fátima de Jesús y Jesús Manuel que nunca se han separado de mi lado y son muy especiales. Ambos han sido mis mejores porristas.

Dedico este trabajo y agradezco a mi pareja Christian Alberto Brito Izquierdo, por estar ahí para mí durante todo el programa de Maestría, en esta nueva etapa de nuestra vida juntos, gracias por acompañarme en todo momento.

Los amo incondicionalmente, gracias por todo y por tanto.

Resumen

Las innovaciones tecnológicas han brindado numerosos beneficios en diversas actividades cotidianas y han desempeñado un papel crucial en el sector educativo, especialmente después del impacto de la pandemia COVID-19. No obstante, la transición abrupta hacia el aprendizaje virtual ha generado dificultades de adaptación para los profesores, quienes no han recibido una orientación adecuada en la implementación y uso de estas herramientas. Con el objetivo de abordar esta problemática, esta investigación se enfocó en identificar patrones de conducta entre los profesores de la Universidad Juárez Autónoma de Tabasco frente a las barreras asociadas al uso de innovaciones tecnológicas, utilizando técnicas de minería de datos. La metodología de la investigación se dividió en seis etapas claramente definidas. En primer lugar, se buscó adquirir un dominio profundo de la aplicación tecnológica seleccionada. A continuación, se procedió a la selección de un conjunto de datos relevante para el estudio. Posteriormente, se llevó a cabo la limpieza y transformación de los datos, asegurando su calidad y coherencia. Luego, se realizaron elecciones adecuadas de algoritmos de minería de datos que permitieran obtener información valiosa a partir de los datos procesados. Las técnicas de minería de datos empleadas incluyeron árboles de decisión, agrupamiento y reglas de asociación. Para la etapa siguiente, se aplicó el proceso de descubrimiento de conocimiento en bases de datos (KDD), donde se seleccionaron tres herramientas específicas de minería de datos para identificar patrones relacionados con las barreras en el uso de innovaciones tecnológicas. La investigación sobre innovaciones tecnológicas ha demostrado su utilidad en diversas áreas, como se ha señalado en estudios anteriores que se han llevado a cabo durante la transición de la modalidad presencial a virtual en el contexto de la pandemia. Sin embargo, también se han

identificado desafíos y limitaciones que deben abordarse para mejorar la eficacia y eficiencia de la minería de datos en este campo en futuras investigaciones. Es fundamental superar estos obstáculos para maximizar el potencial de las innovaciones tecnológicas en el ámbito educativo y brindar a los profesores el apoyo necesario en su adopción y uso efectivo. Cabe destacar que, en el contexto específico de esta investigación, se enfocó en profesores de la Universidad Juárez Autónoma de Tabasco, pero los resultados y las conclusiones pueden ser relevantes y aplicables en entornos educativos similares.

Universidad Juárez Autónoma de Tabasco.
México.

Introducción

1 La minería de datos se ha convertido en una disciplina crucial en la era digital, ya que permite a los investigadores explorar grandes conjuntos de datos y descubrir patrones y tendencias que pueden ser utilizadas para mejorar las decisiones y tomar acciones informadas. Esta investigación consistió en la identificación de las barreras presentadas ante las innovaciones tecnológicas por profesores de educación superior, particularmente de la Universidad Autónoma de Tabasco Juárez (UJAT); la cual fue un área de creciente interés en la minería de datos a partir de la pandemia de COVID-19, cuyo proceso de describe en cinco capítulos, un apéndice y dos anexos.

El primer capítulo proporciona una descripción general del estudio, es decir: antecedentes, planteamiento del problema, incluyendo la definición de este, delimitación de la investigación, objetivos generales y específicos; así mismo, se incluye la justificación y finalmente la metodología utilizada, donde se explicó la metodología empleada para el desarrollo de esta investigación y el enfoque que se utilizó.

En el segundo capítulo se encuentra la base teórica de la investigación, que incluye una descripción del marco de referencia, compuesto por información obtenida de estudios científicos y tesis relacionadas con barreras tecnológicas, innovaciones tecnológicas y la minería de datos. Además, se incluye un marco conceptual que presenta conceptos e investigaciones relacionados con técnicas de minería de datos y el proceso KDD, entre otros. En ese mismo capítulo, se describe el marco tecnológico, que incluye información sobre los programas o herramientas utilizados para realizar minería de datos, incluyendo herramientas *framework*, lenguajes de programación, así como servicios en la nube, y finalmente el marco legal.

El tercer capítulo describe la metodología que se utilizó para llevar a cabo la investigación. La primera fase consistió en la revisión de los datos del *dataset*, la segunda fue la selección de las herramientas de minería de datos *RapidMiner*, *SAS Enterprise Miner* y *Python*, finalmente la tercera etapa fue la aplicación de un enfoque basado en el proceso KDD para limpiar y analizar los datos.

El cuarto capítulo detalla las pruebas realizadas con las técnicas de minería de datos en el *dataset* y los resultados obtenidos usando las herramientas *RapidMiner*, *SAS Enterprise Miner* y *Python*. Por último, en el quinto capítulo se presentan las conclusiones obtenidas de la investigación, así como recomendaciones y posibles áreas de investigación futura relacionadas con las barreras ante las innovaciones tecnológicas.

En esta investigación se utilizó el cuestionario denominado “Barreras para la adopción de innovaciones tecnológicas”, el cual fue empleado para encuestar a los profesores de la UJAT, este se describe detalladamente en el anexo A, donde se proporciona información relevante sobre la composición y la estructura de los datos utilizados en el estudio.

Además, también se incluye el Anexo B, que presenta la Ley Federal de Derechos de Autor. Esta ley es de suma importancia en el contexto de la investigación, ya que regula los derechos de autor y protege la propiedad intelectual en el país. Este anexo ofrece un recurso útil para comprender las disposiciones legales relacionadas con los derechos de autor y su aplicación en el estudio.

Asimismo, se adjunta el Anexo C, que contiene la Ley Federal de Protección de Datos Personales en Posesión de Particulares. Esta ley es fundamental para garantizar la privacidad y la protección de los datos personales de los individuos, este proporciona información esencial sobre las regulaciones y los requisitos legales que deben seguirse

al manejar y procesar datos personales en posesión de particulares. Tanto el Anexo A como el Anexo B y C son documentos relevantes y complementarios a la investigación, ya que abordan aspectos legales y normativos que deben considerarse en el marco del estudio.

Universidad Juárez Autónoma de Tabasco.
México.

Índice general

Índice de tablas.....	xxii
Índice de figuras.....	xxiii
Capítulo 1. Generalidades.....	1
1.1 Antecedentes.....	1
1.1.1 Innovaciones tecnológicas.....	1
1.1.2 Barreras tecnológicas.....	2
1.1.3 Minería de datos.....	3
1.2 Planteamiento del problema.....	4
1.2.1 Definición del problema.....	4
1.2.2 Delimitación de la investigación.....	5
1.2.3 Pregunta de investigación.....	5
1.3 Objetivos.....	6
1.3.1 Objetivo general.....	6
1.3.2 Objetivos específicos.....	6
1.4 Justificación.....	7
1.5 Metodología utilizada.....	8
1.6 Enfoque de la investigación.....	9
1.7 Fuentes de investigación.....	10
1.8 Técnicas de recolección de datos.....	10
1.9 Programación por implementar.....	10

Capítulo 2. Marcos de la investigación	11
2.1 Marco referencial.....	11
2.1.1 Análisis del desempeño docente en la modalidad en línea COVID-19.....	11
2.1.2 Uso crítico de tecnologías de profesores universitarios.....	12
2.1.3 Desarrollo de la competencia digital en profesores universitarios	13
2.1.4 Competencias digitales docentes: Experiencia en nivel universitario	13
2.1.5 Educación Superior virtual: Validación de instrumento en la UEPSE	14
2.1.6 Estrategias tecnológicas en el proceso de enseñanza y aprendizaje	14
2.1.7 TIC's aplicadas a la educación superior por docentes de la UMC.....	15
2.1.8 Docentes en América Latina COVID-19 con tecnologías digitales	15
2.1.9 Educación de América Latina en el contexto de COVID-19.....	16
2.1.10 Percepciones docentes frente al cambio de modalidad.....	17
2.1.11 Tecnología para innovar la práctica docente.....	18
2.1.12 La educación en tiempos de COVID-19.....	19
2.1.13 Educación superior: desafíos en tiempos de COVID-19.....	20
2.1.14 Experiencias docentes en tiempos de COVID-19 en América Latina	21
2.2 Marco conceptual	21
2.2.1 Innovación	21
2.2.2 Tecnología	22
2.2.3 Innovación tecnológica.....	22
2.2.4 Minería de datos	23

2.2.5 Reglas de asociación24

2.2.6 Árboles de decisión.....25

2.2.7 Proceso KDD26

2.2.8 Aprendizaje supervisado27

2.2.9 Aprendizaje no supervisado29

2.2.10 ISO/IEC 9126 norma de calidad del *software*.....29

2.3 Marco tecnológico31

2.3.1 Herramientas para minería de datos - *software* libre31

2.3.2 Herramientas para minería de datos - *software* propietario34

2.3.3 Servicios para minería de datos en la nube35

2.3.4 Lenguajes de programación para minería de datos38

2.4 Marco legal41

2.4.1 Licencias de *software* libre41

2.4.2 Licencias de *software* propietario41

2.4.3 Ley Federal de Derechos de Autor44

2.4.4 Ley Federal de Protección de Datos Personales44

Capítulo 3. Aplicación de la metodología y desarrollo45

3.1 Análisis de la información del *dataset*.....45

3.1.1 Cuantificación de la población de estudio46

3.1.2 Depuración de la población47

3.1.3 Descripción del instrumento48

3.1.4	Aplicación del instrumento	50
3.1.5	Visualización y exploración de los datos	50
3.2	Selección de las herramientas para la aplicación de la minería de datos	52
3.2.1	ISO/IEC 9126.....	53
3.2.2	Características de las herramientas seleccionadas.....	59
3.3	Proceso KDD.....	64
3.3.1	Dominio de la aplicación	66
3.3.2	Selección de un conjunto de datos.....	70
3.3.3	Limpieza y transformación de los datos	72
3.3.4	Elección de algoritmos de minería de datos.....	74
3.3.5	Minería de datos	78
3.3.6	Interpretación de los resultados	91
Capítulo 4.	Resultados.....	93
4.1	Pruebas de las técnicas de minería de datos	93
4.1.1	<i>RapidMiner</i>	94
4.1.2	<i>SAS Enterprise Miner</i>	117
4.1.3	<i>Python</i>	139
4.2	Resultados obtenidos en las herramientas de minería de datos	186
4.2.1	Árboles de decisión.....	186
4.2.2	Agrupación.....	197
4.2.3	Reglas de asociación	207

4.3 Patrones descubiertos en árboles de decisión.....	211
4.3.1 <i>RapidMiner</i>	211
4.3.2 <i>SAS Enterprise Miner</i>	219
4.3.3 <i>Python</i>	224
4.4 Patrones descubiertos en agrupación.....	230
4.4.1 Información de los <i>clústeres</i> obtenidos	230
4.5 Patrones descubiertos en reglas de asociación	246
4.5.1 <i>RapidMiner</i>	246
4.5.2 <i>SAS Enterprise Miner</i>	248
4.5.3 <i>Python</i>	250
4.6 Análisis de usabilidad	261
Capítulo 5. Conclusiones, recomendaciones y trabajos futuros	266
5.1 Conclusiones.....	267
5.2 Discusión.....	281
5.3 Recomendaciones.....	287
5.4 Trabajos futuros.....	292
Referencias.....	295
Glosario.....	312
Anexo A. Cuestionario aplicado a los profesores para la recolección de datos.....	313
Anexo B. Ley Federal de Derechos de Autor	325
Anexo C. Ley Federal de Protección de Datos Personales	329

Índice de tablas

Tabla 1. Frecuencia de Divisiones Académicas	47
Tabla 2. Variables de la investigación	49
Tabla 3. Factores y criterios del modelo ISO/IEC 9126	54
Tabla 4. Herramientas de minería de datos clasificadas en base a la ISO/IEC 9126	56
Tabla 5. Selección de herramientas para minería de datos mediante ISO/IEC 9126.....	57
Tabla 6. Criterios totales de las herramientas analizadas.....	58
Tabla 7. Tipos de licencia de RapidMiner.....	60
Tabla 8. Conjunto de datos del dataset	71
Tabla 9. Prueba de árbol de decisión de la pregunta raíz P12	189
Tabla 10. Decisiones representativas de la pregunta raíz P12	193
Tabla 11. Decisiones representativas de la pregunta raíz P12	196
Tabla 12. Carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT	233
Tabla 13. El equipo tecnológico de apoyo a las actividades docentes es escaso.....	234
Tabla 14. La conectividad de la red representa un problema recurrente	235
Tabla 15. La accesibilidad a la red en los espacios universitarios es inestable	236
Tabla 16. Cuando tengo un problema con algún equipo es difícil conseguir apoyo técnico	237
Tabla 17. Que te ayuden a resolver problemas de software en la UJAT resulta complicado	238
Tabla 18. Existe una política institucional poco clara que promueva la adopción de tecnología en el aula	239
Tabla 19. La estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.....	241
Tabla 20. La institución carece de un sistema que diagnostique las competencias tecnológicas de su personal	242
Tabla 21. La institución carece de un sistema que evalúe las competencias tecnológicas de su personal.....	243
Tabla 22. He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente	244
Tabla 23. He invertido pocos recursos económicos en prepararme para usar la tecnología en mi práctica docente	245
Tabla 24. Respuestas en escala Likert.....	247
Tabla 25. Escala de usabilidad SUS utilizada	262
Tabla 26. Puntuación de las herramientas en escala de usabilidad SUS	263
Tabla 27. Resultados obtenidos de la evaluación por herramienta.....	264
Tabla 28. Patrones identificados en la investigación	279

Índice de figuras

Figura 1. Fases de la metodología de la investigación	8
Figura 2. Dataset utilizado en la investigación.....	45
Figura 3. Selección de los datos para su filtrado	50
Figura 4. Búsqueda de la opción para datos especiales.....	51
Figura 5. Selección y localización de celdas en blanco	51
Figura 6. Identificación de datos nulos en el dataset	52
Figura 7. Arquitectura general de SAS Enterprise Miner	61
Figura 8. Algoritmos para minería de datos de Python	63
Figura 9. Metodología creada para la investigación	65
Figura 10. Interfaz inicial de RapidMiner	67
Figura 11. Interfaz inicial de SAS Enterprise Miner	68
Figura 12. Interfaz inicial de Anaconda Navigator	69
Figura 13. Lectura del dataset en RapidMiner	73
Figura 14. Lectura del dataset en SAS Enterprise Miner	73
Figura 15. Lectura del dataset en Python	74
Figura 16. Clasificación de técnicas de minería de datos	75
Figura 17. Tipos de algoritmos de árbol de decisión	76
Figura 18. Tipos de algoritmo de agrupamiento	77
Figura 19. Tipos de algoritmo de reglas de asociación.....	77
Figura 20. Árbol de decisión obtenido en RapidMiner	79
Figura 21. Árbol de decisión a detalle obtenido en RapidMiner	79
Figura 22. Descripción del árbol de decisión obtenido en RapidMiner.....	80
Figura 23. Transformación Z en RapidMiner	81
Figura 24. Resultado de K-Means en RapidMiner	82
Figura 25. Gráfico de K-Means en RapidMiner.....	83
Figura 26. Extensión Text Processing de RapidMiner	84
Figura 27. Reglas de asociación en RapidMiner	84
Figura 28. Gráfico de reglas de asociación en RapidMiner	85
Figura 29. Árbol de decisión obtenido en SAS Enterprise Miner	86
Figura 30. Gráfico de K-Means en SAS Enterprise Miner.....	86
Figura 31. Reglas de asociación en SAS Enterprise Miner.....	87
Figura 32. Lectura del dataset en libreta de Python	88
Figura 33. Árbol de decisión obtenido en Python	89
Figura 34. Gráfico de K-Means en Python.....	90
Figura 35. Reglas de asociación en Python	91
Figura 36. Parámetros de inicio del árbol de decisión	96
Figura 37. Matriz de confusión de árbol de decisión pregunta raíz P10.....	99
Figura 38. Prueba de árbol de decisión de la pregunta raíz P10	100
Figura 39. Matriz de confusión de árbol de decisión pregunta raíz P12.....	101
Figura 40. Prueba de árbol de decisión de la pregunta raíz P12	102

Figura 41. Matriz de confusión árbol de decisión pregunta raíz P19.....	103
Figura 42. Prueba de árbol de decisión de la pregunta raíz P19	104
Figura 43. Ítems que conforman los dos clústeres	105
Figura 44. Kmeans prueba con dos agrupaciones	106
Figura 45. Promedio de distancias centroides con dos clústeres	107
Figura 46. Ítems que conforman los tres clústeres	108
Figura 47. Kmeans prueba con tres agrupaciones	109
Figura 48. Promedio de distancias centroides con tres clústeres	110
Figura 49. Ítems que conforman los cinco clústeres	111
Figura 50. Kmeans prueba con cinco agrupaciones	112
Figura 51. Promedio de distancias centroides con cinco clústeres	113
Figura 52. Reglas de asociación obtenidas con 55% de confianza	115
Figura 53. Reglas de asociación obtenidas con 70% de confianza	116
Figura 54. Reglas de asociación obtenidas con 90% de confianza	117
Figura 55. Árbol de decisión obtenido para la pregunta raíz P10	120
Figura 56. Matriz de confusión del árbol de decisión - pregunta raíz P10.....	121
Figura 57. Matriz de confusión del árbol de decisión - pregunta raíz P12.....	122
Figura 58. Prueba de árbol de decisión de la pregunta raíz P12	123
Figura 59. Matriz de confusión del árbol de decisión - pregunta raíz P19.....	124
Figura 60. Prueba de árbol de decisión de la pregunta raíz P19	125
Figura 61. Kmeans prueba con dos agrupaciones	128
Figura 62. Resumen de clústeres.....	128
Figura 63. Kmeans prueba con tres agrupaciones	130
Figura 64. Resumen de clústeres.....	131
Figura 65. Kmeans prueba con cinco agrupaciones	133
Figura 66. Resumen de clústeres.....	134
Figura 67. Reglas de asociación obtenidas con 55% de confianza	136
Figura 68. Reglas de asociación obtenidas con 70% de confianza	137
Figura 69. Reglas de asociación obtenidas con 90% de confianza	138
Figura 70. Importación de librerías y carga de datos en Python – P10.....	140
Figura 71. Definición del árbol de decisión – P10.....	141
Figura 72. Cálculo de la precisión del modelo – P10	141
Figura 73. Árbol de decisión generado para P10.....	142
Figura 74. Importación de librerías Python – P12.....	143
Figura 75. Carga de datos para el modelo – P12	143
Figura 76. Definición del árbol de decisión – P12.....	144
Figura 77. Cálculo de la precisión del modelo – P12.....	145
Figura 78. Árbol de decisión generado para P12.....	146
Figura 79. Importación de librerías Python – P19.....	146
Figura 80. Carga de datos para el modelo – P19	147
Figura 81. Definición del árbol de decisión – P19.....	148
Figura 82. Cálculo de la precisión del modelo – P19.....	149

Figura 83. Árbol de decisión generado para P19.....	149
Figura 84. Modificaciones al dataset y normalización.....	151
Figura 85. Importación de librerías en Python.....	154
Figura 86. Conjunto de datos leídos en Python.....	154
Figura 87. Encabezado de dataset en Python.....	155
Figura 88. Columnas implementadas en Python.....	155
Figura 89. Valores de ítems en Python.....	156
Figura 90. Normalización de datos en Python.....	157
Figura 91. Función describe en Python.....	158
Figura 92. Cálculo WCSS en Python.....	159
Figura 93. Gráfico de codo de Jambú.....	160
Figura 94. Tabla de clústeres en Python.....	161
Figura 95. Técnica PCA en Python.....	162
Figura 96. Código de gráfico Kmeans en Python.....	163
Figura 97. Gráfico Kmeans obtenido con dos clústeres.....	164
Figura 98. Matriz de confusión para los dos clústeres.....	165
Figura 99. Código de gráfico Kmeans en Python.....	166
Figura 100. Gráfico de Kmeans obtenido con tres clústeres.....	167
Figura 101. Matriz de confusión para los tres clústeres.....	168
Figura 102. Gráfico Kmeans obtenido con cinco clústeres.....	169
Figura 103. Matriz de confusión para los cinco clústeres.....	170
Figura 104. Importación de librerías y carga de datos en Python.....	172
Figura 105. Transformación de los datos.....	173
Figura 106. Configuración de soporte mínimo.....	174
Figura 107. Reglas de asociación generadas.....	175
Figura 108. Generación del archivo PDF.....	176
Figura 109. Reglas de asociación con 55% confianza y 10% SM.....	177
Figura 110. Importación de librerías y carga de datos en Python.....	178
Figura 111. Configuración de algoritmo apriori y soporte mínimo.....	179
Figura 112. Generación del archivo PDF.....	180
Figura 113. Reglas de asociación con 70% confianza y 15% SM.....	181
Figura 114. Importación de librerías y carga de datos en Python.....	182
Figura 115. Configuración de algoritmo apriori y soporte mínimo.....	183
Figura 116. Generación del archivo PDF.....	184
Figura 117. Reglas de asociación con 90% confianza y 20% SM.....	185
Figura 118. Matriz de confusión de árbol de decisión pregunta raíz P12.....	187
Figura 119. Prueba de árbol de decisión de la pregunta raíz P12.....	188
Figura 120. Matriz de confusión árbol de decisión pregunta raíz P12.....	191
Figura 121. Prueba de árbol de decisión de la pregunta raíz P12.....	192
Figura 122. Cálculo de la precisión del modelo – P12.....	194
Figura 123. Árbol de decisión generado para P12.....	195
Figura 124. Ítems que conforman los dos clústeres.....	198

Figura 125. Kmeans prueba con dos agrupaciones199

Figura 126. Promedio de distancias centroides con dos clústeres200

Figura 127. Kmeans prueba con dos agrupaciones202

Figura 128. Resumen de clústeres203

Figura 129. Gráfico de codo de Jambú205

Figura 130. Gráfico Kmeans obtenido primera prueba206

Figura 131. Matriz de confusión de dos clústeres207

Figura 132. Reglas de asociación obtenidas con 90% de confianza208

Figura 133. Reglas de asociación obtenidas con 90% de confianza209

Figura 134. Reglas de asociación con 90% confianza y 20% SM210

Figura 135. Kmeans prueba con dos agrupaciones231

Figura 136. Reglas de asociación en RapidMiner247

Figura 137. Reglas de asociación SAS Enterprise Miner249

Figura 138. Reglas de asociación con 90% confianza y 20% SM250

Figura 139. Diagrama de Venn de las reglas de asociación251

Universidad Juárez Autónoma de Tabasco.

Capítulo 1. Generalidades

1.1 Antecedentes

1.1.1 Innovaciones tecnológicas

1 Las innovaciones tecnológicas son herramientas fundamentales para la formación y su incorporación en el ámbito educativo es enorme, con un auge más notable en el siglo XXI. No obstante, la pandemia denominada COVID-19, sugirió un cambio drástico en las instituciones educativas; debido a que no era normal trabajar por completo con ellas.

En el ámbito educativo, Jara *et al.* (2021, p. 33) expresan que el cambio de la modalidad presencial a la virtual demostró que los profesores y los alumnos no estaban preparados en la utilización de tecnologías para la rutina escolar diaria. Es a partir de este tipo de cambios repentinos que se empezó a hablar de la sociedad de la información, y, en este sentido, los tipos de utilización de las innovaciones se volvieron importantes al ser percibidas como fundamentales (Arancibia *et al.*, 2020).

1 Santiago (2021, p. 149), hace referencia a que: "Desde hace tiempo los grados de interconexión han aumentado considerablemente, el avance de los nuevos datos y los avances en la correspondencia se presentan como un aparato vital que puede mejorar el acceso a la información, pero también la posibilidad de mezclarla con un número mayor de personas", este autor prioriza a la nueva modalidad, debido a que en algún momento se haría esencial el implementar tecnologías no solo como auxiliares en los proyectos, sino como un instrumento diario en todos los ámbitos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

1.1.2 Barreras tecnológicas

1 Las innovaciones tecnológicas son consideradas como las mejores herramientas de la actualidad, sin embargo, una investigación del Instituto Tecnológico y de Estudios Superiores de Monterrey (ITESM) con el Banco Interamericano de Desarrollo (BID) descubrió que los efectos de esta pandemia fueron contraproducentes, creando diversas barreras tecnológicas; para el caso específico de México, se identificó que el 35% de los profesores reportaron baja o ninguna conexión a Internet dentro de sus instalaciones, afectando negativamente en sus actividades académicas. (Instituto Tecnológico y de Estudios Superiores de Monterrey [ITESM] y Banco Interamericano de Desarrollo [BID], 2021, p. 2).

1 En consecuencia, García y Silva (2021, p.36), afirman que existen diversas barreras que enfrentan las instituciones para introducir con éxito la innovación tecnológica en el ámbito educativo, donde se han identificado dos grupos de barreras que ocurren en las implementaciones de TI: extrínsecas e intrínsecas. Las primeras se refieren a los recursos con los que cuentan los académicos para llevar a cabo sus actividades. El segundo grupo se conforma de las reflexiones de los académicos acerca de su relación con las tecnologías. Así mismo, obtuvieron que el 81% de los profesores deseaban más tiempo para aprender a trabajar con plataformas y entornos digitales, lo cual confirmó que existía una necesidad de formarse más en el uso de las innovaciones tecnológicas.

Se han desarrollado diversos estudios e investigaciones respecto a las carencias y necesidades que surgieron a partir de la pandemia, los cuales han recopilado diversa información, así como las perspectivas de profesores y alumnos de las instituciones, pero

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

para conocerlas mejor, han requerido de la implementación de *software* de minería de datos, con los cuales se ha proporcionado la información más oportuna del momento, así como sus presentaciones eficaces, gracias al uso de estas herramientas.

1.1.3 Minería de datos

La minería de datos nació de la idea de aprovechar dos situaciones: las cantidades masivas de datos y el potencial de las nuevas tecnologías que realiza operaciones analíticas sobre estos datos para su procesamiento. La minería de datos –también conocida como *data mining*, por su traducción del inglés– permite encontrar información oculta en los datos que no siempre es evidente.

La minería de datos es la herramienta encargada de analizar, descubrir y preparar datos para extraer información antes no visible, pero de gran utilidad (Belinchón, 2019). Si todos los datos necesarios se leen y analizan, en general, pueden brindar información procesable (futuras tendencias y comportamientos) que ayuden a la toma de decisiones, porque todos los datos en sí, para el administrador del sistema, no son los más apropiados, sino la información contenida en sus ramificaciones o búsquedas específicas, es decir, los datos que dependen de ellos.

Sin embargo, cuando la escala y cantidad de datos, es notablemente difícil de procesar por el hombre, se necesita el apoyo y eficiencia de las innovaciones tecnológicas. Es por ello por lo que se requiere de métodos inteligentes y específicos de análisis de los datos que puedan descubrir información útil a partir de estos, como lo es el proceso KDD. El término KDD, por sus siglas en inglés *Knowledge Discovery in Databases*, propuesto en el año de 1996 por Fayyad, Piatetsky-Shapiro y Smyth, se

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

refiere a todo a la extracción de conocimiento a partir de una base de datos y donde el conocimiento verdaderamente útil es aquel que se encuentra tras la exploración de los datos (Shahbazi & Byun, 2022).

1.2 Planteamiento del problema

1.2.1 Definición del problema

1 En la región sureste del país, en particular para el caso de la Universidad Juárez Autónoma de Tabasco (UJAT), la pandemia obligó a la suspensión de actividades a partir de marzo de 2020, por lo que tuvo que adaptarse a la implementación de tecnologías como apoyo a la modalidad de clases virtuales. Por consiguiente, el presente estudio se desarrollará con el propósito de analizar los datos obtenidos a través un cuestionario denominado “Barreras para la adopción de innovaciones tecnológicas”, proporcionado por García y Silva (2021), el cuál fue aplicado a profesores de la UJAT, pertenecientes a las 12 Divisiones Académicas, quienes impartieron clases durante el período 2020-2021.

1 Debido a la cantidad de datos, se implementarán tres herramientas de *data mining*, con el propósito de determinar cuál de ellas es la más apropiada para analizar e identificar las barreras del uso de innovaciones tecnológicas, esto mediante técnicas de agrupamiento, reglas de asociación y árboles de decisión. Con lo anterior expuesto, también se utilizará el proceso KDD, con el que se explorarán los comportamientos que presentaron los encuestados ante dichas dificultades.

1.2.2 Delimitación de la investigación

Alcances

Dentro de los alcances de este estudio se enlistan los siguientes:

- El estudio se abordó de forma cuantitativa, en el cual se especificaron y analizaron los patrones de conducta de los profesores de educación superior acerca del uso de innovaciones tecnológicas durante la pandemia.
- Utilización del proceso KDD.
- Se utilizaron técnicas de minería de datos.

Limitaciones

Dentro de las limitaciones de esta investigación se consideran las siguientes:

- El estudio se centró en las percepciones de los profesores respecto a las barreras en el uso de tecnologías.
- Las técnicas de minería de datos a utilizar fueron: árboles de decisión, reglas de asociación y agrupamiento.
- Las dimensiones del estudio se abordaron desde la perspectiva de los profesores encuestados.
- El tiempo de aplicación del instrumento fue de abril a septiembre de 2020.

1.2.3 Pregunta de investigación

¿Cuáles fueron las barreras que afrontaron los docentes de la Universidad Juárez Autónoma de Tabasco (UJAT) ante el uso de las innovaciones tecnológicas durante la pandemia COVID-19?

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

1.3 Objetivos

1.3.1 Objetivo general

Identificar los patrones de conducta que se presentaron en los profesores de la UJAT ante el uso de innovaciones tecnológicas mediante técnicas de minería de datos.

1.3.2 Objetivos específicos

- Identificar la herramienta de minería de datos adecuada para el análisis y minado de los datos.
- Aplicar técnicas de árbol de decisión para lograr una representación de los patrones de conducta.
- Utilizar técnicas de agrupamiento para identificar similitudes entre los patrones de conducta de los profesores.
- Implementar técnicas de reglas de asociación para identificar las relaciones que se presentan entre el uso de innovaciones tecnológicas y los profesores de educación superior.
- Interpretar los resultados obtenidos a través del minado de los datos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

1.4 Justificación

1 Las barreras del uso de innovaciones tecnológicas han sido estudiadas a nivel nacional e internacional en diversas investigaciones, también se ha hablado acerca de las carencias y necesidades que surgieron a partir de la pandemia COVID-19 en el entorno de la Educación Superior, las cuales han recopilado diversa información, como las perspectivas de profesores y alumnos de las instituciones, pero para analizarlas mejor, requirieron de la implementación de *software* de minería de datos, con los cuales lograron proporcionar la información más oportuna de aquel momento, así como sus presentaciones eficaces, gracias al uso de estas herramientas.

En esta investigación fue realizada con el propósito de explorar patrones de conducta de los profesores en relación con el uso de tecnologías innovadoras. Para lograr esto, se utilizaron herramientas de minería de datos, que permitieron identificar las barreras que surgieron durante la implementación de estas tecnologías.

1 Por consiguiente, el hecho de identificar los patrones de conducta de los encuestados logró aportar nuevos descubrimientos en esta área de investigación, así mismo brindó información relevante para la toma de decisiones y/o propuestas de ejecución de alternativas para ser implementadas en las instituciones educativas de nivel superior.

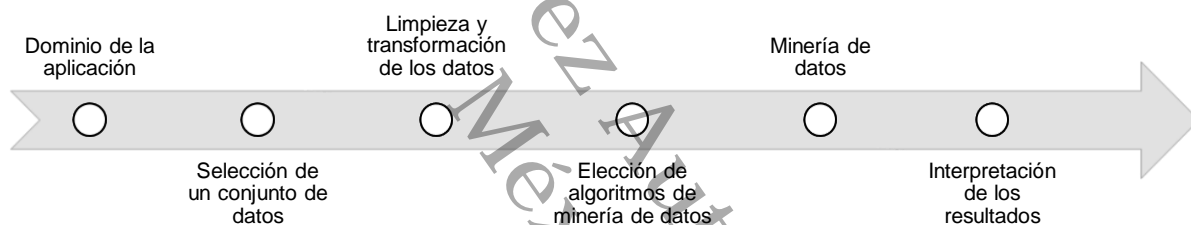
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

1.5 Metodología utilizada

La metodología utilizada en la investigación se basó en el enfoque cuantitativo y estuvo conformada por seis etapas: dominio de la aplicación, selección de un conjunto de datos, limpieza y transformación de los datos, elección de algoritmos de minería de datos, minería de datos e interpretación de los resultados. De lo anteriormente expuesto, estas etapas son representadas por medio de la figura 1 (ver figura 1).

Figura 1.

Fases de la metodología de la investigación



Nota: Elaboración propia.

La primera etapa se trató del aprendizaje de la minería de datos, es decir, se estudiaron conceptos y técnicas claves del tema, así como del manejo y uso de herramientas, para posteriormente reforzar los conocimientos con base en las herramientas que se seleccionarían con fundamento en la ISO/IEC 9126.

En la segunda etapa, selección de un conjunto de datos, se exploraron y analizaron los datos sin utilizar las herramientas de *data mining*, los cuales fueron tomados del *dataset* proporcionado por García y Silva (2021), con un total de 826 registros. La tercera etapa correspondió a la transformación de los datos, donde estos fueron limpiados y

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

organizados con el propósito de identificar las respuestas en blanco o de valor nulo, así como el descarte de *ítems* no necesarios para el minado de datos.

1 Para la cuarta etapa, se seleccionaron correctamente las herramientas de minería de datos, por lo cual se requirió de una base metodológica que se adaptó a las necesidades de esta investigación. Por tal motivo, se utilizó la norma de calidad de *software* ISO/IEC 9126. Se analizaron las características que formaron parte de la calidad y desempeño de las herramientas de minería de datos buscadas con anterioridad, con esto, se elaboró una comparativa basada en el modelo de calidad de *software* ISO/IEC 9126, donde se identificaron las herramientas óptimas a implementar en esta investigación. La quinta etapa fue la minería de datos, por medio del proceso KDD, tomando a consideración los aspectos y resultados que se tenían antes y después del minado de datos del formulario digital.

1 Finalmente, la sexta etapa concluyó con la interpretación de los resultados obtenidos después del análisis de los datos a través de las herramientas de minería de datos. A continuación, se exponen los aspectos relativos a la metodología implementada, el enfoque y las fuentes de investigación, así como los instrumentos y técnicas de recolección de datos y el tipo de programación.

1.6 Enfoque de la investigación

El enfoque de investigación que se implementó fue del tipo cuantitativo. La investigación bajo el enfoque cuantitativo, según Sánchez (2019) se ocupa de los fenómenos que pueden medirse mediante métodos estadísticos de análisis de datos, su

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

objetivo principal es describir, explicar, predecir y controlar objetivamente las causas de cada evento.

1.7 Fuentes de investigación

Las fuentes de información que se utilizaron en esta investigación se dividieron en: primarias y secundarias. Las fuentes primarias empleadas fueron: tesis de maestría y doctorado, artículos en revistas académicas, informes de conferencias y libros académicos digitales de medios nacionales e internacionales. Por otra parte, las fuentes secundarias a utilizar fueron: la lectura e interpretación de artículos académicos, así como el análisis de información obtenida de documentos digitales y sitios *web*, relacionados con los temas de minería de datos, barreras e innovaciones tecnológicas.

1.8 Técnicas de recolección de datos

La técnica de recolección de datos se llevó a cabo a través del formulario digital denominado "Barreras para la adopción de innovaciones tecnológicas", véase en el anexo A; después de seleccionar las herramientas con las que se trabajaría, se analizaron los datos a través de estas; el cual era un cuestionario presentado a profesores de la UJAT, pertenecientes a las 12 Divisiones Académicas, quienes impartieron clases durante el período 2020-2021, conformado por tres variables: sociodemográficas, uso de recursos tecnológicos y recursos disponibles y sistema educativo.

1.9 Programación por implementar

El tipo de programación utilizado fueron las herramientas de minería de datos, *RapidMiner*, *SAS Enterprise Miner* y *Python*. El *dataset* utilizado se encontraba en *Microsoft Excel*.

Capítulo 2. Marcos de la investigación

2.1 Marco referencial

A través de la revisión literaria, varias décadas atrás se empezó a pronunciar el tema del impacto que las herramientas digitales podrían tener en la educación, debido a que los sistemas educativos debían ir más allá de los cambios que facilitaban la formación de nuevas personas y la integración de las innovaciones tecnológicas. Sobre la base de los desafíos que la tecnología presenta a los profesores de educación superior en la actualidad, se han realizado investigaciones en torno a los temas de este estudio:

2.1.1 Análisis del desempeño docente en la modalidad en línea COVID-19

Lugo (2021) en su artículo evaluó el desempeño de los profesores de la Universidad de Guanajuato, con el propósito de tener una retroalimentación de parte de los alumnos, para así identificar los aspectos a fortalecer en la modalidad virtual e implementar estrategias de habilitación docente para mejorar la calidad de enseñanza.

Las dimensiones de la evaluación han sido definidas a partir de las competencias descritas en el Modelo Educativo de la misma institución. El análisis obtuvo los resultados de los instrumentos de evaluación aplicados al final de los ciclos escolares comprendidos del mes de agosto a diciembre de 2020 y de enero a junio del año 2021, durante la modalidad a distancia en línea con motivo de la emergencia sanitaria. En la aplicación de dicho instrumento durante los períodos antes mencionados, se tuvo una participación estudiantil del 66.16% y 56.24% respectivamente.

Finalmente, el estudio concluyó que, en la dimensión de la evaluación docente, se obtuvieron porcentajes superiores al 74%, otorgando dictámenes de Satisfactorio en

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

todas las áreas de la estructura curricular de la institución. Sin embargo, del total de estudiantes inscritos (792) para el período agosto-diciembre 2020, se evaluaron a solo 524. Posteriormente, para el ciclo enero-junio 2021, con 793 estudiantes inscritos, solo 446 fueron evaluaron.

1 Los resultados permitieron conocer el desempeño de los profesores desde la perspectiva de los estudiantes, evidenciando la necesidad de desarrollar capacidades en competencias pedagógicas que favorezcan los diferentes estilos de aprendizaje de los estudiantes, como lo es la modalidad virtual.

2.1.2 Uso crítico de tecnologías de profesores universitarios

Álvarez (2021) tuvo por objetivo el definir los requisitos de formación para la navegación segura en Internet por parte del profesorado universitario, por lo cual, el autor presenta una visión de patrones de comportamiento, niveles de aprendizaje y competencias que se enfocan en el uso seguro e importante de Internet en los docentes.

El estudio fue exploratorio y utilizó una herramienta basada en la estructura digital *DigComp*. La muestra incluyó a 338 profesores pertenecientes a 39 universidades de nueve países de América Latina. Los resultados muestran que las universidades forman a los profesores en habilidades digitales, pero tienen poca formación en seguridad, también tienden a vivir experiencias con consecuencias negativas, pero aumentan significativamente la competencia, por otra parte, el 93% de los encuestados indicó haber participado en los últimos años en actividades educativas y de formación sobre el uso de las tecnologías en la educación, Internet, aplicaciones generales o herramientas multimedia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

2.1.3 Desarrollo de la competencia digital en profesores universitarios

1 Quiñonez *et al.* (2021) basaron su investigación en un análisis del impacto de un curso de formación en el desarrollo de habilidades digitales impartido en una universidad pública del sureste de México a partir del Modelo Integrado de Formación Educativa (MEFI). Se concluyó que las competencias digitales de los docentes mejoraron significativamente, de la misma manera manifestaron en entrevistas que el curso les permitió complementar su formación debido a que conocieron el uso correcto y profundo de las innovaciones tecnológicas en el ámbito educativo.

2.1.4 Competencias digitales docentes: Experiencia en nivel universitario

Sánchez *et al.* (2021) en su investigación expusieron el nivel de competencia para el uso de las tecnologías en cuatro universidades de México. La muestra fue de tipo no probabilística con una población de 375 profesores de nivel superior. La herramienta utilizada para la recopilación de la información fue una escala tipo *Likert* de 20 reactivos. Finalmente, se encontró que el 46.9% de los docentes entrevistados, con poca frecuencia, rara vez o nunca han creado o editado contenido utilizando nuevas tecnologías.

De esta manera, se concluyó que es necesario impulsar el desarrollo de la formación de los profesores respecto a la implementación de plataformas digitales así como las herramientas que fueron necesarias para la modalidad virtual, de este modo se pretende lograr una interacción y diálogos correctos entre profesores - alumnos, esto se debe a que se observó que la poca o nula participación de los académicos en los entornos

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

virtuales, demostró perjudicar en este ámbito, por medio de la desigualdad en el acceso y uso a las innovaciones tecnológicas.

2.1.5 Educación Superior virtual: Validación de instrumento en la UEPSE

Ruiz-Rabasco *et al.* (2021) valoraron las realidades y experiencias del proceso de enseñanza aprendizaje virtual de profesores y alumnos de la universidad, a través de una evaluación de propiedades psicométricas, durante la pandemia COVID-19; valorando la consistencia y fiabilidad de los *items* utilizados.

La muestra fue de un total de 67 docentes y 1819 estudiantes, pertenecientes a todas las carreras y facultades. Se analizó la confiabilidad de los datos mediante Alpha de Cronbach, obteniendo 0.86 (α). Por otro lado, las fichas de evaluación fueron asignadas a diferentes jueces quienes evaluarían: completitud, claridad, pertinencia y consistencia de cada ítem, con una puntuación de confianza de 0.98. Al final, durante la investigación, se desarrolló y ajustó una nueva versión de la herramienta, que resultó ser aún más relevante que la versión anterior.

2.1.6 Estrategias tecnológicas en el proceso de enseñanza y aprendizaje

Sánchez *et al.* (2019) analizaron cómo la pedagogía ha transformado las estrategias en la educación. Este artículo con base en un enfoque cualitativo ha sido utilizado en investigaciones similares en educación.

Se realizó una búsqueda en un total de alrededor de 50 documentos de acuerdo con los criterios de búsqueda: publicaciones desde 2000, excluyendo clásicos y teorías; preferiblemente artículos en inglés, incluidos artículos en español y artículos publicados con mayor frecuencia en revistas reconocidas. Los hallazgos revelaron que la tecnología

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

ha logrado avances significativos en todos los aspectos de la vida diaria en los últimos años. Las conclusiones determinaron que la difusión de estrategias pedagógicas en los procesos educativos mediados por tecnologías e innovaciones ha potenciado significativamente la labor pedagógica.

2.1.7 TIC's aplicadas a la educación superior por docentes de la UMC

Venegas *et al.* (2020) lograron identificar los niveles de conocimiento, preparación y uso de las innovaciones tecnológicas por parte de los docentes en su práctica personal y pedagógica dentro de la UMC.

La investigación fue de tipo descriptiva y asociativa, no experimental; desarrollada por una aplicación de encuesta certificada. La muestra fue de carácter no probabilístico e incluyó a 69 profesores de la UMC y 151 académicos de posgrado. Los datos recopilados permitieron concluir que los docentes de la UMC utilizan las tecnologías de la información en su práctica docente con una falta considerable de conocimiento y formación.

2.1.8 Docentes en América Latina COVID-19 con tecnologías digitales

El estudio realizado por el Laboratorio Latinoamericano y Caribeño de Investigación e Innovación Educativa, SUMMA y el Centro de Estudios para la Transformación de la Educación Social (2021) destacó que Chile fue el país con mayor número de sistemas de educación a distancia, en América Latina y el Caribe, con excelentes condiciones educativas y tecnológicas ante la pandemia COVID-19.

Este estudio se llevó a cabo desde enero de 2021 hasta junio de 2021, donde se examinaron 15 experiencias de docentes, con un enfoque particular en el aprendizaje

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

profesional y sus conexiones con el uso de tecnologías digitales, los cuales pertenecían a Chile, México, Colombia, Honduras y Uruguay. Estos países fueron seleccionados para reflejar la diversidad de entornos y respuestas a la educación continua, utilizando una tipología basada en dos indicadores que representan variables clave para la transición a la educación a distancia: las condiciones tecnológicas iniciales que ya existían, y las previas a la aparición del COVID-19.

Entre los principales resultados presentados, el entorno en el que trabajaban los docentes antes de la pandemia ilustró la diversidad e integración de la tecnología en la práctica educativa y la vida docente. Los aspectos clave incluyeron experiencias de aprendizaje en el uso de la tecnología, equipos y conectividad, modelos curriculares y el papel de la tecnología en el aprendizaje.

También en esta experiencia, el papel del docente y su proactividad resultaron cruciales. Los resultados determinaron que estos se movilaron para encontrar soluciones a los problemas relacionados con el cierre de escuelas, utilizar las TIC para apoyar la enseñanza y el aprendizaje, mantener la comunicación con las familias y facilitar la colaboración con los demás docentes. Sin embargo, se expusieron diversas barreras relacionadas con la conectividad, la industria de las TIC, así como el impacto negativo de la educación a distancia entre estudiantes y profesores.

2.1.9 Educación de América Latina en el contexto de COVID-19

El informe elaborado por la Internacional de la Educación para América Latina [IEAL] (2021) analizó la situación social, laboral y educativa en la que se encontraron los

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

países de la región tras la aparición de la pandemia del COVID-19 y cómo se desarrolló esta situación a partir del 2020.

Con base en la información obtenida en este estudio, se observó que el contexto de la pandemia resultó ser favorable para la promoción de la educación, la cual se conservó en cierta medida en la región. También se detectó que diversos organismos internacionales (como la UNESCO y UNICEF) y organismos financieros internacionales, como el Banco Interamericano de Desarrollo y el Banco Mundial, se aliaron con las empresas privadas de informática y comunicaciones.

Por otro lado, el informe destacó la disparidad entre las escuelas públicas y privadas en la provisión de capacitación para responder a esta emergencia, ya que las escuelas privadas anteriormente tenían claro la inclusión de recursos y herramientas digitales, lo que explicó la brecha de financiamientos para tecnologías equivalentes en cuanto a equipamiento y conocimientos.

2.1.10 Percepciones docentes frente al cambio de modalidad

Briceño y Castellanos (2021) basaron su investigación en una herramienta para examinar las percepciones de los profesores universitarios en ocho categorías, a saber:

a) acceso y conectividad, b) conocimiento de la plataforma y la tecnología, c) hábitos y espacio de trabajo, d) preparación del aula, e) compromiso y motivación de los estudiantes, f) desarrollo de la clase, g) dificultades y actitudes de los docentes frente al cambio, así como h) el proceso de comunicación entre docentes y supervisores.

Esta investigación respaldó la creencia común de que es más probable que los maestros virtuales sean virtuales que los maestros presenciales porque lo eran antes de

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

la pandemia. Asimismo, los resultados demostraron que los problemas que se comparten entre los docentes presenciales y virtuales están relacionados con situaciones que no se presentan en la educación virtual; por ejemplo, los estudiantes continúan teniendo dificultades para acceder a Internet y su disposición a usar la tecnología continúa. La negatividad de los estudiantes es un factor importante por revisar, especialmente en los resultados de los alumnos que participan en clases virtuales de forma común, ya que se asume que el trabajo virtual es normal para ellos, lo que puede estar relacionado con la inclusión experimentada, en apoyo a quienes imparten clases de manera presencial y sufrieron este cambio repentino debido al COVID-19.

2.1.11 Tecnología para innovar la práctica docente

A pesar de que se ha logrado un tremendo progreso en la integración de la tecnología educativa en las últimas décadas, la integración de los maestros ha sido un proceso lento. Antes de la pandemia, un gran número de docentes ya habían manifestado la necesidad de más y mejor capacitación en el uso de herramientas digitales.

Debido a la pandemia y la urgencia del tema, el Grupo de Trabajo de Tecnología e Innovación Educativa (2021) publicó un informe con los resultados de su estudio, de cómo la tecnología logró facilitar las prácticas educativas durante y después de la pandemia. La investigación se centró en tres áreas en las que los maestros utilizaron y seguirán utilizando la tecnología educativa para la enseñanza: gestión de contenidos y entornos de aprendizaje, gestión y evaluación de la instrucción, y desarrollo profesional y comunidades de práctica.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Esta investigación concluyó en que se necesitaron marcos y estándares de competencia digital para proteger el uso de datos y la propiedad intelectual en las plataformas de educación digital, debido a la poca confianza que presentaron los docentes ante el uso de estas, también se presentó la necesidad de mejorar las plataformas digitales: sus repositorios de contenido o capacidades de evaluación del aprendizaje y las habilidades de los docentes para usarlas de manera efectiva.

2.1.12 La educación en tiempos de COVID-19

El informe de la Comisión Económica para América Latina y el Caribe [CEPAL] y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO] (2021) alertó a la comunidad internacional la ampliación de las desigualdades existentes provocadas por la pandemia COVID-19 en términos de acceso a la educación, equidad y calidad, donde se vieron afectados particularmente los grupos más vulnerables.

En este estudio se identificó una necesidad urgente de calcular y establecer prioridades de gasto para los sistemas educativos nacionales, así como garantizar que la educación fuese protegida como uno de los derechos humanos básicos, y su potencial transformador se utilizara no solo para construir sistemas sostenibles, sino también para promover la recuperación social. El estudio llama a prestar especial atención al riesgo de abandono entre los más vulnerables a la pandemia y las consiguientes crisis sanitaria, social y económica.

En conclusión, los desafíos de los sistemas educativos estuvieron relacionados con la preparación ante las crisis, es decir, la resiliencia y adaptabilidad de los sistemas

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

educativos y los beneficios sociales resultantes. Por lo tanto, se enfatizó la necesidad de que las estrategias de recuperación del proceso educativo requirieron coordinación y comunicación entre la planificación e implementación del sector educativo y otros sectores, especialmente en las áreas de salud, nutrición y protección social.

2.1.13 Educación superior: desafíos en tiempos de COVID-19

Con la llegada de la pandemia mundial COVID-19, el sistema universitario español trató de garantizar la calidad del aprendizaje introduciendo nuevos métodos adaptados a los modelos de aprendizaje híbrido y en línea. Con esta premisa en mente, el objetivo de López (2022) fue identificar los principales retos a los que se enfrentó la educación a distancia en la Facultad de Ciencias de la Comunicación de la Universidad de A Coruña y cómo sus docentes se adaptaron al entorno de esta nueva modalidad.

El enfoque de esta investigación fue de tipo descriptivo. La herramienta metódica utilizada para recolectar información se basó en un cuestionario con preguntas abiertas y cerradas dispuestas en cuatro bloques temáticos: recursos metódicos utilizados en el modelo de aprendizaje virtual, ventajas y desventajas de la enseñanza en línea, sugerencias de mejoras encontradas durante la pandemia y aprendizaje en línea y presencial.

En el contexto analizado, las herramientas tecnológicas jugaron un papel importante como promotoras de nuevos enfoques en el proceso de aprendizaje virtual. Los contenidos de apoyo audiovisual o las grabaciones de sesiones teóricas de fácil acceso a través de la plataforma de la institución fueron los recursos más valiosos para los docentes y alumnos. Por otro lado, el 45% del profesorado admitió una falta de

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

competencia más o menos causada por las nuevas tecnologías, y el 47% del profesorado afirmó haber identificado la necesidad de competencias transversales, como la autonomía.

2.1.14 Experiencias docentes en tiempos de COVID-19 en América Latina

Para la Internacional de la Educación en América Latina [IEAL] (2021), la dimensión pedagógica involucrada en todo proceso de aprendizaje siempre ha sido uno de los principales aspectos de la práctica docente dignos de reflexión y formación continua de los docentes. Por ello, entre abril de 2021 y julio de 2021, la IEAL realizó un trabajo de investigación inicial en tres países (Honduras, Costa Rica y Paraguay) para conocer el impacto de la pedagogía educativa al afrontar la pandemia de COVID-19.

Los hallazgos clave de este estudio destacaron que los centros de aprendizaje consultados se enfrentaron a tres retos con la enseñanza virtual: la mayoría de los estudiantes, tienen poco o ningún acceso a Internet, existió falta de conocimiento de los docentes sobre las TIC y/o el uso de las tecnologías en general, especialmente los recursos técnicos utilizados en el proceso de enseñanza, así como la falta de conocimiento de los estudiantes y sus padres sobre las innovaciones tecnológicas.

2.2 Marco conceptual

2.2.1 Innovación

Diversos autores manejan este concepto, a continuación, se ejemplifican algunos:

- El Diccionario de la Real Academia Española (2021) define a la innovación como la modificación de un producto o la invención de uno nunca visto, así como su

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

introducción en un mercado, también plantea la aparición de nuevos y desconocidos en un contexto determinado.

- Según Rincón *et al.* (2022) para las empresas, la innovación es un proceso importante para crear una ventaja competitiva en el mercado, mediante la introducción de productos y servicios nuevos, manteniendo su eficacia.
- Para Soriano y Jiménez (2022) la innovación es un conjunto de actividades vinculadas a los procesos en los que participan: los mercados, los negocios, el trabajo en red y por supuesto la competencia, aprovechando el uso de habilidades organizativas, así como la creatividad de sus elementos.

2.2.2 Tecnología

- Laureano (2022) señala que la tecnología es el medio por el que se intenta crear un orden mundial, en varias ocasiones respaldado por el conocimiento científico.
- La tecnología para Nazar (2022) es el término que se refiere a un conjunto extremadamente diverso de herramientas, instrumentos, máquinas y sistemas.
- La Real Academia Española (2022) lo define como un conjunto de teorías y métodos que permiten poner en práctica el conocimiento científico.

2.2.3 Innovación tecnológica

- Quintero, Orjuela y Escobar (2022) lo entienden como un conjunto de teorías y métodos que permiten poner en práctica el conocimiento científico para un sector o producto en particular, se refiere a un conjunto de herramientas y procedimientos que se consideran necesarios para lograr ciertos objetivos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Para Hidalgo y Gisbert (2022) es el proceso de inventar y utilizar técnicas comerciales de productos o servicios diseñados para generar nuevas ideas y finalmente implementarlas.
- La innovación tecnológica juega un papel importante en las empresas, ayudando a los gerentes a lograr sus objetivos, según Pérez *et al.* (2022) argumenta que todos los factores y actividades de innovación pueden desarrollarse exitosamente junto con una estrategia de innovación.

2.2.4 Minería de datos

- Según Ibañez, Echeverri y Castrillón (2022) la minería de datos es un conjunto de métodos estadísticos para distinguir predictores de mayor o menor importancia a partir de una variable experimental.
- La minería de datos para Castrillón, Castillo y Castaño (2022) es el proceso responsable de la minería no trivial de patrones útiles ocultos implícitamente en los datos, así como la forma más rápida de descubrir grandes cantidades de información.
- Según Pérez *et al.* (2022) la minería de datos es el proceso de encontrar y extraer información útil, patrones y tendencias previamente desconocidos de grandes cantidades de datos que se pueden almacenar en una base de datos.
- Otra de las definiciones existentes establece que la minería de datos es explicada por Gordan *et al.* (2022) como un proceso no trivial para descubrir patrones concebibles ocultos en los datos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

2.2.5 Reglas de asociación

Para IBM (2022a) las reglas de asociación restringen un resultado particular (como la compra de un determinado producto) a un conjunto de condiciones. La ventaja de los algoritmos de reglas de asociación sobre los algoritmos de árboles de decisión más estándar es que pueden existir asociaciones entre cualquier atributo. Los algoritmos de árboles de decisión generarán reglas con una salida, mientras que los algoritmos de asociación intentarán encontrar muchas reglas, cada una de las cuales puede tener resultados diferentes.

Las reglas de asociación son un tipo de patrón de minería de datos que buscan encontrar relaciones o patrones comunes en los datos. Estas reglas se utilizan para descubrir relaciones entre diferentes elementos en un conjunto de datos, como productos que a menudo se compran juntos en un supermercado o eventos que suceden con cierta frecuencia juntos (Diaz *et al.*, 2023).

Según los autores, las reglas de asociación son un enfoque de aprendizaje automático no supervisado que busca encontrar patrones en los datos. Estos patrones son entendidos como reglas de la forma "si A entonces B", donde A y B son elementos o conjuntos de elementos del conjunto de datos.

Por ejemplo, una regla de asociación podría ser "si un cliente compra una botella de vino, entonces es probable que también compre queso". Otra regla podría ser "si un cliente compra una televisión, entonces es probable que también compre un sistema de sonido". Existen varios algoritmos utilizados para encontrar reglas de asociación, como

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

el algoritmo *Apriori*, que busca reglas de asociación sólidas en los datos, y el algoritmo *FP-growth*, que es una técnica eficiente para encontrar patrones frecuentes en los datos.

En resumen, las reglas de asociación son una técnica de minería de datos que busca encontrar patrones comunes en los datos, y se utilizan para descubrir relaciones entre diferentes elementos en un conjunto de datos. Son una herramienta valiosa para entender mejor los datos y tomar decisiones informadas en aplicaciones como la inteligencia de negocios, el marketing y la recomendación de productos.

2.2.6 Árboles de decisión

El árbol de decisión para Riascos y Molina (2015) es una herramienta para seleccionar, estructurar y evaluar problemas en condiciones donde se desconocen varios aspectos. Estos permiten evaluar planes de acción, así como consecuencias, calcular probabilidades y ejecutar simulaciones, través de ellos se presentan diversas alternativas de decisión existentes, como la probabilidad de ocurrencia (Custode & Iacca, 2023).

Los árboles de decisión según Cañete *et al.* (2021) son una técnica de aprendizaje automático supervisado utilizada para la clasificación y la predicción de variables categóricas o continuas. Los árboles de decisión se representan gráficamente como un árbol con ramas y nodos que representan las decisiones y las consecuencias de las decisiones.

Un árbol de decisión comienza con una pregunta raíz, que es una pregunta sobre los datos. A partir de esta pregunta, se toman decisiones y se realizan preguntas adicionales para llegar a una conclusión, que es la clase o el valor predicho. Los nodos internos del árbol representan las preguntas o las decisiones que se toman sobre los

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

datos. Los nodos hoja del árbol representan las respuestas o las clases finales (Nanfack *et al.*, 2022).

Los árboles de decisión son fáciles de entender y de interpretar, y son útiles para problemas de clasificación y regresión. Los árboles de decisión también son utilizados para problemas de selección de características, ya que pueden proporcionar una idea de la importancia relativa de las características en un problema dado.

En resumen, los árboles de decisión son una técnica de aprendizaje automático supervisado que se utilizan para la clasificación y predicción de variables categóricas o continuas. Se representan gráficamente como un árbol con ramas y nodos que representan las decisiones y las consecuencias de las decisiones, son fáciles de entender y interpretar, y son útiles para problemas de clasificación y regresión.

2.2.7 Proceso KDD

Singhal & Himanshu (2022) lo definen como descubrimiento de conocimiento en bases de datos o extracción de conocimiento, el cual es el proceso donde se analizan datos y se descubren algoritmos, dentro de los límites de razonabilidad del desempeño computacional, generando un determinado conjunto de muestras o modelos de datos. Consta de cinco fases, definidas por los autores de la siguiente manera:

- **Integración y recopilación de datos:** esta etapa identifica las fuentes de información que pueden ser de utilidad y dónde obtenerlas.
- **Seleccionar, eliminar y transformar:** incluye la detección de datos faltantes o en su caso perdidos, la creación de nuevos atributos y la enumeración de atributos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- **Minería de datos:** El propósito de esta etapa es adquirir nuevos conocimientos. Para ello se construye un modelo basado en patrones y relaciones entre datos que se pueden utilizar para predecir y comprender mejor los datos.
- **Evaluación e Interpretación:** Existen diferentes métricas para evaluar modelos: precisas, comprensibles e interesantes.
- **Propagación:** una vez que se ha construido y validado un modelo, se puede utilizar para una variedad de propósitos y se debe medir su crecimiento a lo largo del proceso.

El proceso KDD es una herramienta valiosa para extraer información valiosa de grandes conjuntos de datos. Es utilizado en una gran variedad de campos, como el marketing, la inteligencia de negocios, la seguridad de la información y la medicina, para tomar decisiones informadas y mejorar los procesos de negocio (Parlindungan *et al.*, 2022). En resumen, el proceso KDD es un proceso sistemático y iterativo para descubrir patrones útiles en grandes conjuntos de datos, que se compone de varios pasos que van desde la selección de datos hasta la interpretación y presentación de los resultados. Es una herramienta valiosa para la toma de decisiones informadas y la mejora de procesos en una gran variedad de campos.

2.2.8 Aprendizaje supervisado

El aprendizaje supervisado es definido por Wu *et al.* (2022) como un tipo de aprendizaje automático basado en el aprendizaje automático, una forma de convertir ciertos datos de entrada en salida basada en un conjunto de datos de muestra. En el

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

aprendizaje supervisado, el agente toma un par de entradas, así como salidas de ejemplo y aprende una función que especifica la salida en función de entradas específicas.

El aprendizaje supervisado, según Luo *et al.* (2022) es una técnica de aprendizaje automático donde el modelo es entrenado con datos etiquetados, es decir, datos que ya tienen una categoría o etiqueta asignada. El objetivo es que el modelo aprenda a predecir la etiqueta correcta para nuevos datos.

Existen dos tipos principales de aprendizaje supervisado, mencionados por Chiu *et al.* (2022): clasificación y regresión.

- Clasificación: el objetivo es asignar una categoría o clase a un conjunto de datos. Por ejemplo, clasificar un correo electrónico como spam o no spam, o clasificar una imagen como un perro o un gato. Algoritmos como árboles de decisión, Random Forest, Redes Neuronales son ejemplos de algoritmos utilizados en la clasificación.
- Regresión: el objetivo es predecir un valor numérico continuo. Por ejemplo, predecir el precio de una casa en función de sus características, o predecir la probabilidad de que un paciente tenga una enfermedad en función de sus resultados de laboratorio. Algoritmos como la regresión lineal, las redes neuronales son ejemplos de algoritmos utilizados en la regresión.

En resumen, el aprendizaje supervisado se basa en el uso de datos etiquetados para entrenar un modelo de tal forma que este pueda hacer predicciones o clasificaciones correctas para nuevos datos. Es una técnica ampliamente utilizada y muy efectiva para resolver problemas en una gran variedad de campos y aplicaciones.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

2.2.9 Aprendizaje no supervisado

Según Shen *et al.* (2022) este estilo de aprendizaje procesa el conocimiento a partir de datos sin “etiquetas”, por lo que pretende descubrir patrones desconocidos a partir de los propios datos que permitan clasificarlos, este tiene tres áreas de aplicación:

- **Agrupamiento:** cuando el objetivo es dividir el conjunto de datos en grupos lógicamente relacionados, pero previamente desconocidos;
- **Reducción de dimensiones:** cuando se pretende simplificar un conjunto de datos multidimensionales en un espacio de menor dimensión;
- **Detección de anomalías:** donde el objetivo es encontrar observaciones que no coincidan con el patrón esperado.

El aprendizaje no supervisado es una rama del aprendizaje automático donde el modelo es entrenado con datos no etiquetados. Es decir, los datos no tienen etiquetas o categorías predefinidas, sino que el algoritmo debe descubrir patrones y estructuras en los datos de forma autónoma (Montavon *et al.*, 2022).

En general, el aprendizaje no supervisado se utiliza a menudo para explorar y entender los datos, mientras que el aprendizaje supervisado se utiliza para predecir o clasificar etiquetas conocidas.

2.2.10 ISO/IEC 9126 norma de calidad del *software*

1 La ISO bajo el modelo ISO/IEC 9126, ha establecido un estándar internacional para la evaluación de la calidad de productos de *software* el cual fue publicado en 1992, afirman Juárez y Torres (2022); este se dio a conocer con el nombre de “*Information*

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

technology – Software product evaluation: Quality characteristics and guidelines for their use”, en el cual se establecen las características de calidad para productos de *software*.

La ISO/IEC 9126 es aplicable a todo tipo de *software* y está basada en un modelo jerárquico compuesto de tres niveles, los cuales son: características, subcaracterísticas y métricas. En el primer nivel se encuentran seis características principales: funcionalidad, fiabilidad, eficiencia, facilidad de mantenimiento, portabilidad y facilidad de uso. Estas características (o factores) se componen por 27 subcaracterísticas (subfactores) relacionadas con la calidad externa, y 21 subcaracterísticas relacionadas con la calidad interna según Noorman *et al.* (2022).

El estándar se divide en seis categorías de características: funcionalidad, confiabilidad, usabilidad, eficiencia, mantenibilidad y portabilidad. Cada categoría se divide en atributos específicos que se utilizan para evaluar el producto de *software*.

La funcionalidad se refiere a las características y funciones del producto de *software* que cumplen con los requisitos del usuario. La confiabilidad se refiere a la capacidad del producto de *software* para funcionar correctamente y de manera consistente en condiciones normales de uso. La usabilidad se refiere a la facilidad de uso y comprensión del producto de *software*. La eficiencia se refiere a la rapidez y el consumo de recursos del producto de *software*. La mantenibilidad se refiere a la facilidad de modificar y mantener el producto de *software*. Por último, la portabilidad se refiere a la capacidad del producto de *software* para ser ejecutado en diferentes entornos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

2.3 Marco tecnológico

En virtud de la amplia cantidad de los datos a analizar, se requiere a la minería de datos, por lo que se mencionan las herramientas más reconocidas para tratamiento de los datos, después se presenta un comparativo, para facilitar la selección de las más adecuadas, con las que finalmente se desarrolló la investigación.

2.3.1 Herramientas para minería de datos - *software* libre

2.4.1.1 *Weka*

Weka es un acrónimo que proviene de *Waikato Environment for Knowledge Analysis* en la Universidad de Waikato. Incluye varias herramientas de visualización gráfica y varios algoritmos para el análisis de datos y el modelado predictivo. La interfaz de usuario facilita el acceso a muchas funciones (Gurazada *et al.*, 2022). *Weka*, disponible gratuitamente para su uso bajo *GNU*, contiene un conjunto de algoritmos de minería de datos que son más potentes que los algoritmos de filtrado y clasificación, así como algoritmos de formación de asociaciones y agrupación de datos (Bai *et al.*, 2022).

Weka es un *software* de minería de datos y aprendizaje automático que proporciona una plataforma fácil de usar para la exploración y el análisis de datos, así como para la construcción y evaluación de modelos de aprendizaje automático. *Weka* contiene una variedad de herramientas y algoritmos, incluyendo árboles de decisión, regresión, *clustering*, redes neuronales y más (Saleh *et al.*, 2022). Además, también proporciona varios filtros para preprocesar los datos y herramientas para la visualización de resultados.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

2.4.1.2 Orange

Orange es un *software* gratuito de aprendizaje automático para minería de datos y análisis mediante predicciones; escrito en C++, implementa algoritmos de *data mining*, operaciones de procesamiento y representación gráfica. Fue desarrollado en el Laboratorio de Bioinformática de la Facultad de Ciencias de la Computación y Ciencias de la Información de la Universidad de Ljubljana, Eslovenia. Se trata de una aplicación multiplataforma que se distribuye bajo licencia *GPL* y proporciona componentes para: entrada y salida de datos, preprocesamiento de datos, etc. (Hudaefi, 2023).

Orange ofrece una amplia variedad de algoritmos de minería de datos, incluyendo análisis de componentes principales, *k-means*, árboles de decisión, entre otros. También cuenta con una variedad de herramientas para la visualización de datos y la creación de informes. Además, *Orange* se integra con otras herramientas de análisis de datos como *Python* y *R*, permitiendo a los usuarios utilizar algoritmos y librerías específicas. *Orange* es utilizado ampliamente en campos como la ciencia de datos, la investigación, la inteligencia artificial, la educación, entre otros (Ramadhan & Wibowo, 2022).

2.4.1.4 RapidMiner

YALE (Yet Another Learning Environment), es un programa de computadora utilizado para analizar datos. Esto permite el desarrollo de procesos de análisis de datos conectando operadores en un entorno gráfico. Se utiliza en investigación y aplicaciones comerciales. *RapidMiner* ofrece más de 500 operadores orientados a la minería de datos. También le permite utilizar los algoritmos incluidos en *Weka*, que es una especie de código abierto bajo la licencia *AGPL* basada en *Java*. (Fernandes, 2022).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

RapidMiner es un *software* de minería de datos y análisis estadístico que permite a los usuarios importar, limpiar, transformar y modelar datos. Al igual que *KNIME*, también utiliza un enfoque visual para la creación de flujos de trabajo, lo que permite a los usuarios arrastrar y soltar nodos para crear flujos de trabajo complejos. *RapidMiner* según Yuliarina & Hendry (2022) ofrece una amplia variedad de algoritmos de minería de datos, incluyendo regresión, análisis de componentes principales y árboles de decisión, así como también una variedad de herramientas para la visualización de datos y la creación de informes. También se integra con otras herramientas como *R* y *Python*, lo que permite a los usuarios aprovechar las capacidades de estos lenguajes en sus flujos de trabajo de minería de datos. *RapidMiner* es ampliamente utilizado en una variedad de campos, como ciencia de datos, investigación, *marketing*, finanzas, entre otros.

2.4.1.5 Knime

Según Klenzi *et al.* (2018), (*Konstanz Information Miner*) es un entorno de minería de datos construido sobre *Eclipse* y programado en *Java*, desarrollado inicialmente en el Departamento de Bioinformática y Minería de Datos de la Universidad de Konstanz en Alemania. Es una herramienta gráfica para crear modelos en una serie de nodos y flechas que se muestran y vinculan de forma gráfica e interactiva.

KNIME es un *software* libre de minería de datos y análisis estadístico que permite a los usuarios crear flujos de trabajo para análisis de datos, visualización y modelado. A diferencia de otras herramientas de minería de datos, *KNIME* utiliza un enfoque visual para la creación de flujos de trabajo, lo que permite a los usuarios arrastrar y soltar nodos para crear flujos de trabajo complejos, según Maraza *et al.* (2022).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

KNIME ofrece una amplia variedad de nodos para diferentes tareas de minería de datos, como limpieza de datos, análisis estadístico, minería de datos y modelado predictivo.

2.3.2 Herramientas para minería de datos - *software* propietario

2.4.2.1 *Microsoft Excel*

Excel es un *software* de aplicación proporcionado por *Microsoft Office* que se puede usar para realizar operaciones en números organizados en tablas, realizar fórmulas, crear gráficos, resolver cálculos de probabilidad, análisis de datos, estadísticas descriptivas, generación de gráficos y más (Pérez *et al.*, 2022).

Microsoft Excel es una hoja de cálculo de Microsoft que permite a los usuarios crear y editar hojas de cálculo con cálculos y gráficos. Es una herramienta comúnmente utilizada para el análisis de datos, la creación de tablas y gráficos, y la planificación financiera. Excel tiene una interfaz de hoja de cálculo en la que las celdas se pueden utilizar para ingresar y editar datos, y las fórmulas y funciones se pueden utilizar para realizar cálculos (Brusco, 2021). También cuenta con un gran número de herramientas de análisis de datos, incluyendo tablas dinámicas, gráficos y tablas pivotantes, así como también un gran número de funciones y fórmulas para el cálculo automatizado. Excel es utilizado ampliamente en empresas, negocios, instituciones educativas y hogares debido a su versatilidad y facilidad de uso.

2.4.2.3 *SAS Enterprise Miner*

Esta herramienta de mercado para el análisis de datos intelectuales crea modelos de pronóstico precisos y describe una gran cantidad de datos mediante el uso de muchas

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

fuentes diferentes a través de procesos transparentes, lo que permite la cooperación. Incluye una interfaz de usuario visual, incluidos los principios de diseño generales para SAS y herramientas adicionales para trabajar fácilmente (Xu, 2022).

SAS Enterprise Miner es un *software* de minería de datos desarrollado por SAS Institute. Es una herramienta completa para el análisis estadístico y la minería de datos que permite a los usuarios importar, limpiar, transformar y modelar datos. También proporciona una variedad de técnicas de minería de datos, incluyendo regresión, análisis de componentes principales y árboles de decisión según Nath & Anand (2021). Además, *SAS Enterprise Miner* cuenta con una interfaz fácil de usar para la creación de flujos de trabajo y la automatización de tareas, lo que permite a los usuarios analizar grandes conjuntos de datos de forma rápida y eficiente.

2.3.3 Servicios para minería de datos en la nube

2.4.3.1 IBM SPSS Statistics

La plataforma de *software IBM SPSS Statistics*, según IBM (2022b) proporciona análisis estadístico avanzado, cuenta con una biblioteca de algoritmos de aprendizaje automático, análisis de texto, código abierto, integración de datos de gran tamaño y se implementa sin problemas en las aplicaciones. La facilidad de uso, la flexibilidad y la extensibilidad hacen que *SPSS Statistics* sea accesible para usuarios de todos los niveles. También es adecuado para proyectos de cualquier tamaño y complejidad.

IBM SPSS Statistics es un *software* de estadística y minería de datos desarrollado por IBM. El nombre *SPSS* es un acrónimo de "*Statistical Package for the Social Sciences*". El *software* se utiliza para realizar análisis estadísticos y minería de datos en

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

una variedad de campos, incluyendo la investigación social, la investigación de mercado, la investigación de operaciones y el análisis de datos empresariales (Boubker *et al.*, 2022).

IBM SPSS Statistics es ampliamente utilizado en la industria y en la investigación académica para realizar análisis estadísticos y minería de datos y ayudar en la toma de decisiones informadas. El *software* es compatible con diferentes sistemas operativos y está disponible en varios idiomas.

2.4.3.2 Oracle Business Intelligence Server

Oracle Business Intelligence Server es una plataforma única que permite a los clientes descubrir nuevas ideas y tomar decisiones comerciales más rápidas e informadas, ya que ofrece una analítica visual y un descubrimiento de autoservicio ágiles y la mejor solución de analítica corporativa disponible. Servicio móvil instantáneo, cuadros de mando con un gran nivel de interacción, informes operativos potentes, alertas en el momento adecuado, búsqueda de contenido y metadatos, administración de estrategias, acceso nativo a fuentes de *big data*, computación sofisticada en memoria y administración simplificada de sistemas, son algunos de los aspectos que este servicio de *Oracle* ofrece (Oracle, 2022a).

2.4.3.3 Google Cloud Datalab

Según *Google* (2023) *Cloud Datalab* permite explorar, analizar y visualizar datos de forma interactiva y fácil utilizando lenguajes familiares como *Python* y *SQL*. Las plantillas de *Jupyter* preinstaladas, las notas introductorias y los tutoriales incluidos con *Google Cloud Datalab* enseñan a los usuarios cómo: acceder a datos, analizar, rastrear

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

y visualizar datos, usando cuadernos de *Python*, aprendizaje automático de *TensorFlow* y la API de *Google Analytics*, llamada *Google BigQuery*.

2.4.3.4 Apache Hadoop

Apache Hadoop es una plataforma de *software* de código abierto creado y mantenido por una red mundial de desarrolladores, que proporciona una gran capacidad de almacenamiento para cualquier tipo de datos, una enorme potencia informática y de procesamiento de tareas concurrentes prácticamente ilimitadas. La descarga, el uso y la coautoría son gratuitos, aunque cada vez hay más versiones comerciales de *Hadoop* (comúnmente denominadas "distribuciones"). Las distribuciones populares incluyen *Cloudera*, *Hortonworks*, *MapR*, *IBM BigInsights* y *PivotalHD* (SAS Institute Inc., 2022).

Apache Hadoop es un *software* de código abierto para procesamiento de datos distribuidos en *clúster*. Fue diseñado para manejar grandes cantidades de datos de manera eficiente y escalable, y se utiliza comúnmente en aplicaciones de *big data*.

Hadoop consiste en dos componentes principales: el sistema de archivos distribuidos *HDFS* y el motor de procesamiento *MapReduce*. El *HDFS* permite a los usuarios almacenar grandes cantidades de datos en un *clúster* de nodos, mientras que *MapReduce* permite a los usuarios procesar esos datos de manera distribuida (Ganim et al., 2022).

2.4.3.5 Microsoft Azure Machine Learning

Azure Machine Learning es un servicio en la nube el cual permite acelerar y administrar los proyectos de aprendizaje automático. Los profesionales de aprendizaje automático, especialistas de datos e ingenieros pueden usarlo en sus flujos de trabajo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

diarios como parte de su capacitación o creación de nuevos modelos. En *Azure Machine Learning*, se permite el uso de *scripts* de entrenamiento en la nube. A menudo, los clientes incorporan modelos que han creado en código abierto, a fin de ponerlos en marcha en la nube y poder monitorearlos (Microsoft, 2023a).

Microsoft Azure Machine Learning es una plataforma de aprendizaje automático en la nube ofrecida por Microsoft como parte de su suite de productos de inteligencia artificial de *Azure*. Es una herramienta de aprendizaje automático que permite a los usuarios crear, implementar y publicar modelos de aprendizaje automático en la nube de manera fácil y rápida, según Klochko *et al.* (2022).

2.3.4 Lenguajes de programación para minería de datos

2.4.4.1 Java

Java es un lenguaje de programación comercializado por primera vez en el año de 1995 por *Sun Microsystems*. En la actualidad, existen diversas aplicaciones y sitios *web* que no funcionarían si no existiera *Java*. *Java* es un lenguaje seguro, fiable y fácil de aprender. *Java* se encuentra en todas partes: desde las computadoras portátiles hasta los centros de datos, consolas de juegos, supercomputadoras, teléfonos móviles hasta Internet (Java, 2023).

Java para Aung *et al.* (2022) es un lenguaje muy versátil y se utiliza en una amplia gama de aplicaciones, incluyendo desarrollo de aplicaciones *web*, desarrollo de aplicaciones móviles, desarrollo de videojuegos, inteligencia artificial y *big data*. *Java* también es uno de los lenguajes de programación más populares y ampliamente

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

utilizados en todo el mundo, y cuenta con una gran comunidad de desarrolladores y una amplia variedad de bibliotecas y herramientas.

En resumen, *Java* es un lenguaje de programación potente, versátil y ampliamente utilizado que permite a los desarrolladores crear aplicaciones de alta calidad para una amplia variedad de plataformas y dispositivos.

2.4.4.2 Lenguaje R

R es un lenguaje de programación derivado de otro llamado *S*, este cuenta con más de 10 mil paquetes, cubriendo una amplia gama de campos, desde la investigación financiera hasta el análisis de datos Bayesiano y más. Es multiplataforma, permite a los usuarios definir sus propias funciones, tiene excelentes capacidades gráficas y es gratuito (Jiménez, 2019). *R* es un lenguaje de programación y un entorno de desarrollo para análisis estadístico y gráficos. Fue desarrollado por Ross Ihaka y Robert Gentleman en los años 90 en la Universidad de Auckland, Nueva Zelanda. Incluye espejos para servidores, los cuales distribuyen todos los paquetes *R* utilizados para el procesamiento de datos (Grolemund & Wickham, 2017).

R es ampliamente utilizado en diversas áreas de investigación, incluyendo biología, economía, psicología y estadística. Ofrece una amplia variedad de herramientas para análisis estadísticos, modelado y visualización de datos, lo que lo convierte en una herramienta popular para científicos de datos, investigadores y analistas de datos.

Además, los autores Giorgi, *et al.* (2022) expresan que *R* es un lenguaje de programación de código abierto, lo que significa que está disponible de forma gratuita y puede ser utilizado y modificado por cualquier persona. También cuenta con una amplia

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

comunidad de desarrolladores y usuarios que han creado una gran cantidad de paquetes y herramientas adicionales para el análisis de datos.

2.4.4.3 Python

Según *Python* (2022) es un lenguaje de programación informático. Tiene estructuras de datos de alto nivel y cuenta con un sistema de programación orientado a objetos. El intérprete de *Python* y su biblioteca estándar están disponibles gratuitamente en formato fuente y binario en el sitio *web* de *Python*.

Python se puede clasificar como un lenguaje de alto nivel que puede ser interpretado por diferentes sistemas operativos y acepta diferentes métodos de programación; también incluye reglas de estilo en su estructura de código para estandarizar cómo se programa sobre él (Bahit, 2018).

2.4.4.4 Julia

Julia es un lenguaje de programación de propósito general diseñado para ser rápido y eficiente. Fue desarrollado por un grupo de investigadores en el Instituto Tecnológico de Massachusetts (MIT) con el objetivo de crear un lenguaje de programación que combinara la velocidad de los lenguajes compilados con la interactividad y facilidad de uso de los lenguajes interpretados. *Julia* es un lenguaje simple, rápido y de código abierto, según Bezanson *et al.* (2017).

La eficiencia de *Julia* es casi comparable a la de lenguajes de programación estáticos como *C/C++* y *Fortran* (Perkel, 2019). *Julia* se está convirtiendo rápidamente en un lenguaje altamente competitivo en ciencia de datos y computación científica general debido a que es tan fácil de usar como *R* o *Python*.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

2.4 Marco legal

2.4.1 Licencias de *software* libre

2.5.1.1 Licencia AGPL

9 La *Affero GPL (AGPL)* es una licencia libre *copyleft* que pertenece a la *Free Software Foundation*. Es una licencia que nació a partir de la *GPL*, la cual contiene una cláusula que obliga a distribuir el código fuente modificado en aplicaciones que se ejecutan a través de las redes. La *AGPL* solo se aplica si el *software* está diseñado para ofrecer servicios de red, como lo son los servidores *web*, servidores de correo o de juegos multiplataforma (Zhao, 2022). En este segmento, se encuentra *RapidMiner*, cuenta con su versión gratuita, versión de prueba gratis y suscripción (para el método de suscripción se requiere una cotización en caso de ser una empresa que requiera el *software*).

2.5.1.2 Licencia GNU GPL

22 La *GNU GPL* (por su traducción al español Licencia Pública General de *GNU*) es una licencia de *software* libre con *copyleft* que de igual manera pertenece por la *Free Software Foundation*. Los usuarios que utilicen programas con licencia *GPL* son libres de acceder al código fuente, modificarlo y redistribuir los cambios; siempre que compartan el programa completo (con o sin modificaciones) bajo esta misma licencia (GNU 2021). Los *softwares* para minería de datos que cuentan con este tipo de licencia son: *Weka*, *Orange*, *Knime* y *Julia*.

2.4.2 Licencias de *software* propietario

El *software* propietario (o protegido por derechos de autor) es una plataforma actualizable y administrada comercialmente con uso, modificación o redistribución

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

limitados, su código fuente no está disponible o pertenece a una compañía, como es el caso de *Windows* (Díaz, 2018). En esta clasificación, se encuentran las siguientes herramientas para minería de datos: *Microsoft Excel*, *Microsoft Azure Machine Learning*, *SAS Enterprise Miner*, *IBM SPSS*, *Oracle Business Intelligence Server*, *Google Cloud Datalab* y *Apache Hadoop*.

Microsoft Excel se encuentra bajo licencia por medio de suscripciones anuales, teniendo un costo actual de \$3,199 pesos mexicanos, también cuenta con una versión de prueba gratuita durante 30 días (Microsoft, 2023b). *Microsoft Azure Machine Learning*, tiene un costo de \$10 dólares americanos, por área de trabajo al mes, así como el costo extra por hora de experimentación en el entorno (\$1); esta suscripción incluye espacio de almacenamiento ilimitado y *API's web* de producción (Microsoft Azure, 2023).

SAS Enterprise Miner no cuenta con versiones de prueba disponibles para el público en general, la única forma de conocer el costo exacto de su *software* es enviando una solicitud de cotización en su página oficial, donde se detalla que empresa o sector utilizará el *software* y con qué propósito (*SAS Enterprise Miner*, 2022), sin embargo cuenta con la paquetería de *SAS Studio for Academycs*, la cual brinda una licencia de 6 meses gratuitos en todas las plataformas SAS, registrándose con un correo universitario.

IBM SPSS Statics sin embargo, tiene una versión gratuita al año, con una durabilidad de 30 días, incluyendo todos los complementos (estadísticas base, preparación de datos, *Bootstrap*, estadísticas avanzadas, regresión, entre otras); la cual se caracteriza por no contar con soporte técnico ni la versión básica del *SPSS*.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Para el plan de suscripción, a partir de \$99,00 dólares americanos por usuario autorizado al mes, se incluye el soporte técnico y tres complementos opcionales disponibles (IBM, 2022c).

Por otra parte, *Oracle Business Intelligence Server* admite un conjunto de opciones de distribución, análisis y acceso a la información, todo dentro de un entorno *web* totalmente integrado. *Oracle Business Intelligence* no se encuentra disponible en versión gratuita y tampoco ofrece pruebas de su *software*. La versión de pago de *Oracle Business Intelligence* está disponible a partir de un costo de \$150,00 dólares americanos por mes (Oracle, 2022b).

Google Cloud Datalab genera costos a partir de los recursos que el usuario utilice, así como el espacio de almacenamiento este vaya generando en la nube, teniendo un costo base de una instancia por \$498.70 pesos mexicanos por mes (Google Cloud, 2023). Finalmente, *Apache Hadoop* se basa en computadoras autoagrupadas y brinda una solución rentable para que sus usuarios puedan almacenar y procesar grandes cantidades de datos sin necesidad de colocarles formatos específicos.

El costo de *clúster* básico para procesamiento es de \$1 dólar americano por hora, dependiendo de las necesidades del usuario (SAS Institute Inc, 2022b). Para el desarrollo de este estudio se utilizará el sistema operativo *MacOs Monterey* Versión 12.6, con el número de licencia C02XL5TXJPXN y el sistema operativo *Windows 11* Versión Pro, con el número de licencia VK7JG-NPHTM-C97JM-9MPGT-3V66T; así mismo se utilizará la licencia de 6 meses con *SAS Studio for Academy*s registrada con el correo 212h19002@alumno.ujat.mx

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

2.4.3 Ley Federal de Derechos de Autor

El marco legal de esta investigación se relacionará con la “Ley Federal de Derechos de Autor”, publicado por el Diario Oficial de la Federación [DOF] (2005); la cual es la ley encargada de proteger los programas de cómputo, véase el anexo A, en donde se proporciona más información sobre las disposiciones generales de esta ley.

2.4.4 Ley Federal de Protección de Datos Personales

En el desarrollo de la investigación, se utilizó un conjunto de datos que contiene las encuestas aplicadas a profesores de educación superior. Es importante destacar que se debe hacer un uso responsable de la información recopilada, ya que existe una legislación en México que regula la protección de los datos personales. En este sentido, se debe tener en cuenta la Ley Federal de Protección de Datos Personales emitida en 2010, publicada de igual manera por el Diario Oficial de la Federación [DOF] (2010); la cual continua vigente y no ha sufrido modificaciones. Por lo tanto, es fundamental que se respeten los derechos de privacidad de los encuestados y se cumpla con las disposiciones legales en la recolección, tratamiento y uso de datos personales en todo momento; para más detalles de esta ley véase el anexo B, en donde se encuentran las disposiciones generales de esta.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

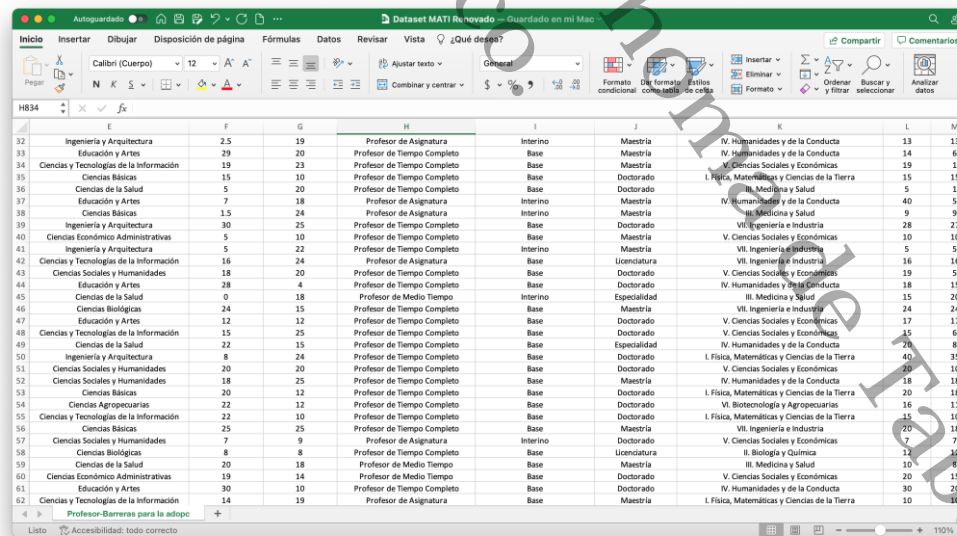
Capítulo 3. Aplicación de la metodología y desarrollo

3.1 Análisis de la información del *dataset*

El *dataset* implementado en esta investigación, fue proporcionado por García y Silva (2021) en el que se diseñaron y validaron originalmente 20 ítems. En la figura 2, se observa la parte inicial del *dataset*, el cual estaba conformado por 826 registros, los cuales eran las respuestas de los profesores en el instrumento diseñado. De igual forma, contenía las variables de género, edad, estado civil, División Académica, antigüedad, número de horas frente a grupo a la semana, categorías de contratación en UJAT, tipo de contratación en UJAT, grado académico, área de formación de su último grado académico y los 20 ítems que conformaron el instrumento aplicado (ver figura 2).

Figura 2.

Dataset utilizado en la investigación



	E	F	G	H	I	J	K	L	M
32	Ingeniería y Arquitectura	2.5	19	Profesor de Asignatura	Interino	Maestría	IV. Humanidades y de la Conducta	13	13
33	Educación y Artes	29	20	Profesor de Tiempo Completo	Base	Maestría	IV. Humanidades y de la Conducta	14	6
34	Ciencias y Tecnologías de la Información	19	23	Profesor de Tiempo Completo	Base	Maestría	V. Ciencias Sociales y Económicas	19	1
35	Ciencias Básicas	15	10	Profesor de Tiempo Completo	Base	Doctorado	I. Física, Matemáticas y Ciencias de la Tierra	15	15
36	Ciencias de la Salud	5	20	Profesor de Tiempo Completo	Base	Doctorado	III. Medicina y Salud	5	1
37	Educación y Artes	7	18	Profesor de Asignatura	Interino	Maestría	IV. Humanidades y de la Conducta	40	5
38	Ciencias Básicas	1.5	24	Profesor de Asignatura	Interino	Maestría	III. Medicina y Salud	9	9
39	Ingeniería y Arquitectura	30	25	Profesor de Tiempo Completo	Base	Doctorado	VII. Ingeniería e Industria	28	27
40	Ciencias Económico-Administrativas	5	10	Profesor de Tiempo Completo	Base	Maestría	V. Ciencias Sociales y Económicas	10	10
41	Ingeniería y Arquitectura	5	22	Profesor de Tiempo Completo	Interino	Maestría	VII. Ingeniería e Industria	5	5
42	Ciencias y Tecnologías de la Información	16	24	Profesor de Asignatura	Base	Licenciatura	VII. Ingeniería e Industria	16	16
43	Ciencias Sociales y Humanidades	18	20	Profesor de Tiempo Completo	Base	Doctorado	V. Ciencias Sociales y Económicas	19	5
44	Educación y Artes	28	4	Profesor de Tiempo Completo	Base	Doctorado	IV. Humanidades y de la Conducta	18	15
45	Ciencias de la Salud	0	18	Profesor de Medio Tiempo	Interino	Especialidad	III. Medicina y Salud	15	20
46	Ciencias Biológicas	24	15	Profesor de Tiempo Completo	Base	Maestría	VII. Ingeniería e Industria	24	24
47	Educación y Artes	12	12	Profesor de Tiempo Completo	Base	Doctorado	V. Ciencias Sociales y Económicas	17	17
48	Ciencias y Tecnologías de la Información	15	25	Profesor de Tiempo Completo	Base	Doctorado	V. Ciencias Sociales y Económicas	15	6
49	Ciencias de la Salud	22	15	Profesor de Tiempo Completo	Base	Especialidad	IV. Humanidades y de la Conducta	20	8
50	Ingeniería y Arquitectura	8	24	Profesor de Tiempo Completo	Base	Doctorado	I. Física, Matemáticas y Ciencias de la Tierra	40	35
51	Ciencias Sociales y Humanidades	20	20	Profesor de Tiempo Completo	Base	Doctorado	V. Ciencias Sociales y Económicas	20	10
52	Ciencias Sociales y Humanidades	18	25	Profesor de Tiempo Completo	Base	Maestría	IV. Humanidades y de la Conducta	18	18
53	Ciencias Básicas	20	12	Profesor de Tiempo Completo	Base	Doctorado	I. Física, Matemáticas y Ciencias de la Tierra	20	18
54	Ciencias Agropecuarias	22	12	Profesor de Tiempo Completo	Base	Doctorado	VI. Biotecnología y Agropecuarias	16	11
55	Ciencias y Tecnologías de la Información	22	10	Profesor de Tiempo Completo	Base	Doctorado	I. Física, Matemáticas y Ciencias de la Tierra	15	10
56	Ciencias Básicas	25	25	Profesor de Tiempo Completo	Base	Maestría	VII. Ingeniería e Industria	20	18
57	Ciencias Sociales y Humanidades	7	9	Profesor de Asignatura	Interino	Doctorado	V. Ciencias Sociales y Económicas	7	7
58	Ciencias Biológicas	8	8	Profesor de Tiempo Completo	Base	Licenciatura	III. Medicina y Salud	12	12
59	Ciencias de la Salud	20	18	Profesor de Medio Tiempo	Base	Maestría	III. Medicina y Salud	10	8
60	Ciencias Económico-Administrativas	19	14	Profesor de Medio Tiempo	Base	Doctorado	V. Ciencias Sociales y Económicas	20	15
61	Educación y Artes	30	10	Profesor de Tiempo Completo	Base	Doctorado	IV. Humanidades y de la Conducta	30	20
62	Ciencias y Tecnologías de la Información	14	19	Profesor de Asignatura	Base	Maestría	I. Física, Matemáticas y Ciencias de la Tierra	10	10

Nota: García y Silva (2021).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3.1.1 Cuantificación de la población de estudio

La cuantificación de la población de estudio en la minería de datos se refiere al proceso de determinar el tamaño y características de la población en la que se basará el análisis de datos. La cuantificación de la población de estudio también es importante para determinar la muestra necesaria para el análisis de datos. Una muestra demasiado pequeña puede no ser representativa de la población, mientras que una muestra demasiado grande puede ser costosa o innecesaria.

La población de estudio de esta investigación fueron profesores de las 12 Divisiones Académicas de la UJAT, los cuales impartieron clases durante el período 2020-2021, utilizando la plataforma institucional MS Teams, debido a la pandemia COVID-19. Se eligió esta población de estudio para determinar las barreras que desarrollaron los docentes ante el uso de las distintas plataformas implementadas por la entidad educativa, a partir de estas, se determinará de qué manera y en qué áreas, afectó el uso de las innovaciones tecnológicas, del mismo modo, diversos profesores manifestaron su satisfacción ante la situación actual, donde se abandonaron los espacios virtuales y se regresó paulatinamente a la interacción directa con los alumnos.

Así mismo, expusieron su descontento ante el uso de las innovaciones tecnológicas, el cual fue otro motivo por el cual se realizó esta investigación; de esta manera se propondrán estrategias que propicien el uso correcto de estas, ante una nueva eventualidad o pandemia.

 Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3.1.2 Depuración de la población

El formulario digital “Barreras para la adopción de innovaciones tecnológicas”, constó de un total de 826 registros, los cuales correspondieron a la muestra de 826 docentes de las 12 Divisiones Académicas de la UJAT, siendo 386 o bien el 47% mujeres y 440 o 53%, hombres, con edades entre 24 y 80 años; la tabla 1 indica la distribución en porcentajes de profesores por cada División Académica (ver tabla 1).

Tabla 1.

Frecuencia de Divisiones Académicas

No.	División Académica	Porcentaje	Frecuencia
1	División Académica de Ciencias Agropecuarias (DACA)	3%	28
2	División Académica de Ciencias Básicas (DACB)	7%	60
3	División Académica de Ciencias Biológicas (DACBIOL)	11%	90
4	División Académica de Ciencias Económico Administrativas (DACEA)	13%	111
5	División Académica de Ciencias de la Salud (DACS)	13%	110
6	División Académica de Ciencias Sociales y Humanidades (DACSYH)	5%	43
7	División Académica de Ciencias y Tecnologías de la Información (DACYTI)	4%	33
8	División Académica de Educación y Artes (DAEA)	9%	72
9	División Académica de Ingeniería y Arquitectura (DAIA)	6%	46
10	División Académica Multidisciplinaria de Comalcalco (DAMC)	12%	99
11	División Académica Multidisciplinaria de Jalpa de Méndez (DAMJM)	7%	56
12	División Académica Multidisciplinaria de los Ríos (DAMR)	9%	78
Total		100%	826

Nota: Elaboración propia en base a García y Silva (2021).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3.1.3 Descripción del instrumento

El instrumento utilizado para obtener la información fue un cuestionario elaborado por García y Silva (2021) donde se implementó la escala *Likert*, el cual se conformó de 20 ítems, los cuales midieron la opinión de los docentes ante el uso de las innovaciones tecnológicas, estos a su vez, se dividen en tres grupos de dimensiones: sociodemográficas, uso de recursos tecnológicos y recursos disponibles y sistema educativo. Unos de los elementos clave de toda investigación son las variables, para Oyola-García (2021, p. 91), la variable: “Es la descripción precisa de las normas y procedimientos que seguirá el investigador para objetivar las variables en su estudio, como resultado de la información obtenida del conocimiento científico previo, así como de su experiencia personal”.

Las dimensiones en la presente investigación fueron tres, la primera fue del tipo sociodemográfico, donde se encontraron las características del profesor y datos tales como su tipo de contratación y edad, la segunda dimensión fue el uso de recursos tecnológicos, en donde se hizo referencia a las tecnologías implementadas por los profesores de la UJAT en sus labores docentes y la tercera dimensión de recursos disponibles con el sistema educativo, que determinó cuáles fueron las problemáticas enfrentadas por los profesores ante la disponibilidad de recursos de parte de la institución.

Por ello, cada pregunta se respondió utilizando la escala tipo *Likert*, donde las respuestas partían de (1) Totalmente en desacuerdo, (2) En desacuerdo, (3) Ni de acuerdo ni en desacuerdo, (4) De acuerdo y (5) Totalmente de acuerdo.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 2.

Variables de la investigación

Dimensiones	Definición operativa	Ítems que integran la dimensión
Sociodemográficas	Conjunto de datos que describen la naturaleza social de los profesores, en esta investigación se consideraron datos de los profesores como el género, edad, estado civil, Divisiones Académicas, grado de estudios, así como tipos de contratación con la UJAT.	<ul style="list-style-type: none"> • Género • Edad • Estado civil • División académica • Antigüedad en la UJAT (años) • Número de horas frente a grupo a la semana • Categorías de contratación en UJAT • Tipo de contratación en UJAT • Grado académico • Área de formación de su último grado académico
Uso de recursos tecnológicos	Implicación de tecnologías en las actividades de los profesores, desde los años en que comenzaron a utilizar computadoras, apoyar sus labores docentes con recursos tecnológicos, así como redes sociales dentro de sus actividades escolares.	<ol style="list-style-type: none"> 1. ¿Aproximadamente hace cuántos años comenzó a usar computadora para apoyar sus actividades docentes? 2. ¿Aproximadamente hace cuántos años comenzó a usar recursos de Internet para apoyar sus actividades docentes? 3. ¿Aproximadamente hace cuántos años usó por primera vez un Ambiente Virtual de Aprendizaje? 4. ¿Aproximadamente desde hace cuántos años usa las redes sociales para trabajar con sus estudiantes? 5. ¿Hace cuántos años que usa WhatsApp para trabajar con sus estudiantes? 6. ¿Hace cuántos años que usa Facebook para trabajar con sus estudiantes? 7. ¿Hace cuántos años que usa Twitter para trabajar con sus estudiantes? 8. ¿Hace cuántos años que usa el correo electrónico para trabajar con sus estudiantes?
Recursos disponibles y sistema educativo	Hace referencia a las problemáticas enfrentadas con la disponibilidad de recursos tecnológicos en la UJAT, tales como la conectividad, accesibilidad a la red, universitaria, así como la estructura administrativa de la institución para establecer el uso eficiente de tecnologías.	<ol style="list-style-type: none"> 9. Existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT 10. El equipo tecnológico de apoyo a las actividades docentes es escaso 11. La conectividad de la red representa un problema recurrente 12. La accesibilidad a la red en los espacios universitarios es inestable 13. Cuando tengo un problema con algún equipo es difícil conseguir apoyo técnico 14. Que te ayuden a resolver problemas de <i>software</i> en la UJAT resulta complicado 15. Existe una política institucional poco clara que promueva la adopción de tecnología en el aula 16. La estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente 17. La institución carece de un sistema que diagnostique las competencias tecnológicas de su personal 18. La institución carece de un sistema que evalúe las competencias tecnológicas de su personal 19. He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente 20. Siempre será mejor la enseñanza tradicional que la virtual

Nota: Elaboración propia en base a García y Silva (2021).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3.1.4 Aplicación del instrumento

El instrumento se aplicó de forma virtual, a través de un formulario electrónico elaborado en Google Forms, el cual fue enviado a los profesores que impartían asignaturas utilizando la plataforma institucional Microsoft Teams dentro de la UJAT, del que se obtuvo una base de datos en Excel, que fue utilizada como dataset para la aplicación de los algoritmos y técnicas de minería de datos en esta investigación.

3.1.5 Visualización y exploración de los datos

El análisis y visualización de los datos del dataset, se realizó a través de una hoja de cálculo en Excel, donde el primer enfoque se basó en la recolección y exploración de los datos en busca de aquellos que se encontrarán sin respuesta, es decir, nulos o vacíos; esto se llevó a cabo mediante un filtro colocado a los datos, donde en primera instancia se seleccionó al dataset por completo, como se puede observar en la figura 3.

Figura 3.

Selección de los datos para su filtrado

Table with columns A through S containing data for various professors, including names, IDs, genders, birth dates, disciplines, and university categories.

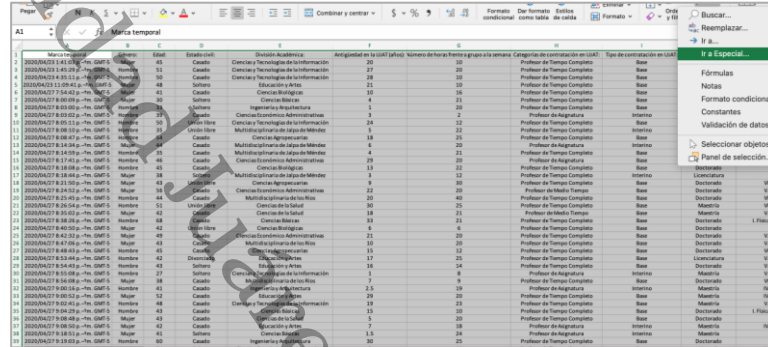
Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Una vez seleccionados todos los datos, se procedió a dar clic en la opción "Buscar y reemplazar", para después seleccionar "Ir a especial", esto se localiza en la ficha Inicio.

Figura 4.

Búsqueda de la opción para datos especiales



Nota: Elaboración propia.

A continuación, se visualizó el cuadro de diálogo "Ir a especial", donde se seleccionó la opción "En blanco" como se aprecia en la figura 5.

Figura 5.

Selección y localización de celdas en blanco



Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Finalmente, la hoja de cálculo no mostró datos nulos, véase la figura 6, donde se observa el dataset con todos los campos llenos.

Figura 6.

Identificación de datos nulos en el dataset

Table with columns: ID, Nombre, Género, Edad, Estado civil, División Académica, Licitación en la UAT (Sexo hombre o mujer), Categoría de contratación en UAT, Tipo de contratación en UAT, Grado académico, Área de formación de su último grado. The table lists 45 individuals with their respective details.

Nota: Elaboración propia.

3.2 Selección de las herramientas para la aplicación de la minería de datos

Las herramientas de minería de datos se utilizan para gestionar los datos e identificar las posibles tendencias y los patrones más significativos. Los programas desarrollados para ello son cada vez más complejos y el abanico de herramientas cada vez mayor, en esta investigación la metodología para analizar y seleccionar las herramientas adecuadas fue el modelo de calidad de software ISO/IEC 9126.

Cabe destacar que se utilizó este modelo debido a que en comparación a la metodología FURPS, éste sí contempla la portabilidad como un factor importante, así mismo, según González (2013) el modelo ISO/IEC 9126 está basado en los modelos de McCall y Boehm.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3.2.1 ISO/IEC 9126

1 La ISO, bajo el modelo ISO/IEC 9126, ha establecido un estándar internacional para la evaluación de la calidad de productos de *software* el cual fue publicado en 1992, afirman Sabpril & Alfin. (2022); este se dio a conocer con el nombre de “*Information technology – Software product evaluation: Quality characteristics and guidelines for their use*”, en el cual se establecen las características de calidad para productos de *software*.

El objetivo principal de la norma ISO/IEC 9126, mencionan Godoy y Avelino (2023), es establecer criterios y métricas para medir diferentes aspectos de la calidad del *software*, incluyendo sus características internas, externas y en uso. Estas características abarcan desde la funcionalidad y confiabilidad del *software* hasta su usabilidad, mantenibilidad y eficiencia.

La norma ISO/IEC 9126 se basa en la idea de que la calidad del *software* no puede ser evaluada únicamente a través de pruebas funcionales, sino que también debe tenerse en cuenta su comportamiento en diferentes contextos de uso y su capacidad para cumplir con las expectativas y necesidades de los usuarios.

9 La ISO/IEC 9126 es aplicable a todo tipo de *software* y está basada en un modelo jerárquico compuesto de tres niveles, los cuales son: características, subcaracterísticas y métricas. En el primer nivel se encuentran seis características principales: funcionalidad, fiabilidad, eficiencia, facilidad de mantenimiento, portabilidad y facilidad de uso. Estos factores se componen por 27 subcaracterísticas relacionadas con la calidad externa, y 21 subcaracterísticas relacionadas con la calidad interna (ver tabla 3).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 3.
Factores y criterios del modelo ISO/IEC 9126

Factores	Criterios
Funcionalidad	<ul style="list-style-type: none"> • Adaptabilidad • Exactitud • Interoperabilidad • Seguridad
Fiabilidad	<ul style="list-style-type: none"> • Madurez • Tolerancia a fallos • Recuperabilidad
Eficiencia	<ul style="list-style-type: none"> • Comportamiento del tiempo • Uso de los recursos
Facilidad de mantenimiento	<ul style="list-style-type: none"> • Análisis • Cambio • Estabilidad • Prueba
Portabilidad	<ul style="list-style-type: none"> • Adaptabilidad • Instalación • Coexistencia • Reemplazo
Facilidad de uso	<ul style="list-style-type: none"> • Comprensibilidad • Aprendizaje • Operatividad • Atractivo

Nota: Elaboración propia.

En base a lo anterior, en el factor de funcionalidad se permite calificar si un *software* maneja adecuadamente las funciones que deberán satisfacer a las necesidades por las cuales fue diseñado. En la fiabilidad se encuentra un conjunto de atributos los cuales se refieren a la capacidad del *software* de mantener su nivel de ejecución bajo condiciones normales en un período de tiempo específico. La eficiencia por su parte permite evaluar la relación entre el nivel de funcionamiento del *software* y la cantidad de recursos usados por este (Isnaini *et al.*, 2021).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

El cuarto grupo denominado facilidad de mantenimiento conforme a lo que describen Sulistiani & Nusaeni (2022) se refiere a los atributos que permiten medir el esfuerzo necesario para realizar modificaciones al *software*, ya sea por la corrección de errores o por el incremento de funcionalidad.

La portabilidad hace referencia a la habilidad que tiene el *software* de ser transferido de un sistema operativo a otro y finalmente, la facilidad de uso consiste en evaluar el esfuerzo que deberá invertir el usuario para utilizar el *software*.

En la actualidad existe una gran cantidad de herramientas de minería de datos, utilizadas para gestionar la información e identificar así, los patrones más importantes o significativos y nuevas tendencias; debido a esto, mediante la ISO/IEC 9126 se clasificaron las herramientas encontradas para *data mining*, en base a los factores y criterios que contiene cada una. fueron ordenadas por pertenecer a los grupos de: *software* libre, *software* propietario, servicios para minería de datos en la nube y lenguajes de programación para minería de datos; por otra parte, al seguir los lineamientos de la norma ISO/IEC 9126, las organizaciones y desarrolladores de *software* pueden mejorar la calidad de sus productos, identificar áreas de mejora y establecer criterios claros para la evaluación y selección de *software* (ver tabla 4).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 4.

Herramientas de minería de datos clasificadas en base a la ISO/IEC 9126

Clasificación	Herramientas de minería de datos
Software libre	<i>Weka</i> <i>Orange</i> <i>RapidMiner</i> <i>Knime</i>
Software propietario	<i>Microsoft Excel</i> <i>SAS Enterprise Miner</i>
Servicios para minería de datos en la nube	<i>IBM SPSS Statistics</i> <i>Oracle Business Intelligence Server</i> <i>Google Cloud Datalab</i> <i>Apache Hadoop</i> <i>Microsoft Azure Machine Learning</i>
Lenguajes de programación para minería de datos	<i>Java</i> <i>Lenguaje R</i> <i>Python</i> <i>Julia</i>

Nota: Elaboración propia.

En la tabla 5 se presenta el análisis de las herramientas de minería de datos, considerando los factores y criterios establecidos en la norma ISO/IEC 9126. Las aplicaciones fueron clasificadas en función de su funcionalidad, fiabilidad, eficiencia, facilidad de mantenimiento, portabilidad y facilidad de uso. Se evaluaron un total de 16 herramientas, que incluyen: *Weka*, *Orange*, *RapidMiner*, *Knime*, *Microsoft Excel*, *Microsoft Azure Machine Learning*, *SAS Enterprise Miner*, *IBM SPSS Statistics*, *Oracle Business Intelligence*, *Google Cloud Datalab*, *Apache Hadoop*, *Java*, *lenguaje R*, *Python* y *Julia* (ver tabla 5), es necesario destacar que el puntaje asignado a las herramientas es conforme al criterio, experiencias y opiniones que tuvieron los investigadores durante el desarrollo de este tema, para cada aplicación de minería de datos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 5.

Selección de herramientas para minería de datos mediante ISO/IEC 9126

Selección de herramientas para minería de datos mediante la ISO/IEC 9126																					
Factores y criterios en base a la ISO/IEC 9126	Funcionalidad			Fiabilidad			Eficiencia		Facilidad de mantenimiento				Portabilidad			Facilidad de uso					
	Adaptabilidad	Exactitud	Interoperabilidad	Seguridad	Madurez	Tolerancia a fallos	Recuperabilidad	Comportamiento del tiempo	Uso de los recursos	Análisis	Cambio	Estabilidad	Prueba	Adaptabilidad	Instalación	Coexistencia	Reemplazo	Comprensibilidad	Aprendizaje	Operatividad	Atractivo
Herramientas para minería de datos																					
<i>Weka</i>																					
<i>Orange</i>																					
<i>RapidMiner</i>																					
<i>Knime</i>																					
<i>Microsoft Excel</i>																					
<i>Microsoft Azure M. L.</i>																					
<i>SAS Enterprise Miner</i>																					
<i>IBM SPSS St.</i>																					
<i>Oracle B.I.</i>																					
<i>Google Cloud Datalab</i>																					
<i>Apache Hadoop</i>																					
<i>Java</i>																					
<i>Lenguaje R</i>																					
<i>Python</i>																					
<i>Julia</i>																					

Nota: Elaboración propia.

En la tabla anterior, se marcó la casilla en color gris, de aquellas herramientas que cumplían con los criterios establecidos en la norma ISO/IEC 9126, lo cual proporciona una referencia clara sobre el grado de cumplimiento de cada herramienta en relación con los diferentes factores de calidad. Después de realizar el análisis de las herramientas mediante la ISO/IEC 9126 por medio del cumplimiento de los criterios de esta, se realizó

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

la comparativa del total de criterios con los que cuenta cada una, para así determinar cuáles obtuvieron el mejor resultado (ver tabla 6).

Tabla 6.

Criterios totales de las herramientas analizadas

Factores en base a la ISO/IEC 9126							
Herramientas	Funcionalidad	Fiabilidad	Eficiencia	Facilidad de mantenimiento	Portabilidad	Facilidad de uso	Total
<i>Weka</i>	3	2	2	3	3	2	15
<i>Orange</i>	3	2	1	3	4	4	17
<i>RapidMiner</i>	4	3	2	4	4	4	21
<i>Knime</i>	3	3	2	2	2	1	13
<i>Microsoft Excel</i>	3	2	1	2	2	4	14
<i>Microsoft Azure M. L.</i>	4	3	1	3	3	3	17
<i>SAS Enterprise Miner</i>	4	3	2	4	4	4	21
<i>IBM SPSS St.</i>	3	2	1	3	2	2	13
<i>Oracle B.I.</i>	3	3	1	2	1	3	13
<i>Google Cloud Datalab</i>	3	2	1	2	3	3	14
<i>Apache Hadoop</i>	2	1	1	3	2	2	11
<i>Java</i>	3	2	1	2	2	2	12
<i>Lenguaje R</i>	3	2	2	2	2	2	13
<i>Python</i>	4	3	2	4	4	4	21
<i>Julia</i>	2	2	1	2	2	1	10

Nota: Elaboración propia.

1 En la tabla anterior se obtuvo que las aplicaciones de minería de datos más completas en cuanto criterios basados en la ISO/IEC 9126, son: *RapidMiner*, *SAS Enterprise Miner* y *Python*, de esta manera se determinó que estas herramientas eran las que se implementarían para el desarrollo de esta investigación.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

10

3.2.2 Características de las herramientas seleccionadas

3.2.2.1 *RapidMiner*

RapidMiner Studio fue desarrollado inicialmente por la Universidad de Dortmund, con sede en Massachusetts, Estados Unidos. *RapidMiner* es una herramienta que posee un potente entorno de programación visual, aplica técnicas de minería de datos crear flujos de trabajo analíticos predictivos, extrae estadísticas e información clave, construye y entrega mejores modelos, según el portal internacional de minería de datos "*Kdnuggets*" (Hernández, 2018), en una encuesta realizada en 2015 *RapidMiner* obtuvo el segundo lugar como herramienta más utilizada.

Esta herramienta ofrece una variedad de algoritmos de preparación de datos y de aprendizaje automático que permite maximizar la productividad de la ciencia de los datos, agilizando la transformación, el desarrollo y la validación de los datos, además permite incluir algoritmos pertenecientes a la herramienta *WEKA* para apoyar a los proyectos en los que se requiera procesar grandes volúmenes de datos, siempre y cuando se considere el costo de una de las licencias de esta herramienta que permita una mayor cantidad de datos a los habituales, esto debido a que la licencia gratuita limita al usuario con solo 10,000 registros permitidos (*RapidMiner, Inc. 2022*).

Las características principales de *RapidMiner* que describen Andry *et al.* (2023) es que está desarrollado en Java y es multiplataforma, actualmente está soportado en *Windows, Linux y MacOs*, cuenta con representación interna de los procesos de análisis de datos en ficheros *XML*, incluidos *SAS, ARFF, Stata* y vía *URL*, además Nigsih & Haryanto (2022), añaden que *RapidMiner* incluye más de 60 tipos de archivos y formatos

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

para datos estructurados o no estructurados, puede usarse de diversas maneras: a través de un *GUI*, en línea de comandos, en *batch* (lotes) o desde otros programas, a través de llamadas a sus bibliotecas. También permite conexiones de bases de datos *JDBC*, incluyendo *Oracle*, *IBM DB2*, *Microsoft SQL Server*, *MySQL*, *Postgres*, entre otras, e incluye gráficos y herramientas de visualización de datos.

Esta investigación se desarrolló con una licencia de tipo educativa, la cual fue totalmente gratis y tenía una durabilidad de 12 meses, sin embargo, se muestran los distintos tipos de licencia que maneja este *software* en 2023 (ver tabla 7).

Tabla 7.

Tipos de licencia de RapidMiner

	Educativo	Gratis	Pequeña	Mediana	Grande
Número de filas de datos	Ilimitado	10,000	100,000	1,000,000	Ilimitado
Caducidad de la licencia	Después de 12 meses	Nunca	Después de 12 meses	Después de 12 meses	Después de 12 meses
Atención al cliente	Comunidad	Comunidad	Empresa	Empresa	Empresa
Cargo anual de licencia	Gratis	Gratis	\$2,500 MXN anuales	\$5,000 MXN anuales	\$10,000 MXN anuales

Nota: Elaboración propia basado en el sitio oficial de *RapidMiner* (2023).

3.2.2.2 SAS Enterprise Miner

SAS Enterprise Miner es una herramienta de minería de datos desarrollada por SAS Institute. Es ampliamente utilizada en la industria y en el ámbito académico para descubrir patrones, realizar análisis predictivos y construir modelos estadísticos (Lim, 2023).

Stanley *et al.* (2023) mencionan que *SAS Enterprise Miner* proporciona una interfaz visual intuitiva que permite a los usuarios diseñar y desarrollar flujos de trabajo

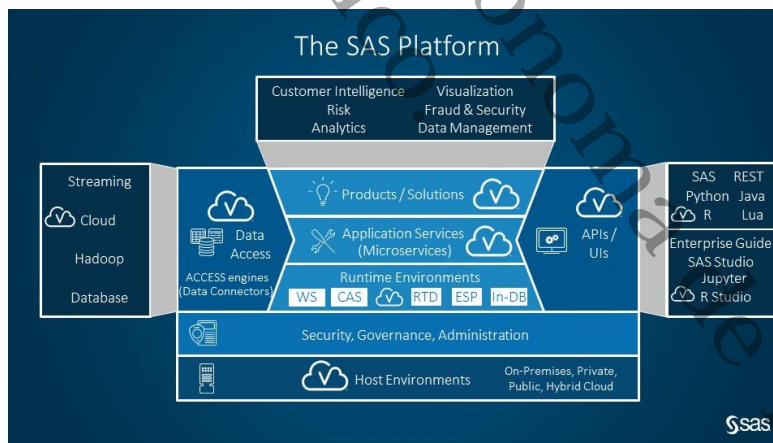
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

de minería de datos de manera eficiente. Utilizando técnicas avanzadas de análisis estadístico y aprendizaje automático, la herramienta permite explorar grandes conjuntos de datos, identificar relaciones y tendencias ocultas, y generar modelos predictivos y descriptivos.

Una de las fortalezas de *SAS Enterprise Miner* según Vijay *et al.* (2014) radica en su capacidad para trabajar con datos estructurados y no estructurados, lo que le permite aprovechar fuentes de información diversas, como bases de datos, archivos planos, documentos de texto, entre otros. La figura 7 presenta un diagrama referente a la arquitectura general de SAS®, es necesario destacar que *SAS Enterprise Miner* forma parte de este (ver figura 7).

Figura 7.

Arquitectura general de SAS Enterprise Miner



Nota: SAS® (2023).

Otra característica destacada de *SAS Enterprise Miner* es su enfoque en la interpretación y explicación de los resultados obtenidos. La herramienta proporciona

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

funcionalidades para evaluar la calidad de los modelos generados, realizar pruebas de validez y analizar las contribuciones de las variables predictoras en el modelo. Esto ayuda a los usuarios a comprender y comunicar de manera efectiva los hallazgos obtenidos a partir de los datos.

Además, *SAS Enterprise Miner* cuenta con capacidades de automatización y programación avanzada, lo que permite a los usuarios más experimentados personalizar y optimizar sus flujos de trabajo, así como integrar la herramienta con otros sistemas y procesos de análisis de datos.

SAS Enterprise Miner presenta algunas desventajas a considerar, mencionan Gunandi *et al.* (2023): en primer lugar, su licencia puede resultar costosa, lo que limita su accesibilidad para usuarios con presupuestos ajustados. Además, requiere tiempo y esfuerzo para dominar todas sus funcionalidades, especialmente para aquellos que son nuevos en minería de datos. Por otra parte, Gümüs *et al.* (2023) mencionan que también puede requerir una infraestructura de hardware adecuada y tiene una dependencia de otros componentes del *software* de SAS. Algunos usuarios pueden encontrar limitaciones en la personalización y adaptación a necesidades específicas. A pesar de estas desventajas, es importante evaluarlas en función de las necesidades y recursos de cada usuario.

3.2.2.3 Python

Python es un lenguaje de programación de alto nivel que se utiliza para desarrollar aplicaciones de todo tipo. A diferencia de otros lenguajes como *Java* o *.NET*, se trata de un lenguaje interpretado, es decir, que no es necesario compilarlo para ejecutar las

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos aplicaciones escritas en *Python*, sino que se ejecutan directamente por el ordenador utilizando un programa denominado interpretador, por lo que no es necesario “traducirlo” a lenguaje máquina (*Python*, 2023).

Según Santander (2022), *Python* es un lenguaje fácil de leer y escribir porque es muy similar al lenguaje del hombre. Además, es de código abierto, multiplataforma y por lo tanto gratuito, lo cual permite el desarrollo de *software* sin limitaciones.

Cuenta con una amplia variedad de bibliotecas y herramientas especializadas en análisis de datos, como *Numpy*, *Pandas*, *Matplotlib* y *Scikit-learn*, que permiten realizar tareas de minería de datos de manera eficiente. *Python* es conocido por su versatilidad, ya que se puede utilizar en diversos entornos y sistemas operativos.

La figura 8 presenta un diagrama referente a los algoritmos de minería de datos que ofrece *Python* actualmente (ver figura 8).

Figura 8.

Algoritmos para minería de datos de Python

The image is a 'Python For Data Science Cheat Sheet' from DataCamp. It is organized into several columns and rows, each with a title and a list of code snippets. The sections include:

- Scikit-learn**: Introduction to the library and a basic example of training a linear regression model.
- Loading The Data**: Code for loading data from CSV files using pandas.
- Training And Test Data**: Code for splitting data into training and testing sets.
- Preprocessing The Data**: Sub-sections for Standardization, Normalization, and Binarization.
- Create Your Model**: Sub-sections for Supervised Learning Estimators (Linear Regression, SVM, Naive Bayes, KNN), Unsupervised Learning Estimators (PCA, K-Means), Model Fitting, and Prediction.
- Evaluate Your Model's Performance**: Sub-sections for Classification Metrics (Accuracy Score, Confusion Matrix, Regression Metrics), Clustering Metrics (Adjusted Rand Index, Homogeneity, V-measure), Cross-Validation, and Grid Search.
- Encoding Categorical Features**: Code for LabelEncoder and OneHotEncoder.
- Imputing Missing Values**: Code for SimpleImputer.
- Generating Polynomial Features**: Code for PolynomialFeatures.

Nota: DataCamp (2022).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Como indican Aguilera *et al.* (2023), *Python* un lenguaje de código abierto y cuenta con una gran comunidad de desarrolladores que contribuyen con la creación de nuevas bibliotecas y la resolución de problemas. Sin embargo, a pesar de todas sus ventajas, *Python* puede ser más lento en comparación con otros lenguajes de programación, especialmente en tareas que requieren un alto rendimiento computacional.

Maturana *et al.* (2022) mencionan que también puede presentar desafíos en cuanto a la gestión de la memoria y el rendimiento en grandes conjuntos de datos. A pesar de estas limitaciones, *Python* sigue siendo una opción popular y poderosa para realizar tareas de minería de datos y análisis de datos.

3.3 Proceso KDD

El término de “descubrimiento de conocimiento en bases de datos”, (por sus siglas en inglés, KDD) fue acuñado en 1996 por Fayyad, Piatetsky-Shapiro y Smyth, mencionan Smedberg & Bandaru (2023), para denotar el amplio proceso de descubrir conocimiento en los datos y enfatizar las aplicaciones "vanguardistas" de técnicas específicas de minería de datos. KDD es esencialmente un proceso automatizado que combina descubrimiento y análisis.

El proceso consiste en extraer patrones de los datos en forma de reglas o funciones para que el usuario pueda analizarlos. Esta tarea generalmente implica el procesamiento previo de datos, la extracción de datos y la presentación de resultados (Shu & Ye, 2023).

KDD se puede utilizar en varios campos, por ejemplo, para determinar el perfil de clientes fraudulentos (evasión fiscal), para revelar relaciones indirectas que existen entre

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

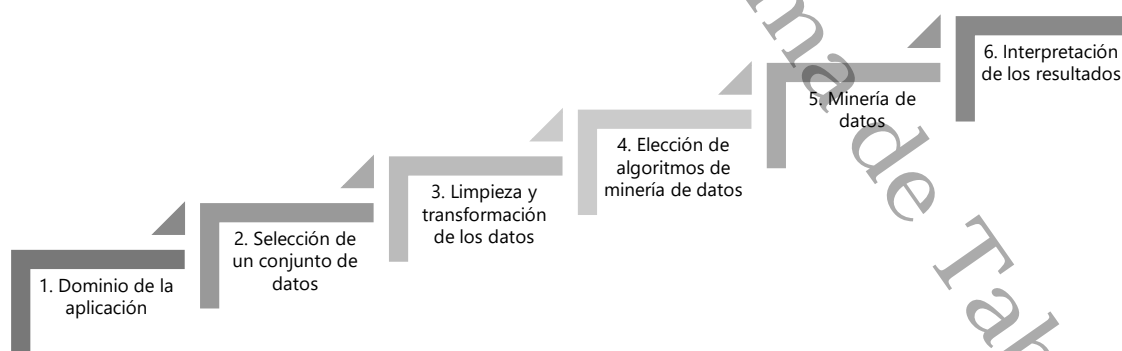
síntomas y enfermedades, entre parámetros técnicos, así como diagnósticos de condiciones de equipos y máquinas.

La minería de datos está directamente relacionada con el proceso KDD, que se refiere al proceso de descubrir conocimiento no trivial e información potencialmente útil en los datos de un almacén de datos, un proceso iterativo que explora exhaustivamente una gran cantidad de datos para determinar las relaciones entre ellos (Joyanes, 2019).

El proceso KDD es interactivo e iterativo, es por ello por lo que los analistas de datos necesitan un amplio conocimiento profesional ya que, en cada etapa de ejecución, los resultados de este proceso serán un aporte importante para la toma de decisiones; se construyó una nueva metodología basada en Yang *et al.* (2023, p.55), que los autores diseñaron a partir del proceso KDD original, para esta investigación se modificó y finalmente constó de seis fases, indicadas a continuación (véase la figura 9).

Figura 9.

Metodología creada para la investigación



Nota: Elaboración propia basada en Yang *et al.* (2023, p.55).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

El proceso KDD es utilizado para extraer conocimiento automáticamente de grandes cantidades de datos, por lo que su naturaleza es iterativa, es decir, que puede usarse tantas veces como sea necesario hasta obtener la información requerida o descubrir verdadero conocimiento nuevo.

Por todo lo anterior descrito se decidió utilizar esta metodología ya que es una las primeras en haber sido utilizadas para el descubrimiento de conocimiento y la cual contempla a la minería de datos, que será aplicada a través de las herramientas: *RapidMiner*, *SAS Enterprise Miner* y *Python*.

3.3.1 Dominio de la aplicación

La primera etapa de esta metodología consistió en adquirir conocimiento previo antes de utilizar las herramientas digitales para aplicar la minería de datos, para lograr el uso óptimo y eficiente de éstas, así como comprender los objetivos de cada aplicación conforme a los algoritmos que contienen.

En este paso, primero se instalaron las tres herramientas de minería de datos seleccionadas: *RapidMiner*, *SAS Enterprise Miner* y *Python* para así conocer sus interfaces, herramientas y elementos que las conforman. *RapidMiner* fue la primera herramienta utilizada, la cual contiene un entorno de trabajo integrado o *IDE*, que facilitará el manejo de datos, codificación y visualización a través de los algoritmos de minería de datos.

3.3.1.1 *RapidMiner*

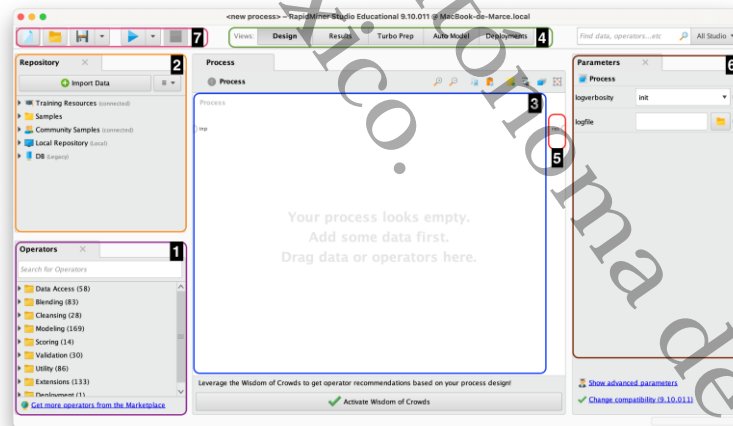
La figura 10 muestra la pantalla principal de *RapidMiner*, el cual emplea una interfaz gráfica de tipo arrastrar y soltar para diseñar procesos de análisis. La interfaz de

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

esta aplicación se compone de siete elementos principales, los cuales son: 1) Operadores, estos son bloques de construcción utilizados para crear procesos; 2) Depósitos, están destinados para conformar el almacenamiento de *RapidMiner*; 3) Panel de proceso, es el área de trabajo para procesos en desarrollo; 4) Vistas, son las áreas de trabajo para acceder a funciones específicas de visualización; 5) Puertos, mecanismos de entrada así como de salida para operadores y procesos; 6) Parámetros, ajustes que modifican el comportamiento de cada operador y 7) Herramientas, en esta sección se guardan los cambios, se exportan *datasets*, así como otros procesos básicos de *RapidMiner* (ver figura 10).

Figura 10.

Interfaz inicial de *RapidMiner*



Nota: Elaboración propia.

3.3.1.2 SAS Enterprise Miner

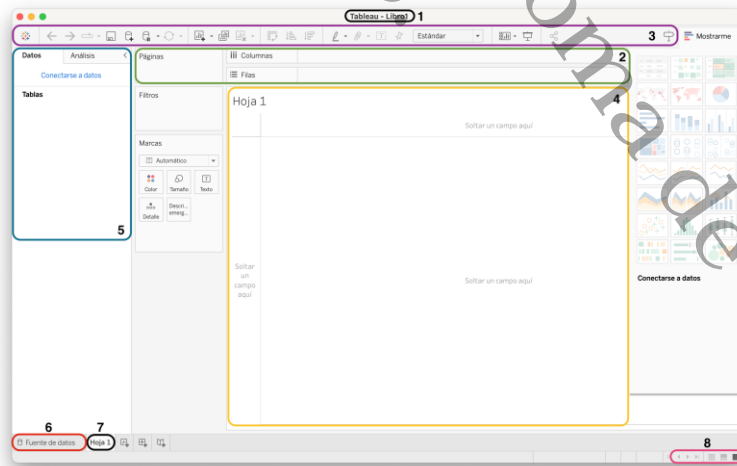
La segunda herramienta que se instaló para desarrollar esta investigación fue SAS *Enterprise Miner*, en la figura 11 se observa la interfaz de esta aplicación, donde se

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

muestran las opciones de: 1) Nombre de libro de trabajo, el cual contiene el nombre predeterminado de “Libro 1”, hasta que el usuario desee modificarlo; 2) Tarjetas y estantes, es posible arrastrar los campos a las tarjetas y los estantes del espacio de trabajo para añadir datos a la vista; 3) Barra de herramientas, esta se utiliza para acceder a comandos y herramientas de análisis y navegación; 4) Vista, este es el lienzo para el área de trabajo, es decir, donde se observa el mismo; 5) Barra lateral, en una hoja de trabajo, esta área contiene el panel de datos y análisis que se pueden implementar en el proyecto; 6) Fuente de datos, al hacer clic en esta pestaña se accede a la fuente de datos y así poder visualizarlos; 7) Pestaña de hojas, las pestañas representan las hojas del libro de trabajo, pueden ser hojas de trabajo, *dashboards* o historias; y finalmente 8) Barra de estado, donde se muestra la información sobre la vista actual (ver figura 11).

Figura 11.

Interfaz inicial de SAS Enterprise Miner



Nota: Elaboración propia.

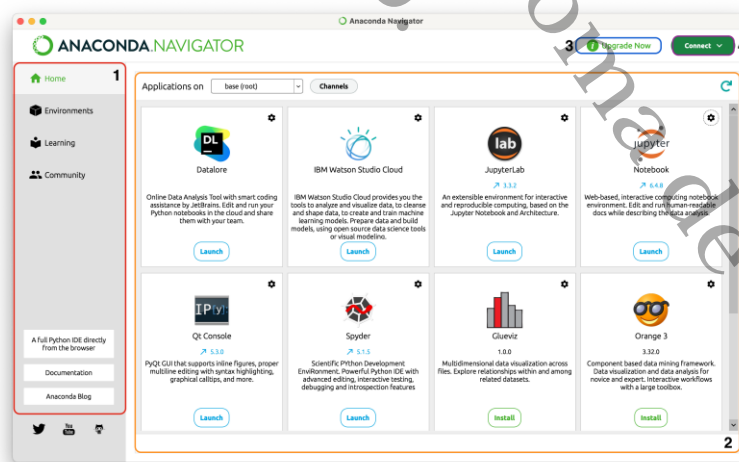
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3.3.1.3 Python

La última herramienta utilizada fue *Python*, el cual se trabajó a través de la plataforma *Anaconda Navigator*, la cual es una distribución libre y abierta, utilizada en ciencia de datos y aprendizaje automático, en la figura 12 se observa la interfaz de esta distribución, donde se encuentra el complemento llamado *Jupyter*, donde se implementará la minería de datos: 1) Menú, donde se encuentran distintas opciones para navegar en la aplicación; 2) Aplicaciones, en esta sección se muestran los complementos que se pueden utilizar a través de *Anaconda Navigator*; 3) Mejorar, este es un botón que muestra los planes que se pueden adquirir monetariamente, con mejoras en *Anaconda* y finalmente 4) Conectar, donde se puede conectar y desconectar a Internet el entorno (ver figura 12).

Figura 12.

Interfaz inicial de Anaconda Navigator



Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3.3.2 Selección de un conjunto de datos

Generalmente, a lo largo de una investigación se recopilan conjuntos de datos de diverso tipo, que pueden llegar a tener grandes dimensiones, a pesar de que los costos de almacenamiento disminuyen cada día, existen motivos que hacen necesario decidir qué datos preservar (más allá de los límites de la investigación) y qué datos eliminar.

El segundo paso de la metodología de esta investigación es la selección de un conjunto de datos, esta fase corresponde a la selección de la fuente de datos a utilizar, después se crea el conjunto de estos a partir del usuario final, es donde se comienza a seleccionar el conjunto de datos o un subconjunto de diversas tablas, variables o muestras de datos, los cuales son los que se utilizan para realizar el análisis de los datos.

El *dataset* a implementar en esta investigación, fue proporcionado por García y Silva (2021) en el que se diseñaron y validaron originalmente 20 *ítems*, con un total de 826 registros. De igual forma, contiene las variables de género, edad, estado civil, División Académica, antigüedad, número de horas frente a grupo a la semana, categorías de contratación en UJAT, tipo de contratación en UJAT, grado académico, área de formación de su último grado académico y los 20 *ítems* que conforman el instrumento aplicado.

Se verificó la calidad de los datos, se realizaron análisis de integridad y consistencia, y se aplicaron técnicas de limpieza y preprocesamiento para asegurar la fiabilidad de los datos utilizados en la investigación. En la tabla 8 se muestran las variables enumeradas, las cuales conforman el conjunto de datos de destino y estas a su vez fueron utilizadas para la minería de datos, cabe destacar que el conjunto de datos contiene datos numéricos y no numéricos (ver tabla 8).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 8.

Conjunto de datos del dataset

Número	Variables
1	Género
2	Edad
3	Estado civil
4	División académica
5	Antigüedad en la UJAT (años)
6	Número de horas frente a grupo a la semana
7	Categorías de contratación en UJAT
8	Tipo de contratación en UJAT
9	Grado académico
10	Área de formación de su último grado académico
11	¿Aproximadamente hace cuántos años comenzó a usar computadora para apoyar sus actividades docentes?
12	¿Aproximadamente hace cuántos años comenzó a usar recursos de Internet para apoyar sus actividades docentes?
13	¿Aproximadamente hace cuántos años usó por primera vez un Ambiente Virtual de Aprendizaje?
14	¿Hace cuántos años que usa WhatsApp para trabajar con sus estudiantes?
15	¿Hace cuántos años que usa Facebook para trabajar con sus estudiantes?
16	¿Hace cuántos años que usa Twitter para trabajar con sus estudiantes?
17	¿Hace cuántos años que usa el correo electrónico para trabajar con sus estudiantes?
18	¿Qué otras redes sociales o recursos emplean para trabajar con sus estudiantes?
19	Existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT
20	El equipo tecnológico de apoyo a las actividades docentes es escaso
21	La conectividad de la red representa un problema recurrente
22	La accesibilidad a la red en los espacios universitarios es inestable
23	Cuando tengo un problema con algún equipo es difícil conseguir apoyo técnico
24	Que te ayuden a resolver problemas de <i>software</i> en la UJAT resulta complicado
25	Existe una política institucional poco clara que promueva la adopción de tecnología en el aula
26	La estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente
27	La institución carece de un sistema que diagnostique las competencias tecnológicas de su personal
28	La institución carece de un sistema que evalúe las competencias tecnológicas de su personal
29	He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente
30	Siempre será mejor la enseñanza tradicional que la virtual

Nota: Elaboración propia.

3.3.3 Limpieza y transformación de los datos

En esta fase se analizó la calidad de los datos disponibles para el análisis, seguido del proceso de preparación y limpieza de datos. Se trató de preprocesamiento de datos, transformación y generación de nuevas funcionalidades a partir de funcionalidades existentes según Shaheen & Khan (2023). Las técnicas de reducción o transformación de dimensionalidad se utilizaron para reducir el número efectivo de variables bajo consideración o para encontrar una representación invariable de los datos.

Se analizó cada uno de los datos y variables del *dataset*, para posteriormente proceder a la limpieza de estos y prepararlos para el preprocesamiento previo debido a que fueron ingresados en las herramientas. Durante la etapa de limpieza de los datos, no se detectaron datos nulos, tampoco inconsistencias en ellos.

Para la implementación del *dataset* en las herramientas de minería de datos no se requirió modificar el tipo de archivo, es decir, este se mantuvo siempre en formato *Excel* o *XLSX*, debido a que las aplicaciones seleccionadas contienen lectura de archivos en ambos tipos. A continuación, se realizó el preprocesamiento de los datos en las tres herramientas, para verificar que la lectura y visualización del *dataset* fuera óptima.

En la figura 13 se observa la herramienta *RapidMiner*, donde se obtuvo el *dataset* después de su lectura con un total de 826 registros y 0 errores, tanto de tipos de datos como de campos (ver figura 13).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 13.

Lectura del dataset en RapidMiner

Género	Edad	Estado civil	División Académica	Antigüedad en la UJAT (años)	Número de horas frente a grupo a la semana	Categorías de contratos
Mujer	45	Casado	Ciencias y Tecnologías de ...	20	10	Profesor de Tiempo Com
Hombre	51	Casado	Ciencias y Tecnologías de ...	27	20	Profesor de Tiempo Com
Hombre	50	Casado	Ciencias y Tecnologías de ...	28	10	Profesor de Tiempo Com
Mujer	48	Soltero	Educación y Artes	21	10	Profesor de Tiempo Com
Mujer	41	Casado	Ciencias Biológicas	10	16	Profesor de Tiempo Com
Mujer	30	Soltero	Ciencias Básicas	4	21	Profesor de Tiempo Com
Hombre	33	Soltero	Ingeniería y Arquitectura	1	20	Profesor de Tiempo Com
Hombre	50	Unión libre	Ciencias y Tecnologías de ...	24	12	Profesor de Tiempo Com
Hombre	35	Unión libre	Multidisciplinaria de Jalpa ...	5	22	Profesor de Tiempo Com
Hombre	54	Casado	Ciencias Agropecuarias	18	25	Profesor de Tiempo Com
Hombre	35	Casado	Multidisciplinaria de Jalpa ...	4	21	Profesor de Tiempo Com
Hombre	45	Casado	Ciencias Biológicas	13	22	Profesor de Tiempo Com
Mujer	38	Soltero	Multidisciplinaria de Jalpa ...	3	12	Profesor de Tiempo Com
Mujer	43	Unión libre	Ciencias Agropecuarias	9	30	Profesor de Tiempo Com
Hombre	44	Casado	Multidisciplinaria de los Ríos	20	40	Profesor de Tiempo Com
Hombre	51	Unión libre	Ciencias de la Salud	10	25	Profesor de Tiempo Com

Nota: Elaboración propia.

La segunda lectura se realizó en la aplicación *SAS Enterprise Miner*, donde se cargó el archivo en su formato original *XLSX*, y se visualizó como se aprecia en la figura 14 (ver figura 14), obteniendo 826 registros y 0 errores.

Figura 14.

Lectura del dataset en SAS Enterprise Miner

Nombre	Género	Edad	Estado civil	División Académica	Antigüedad en la UJAT (años)	Número de horas frente a grupo a la semana
Profesor-Barreras para la adop...	Mujer	45	Casado	Ciencias y Tecnologías de la U...	20	10
Profesor-Barreras para la adop...	Hombre	51	Casado	Ciencias y Tecnologías de la U...	27	20
Profesor-Barreras para la adop...	Hombre	50	Casado	Ciencias y Tecnologías de la U...	28	10
Profesor-Barreras para la adop...	Mujer	48	Soltero	Educación y Artes	21	10
Profesor-Barreras para la adop...	Mujer	41	Casado	Ciencias Biológicas	10	16
Profesor-Barreras para la adop...	Mujer	30	Soltero	Ciencias Básicas	4	21
Profesor-Barreras para la adop...	Hombre	33	Soltero	Ingeniería y Arquitectura	1	20
Profesor-Barreras para la adop...	Hombre	50	Unión libre	Ciencias y Tecnologías de la U...	24	12
Profesor-Barreras para la adop...	Hombre	35	Unión libre	Multidisciplinaria de Jalpa de ...	5	22
Profesor-Barreras para la adop...	Hombre	54	Casado	Ciencias Agropecuarias	18	25
Profesor-Barreras para la adop...	Hombre	35	Casado	Multidisciplinaria de Jalpa de ...	4	21
Profesor-Barreras para la adop...	Hombre	45	Casado	Ciencias Biológicas	13	22
Profesor-Barreras para la adop...	Mujer	38	Soltero	Multidisciplinaria de Jalpa de ...	3	12
Profesor-Barreras para la adop...	Mujer	43	Unión libre	Ciencias Agropecuarias	9	30
Profesor-Barreras para la adop...	Hombre	44	Casado	Multidisciplinaria de los Ríos	20	40
Profesor-Barreras para la adop...	Hombre	51	Unión libre	Ciencias de la Salud	10	25

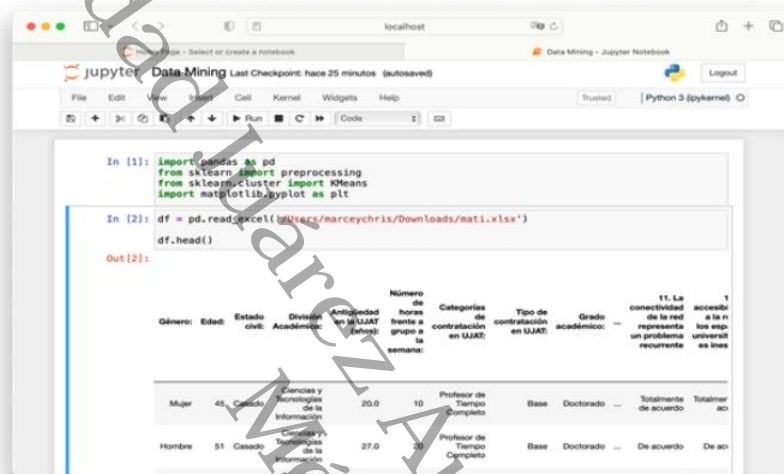
Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

La tercera herramienta fue *Python*, de la misma manera se obtuvieron 0 errores y 826 registros leídos correctamente (ver figura 15).

Figura 15.

Lectura del dataset en Python



```
In [1]: import pandas as pd
from sklearn import preprocessing
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

In [2]: df = pd.read_excel('https://marceyChris/Downloads/mat1.xlsx')
df.head()
```

Out [2]:

Género:	Edad:	Estado civil:	División Académica:	Antigüedad en la UAMT (años):	Número de horas trabajadas por la semana:	Categorías de contratación en UAMT:	Tipo de contratación en UAMT:	Grado académico:	11. La conectividad de la red representa un problema recurrente en las	1
Mujer	45	Casada	Ciencias y Tecnologías de la Información	20.0	10	Profesor de Tiempo Completo	Base	Doctorado	De acuerdo	Totalmente de acuerdo
Hombre	51	Casado	Tecnologías de la Información	27.0	20	Profesor de Tiempo Completo	Base	Doctorado	De acuerdo	De acuerdo

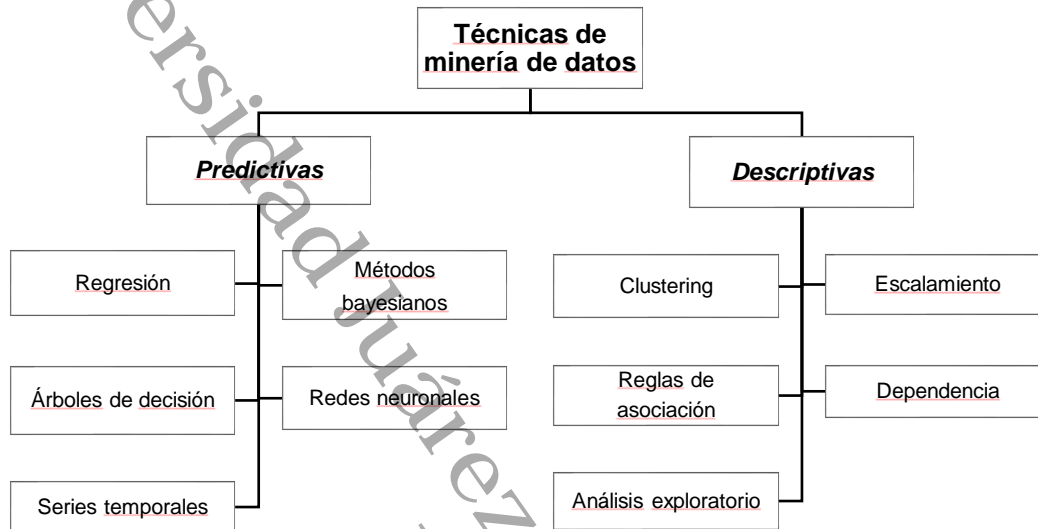
Nota: Elaboración propia.

3.3.4 Elección de algoritmos de minería de datos

La elección de los mejores algoritmos para una tarea analítica específica puede ser un desafío. Aunque se pueden usar diferentes algoritmos para realizar la misma tarea, cada uno de ellos genera un resultado diferente, y algunos pueden generar más de un tipo de resultado. En este paso se seleccionaron y describieron los algoritmos a implementar en la minería de datos de esta investigación, obsérvese en la figura 16 la clasificación de las técnicas que existen para el minado de datos (ver figura 16).

Figura 16.

Clasificación de técnicas de minería de datos



Nota: Elaboración propia.

En este paso del método se analizan diversas técnicas y funciones de minería de datos o conocimiento, pero previamente se seleccionan las técnicas de minería de datos a utilizar, que son: árboles de decisión, agrupamiento y reglas de asociación.

La razón de elegir árboles de decisión es que permiten la visualización de modelos jerárquicos donde cada elemento representa una o más clases extendidas por nodos. Por otro lado, la técnica de *clustering* consiste en realizar análisis de conglomerados y agruparlos en diferentes categorías, clases o grupos dependiendo de la clase de que se trate. Finalmente, las reglas de asociación definen reglas de clasificación basadas en los elementos y variables utilizadas. Por el contrario, una sola regla de asociación puede predecir los valores de múltiples atributos o variables. Por estas razones, y porque son

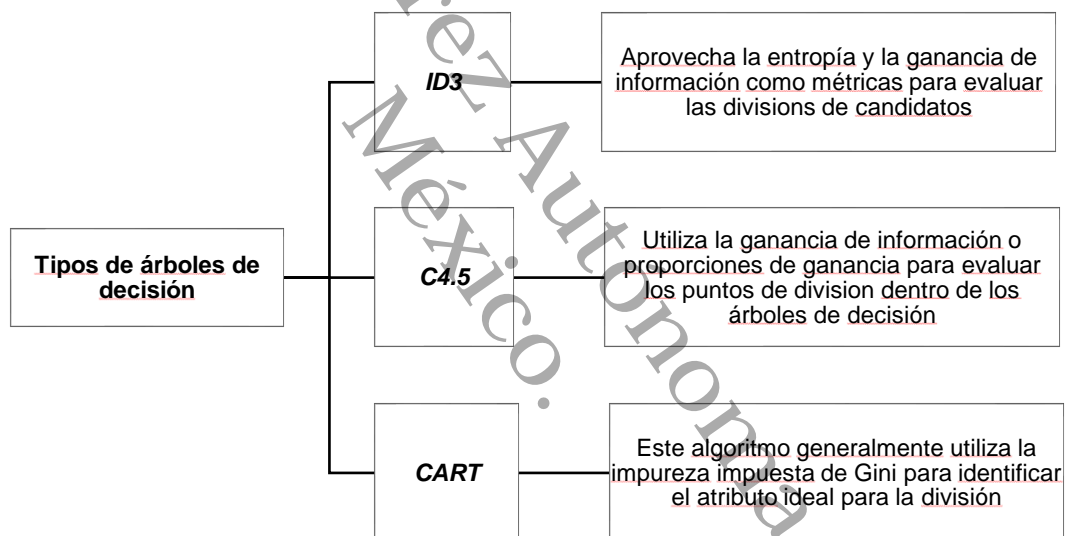
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

las técnicas más comunes y populares para analizar datos y descubrir conocimiento a partir de datos, se eligieron estas técnicas.

Los árboles de decisión consideran diferentes algoritmos de generación, estos incluyen el algoritmo C4.5, el cual es el algoritmo utilizado en esta investigación. En la figura 17 se observan los tipos de algoritmos de árbol de decisión que existen (ver figura 17).

Figura 17.

Tipos de algoritmos de árbol de decisión

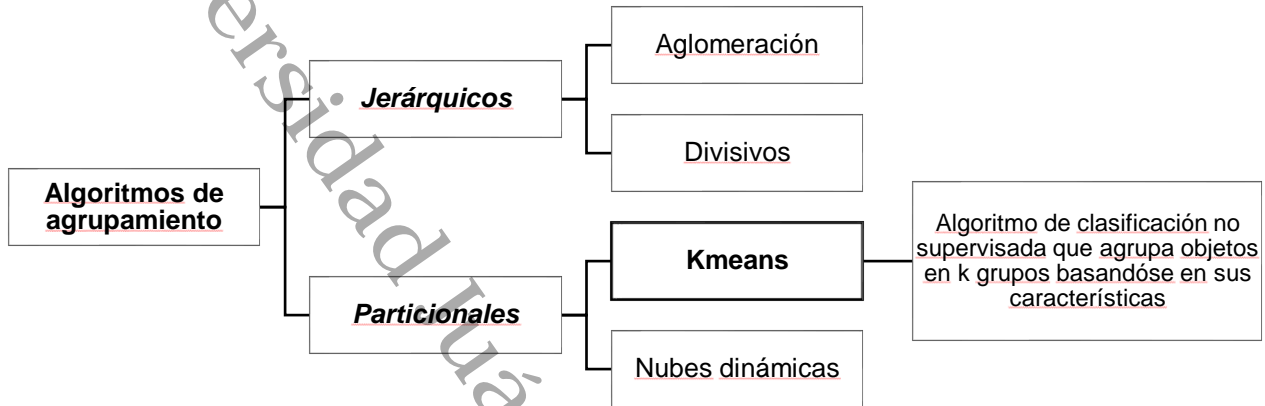


Nota: Elaboración propia con base en Amarda *et al.* (2022).

Entre los métodos de agrupamiento previamente seleccionados, se pueden utilizar diferentes algoritmos, entre los cuales se eligió el algoritmo *Kmeans*, la figura 18 muestra los tipos de algoritmos de agrupamiento y los detalles de este (ver figura 18).

Figura 18.

Tipos de algoritmo de agrupamiento

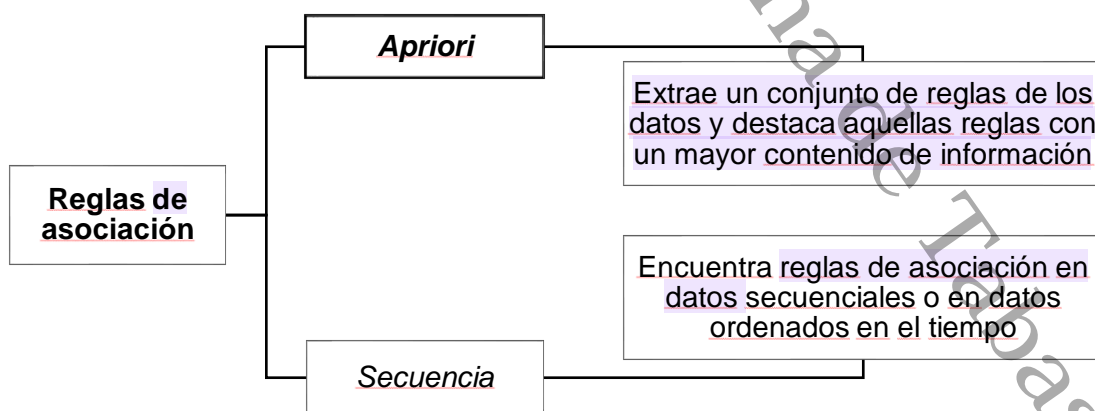


Nota: Elaboración propia con base en Haji et al. (2023).

El tercer algoritmo elegido fue el algoritmo *apriori*, que pertenece a las reglas de asociación. La figura 19 muestra los tipos de algoritmos para esta técnica y detalles sobre el algoritmo *apriori* (ver figura 19).

Figura 19.

Tipos de algoritmo de reglas de asociación



Nota: Elaboración propia con base en Zhang et al. (2023).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3.3.5 Minería de datos

En esta fase se utilizaron modelos para descubrir patrones y comportamientos en datos previamente desconocidos, donde se podían utilizar modelos de clasificación, regresión, *clustering*, asociación, entre otros.

Este paso comprendió el análisis de los datos a través de las herramientas de minería de datos, donde se implementó el *dataset* previamente preprocesado, para continuar con los algoritmos seleccionados dentro de cada aplicación; a continuación, se divide en tres partes este punto, debido a que se explica el proceso de minería de datos llevado a cabo en cada herramienta conforme a los tres algoritmos seleccionados.

RapidMiner

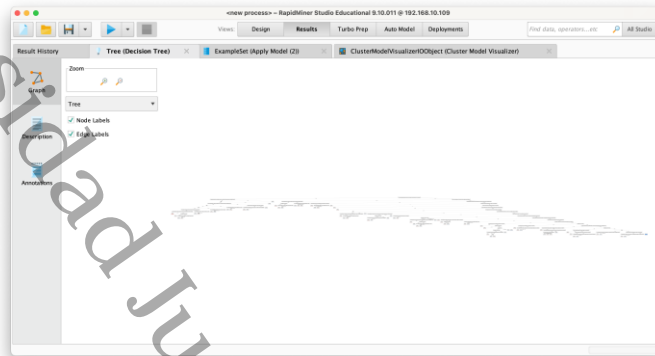
La primera herramienta de minería de datos utilizada fue *RapidMiner*, la cual contiene una interfaz gráfica dinámica e intuitiva. Una vez realizada la lectura del *dataset* y comprobar los tipos de datos de cada variable, se procedió a implementar el primer algoritmo de minería de datos, los árboles de decisión; donde se arrastraron los operadores de lectura de archivo *Excel* y árbol de decisión. Cada operador puede requerir algunas entradas y entrega algunas salidas. Estos tipos de entrada y salida se pasan entre los operadores.

Después de normalizar se comenzó la búsqueda de los grupos, por ello se utilizó el operador de *clustering* y se arrastró en la hoja de trabajo, utilizando un criterio de conveniencia, es decir, 2. Finalmente se añadió el operador de *decision tree*, el cual desarrolla gráficamente el árbol de decisión que se requiere, como se muestra en la figura 20, considérese que el árbol obtenido es grande (ver figura 20).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 20.

Árbol de decisión obtenido en RapidMiner

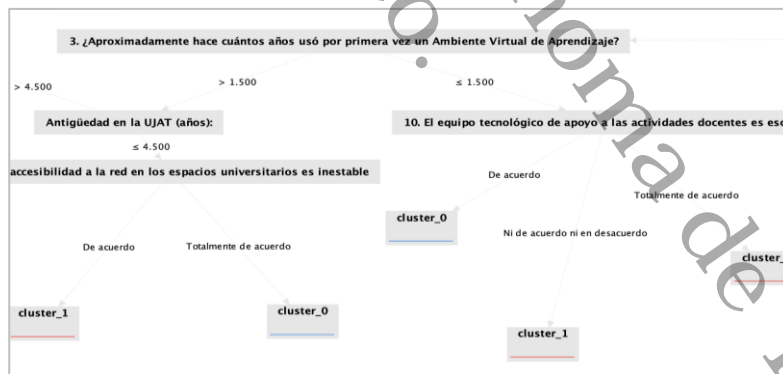


Nota: Elaboración propia.

Para observar con más detalle, en la figura 21 se muestra parte del árbol de decisión (ver figura 21).

Figura 21.

Árbol de decisión a detalle obtenido en RapidMiner



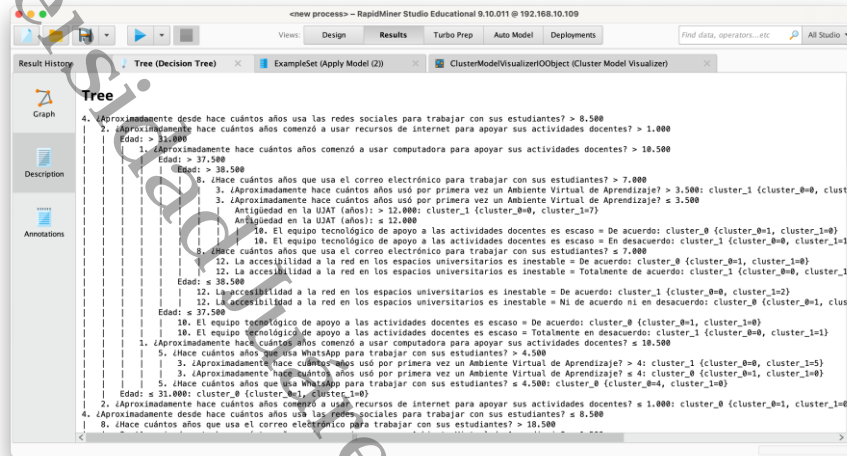
1 Nota: Elaboración propia.

Por otra parte, también fue posible visualizar la descripción del árbol de decisiones en RapidMiner, como se muestra en la figura 22 (ver figura 22).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 22.

Descripción del árbol de decisión obtenido en RapidMiner



Nota: Elaboración propia.

El segundo algoritmo por implementar fue *K-Means* o *clustering*. El algoritmo *K-means* es un algoritmo de agrupamiento utilizado para dividir un conjunto de puntos en *K* grupos (donde *K* es un número específico dado) de manera que los puntos en un grupo sean lo más similares entre sí posible. El algoritmo según Ikotun *et al.* (2023) se basa en la idea de elegir *K* puntos iniciales aleatorios en el conjunto de datos como centroides de los grupos, y luego asignar cada punto restante al grupo cuyo centroide está más cerca, finalmente se calcula un nuevo centroide para cada grupo y se vuelven a asignar los puntos, y este proceso se repite hasta que no se producen más cambios en las asignaciones de grupo.

De la misma manera que en árboles de decisión, se arrastraron los distintos operadores a la hoja de trabajo de *RapidMiner*: *retrieve* que es un operador para leer datos; *normalize* que es un operador para normalizar el *dataset*; *clustering* el cual es un

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

operador para estimar el modelo de agrupamiento y un último operador llamado *performance*, que sirve para evaluar el desempeño del modelo de agrupamiento.

K-means es un algoritmo basado en distancias, lo que se buscó fue que las distintas variables compitieran en los mismos términos, es decir, que no fueran dominados por las medidas más grandes, por ello se utilizó el operador de *normalize*. Por consiguiente, se implementó el método de *Z-transformation*, el cual consiste en restar la media y dividir entre la desviación estándar, como se muestra en la figura 23 (ver figura 23).

Figura 23.

Transformación Z en RapidMiner

Row No.	Edad:	Antigüeda...	Número de...	1. ¿Aproxi...	2. ¿Aproxi...	3. ¿Aproxi...	4. ¿Aproxi...	5. ¿Hace cu...	6. ¿Hace cu...	7. ¿Hace cu...
1	-0.067	0.559	-1.239	0.679	0.451	-0.266	0.038	0.400	-0.665	-0.340
2	0.517	1.233	0.137	1.390	0.451	3.341	-0.651	-0.497	-0.665	-0.340
3	0.420	1.330	-1.239	1.628	1.557	2.439	2.333	0.400	-0.665	-0.340
4	0.225	0.655	-1.239	-0.517	-0.931	-1.167	0.038	0.400	-0.665	-0.340
5	-0.456	-0.405	-0.414	-0.517	-0.240	-0.807	-0.192	-1.094	0.474	-0.340
6	-1.527	-0.984	0.274	-1.232	-1.069	-1.167	-1.110	-0.198	-0.665	-0.340
7	-1.235	-1.273	0.137	0.079	-1.269	-0.626	-1.110	-1.094	-0.665	-0.340
8	-0.651	-1.080	-2.340	1.271	-1.207	0.636	-0.651	-0.657	-0.665	-0.340
9	0.420	0.944	-0.964	-0.517	-1.484	-1.167	-1.110	-1.094	-0.665	-0.340
10	-1.040	-0.888	0.412	-1.113	-0.931	-0.626	-0.192	0.101	0.474	-0.340
11	0.809	0.366	0.825	1.866	2.525	1.538	0.038	0.101	-0.665	-0.340
12	-0.164	-0.791	0.137	0.675	1.143	-0.266	0.038	0.101	-0.665	-0.340
13	-1.040	-0.984	0.274	0.079	0.451	0.095	-1.110	-1.094	-0.665	-0.340
14	0.030	1.426	0.137	2.105	2.386	2.439	0.726	1.297	1.612	-0.340
15	-0.067	-0.116	0.412	-0.278	0.037	-0.987	-0.651	-0.795	-0.665	-0.340
16	-0.748	-1.080	-0.964	-1.113	-0.654	-1.167	-0.421	-0.198	-0.665	-0.340
17	-0.262	-0.502	1.512	-0.636	-0.378	-0.626	-0.651	-0.795	-0.665	-0.340

Nota: Elaboración propia.

1

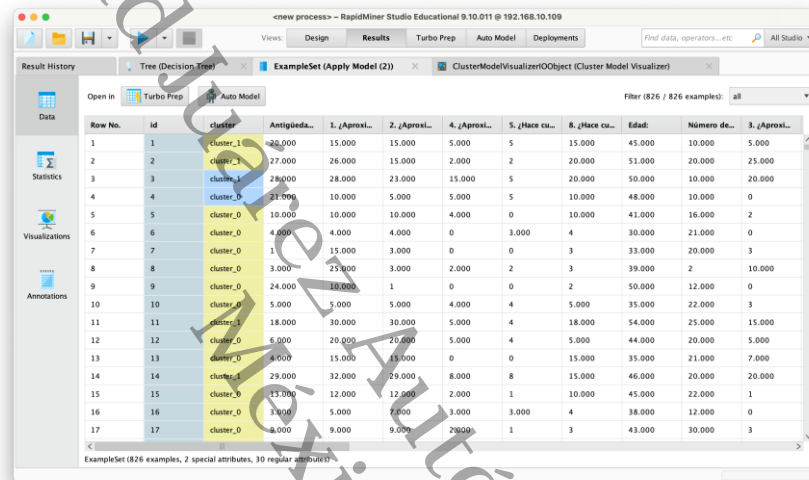
Después de normalizar se comenzó la búsqueda de los grupos, por ello se utilizó el operador de *clustering* y se arrastró en la hoja de trabajo, modificando a $k = 2$, utilizando un criterio de conveniencia. Una vez complementado el algoritmo *K-means* de manera

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

gráfica en *RapidMiner*, en la figura 24 se observa cómo se obtuvo el *dataset* con una nueva columna denominada *clúster*, la cual indica a que grupo pertenece cada uno de los 826 registros (ver figura 24).

Figura 24.

Resultado de K-Means en RapidMiner



Row No.	Id	cluster	Antigüeda...	1. ¿Aproxi...	2. ¿Aproxi...	4. ¿Aproxi...	5. ¿Hace cu...	8. ¿Hace cu...	Edad:	Número de...	3. ¿Aproxi...
1	1	cluster_1	28.000	15.000	15.000	5.000	5	15.000	45.000	10.000	5.000
2	2	cluster_1	27.000	26.000	15.000	2.000	2	20.000	51.000	20.000	25.000
3	3	cluster_1	28.000	28.000	23.000	15.000	5	20.000	50.000	10.000	20.000
4	4	cluster_0	21.000	10.000	5.000	5.000	5	10.000	48.000	10.000	0
5	5	cluster_0	10.000	10.000	10.000	4.000	0	10.000	41.000	16.000	2
6	6	cluster_0	4.000	4.000	4.000	0	3.000	4	30.000	21.000	0
7	7	cluster_0	15.000	3.000	3.000	0	0	3	33.000	20.000	0
8	8	cluster_0	3.000	25.000	3.000	2.000	2	3	39.000	2	10.000
9	9	cluster_0	24.000	10.000	1	0	0	2	50.000	12.000	0
10	10	cluster_0	5.000	5.000	5.000	4.000	4	5.000	35.000	22.000	3
11	11	cluster_0	18.000	30.000	30.000	5.000	4	18.000	54.000	25.000	15.000
12	12	cluster_0	6.000	20.000	20.000	5.000	4	5.000	44.000	20.000	5.000
13	13	cluster_0	4.000	15.000	15.000	0	0	15.000	35.000	21.000	7.000
14	14	cluster_1	29.000	32.000	29.000	8.000	8	15.000	46.000	20.000	20.000
15	15	cluster_0	13.000	12.000	12.000	2.000	1	10.000	45.000	22.000	1
16	16	cluster_0	3.000	5.000	7.000	3.000	3.000	4	38.000	12.000	0
17	17	cluster_0	4.000	9.000	9.000	2.000	1	3	43.000	30.000	3

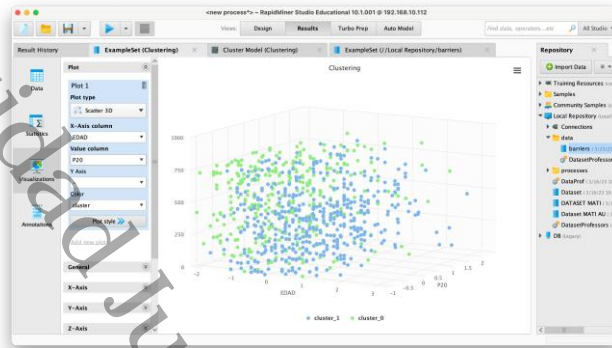
Nota: Elaboración propia.

1 *RapidMiner* obtuvo que el *clúster* 0 se conformó de 541 datos y el *clúster* 1 contenía 285 datos, de esta manera, el gráfico de *Kmeans* o de puntos, demostró la correlación de ambos *clústeres*, como se observa en la figura 25, esto es para visualizar la revisión de agrupamiento de los datos de dos variables en función de los *clústeres* creados con la técnica de agrupamiento de minería de datos, esta revisión se puede realizar con cada una de las variables del *dataset* (ver figura 25).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 25.

Gráfico de K-Means en RapidMiner



Nota: Elaboración propia.

De la misma manera, *RapidMiner* es capaz de comparar dos variables distintas de todo el *dataset* haciendo posible su visibilidad en la pestaña de gráficos como se muestra en la figura 25, obteniendo así de una manera más dinámica los resultados y descubriendo nuevas correlaciones no contempladas anteriormente.

El tercer algoritmo implementado fueron las reglas de asociación, estas se definen como un conjunto de técnicas que permiten establecer relaciones de interés con la finalidad de descubrir hechos que aporten valor dentro de las variables que facilitan los datos que son enormes, al igual que en los algoritmos anteriores, estos operadores deben añadirse al lector del archivo, seleccionando los atributos a analizarse.

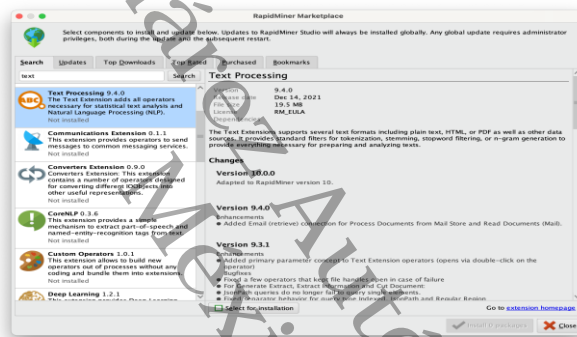
Sin embargo, *RapidMiner* es capaz de desarrollar reglas de asociación a partir de texto, por lo cual, se requirió de la extensión *Text Processing*, la cual se descargó de la pestaña *Extensions* (ver figura 26).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

La extensión de Text Processing en RapidMiner se encarga de brindar herramientas y técnicas específicas para el procesamiento y análisis de datos textuales. Permite realizar tareas como la limpieza y preprocesamiento de textos, la extracción de características relevantes, la minería de opiniones y sentimientos, la clasificación de textos, el análisis de tópicos y la generación de modelos de lenguaje.

Figura 26.

Extensión Text Processing de RapidMiner



Nota: Elaboración propia.

En la figura 27 se observan las reglas de asociación que RapidMiner encontró en el dataset preprocesado (ver figura 27).

Figura 27.

Reglas de asociación en RapidMiner

Rule No.	Item	support	confidence	apriori	equilibrium	lift	significance	best	biology
1	major ciencia ciencias tecnologías de la información prof...	1	0	0	0	0	0	1	0
2	hombre casual ciencias tecnologías de la información prof...	1	0	0	0	0	0	1	0
3	hombre casual ciencias tecnologías de la información prof...	1	0	0	0	0	0	1	0
4	major admi ciencias económicas prof...	1	0	0	0	1	0	1	0
5	major admi ciencias biología profesor de tiempo com...	1	0	1	0	0	0	1	0
6	major admi ciencias biología profesor de tiempo complet...	1	0	0	0	0	0	1	0
7	hombre casual ciencias económicas profesor de tiempo...	1	0	0	1	0	0	1	0
8	hombre casual ciencias económicas administrativas profes...	1	1	0	0	0	1	1	0
9	hombre casual ciencias tecnologías de la información...	1	1	0	0	0	1	1	0
10	hombre casual con multilingüismo de países de habla...	1	0	0	0	0	0	0	0
11	hombre casual ciencias agronomías profesor de tiempo...	1	0	1	0	0	0	1	0
12	major casual multilingüismo de países de habla prof...	1	0	1	0	0	0	1	0
13	hombre casual multilingüismo de países de habla prof...	1	0	0	0	0	0	1	0
14	hombre casual ciencias económicas administrativas profes...	1	1	0	0	0	1	1	0
15	hombre casual ciencias biología profesor de tiempo...	1	0	0	0	0	0	1	1

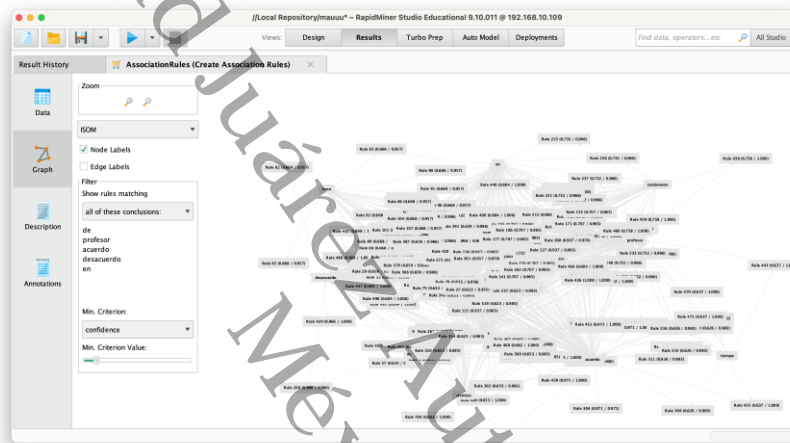
Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Al colocar el soporte mínimo en 0.95 con el operador *FP-Growth*, *RapidMiner* automáticamente lanza el gráfico que se observa en la figura 28, con un total de 947 reglas de asociación (ver figura 28).

Figura 28.

Gráfico de reglas de asociación en *RapidMiner*



1 Nota: Elaboración propia.

SAS Enterprise Miner

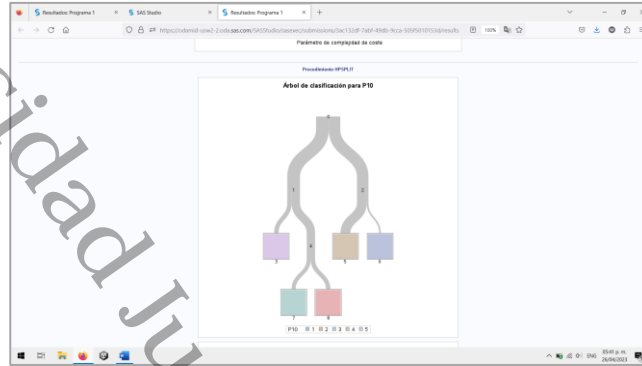
La segunda herramienta de minería de datos utilizada fue *SAS Enterprise Miner*, la cual contiene una interfaz gráfica sencilla y muy ágil. Una vez realizada la lectura del *dataset* y comprobar los tipos de datos de cada variable, se procedió a implementar los árboles de decisión.

Se “conectó” al *dataset* en formato *Excel*, una vez cargado el archivo, se mostró en la interfaz de *SAS Enterprise Miner*, una vez allí se realizó el árbol de decisión añadiendo los parámetros para la comparación de los datos, el agrupamiento y la ramificación (ver figura 29).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 29.

Árbol de decisión obtenido en SAS Enterprise Miner

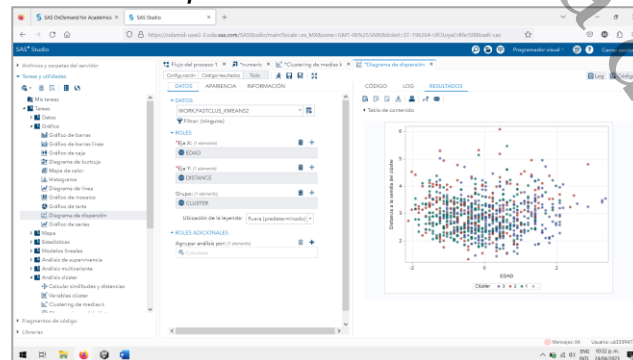


Nota: Elaboración propia.

Para el segundo algoritmo, denominado *Kmeans*, se realizaron pasos similares a *RapidMiner*, sin embargo, para encontrar la función de *clústeres*, se navegó a través de la interfaz de *SAS Enterprise Miner* en la pestaña Análisis, donde en automático el *software* solicita el número de *clústeres*, que para esta investigación es 2 y proporciona el gráfico de círculos que se muestra en la figura 30, con ambos *clústeres* (ver figura 30).

Figura 30.

Gráfico de K-Means en SAS Enterprise Miner



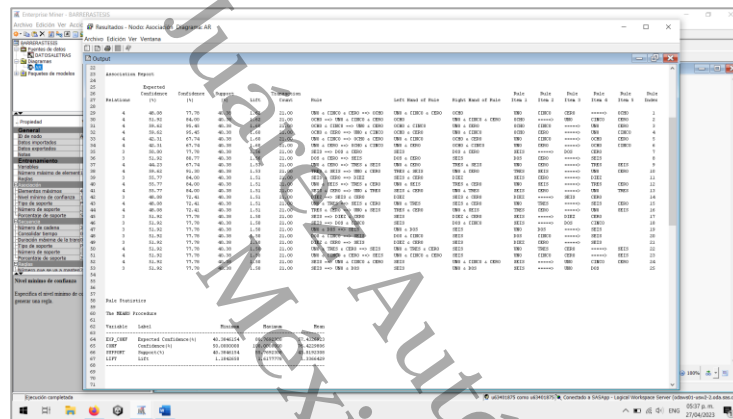
Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

El tercer algoritmo implementado fue la regla de asociación, de la cual se trabajó con el dataset sin realizar modificaciones, debido a que SAS Enterprise Miner acepta valores numéricos y texto, se obtuvieron distintos valores debido a la cantidad de ítems del dataset, como se aprecia en la figura 31 (ver figura 31).

Figura 31.

Reglas de asociación en SAS Enterprise Miner



Nota: Elaboración propia.

Python

La tercera herramienta de minería de datos a utilizar fue Python, el cual se maneja por código a través de diversos softwares que apoyan en la fácil visibilidad de su interfaz, por ello, como anteriormente se explicó, Python fue utilizado por medio de Anaconda Navigator.

El primer algoritmo implementado fueron los árboles de decisión, que son representaciones gráficas de posibles soluciones a una decisión basadas en ciertas condiciones, es uno de los algoritmos de aprendizaje supervisado más utilizados en

1

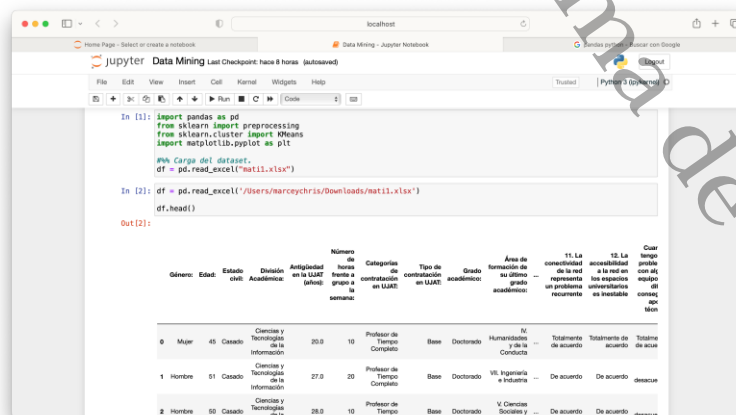
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Machine Learning y pueden realizar tareas de clasificación o regresión. En la etapa de dominio de la aplicación se explicó que *Python* cuenta con diversas paqueterías que hacen posible la implementación de los algoritmos para minería de datos. *Anaconda Navigator* que es el entorno gráfico utilizado, las incluye por defecto.

1 Una vez abierta la libreta de *Python*, se requirió de la importación del *dataset* en la herramienta, la cual permite la lectura de archivos en *Excel*, *CSV*, *TXT*, entre otros. Se comenzó con la lectura del *dataset*, a través de la librería denominada *pandas*, como se puede observar en la figura 32, la cual es una biblioteca de *software* para la manipulación y el análisis de datos. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series de tiempo; una vez cargada la librería, la libreta de *Python* carga parte del *dataset*, para que usuario pueda visualizar que los datos se han cargado de manera correcta y sin errores (ver figura 32).

Figura 32.

Lectura del *dataset* en libreta de *Python*



```
In [1]: import pandas as pd
from sklearn import preprocessing
from sklearn.cluster import KMeans
import matplotlib.pyplot as plt

#Carga del dataset:
df = pd.read_excel('mat11.xlsx')

In [2]: df = pd.read_excel('Users/narceyChris/Downloads/mat11.xlsx')
df.head()

Out[2]:
```

	Nombre de horas en la UAMT (semana)	Categoría de contratación en UAMT	Tipo de contrato en UAMT	Grado académico	Área de formación de su último grado académico	11. La conectividad de la red representa un problema recurrente en su institución	12. La accesibilidad a la red en los espacios universitarios es adecuada	Cuor tiempo promedio con el equipo de conexión de la red
0	Mujer 45 Casado Ciencias y Tecnología de la Información 20.0 10 Profesor de Tiempo Completo Base Doctorado NI Humanidades y de la Conducta Totalmente De acuerdo Totalmente De acuerdo Totalmente De acuerdo							
1	Hombre 51 Casado Ciencias y Tecnología de la Información 27.0 20 Profesor de Tiempo Completo Base Doctorado VII Ingeniería e Informática De acuerdo De acuerdo De acuerdo							
2	Hombre 50 Casado Ciencias y Tecnología de la Información 28.0 10 Profesor de Tiempo Completo Base Doctorado V Ciencias Sociales y de la Conducta De acuerdo De acuerdo De acuerdo							

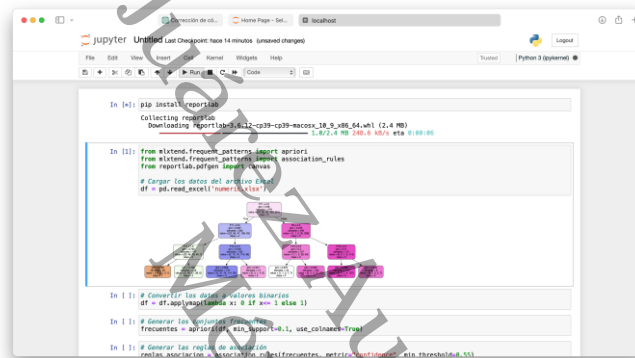
Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Una vez cargado el archivo, se codificó en la interfaz de *Python* a través de las librerías de *pandas* para crear cada uno de los árboles de decisión, como el que se muestra en la figura 33, el cual fue el primer árbol de decisión resultante o de prueba para esta investigación (ver figura 33).

Figura 33.

Árbol de decisión obtenido en Python



```
In [1]: pip install reportlab
Collecting reportlab
  Downloading reportlab-3.6.12-cp39-cp39-macosx_10_9_universal2.whl (2.4 MB)
  1.872.4 MB 248.6 kB/s eta 9:00:00

In [2]: from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
from reportlab.pdfgen import canvas
# Cargar los datos del archivo Excel
df = pd.read_excel('mineria_datos1.xlsx')

In [3]: # Convertir los datos a columnas binarias
df = df.apply(lambda x: x if x == 1 else 0)

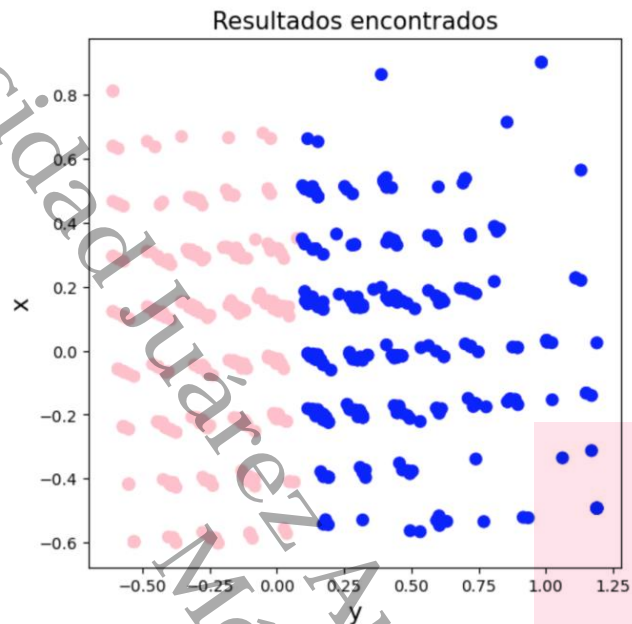
In [4]: # Generar las reglas de asociación
frecuentes = apriori(df, support=0.1, use_colnames=True)

In [5]: # Generar las reglas de asociación
rules = association_rules(frecuentes, metric='lift', min_lift=0.5)
```

Nota: Elaboración propia.

1 En el segundo algoritmo utilizado, conocido como *Kmeans*, se siguieron pasos similares a los realizados en *RapidMiner* y *SAS Enterprise Miner*. Sin embargo, para determinar la función de *clústeres*, se implementó un código en *Python*. En este caso, se estableció de forma predeterminada un valor de 2 para el tamaño de los *clústeres* a crear. El resultado de este proceso se visualizó en un gráfico que se muestra en la figura 34, donde se pueden observar los dos *clústeres* generados.

Esta técnica de agrupación permitió identificar patrones y segmentar los datos en grupos homogéneos, lo cual facilita un análisis más detallado y comprensión de los resultados obtenidos (ver figura 34).

Figura 34.Gráfico de *K-Means* en *Python*

Nota: Elaboración propia.

El tercer algoritmo utilizado en la investigación fue el de reglas de asociación, el cual se aplicó directamente al conjunto de datos sin realizar modificaciones previas. Dado que *Python* tiene la capacidad de manejar tanto valores numéricos como textuales, se obtuvieron resultados diversos teniendo en cuenta la amplia variedad de *ítems* presentes en el *dataset*. Los resultados de las asociaciones encontradas entre los diferentes elementos del conjunto de datos se pueden apreciar en la figura 35 (ver figura 35).

Esta técnica permitió identificar patrones de co-ocurrencia entre los elementos, lo cual brinda información valiosa sobre las relaciones existentes en los datos analizados.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 35.

Reglas de asociación en Python

```
In [5]: # Generar reglas de asociación con confianza del 55%
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.55)

In [6]: print(rules)
```

	antecedents \		
0	(CINCO)		
1	(CINCO)		
2	(CINCO)		
3	(CINCO)		
4	(CINCO)		
...			
9168	(POUNCE De acuerdo, PDIECISIETE De acuerdo, B...		
9169	(POUNCE De acuerdo, CERO, PDIECIOCHO De acuerdo)		
9170	(POUNCE De acuerdo, Profesor de Tiempo Comple...		
9171	(POUNCE De acuerdo, Profesor de Tiempo Comple...		
9172	(POUNCE De acuerdo, Base, PDIECIOCHO De acuerdo)		
	consequents	antecedent support \	
0	(DIEZ)	0.445755	
1	(Base)	0.445755	
2	(CERO)	0.445755	
3	(Casado)	0.445755	
4	(Hombre)	0.445755	
...			
9168	(Profesor de Tiempo Completo, CERO, PDIECIOCHO...	0.176887	
9169	(PDIECISIETE De acuerdo, Base, Profesor de Tie...	0.180425	
9170	(PDIECISIETE De acuerdo, PDIECIOCHO De acuerdo...	0.180425	
9171	(PDIECISIETE De acuerdo, CERO, Base)	0.143868	
9172	(PDIECISIETE De acuerdo, CERO, Profesor de Tie...	0.182783	

Nota: Elaboración propia.

Los modelos presentados en este estudio requirieron ajustes y modificaciones en el *dataset* para lograr una visualización correcta y comprensible de los resultados. Estas modificaciones incluyeron la limpieza de datos, la selección de variables relevantes, la normalización de datos y la aplicación de técnicas de procesamiento de datos. Estos ajustes fueron necesarios para garantizar la precisión y la interpretación adecuada de los modelos utilizados en el análisis de los datos.

3.3.6 Interpretación de los resultados

En esta fase, la interpretación o lectura de los patrones que fueron descubiertos en la fase anterior se utilizarán como entrada para las operaciones posteriores o bien, pueden retroalimentarse a la fase anterior para iteraciones posteriores, siendo esta la última fase del proceso original KDD. Según Báez *et al.* (2022), el propósito de este paso es interpretar el conocimiento descubierto a partir de los datos para mejorar el

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

rendimiento del sistema de rendimiento del proceso y para resolver posibles conflictos con el conocimiento previo de los datos.

Con base en lo anterior, se registraron las pruebas realizadas a los modelos desarrollados y creados, así como los resultados y sus interpretaciones, con el fin de comprender e identificar los patrones de comportamiento de los profesores de educación superior de la UJAT ante las innovaciones tecnológicas.

Descubrir dicho conocimiento e implementar la minería de datos puede llevar a los expertos o partes interesadas en el campo de la educación a tomar decisiones creando o cambiando las medidas y regulaciones existentes para que el nivel educativo sea más alto en un futuro, así como las posibles precauciones que se tendrían contempladas ante una nueva situación mundial.

Capítulo 4. Resultados

En este capítulo se presentan los resultados de las pruebas realizadas con el conjunto de datos para la minería de datos utilizando diferentes técnicas y algoritmos con tres herramientas. El capítulo está estructurado en cinco secciones, la primera de ellas describe las pruebas realizadas en las tres herramientas de minería de datos: *RapidMiner*, *SAS Enterprise Miner* y *Python*. El segundo apartado presenta los resultados obtenidos en estas herramientas. La tercera sección explica la interpretación de los resultados obtenidos, mientras que en el cuarto apartado se describen los patrones descubiertos a través de la minería de datos. Por último, el quinto apartado incluye un análisis de la usabilidad de las herramientas previamente seleccionadas.

4.1 Pruebas de las técnicas de minería de datos

En esta sección, se describen detalladamente las pruebas realizadas para la construcción de modelos de árboles de decisión, agrupamiento y reglas de asociación. Cada algoritmo se presenta en el mismo orden y se presentan tres pruebas para cada uno de ellos. Estas pruebas fueron diseñadas para evaluar la efectividad y el rendimiento de los modelos de minería de datos desarrollados. Cada prueba se llevó a cabo utilizando diferentes parámetros y configuraciones, con el objetivo de obtener resultados comparativos y comprender mejor el comportamiento de los algoritmos en el contexto específico de los datos analizados. A continuación, se presentan los detalles y los análisis de cada prueba realizada.

4.1.1 *RapidMiner*

Se comienza presentando la herramienta *RapidMiner*, para la cual se llevaron a cabo diversas pruebas de usabilidad con el objetivo de familiarizarse con su funcionamiento y de utilizar las técnicas de clasificación, agrupamiento y reglas de asociación mediante distintos algoritmos y visualizadores gráficos.

Se realizaron pruebas con diferentes algoritmos para los modelos de árboles de decisión, agrupamiento y reglas de asociación. Para los árboles de decisión, se evaluaron las funciones de *Decision Tree*, *Random Forest* y *Random Tree*, cada una con su correspondiente representación visual ya incluidas, destacando por su fácil manejo e implementación la función de *Decision Tree*. En cuanto al modelo de agrupamiento, se probaron los algoritmos *K-Means*, *K-Means(fast)* y *X-Means*, todos con su respectiva visualización.

Sin embargo, conforme a la experiencia de uso se determinó que el algoritmo *Kmeans* era el óptimo para las pruebas de las técnicas de minería de datos, debido a que es uno de los algoritmos más utilizados para el análisis de datos y es muy eficiente en términos de tiempo y recursos computacionales. Para las reglas de asociación, se probaron el algoritmo *create association rules*, el elemento *FP-Growth* y el algoritmo *Apriori*, que encuentra conjuntos de elementos frecuentes para calcular reglas de asociación; determinando que el algoritmo *A priori* era el indicado para esta investigación, debido a que es muy útil para identificar patrones y relaciones en grandes conjuntos de datos, de manera óptima y rápida.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Árboles de decisión

Se llevaron a cabo diversas pruebas en la herramienta *RapidMiner* para seleccionar la pregunta principal del árbol de decisión. Se priorizaron los *ítems* con mayor ganancia de información. En términos simples, la ganancia de información es una métrica que se utiliza en minería de datos y aprendizaje automático para seleccionar las características o variables más importantes para un modelo de predicción.

Esta métrica se calcula midiendo la reducción de la entropía (medida de la incertidumbre) en los datos al dividirlos según una característica en particular, lo que significa que se busca aquella característica que nos proporciona la mayor reducción de incertidumbre en los datos. Por lo tanto, las preguntas raíz, que son aquellas preguntas esenciales para comprender un problema o fenómeno en particular, pueden ser utilizadas para dividir los datos según una característica en particular y, de esta manera, calcular la ganancia de información para seleccionar las características más importantes para un modelo de predicción.

Conforme a los conceptos anteriores, se seleccionaron tres *ítems* a manera de preguntas raíz para las pruebas a las que se sometió el *dataset*, siendo estas las que obtuvieron mejores resultados en las pruebas de entrenamiento y evaluación:

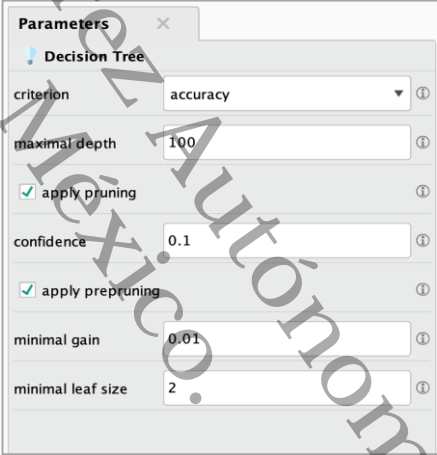
- 10. El equipo tecnológico de apoyo a las actividades docentes es escaso (P10),
- 12. La accesibilidad a la red en los espacios universitarios es inestable (P12),
- 19. He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente (P19).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

La selección de preguntas raíz se consideró como una tarea crítica en la investigación y el análisis de las barreras ante el uso de innovaciones tecnológicas en los profesores de educación superior. Para seleccionar las preguntas raíz, como se observa en la figura 36, fue importante considerar diversos factores, como el contexto del problema, los objetivos de la investigación y el análisis, así como las limitaciones de los datos disponibles (ver figura 36).

Figura 36.

Parámetros de inicio del árbol de decisión



Parameter	Value
criterion	accuracy
maximal depth	100
apply pruning	<input checked="" type="checkbox"/>
confidence	0.1
apply prepruning	<input checked="" type="checkbox"/>
minimal gain	0.01
minimal leaf size	2

Nota: Elaboración propia.

Las preguntas raíz seleccionadas se consideraron relevantes y específicas contemplando las dimensiones de las que está conformada la investigación. Después de la selección de las preguntas raíz se realizaron configuraciones iniciales para generar el árbol de decisión de la primera prueba. Se dejó por defecto el criterio establecido en frecuencias para seleccionar los atributos para la división, lo que maximiza la precisión de todo el árbol. En este caso la configuración que quedó por defecto fue el criterio

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

(*criterion*) establecido en frecuencias (*accuracy*), que seleccionó los elementos según los cuales se escogieron los atributos para la división, en este caso la frecuencia, la cual seleccionó un atributo para dividirlo, lo que maximizó la precisión de todo el árbol.

Por otra parte, para la profundidad del árbol (*maximal depth*) se estableció el valor por defecto, 1. Este parámetro se utilizó para restringir la profundidad del árbol de decisión. Se intentó colocar el valor en '-1', el parámetro de profundidad máxima, sin embargo, *RapidMiner* no estableció un límite a la profundidad del árbol y este nunca detuvo su generación. En este caso, el árbol se construyó hasta que se cumplieron otros criterios de parada.

Seguido de esto, también se aplicó la poda (*apply pruning*), debido a que el modelo de árbol de decisión se puede podar después de la generación. Algunas ramas se reemplazaron por hojas de acuerdo con el parámetro de confianza, el cual se estableció en 1.0, que en *RapidMiner* equivale al 100%, este parámetro especificó el nivel de confianza utilizado para el cálculo del error pesimista de poda.

De la misma manera, se habilitó la casilla de prepoda (*apply prepruning*), el cual es un parámetro que especifica si se deben usar más criterios de parada que la profundidad máxima durante la generación del modelo de árbol de decisión. Como se marcó, los parámetros ganancia mínima, tamaño mínimo de hoja, tamaño mínimo para división y número de alternativas de prepoda se utilizaron como criterios de parada.

Es importante destacar que el nodo solo se divide si su ganancia es mayor que la ganancia mínima. Un valor más alto de ganancia mínima daría como resultado menos

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

divisiones y, por lo tanto, un árbol más pequeño, sin embargo, un valor demasiado alto evitaría por completo la división y se generaría un árbol con un solo nodo.

Finalmente, se estableció un valor de 2 para el tamaño mínimo de hoja (minimal leaf size). El tamaño de una hoja se refiere al número de ejemplos que se encuentran en su subconjunto. El árbol de decisión se construyó de manera que cada hoja contenga al menos el número mínimo de ejemplos requeridos por el tamaño de hoja establecido. Es importante destacar que los parámetros utilizados fueron ajustados a medida que se realizaron las tres pruebas correspondientes.

Esta adaptación de parámetros permitió optimizar el rendimiento y la precisión de los resultados obtenidos en la aplicación del algoritmo de árbol de decisión. La primera prueba realizada fue con la variable objetivo P10, donde la precisión de clasificación se pudo interpretar en porcentajes, obteniendo un 96.19% de exactitud o accuracy. En cuanto a la correcta clasificación de las variables se logró evaluar con la matriz de confusión que se presenta en la figura 35, donde se muestran los porcentajes que identifican a los datos clasificados correctamente, en este caso para el ítem "Totalmente en desacuerdo" fue en un 92.22%, para el ítem "En desacuerdo" fue de 92.72%, el ítem "Ni de acuerdo ni en desacuerdo" obtuvo un 96.05%, el ítem "De acuerdo" resultó con un 90.00% y para el ítem "Totalmente de acuerdo", se obtuvo un 90.91%.

Por consiguiente, con estos resultados se interpreta que el modelo construido con la variable P10 es relativamente muy acertado debido a que la precisión de clasificación fue del 96.19%, lo que significa que el modelo logró clasificar correctamente el 96.19%

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

de los datos. Esta alta precisión es un indicador de la efectividad del modelo en la tarea de clasificación.

Además, se comprobó este porcentaje obtenido utilizando una matriz de confusión, en la que se muestran los porcentajes de datos clasificados correctamente para cada uno de los *ítems*. En general, se obtuvieron porcentajes altos de clasificación correcta para todos los *ítems*, con un mínimo de 90% y un máximo de 96.05%. Por lo tanto, se concluye que el modelo construido con la variable P10 es relativamente muy acertado, lo que sugiere que es una variable importante para el análisis en cuestión. (ver figura 37).

Figura 37.

Matriz de confusión de árbol de decisión pregunta raíz P10

accuracy: 96.19% +/- 4.31% (micro average: 96.20%)

	true Totalmente de ...	true Ni de acuerdo n...	true De acuerdo	true En desacuerdo	true Totalmente en ...	class precision
pred. Totalmente de...	166	0	13	1	0	92.22%
pred. Ni de acuerdo...	0	158	21	11	1	92.72%
pred. De acuerdo	13	22	290	11	1	96.05%
pred. En desacuerdo	1	12	0	68	4	90.00%
pred. Totalmente en...	0	0	1	2	30	90.91%
class recall	92.22%	92.29%	99.23%	93.12%	93.33%	

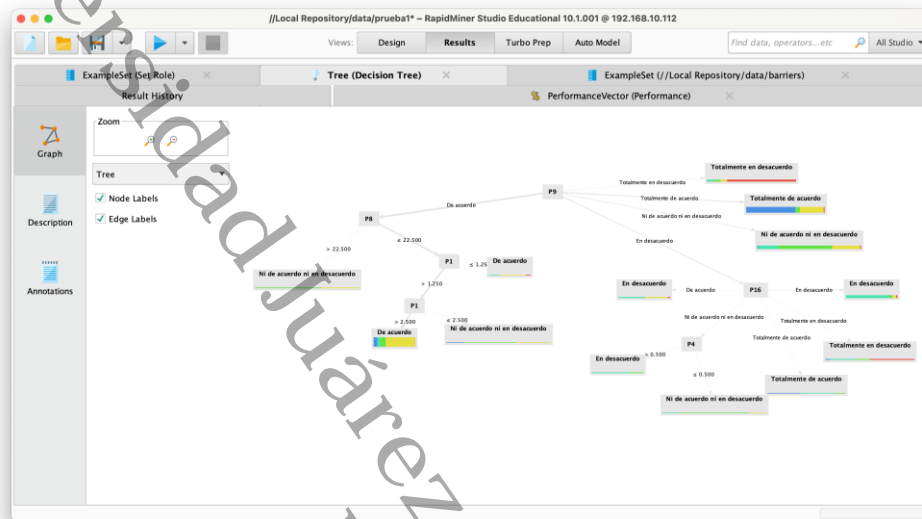
Nota: Elaboración propia.

Después de evaluar esta primera prueba de entrenamiento, la figura 38 muestra al árbol de decisión que generó con la variable P10, el cual se estructuró con 52 niveles y 166 nodos, cabe mencionar que este diagrama se encuentra clasificado de acuerdo con los cinco *ítems* mencionados con anterioridad, debido a que eran parte del clasificador objetivo en la configuración de *RapidMiner* (ver figura 38).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 38.

Prueba de árbol de decisión de la pregunta raíz P10



Nota: Elaboración propia.

RapidMiner brinda la opción de configurar los niveles para una representación visual más adecuada, lo que permite al usuario elegir si desea visualizar todas las ramificaciones posibles o solo algunos niveles. En la figura 36, se observa un diagrama que representa los primeros 6 niveles de profundidad.

Los niveles no mostrados en la imagen no se eliminaron, sino que se ocultaron mediante la configuración del software. La segunda prueba con árboles de decisión se enfocó en la variable P12. Al igual que en el primer ejemplo, no se modificaron los parámetros de inicio y se utilizó el 70% de los datos para entrenar el modelo y el 30% de prueba.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Se presenta a continuación la matriz de confusión para determinar el porcentaje de clasificación correcta. En la figura 39 se observa que los porcentajes de datos correctamente clasificados continúan siendo superiores al 90%, con porcentajes de 91.47%, 94.78%, 92.14%, 99.44%, y 92.50% para los ítems "Totalmente en desacuerdo", "En desacuerdo", "Ni de acuerdo ni en desacuerdo", "De acuerdo" y "Totalmente de acuerdo", respectivamente.

La precisión de clasificación obtenida con la variable P12 fue de 98.16%, en comparación con la variable P10, que obtuvo una exactitud del 96.19%.

En contraste con la variable P10, en esta prueba, el ítem "En desacuerdo" obtuvo el mayor porcentaje de clasificación (99.44%). Con esta información, se logró interpretar que el uso de la variable P12 como pregunta raíz también generó un modelo de clasificación correcto y que no se presentaron cambios significativos en comparación con la prueba anterior (ver figura 39).

Figura 39.

Matriz de confusión de árbol de decisión pregunta raíz P12

accuracy: 98.16% +/- 3.75% (micro average: 98.16%)

	true Totalmente de ...	true De acuerdo	true Ni de acuerdo n...	true En desacuerdo	true Totalmente en ...	class precision
pred. Totalmente de...	277	55	4	3	1	91.47%
pred. De acuerdo	44	195	42	19	1	94.78%
pred. Ni de acuerdo...	2	24	18	12	0	92.14%
pred. En desacuerdo	3	20	15	44	7	99.44%
pred. Totalmente en...	1	3	3	4	29	92.50%
class recall	94.71%	95.66%	91.95%	93.66%	96.32%	

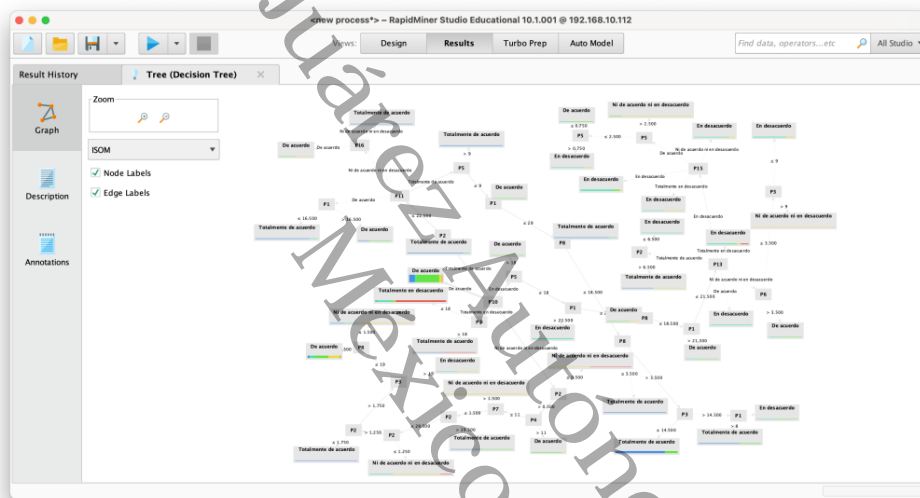
Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tras el análisis de los datos, se visualizó el árbol de decisiones resultante, el cual tuvo un número de niveles de 57 con 208 nodos. Esto indica que el nivel de detalle en este árbol fue aún mayor que en la primera prueba, como se puede apreciar en la figura 40 (ver figura 40).

Figura 40.

Prueba de árbol de decisión de la pregunta raíz P12



Nota: Elaboración propia.

De la misma manera que en la prueba anterior de la pregunta raíz P10, se optó por ocultar los niveles subsecuentes del árbol de decisión debido a su gran extensión, lo cual dificultaba su correcta visualización, debido a esto, en la figura 38 solo se observan los primeros treinta nodos del diagrama generado.

La tercera prueba realizada con los árboles de decisión utilizó la variable P19 como pregunta raíz y se realizó la prueba de entrenamiento con el 70% de los datos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Los resultados de la evaluación se presentan en la figura 41, donde se observa que los porcentajes de datos correctamente clasificados continúan siendo superiores al 90%, con porcentajes de 96.64% para la variable "Totalmente en desacuerdo", 90.20% para la respuesta "En desacuerdo", 94.93% para el ítem "Ni de acuerdo ni en desacuerdo", 92.54% para "De acuerdo" y 91.17% para "Totalmente de acuerdo" (ver figura 41). La precisión de clasificación con la variable P10 fue de 96.19%, sin embargo, la variable P12 obtuvo un porcentaje de 98.16%, finalmente P19 resultó con un 90.10%, siendo este el porcentaje más bajo de las tres preguntas raíz.

Figura 41.

Matriz de confusión árbol de decisión pregunta raíz P19

accuracy: 90.10% +/- 8.62 (micro average: 90.10%)

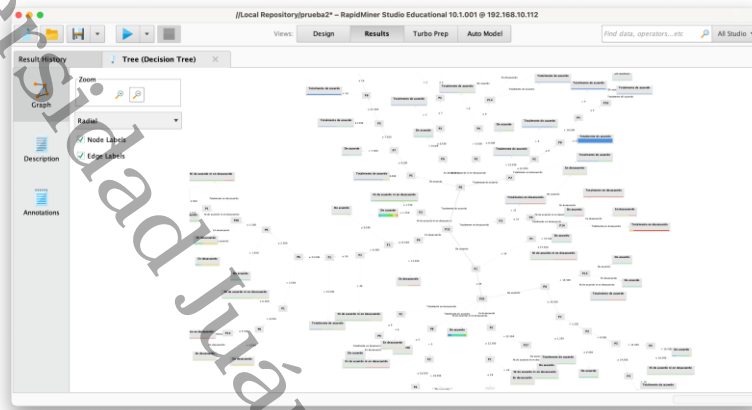
	true Totalment...	true Ni de acu...	true De acuerdo	true Totalment...	true En desacu...	class precision
pred. Totalme...	187	13	6	2	36	96.64%
pred. Ni de ac...	4	59	13	1	21	90.20%
pred. De acuer...	2	21	87	6	18	94.93%
pred. Totalme...	2	0	4	9	2	92.54%
pred. En desac...	32	29	33	2	237	91.17%
class recall	92.38%	98.36%	90.84%	95.00%	95.48%	

Nota: Elaboración propia.

En la tercera prueba realizada, se generó el diagrama de árbol correspondiente utilizando la variable P19 como pregunta raíz. La figura 42 muestra el resultado obtenido, el cual cuenta con un total de 287 nodos y 60 niveles, convirtiéndolo en el diagrama con mayor flujo de información entre las pruebas presentadas en este apartado (ver figura 42).

Figura 42.

Prueba de árbol de decisión de la pregunta raíz P19



Nota: Elaboración propia.

En la figura 42 se puede ver que, aunque el diagrama no se creó con una clasificación específica, la mayoría de los datos en la variable P19 están clasificados como "Totalmente en desacuerdo", como se muestra en la matriz de confusión. Esto influye en la creación del flujo de decisiones. Es importante tener en cuenta que, aunque las tres variables utilizadas para las pruebas dieron resultados favorables en las métricas, los diagramas generados son diferentes.

Por lo tanto, elegir el nodo raíz para construir el modelo es crucial para identificar patrones en esta investigación. Asimismo, los resultados obtenidos en las pruebas pueden estar influenciados por diversos factores, tales como la calidad de los datos, la cantidad de datos disponibles y el método utilizado para la evaluación del modelo. Por lo tanto, es importante realizar una evaluación rigurosa de los resultados obtenidos y considerar su validez en el contexto de la investigación, donde se obtenga como mínimo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

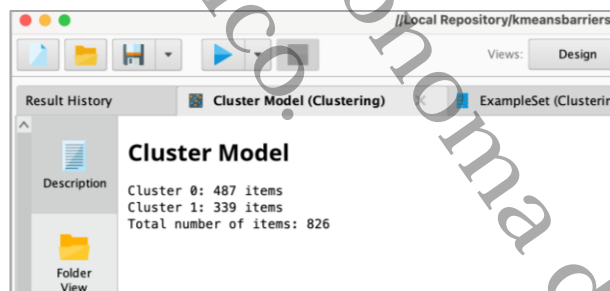
un 85% de exactitud en el análisis de los datos a través de los árboles de decisión, este valor se basó en las diversas investigaciones referenciadas en esta investigación.

Agrupación

Durante las pruebas de minería de datos, se utilizó el modelo de agrupación para encontrar el número óptimo de *clústeres* que mejor representara los datos. El algoritmo *Kmeans* se utilizó en diferentes pruebas, pero este apartado se centra en tres pruebas específicas con diferentes agrupaciones que se descartaron en el proceso, como se puede observar en la figura 43, al realizar la prueba con dos *clústeres*, se obtuvo que el *Clúster 0* se conformó por 487 *ítems*, mientras que el *Clúster 1* se integró por 339 *ítems* (ver figura 43).

Figura 43.

Ítems que conforman los dos clústeres



Nota: Elaboración propia.

A continuación, en la figura 44 se consideraron las variables de edad y *clúster*. Se puede observar que el *clúster* azul representa a los profesores con edades comprendidas entre 24 a 52 años, mientras que el *clúster* verde representa a docentes con edades entre más de 52 y 80 años. Sin embargo, en el *clúster* verde, algunos puntos se encuentran

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

lejos del centroide, especialmente en el caso de profesores de 24 a 28 años. Del mismo modo, en el *cluster* azul, hay puntos distantes en el rango de edad de 75 a 80 años (ver figura 44).

Figura 44.

Kmeans prueba con dos agrupaciones



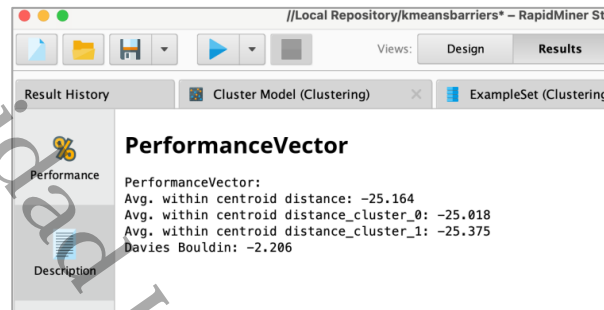
Nota: Elaboración propia.

En el caso de *RapidMiner*, aunque la herramienta puede calcular la medida de la silueta para evaluar la calidad del modelo de *Kmeans*, esta medida no proporciona una evaluación completa de la calidad del modelo debido a la subjetividad de la técnica.

Además, la interpretación de los resultados de la medida de la silueta también requiere un conocimiento profundo del conjunto de datos y del problema que se está abordando; sin embargo, a partir de la versión 9.2 la medida de silueta fue eliminada de *RapidMiner*, siendo suplantada por las *average* o promedio por su traducción al español, obteniendo los siguientes resultados para la prueba de dos *clústeres* (ver figura 45).

Figura 45.

Promedio de distancias centroides con dos clústeres



Nota: Elaboración propia.

Estos resultados sugirieron que el *clustering* ha producido grupos (*clústeres*) altamente cohesivos y bien separados, indicado por el valor negativo de la distancia media dentro de los centroides (-25.164), lo que significa que los puntos dentro de cada *clúster* están muy cerca de sus centroides correspondientes.

Además, se pudo observar que la distancia media dentro de cada uno de los *clústeres* es similar, ya que los valores para "Avg. within centroid distance_cluster_0" y "Avg. within centroid distance_cluster_1" son casi idénticos, lo que sugirió que los dos *clústeres* tuvieron la misma cohesión. El valor de la medida Davies Bouldin de -2.206 mostró que la separación entre los *clústeres* es muy buena, ya que esta medida cuantifica la relación entre la dispersión dentro de los *clústeres* y la separación entre los *clústeres*. Un valor negativo de la medida Davies Boulding indica una excelente separación entre los *clústeres*, por lo tanto, un valor de -2.206 en la medida Davies Bouldin sugiere que la separación entre los *clústeres* es excelente, lo que significa que los grupos son muy cohesivos y están bien separados. En resumen, estos resultados sugirieron que el

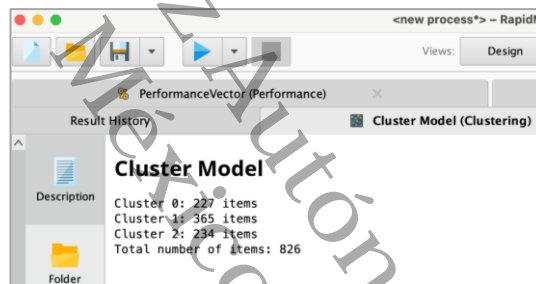
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

análisis de *clustering* produjo grupos altamente cohesivos y bien separados, con una separación global entre los *clústeres* muy adecuada.

Sin embargo, se realizaron otras dos pruebas con más agrupaciones, utilizando la variable Edad, ya que se determinó que con esta variable las agrupaciones en *RapidMiner* son más claras. En la segunda prueba, se utilizaron tres agrupaciones, como se muestra en la figura 46, donde se obtuvo que los *clústeres* 0, 1 y 2 se conformaron por 227, 365 y 234 *ítems*, respectivamente (ver figura 46).

Figura 46.

Ítems que conforman los tres clústeres



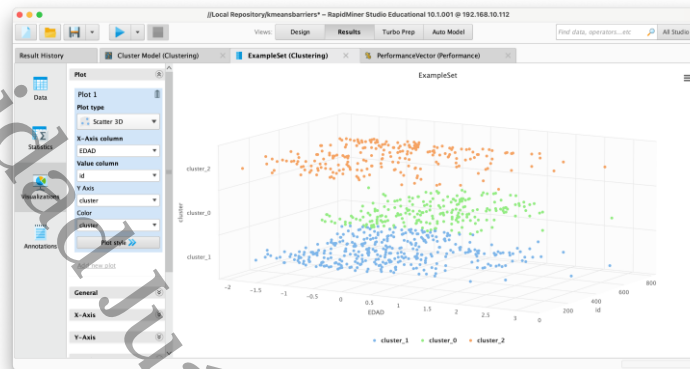
Nota: Elaboración propia.

En el gráfico interactivo de *Kmeans*, el primer *clúster* en color azul representa a profesores con edades comprendidas entre 24 y 32 años, el *clúster* verde está compuesto por profesores de más de 32 a 50 años, y el *clúster* naranja incluye a profesores con edades entre más de 50 y 80 años (ver figura 47).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 47.

Kmeans prueba con tres agrupaciones



Nota: Elaboración propia.

El método de la silueta es una técnica de evaluación de *clustering* que permite medir la calidad de los *clústeres* formados. Esta técnica no está disponible en *RapidMiner* debido a limitaciones de desarrollo que presentó la aplicación hasta la versión 9.2.

Como alternativa, *RapidMiner* utiliza la medida de distancia promedio dentro del *clúster* y la medida de índice de Davies Bouldin para evaluar la calidad de los *clústeres*. La distancia promedio dentro del *clúster* indica qué tan similares son los puntos dentro de cada *clúster*, mientras que el índice de Davies Bouldin mide la calidad de la agrupación en términos de la separación entre los *clústeres* y la varianza dentro de ellos.

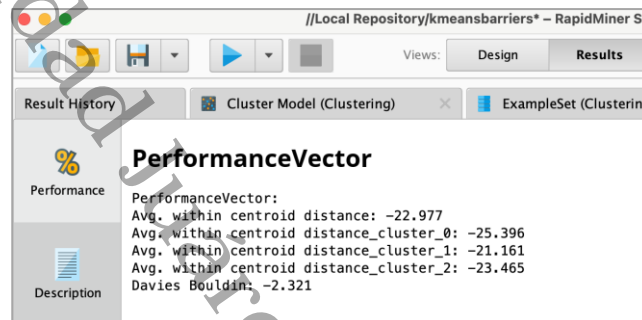
Aunque el método de la silueta es una técnica precisa para la evaluación de *clustering*, el uso de la distancia promedio dentro del *clúster* y el índice de Davies Bouldin puede proporcionar una evaluación más útil y confiable de los *clústeres* formados en *RapidMiner*, debido a sus representaciones con datos numéricos precisos, *RapidMiner*

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

arrojó los resultados de promedio de distancias entre centroides para el análisis de tres *clústeres* (ver figura 48).

Figura 48.

Promedio de distancias centroides con tres clústeres



Nota: Elaboración propia.

Los resultados demostraron que se realizó un análisis de *clústeres* y se evaluaron los resultados utilizando dos métricas: la distancia promedio dentro del *clúster* y el índice de Davies Bouldin. La distancia promedio dentro de los *clústeres* indica qué tan similares son los puntos dentro de cada *clúster*, mientras que el índice de Davies Bouldin mide la calidad de la agrupación en términos de la separación entre los *clústeres* y la varianza dentro de ellos.

En este caso, los resultados indicaron que el modelo de *clústeres* logró agrupar los puntos de manera efectiva, ya que la distancia promedio dentro de cada *clúster* es relativamente baja. Además, el índice de Davies Bouldin es también bajo con promedio de -2.321, lo que sugiere que los *clústeres* están bien separados y son cohesivos.

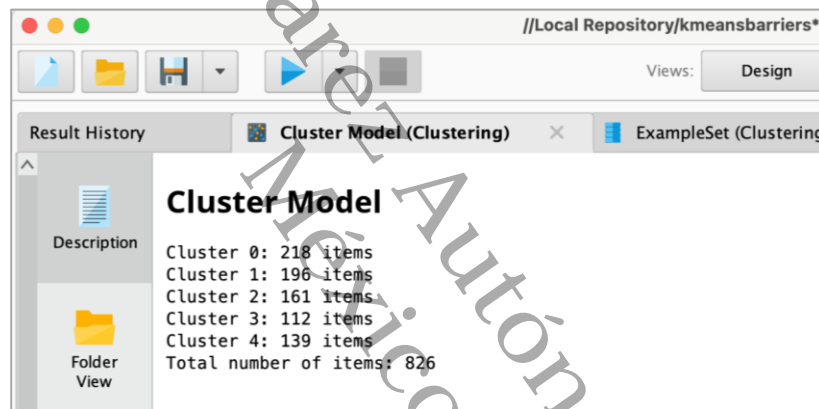
En particular, el *clúster* 1 parece haber sido el más cohesivo de todos, con la distancia promedio más baja entre sus puntos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Hasta ahora se han presentado dos pruebas de agrupamiento con dos y tres *clústeres* respectivamente. Sin embargo, queda por mostrar una tercera prueba con la herramienta *RapidMiner* utilizando cinco grupos. En este ejemplo, se utilizó la función de *Kmeans* interactivo y se presentan los nuevos grupos conformados por 218, 196, 161, 112 y 139 *ítems*, respectivamente, como se puede apreciar en la figura (ver figura 49).

Figura 49.

Ítems que conforman los cinco clústeres



Nota: Elaboración propia.

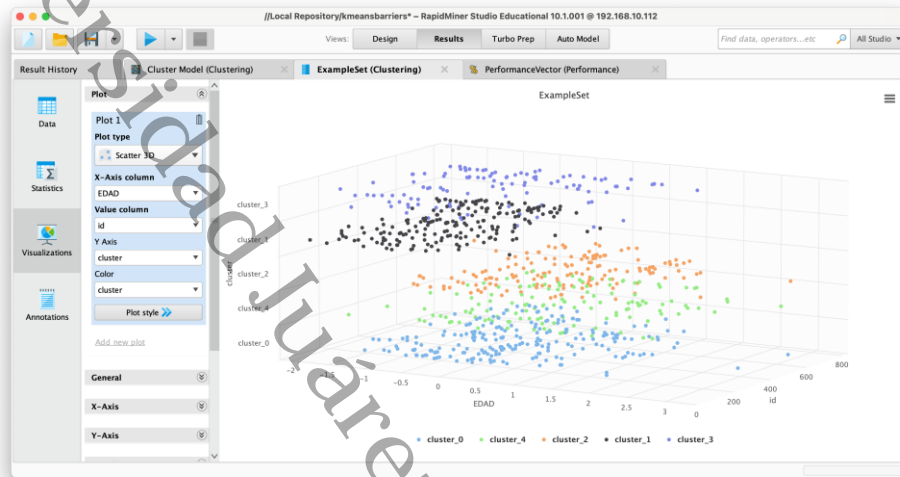
En la figura 50 se muestran los nuevos grupos creados, sin embargo, se observa que los grupos negro y morado están muy cerca entre sí y algunos datos se superponen. Del mismo modo, la segmentación entre los grupos naranja, verde y azul es confusa.

En resumen, el número de *clústeres* utilizado en este análisis no resultó adecuado, como se puede apreciar en la figura siguiente (ver figura 50).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 50.

Kmeans prueba con cinco agrupaciones



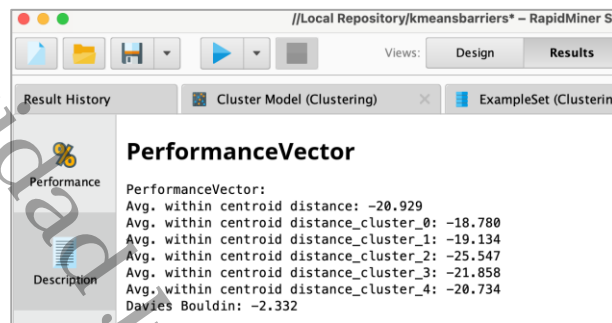
Nota: Elaboración propia.

Después de las pruebas realizadas con el algoritmo de agrupamiento, se pudo observar que se presentaron similitudes en los grupos, especialmente en los que oscilaban entre los 30 y 50 años, los cuales tendieron a relacionarse y dificultaron una separación adecuada de los *clústeres*. Es importante tener en cuenta este aspecto al elegir el número óptimo de agrupaciones.

Para esta última prueba se comprobaron los resultados obtenidos a través de la medición de promedios de *RapidMiner*, obteniendo los datos especificados en la siguiente figura (ver figura 51).

Figura 51.

Promedio de distancias centroides con cinco clústeres



Nota: Elaboración propia.

Estos resultados indican que se ha realizado un análisis de *clústeres* y se han utilizado dos métricas para evaluar la calidad de la agrupación. La medida "Avg. within centroid distance" indica la distancia promedio de los puntos dentro de cada *clúster* al centroide correspondiente, mientras que "Davies Bouldin" mide la calidad general de los *clústeres* formados, que para este caso fue de -2.332.

Los resultados específicos indican que los *clústeres* 0, 1, 3 y 4 tienen valores de "Avg. within centroid distance" similares entre sí, con valores de alrededor de -19 a -21, lo que sugiere que los puntos dentro de estos *clústeres* son relativamente similares entre sí en términos de sus características. Sin embargo, el *clúster* 2 tiene un valor de "Avg. within centroid distance" mucho más alto, de -25.547, lo que sugiere que los puntos dentro de ese *clúster* son muy diferentes entre sí en comparación con los otros *clústeres*.

Para finalizar, se debe considerar que a medida que se aumentó el número de grupos, la métrica de evaluación de calidad disminuyó, así mismo, es importante destacar que elegir un número de *clústeres* demasiado grande puede resultar en una sobrecarga

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

de información y en agrupaciones con muy pocos puntos, lo que dificulta la interpretación y utilidad de los resultados.

Por otro lado, si se elige un número muy pequeño de *clústeres*, se puede perder información importante y los grupos resultantes pueden no ser representativos de los datos subyacentes. Por lo tanto, es importante encontrar un equilibrio y elegir el número óptimo de *clústeres* que permita una buena representación de los datos y una interpretación clara y útil de los resultados.

Reglas de asociación

En esta sección se detallan las pruebas realizadas utilizando la técnica de reglas de asociación, empleando la herramienta *RapidMiner*. El objetivo de estas pruebas fue evaluar la usabilidad de los parámetros y mejorar la eficiencia en la búsqueda de las reglas más relevantes, con el fin de obtener resultados más claros y fácilmente interpretables.

En la primera prueba, se estableció un soporte mínimo del 10% y un mínimo de confianza del 55%. Estos valores fueron seleccionados para identificar las reglas que tuvieran un respaldo significativo en los datos y una confianza aceptable. Los resultados obtenidos se presentan en la figura 52, donde se muestran las reglas encontradas y sus respectivos valores de soporte y confianza (ver figura 52).

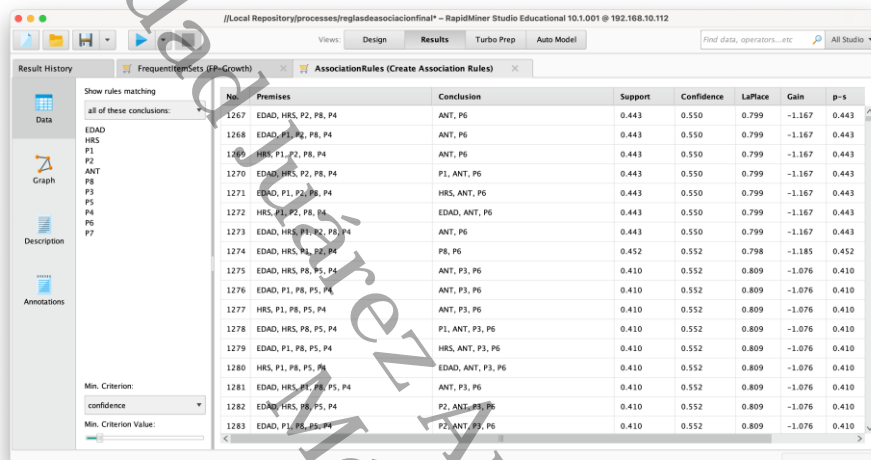
Esta etapa de prueba permitió ajustar y optimizar los parámetros de búsqueda, buscando encontrar un equilibrio entre la cantidad de reglas generadas y su relevancia. A partir de estos resultados, se realizó un análisis más detallado y se establecieron los

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

criterios para la selección de las reglas más relevantes que se utilizarían en el posterior análisis e interpretación de los datos.

Figura 52.

Reglas de asociación obtenidas con 55% de confianza



No.	Premises	Conclusion	Support	Confidence	Lift	Gain	p-s
1267	EDAD, HRS, P2, P8, P4	ANT, P6	0.443	0.550	0.799	-1.167	0.443
1268	EDAD, P1, P2, P8, P4	ANT, P6	0.443	0.550	0.799	-1.167	0.443
1269	HRS, P1, P2, P8, P4	ANT, P6	0.443	0.550	0.799	-1.167	0.443
1270	EDAD, HRS, P2, P8, P4	P1, ANT, P6	0.443	0.550	0.799	-1.167	0.443
1271	EDAD, P1, P2, P8, P4	HRS, ANT, P6	0.443	0.550	0.799	-1.167	0.443
1272	HRS, P1, P2, P8, P4	EDAD, ANT, P6	0.443	0.550	0.799	-1.167	0.443
1273	EDAD, HRS, P1, P2, P8, P4	ANT, P6	0.443	0.550	0.799	-1.167	0.443
1274	EDAD, HRS, P1, P8, P4	P8, P6	0.452	0.552	0.798	-1.185	0.452
1275	EDAD, HRS, P8, P5, P4	ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410
1276	EDAD, P1, P8, P5, P4	ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410
1277	HRS, P1, P8, P5, P4	ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410
1278	EDAD, HRS, P8, P5, P4	P1, ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410
1279	EDAD, P1, P8, P5, P4	HRS, ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410
1280	HRS, P1, P8, P5, P4	EDAD, ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410
1281	EDAD, HRS, P1, P8, P5, P4	ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410
1282	EDAD, HRS, P8, P5, P4	P2, ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410
1283	EDAD, P1, P8, P5, P4	P2, ANT, P3, P6	0.410	0.552	0.809	-1.076	0.410

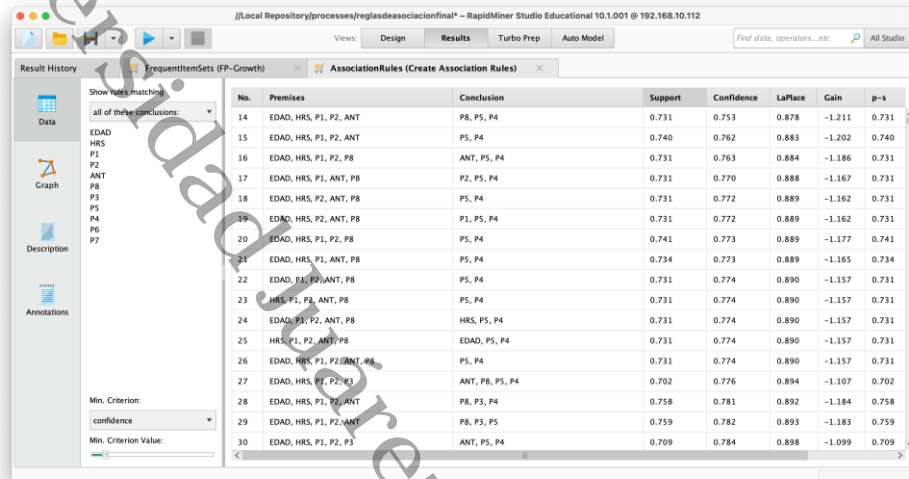
Nota: Elaboración propia.

Con base en los parámetros seleccionados en la figura 50, se obtuvieron 15,698 reglas con un *lift* mayor a uno, lo que indica su repetición dentro del conjunto de datos. Sin embargo, al disminuir el porcentaje de confianza, el parámetro *gain* mostró métricas negativas, disminuyendo así la calidad de las reglas y ofreciendo información poco confiable. Esta primera prueba generó una cantidad excesiva de reglas, y no todas ofrecen información relevante. La herramienta por defecto está configurada para generar todas las reglas posibles sin necesidad de filtros, lo que hace necesario realizar diferentes pruebas. La figura 53 presenta un nuevo conjunto de reglas generadas con un soporte mínimo del 15% y una confianza mínima del 70% (ver figura 53).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 53.

Reglas de asociación obtenidas con 70% de confianza



The screenshot shows the 'AssociationRules (Create Association Rules)' window in RapidMiner Studio. The table displays 30 association rules with columns for No., Premises, Conclusion, Support, Confidence, LaPlace, Gain, and p-s. The confidence values are all 0.731, 0.772, 0.773, 0.774, 0.776, 0.781, 0.782, 0.784, or 0.702, indicating a 70% confidence threshold.

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s
14	EDAD, HRS, P1, P2, ANT	P8, P5, P4	0.731	0.753	0.878	-1.211	0.731
15	EDAD, HRS, P1, P2, ANT	P5, P4	0.740	0.762	0.883	-1.202	0.740
16	EDAD, HRS, P1, P2, P8	ANT, P5, P4	0.731	0.763	0.884	-1.186	0.731
17	EDAD, HRS, P1, ANT, P8	P2, P5, P4	0.731	0.770	0.888	-1.167	0.731
18	EDAD, HRS, P2, ANT, P8	P5, P4	0.731	0.772	0.889	-1.162	0.731
19	EDAD, HRS, P2, ANT, P8	P1, P5, P4	0.731	0.772	0.889	-1.162	0.731
20	EDAD, HRS, P1, P2, P8	P5, P4	0.741	0.773	0.889	-1.177	0.741
21	EDAD, HRS, P1, ANT, P8	P5, P4	0.734	0.773	0.889	-1.165	0.734
22	EDAD, P1, P2, ANT, P8	P5, P4	0.731	0.774	0.890	-1.157	0.731
23	HRS, P1, P2, ANT, P8	P5, P4	0.731	0.774	0.890	-1.157	0.731
24	EDAD, P1, P2, ANT, P8	HRS, P5, P4	0.731	0.774	0.890	-1.157	0.731
25	HRS, P1, P2, ANT, P8	EDAD, P5, P4	0.731	0.774	0.890	-1.157	0.731
26	EDAD, HRS, P1, P2, ANT, P8	P5, P4	0.731	0.774	0.890	-1.157	0.731
27	EDAD, HRS, P1, P2, P8	ANT, P8, P5, P4	0.702	0.776	0.894	-1.107	0.702
28	EDAD, HRS, P1, P2, ANT	P8, P3, P4	0.758	0.781	0.892	-1.184	0.758
29	EDAD, HRS, P1, P2, ANT	P8, P3, P5	0.759	0.782	0.893	-1.183	0.759
30	EDAD, HRS, P1, P2, P3	ANT, P5, P4	0.709	0.784	0.898	-1.099	0.709

Nota: Elaboración propia.

En este segundo intento, se agregó el requisito de que las reglas tuvieran un mínimo de tres ítems en los datos antecesores, lo que proporciona más información para la detección de patrones. Con esta configuración se generaron un total de 3,545 reglas, lo que indica que en esta segunda prueba se pudo identificar información valiosa para la detección de patrones en los profesores. No obstante, al aumentar los parámetros de soporte y confianza, se genera un número aún menor de reglas, como se muestra en la figura 54 que representa la tercera prueba (ver figura 54).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 54.

Reglas de asociación obtenidas con 90% de confianza

No.	Premises	Conclusion	Support	Confidence	Lift	Gain	B-S
1	EDAD, P1, P2, ANT, P8	HRS	0.944	1	1	-0.944	0.944
2	HRS, P1, P2, ANT, P8	EDAD	0.944	1	1	-0.944	0.944
3	EDAD, HRS, P1, ANT, P3	P2	0.891	1	1	-0.891	0.891
4	EDAD, P1, P2, ANT, P3	HRS	0.891	1	1	-0.891	0.891
5	HRS, P1, P2, ANT, P3	EDAD	0.891	1	1	-0.891	0.891
6	EDAD, HRS, P2, ANT, P5	P1	0.815	1	1	-0.815	0.815
7	EDAD, P1, P2, ANT, P5	HRS	0.815	1	1	-0.815	0.815
8	HRS, P1, P2, ANT, P5	EDAD	0.815	1	1	-0.815	0.815
9	EDAD, HRS, P2, ANT, P4	P1	0.806	1	1	-0.806	0.806
10	EDAD, P1, P2, ANT, P4	HRS	0.806	1	1	-0.806	0.806
11	HRS, P1, P2, ANT, P4	EDAD	0.806	1	1	-0.806	0.806
12	EDAD, HRS, P1, ANT, P6	P2	0.452	1	1	-0.452	0.452
13	EDAD, HRS, P2, ANT, P6	P1	0.452	1	1	-0.452	0.452
14	EDAD, P1, P2, ANT, P6	HRS	0.452	1	1	-0.452	0.452
15	HRS, P1, P2, ANT, P6	EDAD	0.452	1	1	-0.452	0.452
16	EDAD, HRS, P1, ANT, P7	P2	0.154	1	1	-0.154	0.154
17	EDAD, HRS, P2, ANT, P7	P1	0.154	1	1	-0.154	0.154

Nota: Elaboración propia.

En la figura 54 se puede apreciar que se estableció un soporte mínimo del 20% y un mínimo de confianza del 90%, además de requerir un mínimo de cinco *items* en los datos antecesores. Como resultado, se obtuvieron 31 reglas con un *lift* mayor a uno que pueden ser útiles para la interpretación.

Estos ejemplos demuestran la importancia de reducir el número de reglas para obtener información eficiente. A partir de un total inicial de 15,698 reglas, se logró obtener un conjunto más reducido y relevante de 31 reglas. Sin embargo, es importante tener en cuenta que una reducción excesiva puede llevar a la pérdida de información valiosa.

4.1.2 SAS Enterprise Miner

En esta investigación se utilizó la herramienta *SAS Enterprise Miner* para aplicar técnicas de minería de datos, específicamente en las tareas de árboles de decisión,

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

agrupación y reglas de asociación. Se llevaron a cabo diversas pruebas de usabilidad para familiarizarse con su funcionamiento y utilizar los algoritmos adecuados para cada tarea. En la tarea de Árboles de decisión, se evaluaron diferentes algoritmos como el CART, ID3 y el C4.5, y se determinó que el algoritmo CART era el más fácil de implementar y utilizar para este tipo de análisis.

En cuanto a la tarea de agrupación se evaluaron diferentes algoritmos como el *Kmeans*, el *Clustering* Hierárquico y el DBSCAN, determinando que el algoritmo *Kmeans* era el más eficiente y preciso para esta tarea en particular. Para la tarea de diagrama de dispersión (visualización de *clústeres*) se utilizaron algoritmos como el Análisis Factorial y el Análisis de Componentes Principales, y se determinó que el Análisis de Componentes Principales era el más adecuado para esta tarea en específico.

Por último, para la tarea de Reglas de Asociación se probaron diferentes algoritmos como el *Apriori*, el FP-Growth y el Algoritmo de Árbol de Decisión, y se determinó que el algoritmo *Apriori* era el más eficiente y rápido para identificar patrones y relaciones en grandes conjuntos de datos.

Árboles de decisión

Se llevaron a cabo pruebas en la herramienta *RapidMiner* para seleccionar la pregunta principal del árbol de decisión en *SAS Enterprise Miner*. Se priorizaron los *ítems* con mayor ganancia de información, que es una métrica utilizada en minería de datos y aprendizaje automático para identificar las características más importantes de un modelo de predicción. La ganancia de información se calcula midiendo la reducción de la incertidumbre en los datos al dividirlos según una característica en particular. Se

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

seleccionaron tres preguntas raíz relevantes y específicas para el análisis de las barreras ante el uso de innovaciones tecnológicas en los profesores de educación superior:

- 10. El equipo tecnológico de apoyo a las actividades docentes es escaso (P10),
- 12. La accesibilidad a la red en los espacios universitarios es inestable (P12),
- 19. He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente (P19).

Después de la selección de las preguntas raíz, se generó el árbol de decisión en *SAS Enterprise Miner* utilizando el criterio de frecuencia por defecto para seleccionar los atributos para la división y maximizar la precisión del árbol, antes de visualizar los gráficos es importante mencionar que un árbol de decisión es una herramienta de análisis y visualización de datos que permite tomar decisiones basadas en una serie de condiciones y resultados posibles. En esencia, es una estructura de ramificación que comienza en un nodo raíz y se ramifica hacia abajo en varios nodos secundarios, cada uno de los cuales representa una posible decisión o resultado.

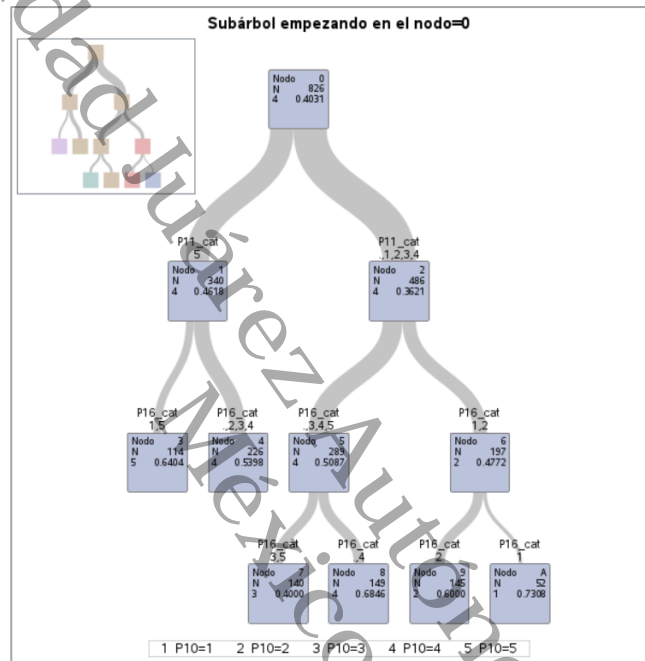
Estos árboles son ampliamente utilizados en campos como la inteligencia artificial, la minería de datos y el aprendizaje automático, y son especialmente útiles cuando se trata de problemas de clasificación o predicción, a continuación en la figura 55, se observa el primer árbol obtenido a partir de la pregunta raíz P10, el cual tuvo una profundidad de 10 y 238 hojas antes de la poda, sin embargo, después de esta, se logró una profundidad de 3 niveles y 6 hojas en el árbol de decisión, en esta aplicación no fue necesario ocultar ningún nivel del árbol resultante, debido a que *SAS Enterprise Miner* produce la versión

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

más óptima de este, es importante destacar que se utilizó el 70% de los datos para entrenar el modelo y el 30% como prueba (ver figura 55).

Figura 55.

Árbol de decisión obtenido para la pregunta raíz P10



Nota: Elaboración propia.

La primera evaluación se llevó a cabo utilizando la variable P10. La tasa de clasificación precisa se pudo interpretar en forma de porcentajes, logrando un accuracy del 95.62%. La clasificación correcta de las variables se evaluó utilizando una matriz de confusión, la cual se muestra en la figura 56. Se observa que el porcentaje de datos clasificados correctamente para cada ítem fue de 96.56% - "Totalmente en desacuerdo", 95.88% - "En desacuerdo", 94.09% - "Ni de acuerdo ni en desacuerdo", 96.73% - "De acuerdo" y 94.87% - "Totalmente de acuerdo".

 Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Por lo tanto, el modelo construido con la variable P10 es altamente preciso, ya que la tasa de clasificación fue del 95.62%, lo que significa que el modelo pudo clasificar la mayoría de los datos. Además, se puede concluir que la variable P10 es importante para el análisis en cuestión, según los resultados de la matriz de confusión que se presentan en la figura 56 (ver figura 56).

Figura 56.

Matriz de confusión del árbol de decisión - pregunta raíz P10

Procedimiento HPSPLIT						
Matriz de confusión basada en el modelo						
Real	Predichos					Tasa de error
	1	2	3	4	5	
1	38	10	5	5	0	0.3448
2	7	87	24	30	0	0.4122
3	5	21	56	49	6	0.5912
4	2	27	45	224	35	0.3273
5	0	0	10	67	73	0.5133

Nota: Elaboración propia.

La segunda prueba con árboles de decisión se enfocó en la variable P12. Al igual que en el primer ejemplo, no se modificaron los parámetros de inicio y se utilizó el 70% de los datos para entrenar el modelo.

Se presenta a continuación la matriz de confusión para determinar el porcentaje de clasificación correcta. En la figura 57 se observa que los porcentajes de datos correctamente clasificados continúan siendo superiores al 90%, con porcentajes de error de 2.36%, 2.56%, 7.92%, 3.40% y 1.52% para los ítems "Totalmente en desacuerdo" (97.64%), "En desacuerdo" (97.44%), "Ni de acuerdo ni en desacuerdo" (92.08%), "De

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

acuerdo" (96.6%) y "Totalmente de acuerdo" (98.48%), respectivamente. La precisión de clasificación obtenida con la variable P12 fue de 96.44%, en comparación con la variable P10, que obtuvo una exactitud del 95.62% (ver figura 57).

Figura 57.

Matriz de confusión del árbol de decisión - pregunta raíz P12

Procedimiento HPSPLIT						
Matriz de confusión basada en el modelo						
Real	Predichos					Tasa de error
	1	2	3	4	5	
1	187	36	13	8	0	0.2336
2	32	237	29	35	0	0.2883
3	5	21	60	14	0	0.4000
4	1	18	20	92	0	0.2977
5	2	2	0	14	0	1.0000

Nota: Elaboración propia.

En contraste con la variable P10, en esta prueba, el ítem "De acuerdo" obtuvo el mayor porcentaje de clasificación (96.73%). Con esta información, se logró interpretar que el uso de la variable P12 como pregunta raíz también generó un modelo de clasificación correcto y que no se presentaron cambios significativos en comparación con la prueba anterior. Tras el análisis de los datos, se visualizó el árbol de decisiones resultante, el cual tuvo una profundidad máxima de 10 niveles y 194 hojas, sin embargo, después de la poda se obtuvo una profundidad de árbol de 5 y un número de hojas de 7 después de la poda. Esto indica que el nivel de detalle en este árbol fue similar que la primera prueba, como se puede apreciar en la figura 58 (ver figura 58).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

respuesta "En desacuerdo", 96.00% para el ítem "Ni de acuerdo ni en desacuerdo", 97.03% para "De acuerdo" y 99.00% para "Totalmente de acuerdo" (ver figura 59).

La precisión de clasificación con la variable P10 fue de 95.62%, sin embargo, la variable P12 obtuvo un porcentaje de 96.44%, finalmente P19 resultó con un 97.33%, siendo este el porcentaje más alto de las tres preguntas raíz.

Figura 59.

Matriz de confusión del árbol de decisión - pregunta raíz P19

Procedimiento HPSPLIT						
Matriz de confusión basada en el modelo						
Real	Predichos					Tasa de error
	1	2	3	4	5	
1	187	36	13	8	0	0.2336
2	32	237	29	35	0	0.2883
3	5	21	60	14	0	0.4000
4	1	18	20	92	0	0.2977
5	2	2	0	14	0	1.0000

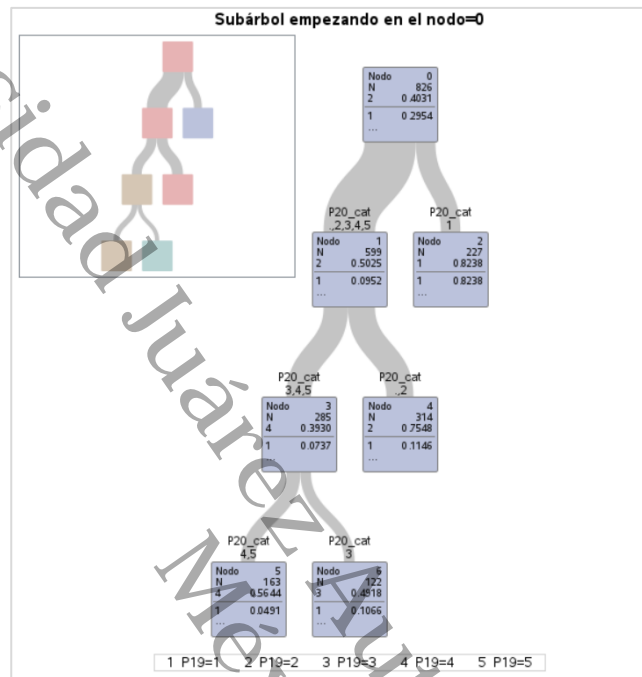
Nota: Elaboración propia.

En la tercera prueba realizada, se generó el diagrama de árbol correspondiente utilizando la variable P19 como pregunta raíz. La figura 60 muestra el resultado obtenido, el cual cuenta con un total de 215 nodos y 10 niveles, convirtiéndolo en el diagrama con mayor flujo de información entre las pruebas presentadas en este apartado, sin embargo, tras la poda del árbol, este finalmente resultó en 3 niveles y 4 nodos (ver figura 60).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 60.

Prueba de árbol de decisión de la pregunta raíz P19



Nota: Elaboración propia.

En la figura 60 se puede observar que, aunque el diagrama no fue creado con una clasificación específica, la mayoría de los datos en la variable P19 se clasifican como "Totalmente de acuerdo", tal como se muestra en la matriz de confusión. Esto tiene una gran influencia en la creación del flujo de decisiones.

Es crucial tener en cuenta que, aunque las tres variables utilizadas en las pruebas arrojaron resultados favorables en las métricas, los diagramas generados son diferentes. Por tanto, la elección del nodo raíz para construir el modelo es fundamental para identificar patrones en esta investigación. Asimismo, los resultados de las pruebas

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

pueden estar influenciados por diversos factores, como la calidad de los datos, la cantidad de datos disponibles y el método utilizado para evaluar el modelo.

Sin embargo, es esencial realizar una evaluación rigurosa de los resultados obtenidos y considerar su validez en el contexto de la investigación, donde se espera alcanzar una precisión mínima del 85% en el análisis de los datos mediante árboles de decisión. En general, se espera que un árbol de decisión tenga una alta precisión en la clasificación de los datos. La precisión se mide utilizando la matriz de confusión, que muestra la cantidad de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. La precisión se calcula como la proporción de la suma de los verdaderos positivos y los verdaderos negativos sobre el total de los datos.

En muchos casos, un porcentaje de precisión de alrededor del 85% se considera adecuado para un árbol de decisión, aunque esto dependerá de la aplicación específica. En cualquier caso, es importante evaluar cuidadosamente la precisión del árbol de decisión y considerar factores como la calidad de los datos y la complejidad del modelo antes de tomar decisiones basadas en sus resultados.

Agrupación

En esta sección, se presentan los resultados obtenidos a través de las pruebas de minería de datos utilizando el algoritmo de agrupación *Kmeans*. El objetivo principal fue determinar el número óptimo de *clústeres* que mejor representara los datos y analizar la distribución de profesores en diferentes rangos de edad. Se llevaron a cabo tres pruebas específicas, centrándose en la clasificación de los datos en dos *clústeres*. Además, se examinó la relación entre la variable de edad y la pertenencia al *clúster* correspondiente.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Los resultados revelaron patrones interesantes, aunque también se observaron casos atípicos que se alejaban de la tendencia general del *clúster*.

Durante las pruebas de minería de datos, se empleó el algoritmo de agrupación *Kmeans* para determinar el número óptimo de *clústeres* que mejor representara los datos. En este apartado, se presentan los resultados de tres pruebas específicas.

En la primera prueba, se utilizó un valor predeterminado de dos *clústeres*. Los resultados mostraron que el *Clúster 1* estaba compuesto por 487 *ítems*, mientras que el *Clúster 2* incluía un total de 339 *ítems*. Es importante destacar que en esta prueba se obtuvo exactamente el mismo número de *ítems* por *clúster*, al igual que en *RapidMiner*.

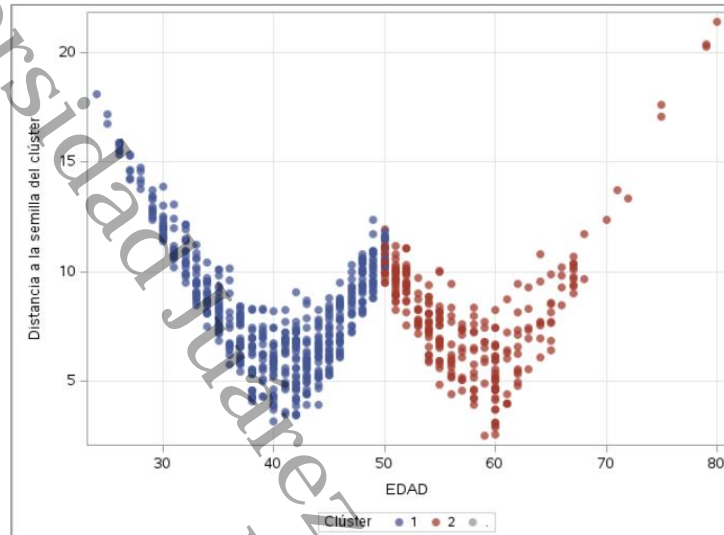
A continuación, en la figura 61, se consideraron las variables de edad y *clúster*. Se observa que el *clúster* azul representa a los profesores con edades comprendidas entre 24 y 52 años, mientras que el *clúster* rojo representa a los docentes con edades superiores a 52 y hasta 80 años. Sin embargo, en el *clúster* azul, se identifican algunos puntos alejados del centroide, especialmente en el caso de profesores de 24 a 25 años.

Del mismo modo, en el *clúster* rojo, se encuentran puntos distantes en el rango de edad de 70 a 80 años. Estos resultados sugieren que, con dos agrupaciones, los datos se clasifican de manera adecuada en su totalidad (ver figura 61).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 61.

Kmeans prueba con dos agrupaciones



Nota: Elaboración propia.

En el caso de SAS Enterprise Miner, como se puede observar en la figura 62, se proporciona un resumen completo de los clústeres generados, el cual contiene la frecuencia de ítems en cada clúster, la desviación estándar RMS obtenida, la distancia máxima desde la semilla hasta la observación, el clúster más cercano a cada dato y la distancia entre centroides del clúster (ver figura 62).

Figura 62.

Resumen de clústeres

Resumen de clúster						
Clúster	Frecuencia	Desviación estándar RMS	Distancia máxima desde semilla hasta la observación	Radio sobrepasado	Clúster más cercano	Distancia entre Centroides del clúster
1	487	1.8982	18.0812		2	17.5218
2	339	1.8306	21.4122		1	17.5218

Nota: Elaboración propia.

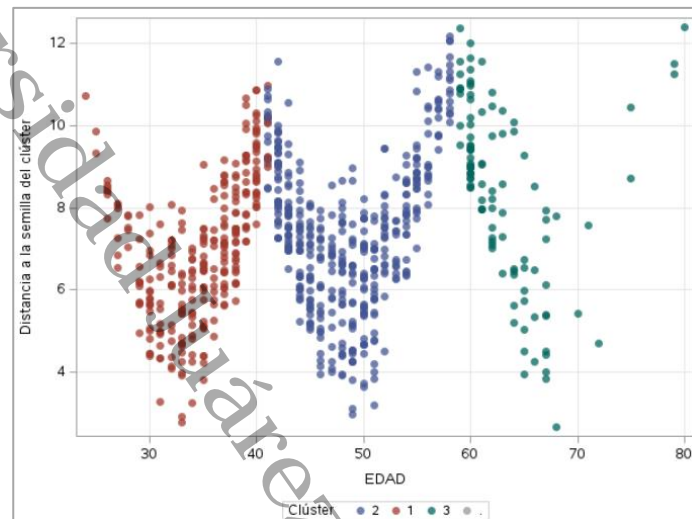
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Estos resultados sugirieron que el *clustering* ha producido dos grupos (*clústeres*) altamente cohesivos y bien separados, indicado por el valor de la distancia entre los centroides del *clúster* (17.52), lo que significa que los puntos dentro de cada *clúster* están muy cerca de sus centroides correspondientes.

Un valor alto de la distancia entre los centroides indica una buena separación entre los *clústeres*, lo que significa que los puntos dentro de cada *clúster* están cercanos entre sí y muy alejados de los puntos de los otros *clústeres*. Esto indica que el *clustering* ha logrado agrupar objetos similares juntos y separarlos fue exitoso en la identificación de grupos bien definidos y homogéneos.

Sin embargo, se realizaron otras dos pruebas con más agrupaciones, utilizando la variable Edad, ya que se determinó que con esta variable las agrupaciones en *SAS Enterprise Miner* son más claras. En la segunda prueba, se utilizaron tres agrupaciones, como se muestra en la figura 63, donde se obtuvo que los *clústeres* 1, 2 y 3 se conformaron por 283, 435 y 108 *ítems*, respectivamente (ver figura 63).

En el gráfico interactivo de *Kmeans*, el primer *clúster* en color rojo representa a profesores con edades comprendidas entre 24 y 32 años, el *clúster* azul está compuesto por profesores de más de 32 a 50 años, y el *clúster* verde incluye a profesores con edades entre más de 50 y máximo 80 años (ver figura 63).

Figura 63.*Kmeans prueba con tres agrupaciones*

Nota: Elaboración propia.

Como anteriormente se ha descrito, *SAS Enterprise Miner* es una herramienta de minería de datos que ofrece diversas técnicas de análisis para la segmentación de datos, incluyendo el algoritmo de *clustering Kmeans*. Al aplicar este método de agrupamiento, el *software* genera un conjunto de *clústeres* o grupos que representan patrones o características similares en los datos.

Al finalizar el proceso, *SAS Enterprise Miner* proporciona un resumen de los *clústeres* resultantes, que permite al usuario entender mejor las estructuras latentes en los datos y realizar análisis posteriores. Este resumen incluye información como el tamaño y la forma de cada *clúster*, así como las variables que contribuyen más significativamente a la formación de cada grupo.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

En este sentido, el resumen de *clústeres* es una herramienta útil para el análisis exploratorio de datos y la toma de decisiones basada en la segmentación de estos, para el caso de la generación de tres *clústeres*, *SAS Enterprise Miner* arrojó los siguientes resultados para el análisis de los tres *clústeres* (ver figura 64).

Figura 64.

Resumen de clústeres

Resumen de clúster						
Clúster	Frecuencia	Desviación estándar RMS	Distancia máxima desde semilla hasta la observación	Radio sobrepasado	Clúster más cercano	Distancia entre Centroides del clúster
1	283	1.5469	10.9641		2	13.8855
2	435	1.6985	12.1904		1	13.8855
3	108	1.5878	12.4030		2	14.7036

Nota: Elaboración propia.

Los resultados de esta prueba presentan un análisis de *clústeres* y la evaluación de los resultados mediante tres métricas comúnmente utilizadas: la desviación estándar RMS, la distancia máxima desde la semilla hasta la observación y la distancia entre centroides del *clúster*.

La desviación estándar RMS es una métrica que mide la dispersión de los datos dentro de un *clúster*, proporcionando una idea de qué tan similares son las observaciones dentro del mismo grupo. Una desviación estándar RMS baja indica que los datos dentro del *clúster* están muy cercanos entre sí, lo que sugiere que la selección de variables o atributos utilizados para crear el *clúster* fue adecuada y efectiva.

La distancia máxima desde la semilla hasta la observación es una medida de la distancia más larga entre cualquier observación y la semilla del *clúster* al que pertenece.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

La distancia entre centroides del *clúster* mide la distancia entre los centroides de dos *clústeres* diferentes. Cuanto mayor sea la distancia entre los centroides, mayor será la diferencia entre los *clústeres*, lo que sugiere que la selección de variables o atributos utilizados para la creación del *clúster* fue efectiva y que los grupos creados son significativamente diferentes entre sí.

En este caso, los resultados indicaron que el modelo de *clústeres* logró agrupar los puntos de manera efectiva, ya que la distancia promedio dentro de cada *clúster* es relativamente baja, siendo de 13.88. Es necesario destacar que un valor bajo de la distancia promedio dentro de cada *clúster* indica que los puntos dentro de cada *clúster* están cercanos entre sí. Por lo tanto, se puede decir que el modelo de *clustering* ha sido efectivo en agrupar los puntos similares juntos y separarlos de los puntos que son diferentes. Además, la distancia máxima desde las semillas hasta la observación es también baja con promedio de 11.85, lo que sugiere que los *clústeres* están bien separados y son cohesivos. En particular, el *clúster* 1 parece haber sido el más cohesivo de todos, con la distancia promedio más baja entre sus puntos.

Hasta ahora se han presentado dos pruebas de agrupamiento con dos y tres *clústeres* respectivamente. Sin embargo, queda por mostrar una tercera prueba con la herramienta *SAS Enterprise Miner* utilizando cinco grupos. En este ejemplo, se utilizó la "tarea" de *Kmeans* y se presentan los nuevos grupos conformados por 294, 106, 186, 227 y 13 *ítems*, respectivamente.

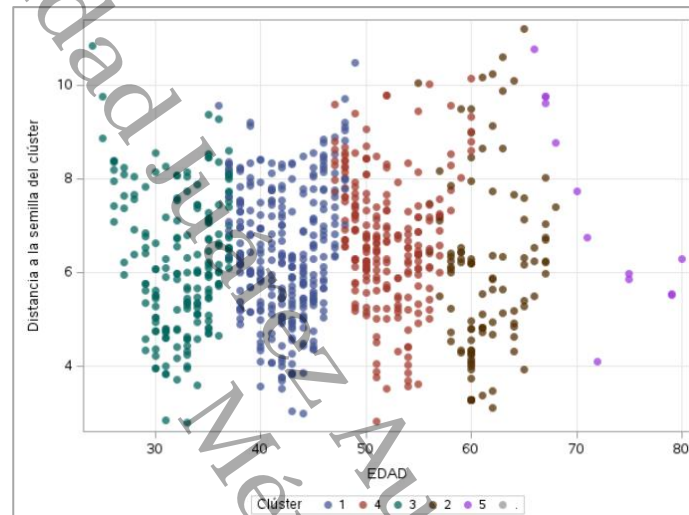
En la figura 65 se muestran los nuevos grupos creados, sin embargo, se observa que los grupos verde, azul y rojo están muy cerca entre sí y algunos datos se superponen,

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

por lo tanto, el número de *clústeres* utilizado en este análisis no resultó adecuado, como se puede apreciar en la figura siguiente (ver figura 65).

Figura 65.

Kmeans prueba con cinco agrupaciones



Nota: Elaboración propia.

Después de las pruebas realizadas con el algoritmo de agrupamiento, se pudo observar que se presentaron similitudes en los grupos, especialmente en los que oscilaban entre los 30 y 50 años, los cuales tendieron a relacionarse y dificultaron una separación adecuada de los *clústeres*. Es importante tener en cuenta este aspecto al elegir el número óptimo de agrupaciones.

Para esta última prueba se comprobaron los resultados obtenidos a través de la tabla de resumen de *clústeres* que proporciona *SAS Enterprise Miner* obteniendo los datos especificados en la figura 66 (ver figura 66).

Figura 66.*Resumen de clústeres*

Resumen de clúster						
Clúster	Frecuencia	Desviación estándar RMS	Distancia máxima desde semilla hasta la observación	Radio sobrepasado	Clúster más cercano	Distancia entre Centroides del clúster
1	294	1.4396	10.4902		4	9.6970
2	106	1.3622	11.2168		4	9.5733
3	186	1.4218	10.8500		1	9.9498
4	227	1.5025	10.1485		2	9.5733
5	13	1.6314	10.7841		2	11.5150

Nota: Elaboración propia.

Estos resultados indican que se ha realizado un análisis de *clústeres* y se han utilizado tres métricas para evaluar la calidad de la agrupación. Como anteriormente se mencionó, la desviación estándar RMS es una métrica que mide la dispersión de los datos dentro de cada *clúster*, es decir, qué tan alejados están los puntos de datos del centroide de cada *clúster*.

Por otra parte, la distancia máxima desde la semilla hasta la observación es una métrica que mide la distancia entre cada punto de datos y su semilla correspondiente, lo que indica qué tan lejos están los puntos de datos de la semilla inicial en cada *clúster*. Por último, la distancia entre centroides mide la distancia entre los centroides de cada *clúster*, lo que indica qué tan diferentes son los *clústeres* entre sí

Los resultados específicos que anteriormente se mostraron, indican que los *clústeres* 1, 2, 3 y 4 tienen valores de distancias entre centroides del *clúster* similares entre sí, con valores de alrededor de 9.57 a 9.94, lo que sugiere que los puntos dentro de estos *clústeres* son relativamente similares entre sí en términos de sus características.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Sin embargo, el *clúster* 5 tiene una distancia entre centroides del *clúster* mucho más alta, de 11.51, lo que sugiere que los puntos dentro de ese *clúster* son muy diferentes entre sí en comparación con los otros *clústeres*.

El análisis de *clústeres* se evaluó con tres métricas a través de *SAS Enterprise Miner*: desviación estándar RMS, distancia máxima desde la semilla y distancia entre centroides del *clúster*. A medida que se aumentó el número de grupos, la métrica de calidad disminuyó al generar cinco grupos. Elegir un número demasiado grande puede generar agrupaciones con pocos puntos y dificultar la interpretación de los resultados, mientras que un número muy pequeño puede perder información importante.

Es importante encontrar un equilibrio para elegir el número óptimo de *clústeres* que permita una buena representación de los datos y una interpretación clara y útil de los resultados.

Reglas de asociación

Se presentan en esta sección las pruebas realizadas utilizando la técnica de reglas de asociación con la herramienta *SAS Enterprise Miner*. El objetivo fue evaluar la usabilidad de los parámetros y mejorar la eficiencia en la búsqueda de las principales reglas para una interpretación más clara de los resultados. En la primera prueba se utilizó un soporte mínimo del 10% y un mínimo de confianza del 55%, como se puede observar en la figura 67 (ver figura 67).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 67.

Reglas de asociación obtenidas con 55% de confianza

Association Report															
Relations	Expected Confidence (%)	Confidence (%)	Support (%)	Lift	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	Rule Item 2	Rule Item 3	Rule Item 4	Rule Item 5	Rule	Index
4	5.77	66.67	3.85	11.86	2.00	VEINTE & SEIS => NUEVE & DOCE	VEINTE & SEIS	NUEVE & DOCE	VEINTE	SEIS	*****	NUEVE	DOCE	1	
4	5.77	66.67	3.85	11.86	2.00	NUEVE & DOCE => VEINTE & SEIS	NUEVE & DOCE	VEINTE & SEIS	NUEVE	DOCE	*****	VEINTE	SEIS	2	
4	3.85	33.33	3.85	8.67	2.00	SEIS & QUINCE & CUATRO => DIECIN	SEIS & QUINCE & CUATRO	DIECIN	SEIS	QUINCE	CUATRO	*****	DIECIN	3	
4	5.77	50.00	3.85	8.67	2.00	VEINTE & TRES => NUEVE & DOCE	VEINTE & TRES	NUEVE & DOCE	VEINTE	TRES	*****	NUEVE	DOCE	4	
4	5.77	50.00	3.85	8.67	2.00	VEINTE & OCHO => NUEVE & DOCE	VEINTE & OCHO	NUEVE & DOCE	VEINTE	OCHO	*****	NUEVE	DOCE	5	
4	5.77	50.00	3.85	8.67	2.00	VEINTE & DOS => NUEVE & DOCE	VEINTE & DOS	NUEVE & DOCE	VEINTE	DOS	*****	NUEVE	DOCE	6	
4	3.85	33.33	3.85	8.67	2.00	VEINTE => NUEVE & DOCE & CINCO	VEINTE	NUEVE & DOCE & CINCO	VEINTE	*****	NUEVE	DOCE	CINCO	7	
4	11.54	100.00	3.85	8.67	2.00	DIECIN => SEIS & QUINCE & CUATRO	DIECIN	SEIS & QUINCE & CUATRO	DIECIN	*****	SEIS	QUINCE	CUATRO	8	
4	11.54	100.00	3.85	8.67	2.00	NUEVE & DOCE & CINCO => VEINTE	NUEVE & DOCE & CINCO	VEINTE	NUEVE	DOCE	CINCO	*****	VEINTE	9	
4	7.69	66.67	3.85	8.67	2.00	NUEVE & DOCE => VEINTE & DOS	NUEVE & DOCE	VEINTE & DOS	NUEVE	DOCE	*****	VEINTE	DOS	10	
4	7.69	66.67	3.85	8.67	2.00	NUEVE & DOCE => VEINTE & OCHO	NUEVE & DOCE	VEINTE & OCHO	NUEVE	DOCE	*****	VEINTE	OCHO	11	
4	7.69	66.67	3.85	8.67	2.00	NUEVE & DOCE => VEINTE & TRES	NUEVE & DOCE	VEINTE & TRES	NUEVE	DOCE	*****	VEINTE	TRES	12	
4	13.46	100.00	3.85	7.43	2.00	SEIS & ONCE => NUEVE & CUATRO	SEIS & ONCE	NUEVE & CUATRO	SEIS	ONCE	*****	NUEVE	CUATRO	13	
3	13.46	100.00	3.85	7.43	2.00	DIECIN => SEIS & QUINCE	DIECIN	SEIS & QUINCE	DIECIN	*****	SEIS	QUINCE	14		
4	13.46	100.00	3.85	7.43	2.00	UNO & DIECIN => SEIS & QUINCE	UNO & DIECIN	SEIS & QUINCE	UNO	DIECIN	*****	SEIS	QUINCE	15	
4	13.46	100.00	3.85	7.43	2.00	DIEZ & DIECIN => SEIS & QUINCE	DIEZ & DIECIN	SEIS & QUINCE	DIEZ	DIECIN	*****	SEIS	QUINCE	16	
4	13.46	100.00	3.85	7.43	2.00	DIECIN & CUATRO => SEIS & QUINCE	DIECIN & CUATRO	SEIS & QUINCE	DIECIN	CUATRO	*****	SEIS	QUINCE	17	
4	13.46	100.00	3.85	7.43	2.00	DIECIN & CINCO => SEIS & QUINCE	DIECIN & CINCO	SEIS & QUINCE	DIECIN	CINCO	*****	SEIS	QUINCE	18	
4	13.46	100.00	3.85	7.43	2.00	DIECIN & CERO => SEIS & QUINCE	DIECIN & CERO	SEIS & QUINCE	DIECIN	CERO	*****	SEIS	QUINCE	19	
4	13.46	100.00	3.85	7.43	2.00	DIECIN => SEIS & QUINCE & CERO	DIECIN	SEIS & QUINCE & CERO	DIECIN	*****	SEIS	QUINCE	CERO	20	
4	13.46	100.00	3.85	7.43	2.00	DIECIN => SEIS & QUINCE & CINCO	DIECIN	SEIS & QUINCE & CINCO	DIECIN	*****	SEIS	QUINCE	CINCO	21	
4	13.46	100.00	3.85	7.43	2.00	DIECIN => SEIS & QUINCE & DIEZ	DIECIN	SEIS & QUINCE & DIEZ	DIECIN	*****	SEIS	QUINCE	DIEZ	22	
4	13.46	100.00	3.85	7.43	2.00	DIECIN => UNO & SEIS & QUINCE	DIECIN	UNO & SEIS & QUINCE	DIECIN	*****	UNO	SEIS	QUINCE	23	
3	3.85	28.57	3.85	7.43	2.00	SEIS & QUINCE => DIECIN	SEIS & QUINCE	DIECIN	SEIS	QUINCE	*****	DIECIN	24		
4	3.85	28.57	3.85	7.43	2.00	UNO & SEIS & QUINCE => DIECIN	UNO & SEIS & QUINCE	DIECIN	UNO	SEIS	QUINCE	*****	DIECIN	25	

Nota: Elaboración propia.

Basándonos en los parámetros establecidos en la figura 67, se identificaron 75 reglas con un *lift* superior a uno, lo que sugiere una alta frecuencia de ocurrencia dentro del conjunto de datos. Aunque se obtuvieron 75 reglas a partir de esta primera prueba, no todas ellas ofrecen información útil o relevante.

La herramienta por defecto está configurada para generar todas las reglas posibles sin necesidad de filtros, lo que hace necesario realizar diferentes pruebas, sin embargo, SAS Enterprise Miner elimina las reglas de asociación redundantes, repetidas y aquellas que carezcan de información relevante, es decir, que indiquen a un solo ítem. La figura 68 presenta un nuevo conjunto de reglas generadas con un soporte mínimo del 15% y una confianza mínima del 70%.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 68.

Reglas de asociación obtenidas con 70% de confianza

Association Report														
Relations	Expected Confidence (%)	Confidence (%)	Support (%)	Lift	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	Rule Item 2	Rule Item 3	Rule Item 4	Rule Item 5	Rule Index
4	48.08	77.78	40.38	1.62	21.00	UNO & CINCO & CERO ==> OCHO	UNO & CINCO & CERO	OCHO	UNO	CINCO	CERO	====>	OCHO	1
4	51.92	84.00	40.38	1.62	21.00	OCHO ==> UNO & CINCO & CERO	OCHO	UNO & CINCO & CERO	OCHO	====>	UNO	CINCO	CERO	2
4	59.62	95.45	40.38	1.80	21.00	OCHO & CINCO ==> UNO & CERO	OCHO & CINCO	UNO & CERO	OCHO	CINCO	====>	UNO	CERO	3
4	59.62	95.45	40.38	1.80	21.00	OCHO & CERO ==> UNO & CINCO	OCHO & CERO	UNO & CINCO	OCHO	CERO	====>	UNO	CINCO	4
4	42.31	67.74	40.38	1.60	21.00	UNO & CINCO ==> OCHO & CERO	UNO & CINCO	OCHO & CERO	UNO	CINCO	====>	OCHO	CERO	5
4	42.31	67.74	40.38	1.60	21.00	UNO & CERO ==> OCHO & CINCO	UNO & CERO	OCHO & CINCO	UNO	CERO	====>	OCHO	CINCO	6
3	50.00	77.78	40.38	1.56	21.00	SEIS ==> DOS & CERO	SEIS	DOS & CERO	SEIS	====>	DOS	CERO		7
3	51.92	80.77	40.38	1.58	21.00	DOS & CERO ==> SEIS	DOS & CERO	SEIS	DOS	CERO	====>	SEIS		8
4	44.23	67.74	40.38	1.83	21.00	UNO & CERO ==> TRES & SEIS	UNO & CERO	TRES & SEIS	UNO	CERO	====>	TRES	SEIS	9
4	59.62	91.30	40.38	1.53	21.00	TRES & SEIS ==> UNO & CERO	TRES & SEIS	UNO & CERO	TRES	SEIS	====>	UNO	CERO	10
3	55.77	84.00	40.38	1.51	21.00	SEIS & CERO ==> DIEZ	SEIS & CERO	DIEZ	SEIS	CERO	====>	DIEZ		11
4	55.77	84.00	40.38	1.51	21.00	UNO & SEIS ==> TRES & CERO	UNO & SEIS	TRES & CERO	UNO	SEIS	====>	TRES	CERO	12
4	55.77	84.00	40.38	1.51	21.00	SEIS & CERO ==> UNO & TRES	SEIS & CERO	UNO & TRES	SEIS	CERO	====>	UNO	TRES	13
3	48.08	72.41	40.38	1.51	21.00	DIEZ ==> SEIS & CERO	DIEZ	SEIS & CERO	DIEZ	====>	SEIS	CERO		14
4	48.08	72.41	40.38	1.51	21.00	UNO & TRES ==> SEIS & CERO	UNO & TRES	SEIS & CERO	UNO	TRES	====>	SEIS	CERO	15
4	48.08	72.41	40.38	1.51	21.00	TRES & CERO ==> UNO & SEIS	TRES & CERO	UNO & SEIS	TRES	CERO	====>	UNO	SEIS	16
3	51.92	77.78	40.38	1.50	21.00	SEIS ==> DIEZ & CERO	SEIS	DIEZ & CERO	SEIS	====>	DIEZ	CERO		17
3	51.92	77.78	40.38	1.50	21.00	SEIS ==> DOS & CINCO	SEIS	DOS & CINCO	SEIS	====>	DOS	CINCO		18
3	51.92	77.78	40.38	1.50	21.00	UNO & DOS ==> SEIS	UNO & DOS	SEIS	UNO	DOS	====>	SEIS		19
3	51.92	77.78	40.38	1.50	21.00	DOS & CINCO ==> SEIS	DOS & CINCO	SEIS	DOS	CINCO	====>	SEIS		20
3	51.92	77.78	40.38	1.50	21.00	DIEZ & CERO ==> SEIS	DIEZ & CERO	SEIS	DIEZ	CERO	====>	SEIS		21
4	51.92	77.78	40.38	1.50	21.00	UNO & TRES & CERO ==> SEIS	UNO & TRES & CERO	SEIS	UNO	TRES	CERO	====>	SEIS	22
4	51.92	77.78	40.38	1.50	21.00	UNO & CINCO & CERO ==> SEIS	UNO & CINCO & CERO	SEIS	UNO	CINCO	CERO	====>	SEIS	23
4	51.92	77.78	40.38	1.50	21.00	SEIS ==> UNO & CINCO & CERO	SEIS	UNO & CINCO & CERO	SEIS	====>	UNO	CINCO	CERO	24
3	51.92	77.78	40.38	1.50	21.00	SEIS ==> UNO & DOS	SEIS	UNO & DOS	SEIS	====>	UNO	DOS		25

Nota: Elaboración propia.

En este segundo intento, no se agregó el requisito de que las reglas tuvieran un mínimo de tres ítems en los datos antecedentes, caso contrario con la aplicación *RapidMiner*. Con esta configuración se generaron un total de 40 reglas, lo que indica que en esta segunda prueba se pudo identificar información valiosa para la detección de patrones en los profesores.

No obstante, al aumentar los parámetros de soporte y confianza, se genera un número aún menor de reglas, como se muestra en la figura 69 que representa la tercera prueba.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 69.

Reglas de asociación obtenidas con 90% de confianza

Association Report														
Relations	Expected Confidence (%)	Confidence (%)	Support (%)	Lift	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	Rule Item 2	Rule Item 3	Rule Item 4	Rule Item 5	Rule Index
2	61.54	76.32	55.77	1.24	29.00	UNO ==> TRES	UNO	TRES	UNO	====>	TRES			1
2	73.08	90.63	55.77	1.24	29.00	TRES ==> UNO	TRES	UNO	TRES	====>	UNO			2
2	73.08	83.78	59.62	1.15	31.00	CINCO ==> UNO	CINCO	UNO	CINCO	====>	UNO			3
2	71.15	81.58	59.62	1.15	31.00	UNO ==> CINCO	UNO	CINCO	UNO	====>	CINCO			4
2	80.77	90.63	55.77	1.12	29.00	TRES ==> CERO	TRES	CERO	TRES	====>	CERO			5
2	61.54	69.05	55.77	1.12	29.00	CERO ==> TRES	CERO	TRES	CERO	====>	TRES			6
2	80.77	86.49	61.54	1.07	32.00	CINCO ==> CERO	CINCO	CERO	CINCO	====>	CERO			7
2	71.15	76.19	61.54	1.07	32.00	CERO ==> CINCO	CERO	CINCO	CERO	====>	CINCO			8
2	73.08	73.81	59.62	1.01	31.00	CERO ==> UNO	CERO	UNO	CERO	====>	UNO			9
2	80.77	81.58	59.62	1.01	31.00	UNO ==> CERO	UNO	CERO	UNO	====>	CERO			10

Nota: Elaboración propia.

Se realizaron ajustes específicos de parámetros, como se muestra en la figura 69, estableciendo un soporte mínimo del 20% y un mínimo de confianza del 90%. Estos ajustes permitieron identificar un total de 25 reglas con un *lift* mayor a uno, las cuales resultan valiosas para la interpretación de los datos. Estas pruebas resaltan la importancia de reducir el número de reglas para obtener información eficiente.

A partir de un conjunto inicial de 75 reglas, se logró obtener un conjunto más reducido pero relevante de 25 reglas. Sin embargo, es fundamental tener precaución al realizar una reducción excesiva, ya que podría implicar la pérdida de información importante. Es necesario encontrar un equilibrio entre la relevancia de las reglas y la cantidad de información que se desea obtener.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

4.1.3 Python

Árboles de decisión

Los árboles de decisión son una herramienta utilizada para tomar decisiones basadas en una serie de condiciones lógicas. En *Python*, es posible utilizar la librería *sklearn* para aplicar el modelo de árboles de decisión a los datos. Una vez que el modelo está entrenado, se pueden utilizar diferentes herramientas para visualizar el árbol y analizar los resultados, como diagramas de árbol y tablas de importancia de variables.

Los árboles de decisión son una herramienta poderosa para la toma de decisiones y la predicción en el análisis de datos. En *Python*, existen diversas librerías que permiten la construcción de árboles de decisión para clasificación y regresión, incluyendo *Scikit-learn*, una de las más populares en el campo del aprendizaje automático. En este contexto, se realizaron tres pruebas en la aplicación de árboles de decisión en *Python* utilizando *Scikit-learn*, con el objetivo de evaluar la eficacia de esta técnica en diferentes escenarios y conjuntos de datos.

En la primera prueba, como se observa en la figura 70, se evaluó el rendimiento del primer árbol de decisión para la clasificación de los patrones de conducta de los profesores de la UJAT, conforme al *dataset* implementado en esta investigación, siendo la pregunta raíz P10. El equipo tecnológico de apoyo a las actividades docentes es escaso. En primera instancia, se definió la variable X como todas las columnas del *Dataframe* df excepto la columna "P10", que era la variable objetivo a predecir. La variable y se definió como la columna "P10" (ver figura 70).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 70.

Importación de librerías y carga de datos en Python – P10

```
In [1]: import pandas as pd
        from sklearn.tree import DecisionTreeClassifier
        from sklearn.model_selection import train_test_split
        from sklearn import metrics
        from sklearn.tree import export_graphviz
        import graphviz

In [2]: # Cargar los datos del archivo Excel
        df = pd.read_excel('numeric.xlsx')

        # Definir las características (features) y la variable objetivo (target)
        X = df.drop('P10', axis=1)
        y = df['P10']
```

Nota: Elaboración propia.

Una vez cargado el *dataset* en *Python*, se procedió a dividir los datos en dos conjuntos: un conjunto de entrenamiento que fue del 70% y un conjunto de prueba del 30% restante, como se observa en la figura 71; es necesario destacar que el conjunto de entrenamiento se utilizó para entrenar el modelo, mientras que el conjunto de prueba se utilizó para evaluar la precisión del modelo. A continuación, se definió el modelo de árbol de decisión con algunas opciones de configuración, tales como *random_state*, la cual fue una semilla aleatoria para reproducibilidad y *max_depth* la profundidad máxima del árbol.

Una vez definidos los parámetros de entrenamiento y prueba, así como el modelo de árbol de decisión, se procedió con el ajuste del modelo de árbol de decisión utilizando los datos de entrenamiento, en esta etapa se utilizaron los datos de prueba para hacer predicciones utilizando el modelo de árbol de decisión (ver figura 71).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 71.

Definición del árbol de decisión – P10

```
In [3]: # Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

# Definir el modelo del árbol de decisión
model = DecisionTreeClassifier(random_state=42, max_depth=3)

# Ajustar el modelo con los datos de entrenamiento
model.fit(X_train, y_train)

Out[3]:
DecisionTreeClassifier
DecisionTreeClassifier(max_depth=3, random_state=42)
```

Nota: Elaboración propia.

Para los árboles de decisión existe el cálculo de precisión, con el que se comprueba la precisión del modelo entrenado, por ello, la precisión es importante porque permite saber cuánto se puede confiar en las predicciones realizadas por el modelo. Una precisión alta significa que el modelo es capaz de hacer predicciones precisas, mientras que una precisión baja significa que el modelo es menos confiable y puede requerir ajustes o mejoras. Debido a lo anterior, se calculó la precisión del modelo comparando las predicciones realizadas en el conjunto de prueba con los valores reales de la variable objetivo, como se puede observar en la figura 72. La precisión del modelo indicó qué tan bien se desempeñó el modelo en la tarea de clasificación (ver figura 72).

Figura 72.

Cálculo de la precisión del modelo – P10

```
In [4]: # Realizar predicciones en el conjunto de prueba
y_pred = model.predict(X_test)

# Calcular la precisión del modelo
accuracy = metrics.accuracy_score(y_test, y_pred)
print("Precisión:", accuracy)

Precisión: 0.9370967741935484
```

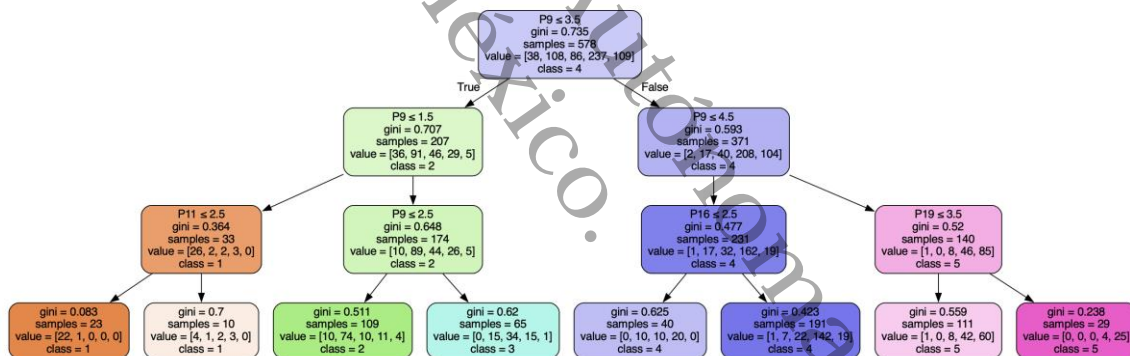
Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Una precisión de 0.9370 o 93.70% significa que el modelo es altamente preciso en sus predicciones, siendo una cantidad muy cercana a 1 o 100% su equivalencia en *Python*. Esto sugirió que el modelo aprendió patrones importantes en los datos de entrenamiento y logró generalizar bien en los datos de prueba. En otras palabras, el modelo fue capaz de predecir con alta precisión la respuesta de los encuestados a la pregunta P10 basado en las características de estos. Por lo tanto, se concluyó que el modelo resultó muy útil para predecir las respuestas a la pregunta P10; a continuación, se presenta el árbol de decisiones que a partir de esta primera prueba (ver figura 73).

Figura 73.

Árbol de decisión generado para P10



Nota: Elaboración propia.

Para la segunda prueba se evaluó la precisión del modelo de árbol de decisión en la clasificación conforme a la segunda pregunta raíz que es P12: 12. La accesibilidad a la red en los espacios universitarios es inestable, donde se inicializó *Python* con la importación de las librerías *pandas* necesarias así como *sklearn* y *graphviz*, como se mencionó anteriormente, *pandas* es una biblioteca de *Python* utilizada para la

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

manipulación y análisis de datos, por su parte, *sklearn* es una biblioteca de aprendizaje automático que proporciona una amplia variedad de algoritmos para resolver problemas de clasificación, regresión, agrupamiento, etc. *Graphviz* es una herramienta para visualizar gráficos y diagramas. Juntas estas bibliotecas fueron utilizadas para cargar y procesar los datos, entrenar los modelos de aprendizaje automático y visualizar los resultados en forma de árboles de decisión (ver figura 74).

Figura 74.

Importación de librerías Python – P12

```
In [1]: import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
import graphviz
```

Nota: Elaboración propia.

A continuación, se prosiguió con la definición de la variable X como todas las columnas del *Dataframe* df excepto la columna "P12", que era la variable objetivo a predecir. La variable y se definió como la columna "P12" (ver figura 75).

Figura 75.

Carga de datos para el modelo – P12

```
In [2]: # Cargar los datos del archivo Excel
df = pd.read_excel('numeric.xlsx')

# Definir las características (features) y la variable objetivo (target)
X = df.drop('P12', axis=1)
y = df['P12']
```

Nota: Elaboración propia.

Por otra parte, como se puede observar en la figura 76, se dividieron los datos en dos conjuntos: un conjunto de entrenamiento (70%) y un conjunto de prueba (30%).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Se definió el modelo de árbol de decisión con algunas opciones de configuración: *random_state* que fue una semilla aleatoria para reproducibilidad y *max_depth* como la profundidad máxima del árbol. El parámetro *test_size=0.3* indicó que el 30% de los datos fueron utilizados para prueba y el 70% restante para entrenamiento. La línea de código que se muestra en la figura 76 (*model.fit*), ajustó el modelo de árbol de decisión utilizando los datos de entrenamiento, en ella se utilizaron los datos de prueba para hacer predicciones conforme al modelo de árbol de decisión para P12 (ver figura 76).

Figura 76.

Definición del árbol de decisión – P12

```
In [3]: # Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

# Definir el modelo del árbol de decisión
model = DecisionTreeClassifier(random_state=42, max_depth=3)

# Ajustar el modelo con los datos de entrenamiento
model.fit(X_train, y_train)

Out[3]:
DecisionTreeClassifier
DecisionTreeClassifier(max_depth=3, random_state=42)
```

Nota: Elaboración propia.

Finalmente, se calculó la precisión del modelo comparando las predicciones realizadas en el conjunto de prueba con los valores reales de la variable objetivo. La precisión del modelo indicó qué tan bien se desempeñó el modelo en la tarea de clasificación. En estos casos de árboles de decisión, la precisión es importante porque permite saber cuánto se puede confiar en las predicciones realizadas por el modelo.

Una precisión de 94.51%, como se puede observar en la figura 77, significó que el modelo fue bastante preciso en sus predicciones. Esto sugirió que el modelo logró aprender patrones importantes en los datos de entrenamiento y generalizó bien en los

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

datos de prueba. En otras palabras, el modelo fue capaz de predecir con alta precisión la respuesta de los profesores a la pregunta P12 basado en sus características (ver figura 77).

Figura 77.

Cálculo de la precisión del modelo – P12

```
In [4]: # Realizar predicciones en el conjunto de prueba
y_pred = model.predict(X_test)

# Calcular la precisión del modelo
accuracy = metrics.accuracy_score(y_test, y_pred)
print("Precisión:", accuracy)

Precisión: 0.9451612903225806
```

Nota: Elaboración propia.

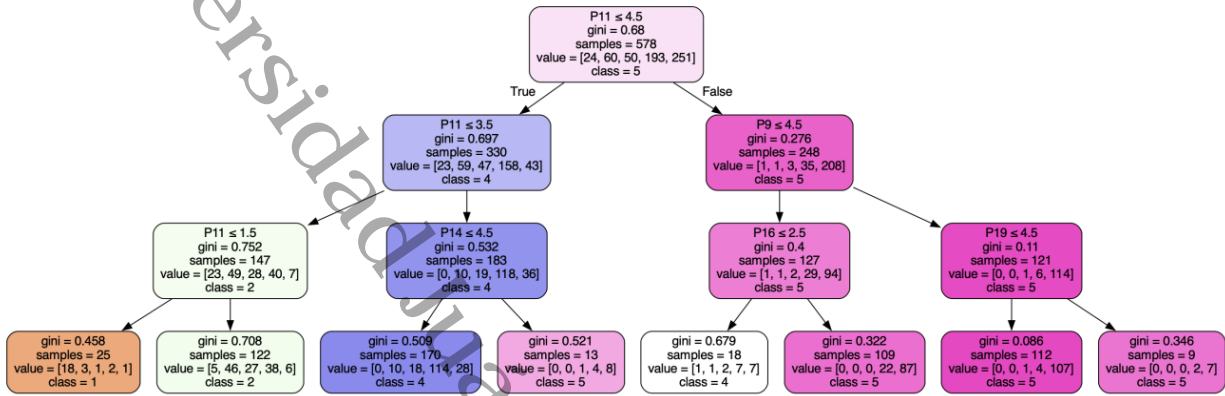
En conclusión, el modelo de árbol de decisiones utilizado demostró ser útil y altamente preciso en la predicción de las respuestas a la pregunta P12 en la población de encuestados. Los resultados obtenidos se visualizan en el árbol de decisiones generado durante la segunda prueba, donde la variable raíz P12 juega un papel fundamental en la estructura del árbol.

Este modelo proporciona una representación clara y fácilmente interpretable de los patrones y relaciones existentes en los datos. Estos hallazgos resaltan la capacidad del modelo de árbol de decisiones para ofrecer *insights* valiosos y apoyar la toma de decisiones informadas en futuras investigaciones y aplicaciones.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 78.

Árbol de decisión generado para P12



Nota: Elaboración propia.

En esta tercera y última prueba, se procedió de la misma manera que las anteriores, donde se obtuvo un árbol de decisión, así como la evaluación de la precisión del modelo conforme a la tercera pregunta raíz denominada P19: He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente; para comenzar se inicializó en *Python* con la importación de las librerías *pandas* necesarias, así como *sklearn* y *graphviz*, nuevamente (ver figura 79).

Figura 79.

Importación de librerías Python – P19

```
In [1]: import pandas as pd
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.tree import export_graphviz
import graphviz
```

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

A continuación, se prosiguió con la definición de la variable X como todas las columnas del *Dataframe* df excepto la columna "P19", que era la variable objetivo a predecir. La variable y se definió como la columna "P19" (ver figura 80).

Figura 80.

Carga de datos para el modelo – P19

```
In [2]: # Cargar los datos del archivo Excel
df = pd.read_excel('numeric.xlsx')

# Definir las características (features) y la variable objetivo (target)
X = df.drop('P19', axis=1)
y = df['P19']
```

Nota: Elaboración propia.

Para llevar a cabo la evaluación del modelo de árbol de decisión, se procedió a dividir los datos en dos conjuntos distintos: un conjunto de entrenamiento y un conjunto de prueba. Se configuraron opciones específicas del modelo, como el *random_state*, que se utilizó como una semilla aleatoria para garantizar la reproducibilidad de los resultados, y el *max_depth*, que estableció la profundidad máxima del árbol.

En la implementación del código, se asignó el 70% de los datos al conjunto de entrenamiento y el 30% restante al conjunto de prueba. El parámetro *test_size=0.3* indicó que se reservó el 30% de los datos para las pruebas, mientras que el 70% restante se utilizó para entrenar el modelo. La línea de código resaltada en la figura 81 (*model.fit*) ajustó el modelo de árbol de decisión utilizando los datos de entrenamiento, y posteriormente se utilizaron los datos de prueba para realizar predicciones basadas en dicho modelo (ver figura 81). Este enfoque permitió evaluar el rendimiento del modelo al comparar las predicciones con los valores reales de los datos de prueba.

Figura 81.*Definición del árbol de decisión – P19*

```
In [3]: # Dividir los datos en conjuntos de entrenamiento y prueba
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1)

# Definir el modelo del árbol de decisión
model = DecisionTreeClassifier(random_state=42, max_depth=3)

# Ajustar el modelo con los datos de entrenamiento
model.fit(X_train, y_train)

Out[3]: DecisionTreeClassifier
DecisionTreeClassifier(max_depth=3, random_state=42)
```

Nota: Elaboración propia.

Para evaluar la precisión del tercer modelo, se compararon las predicciones realizadas en el conjunto de prueba con los valores reales de la variable objetivo. En el caso de los árboles de decisión, la precisión es un indicador crucial, ya que refleja la confiabilidad de las predicciones del modelo.

Una alta precisión implica que el modelo es capaz de realizar predicciones precisas, mientras que una baja precisión indica que el modelo puede ser menos confiable y podría requerir ajustes o mejoras.

En este contexto, se obtuvo una precisión del 96.53%, como se muestra en la figura 82. Esta cifra demuestra que el modelo fue altamente preciso en sus predicciones. Sugiere que el modelo logró capturar patrones importantes en los datos de entrenamiento y pudo generalizar de manera efectiva en los datos de prueba (ver figura 82). Esta alta precisión respalda la capacidad del modelo de tomar decisiones acertadas y proporciona confianza en su capacidad para realizar predicciones precisas en situaciones similares.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 82.

Cálculo de la precisión del modelo – P19

```
In [4]: # Realizar predicciones en el conjunto de prueba
y_pred = model.predict(X_test)

# Calcular la precisión del modelo
accuracy = metrics.accuracy_score(y_test, y_pred)
print("Precisión:", accuracy)

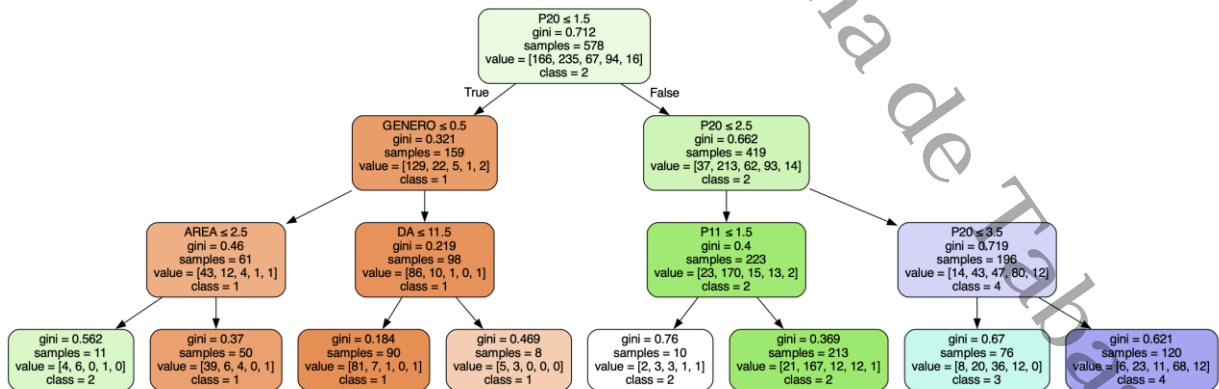
Precisión: 0.9653225806451613
```

Nota: Elaboración propia.

En otras palabras, el modelo obtenido era capaz de predecir con alta precisión la respuesta de los profesores a la pregunta P19 basado en las características de estos. Por lo tanto, se concluyó que el modelo resultó útil para predecir las respuestas a la pregunta P19 en la población de encuestados de la que se obtuvieron los datos; a continuación, se presenta el árbol de decisiones que se obtuvo a partir de esta tercera prueba (ver figura 83).

Figura 83.

Árbol de decisión generado para P19



Nota: Elaboración propia.

Agrupación

El modelo de *Kmeans* es un algoritmo de *clustering* utilizado para agrupar datos similares en diferentes *clústeres*. Una vez que los datos están agrupados en *clústeres*, se pueden visualizar los resultados utilizando diferentes herramientas, como diagramas de dispersión y tablas de frecuencia.

En la presente investigación se realizaron tres pruebas de *clustering* utilizando el algoritmo *Kmeans* en *Python*, con el objetivo de identificar las barreras ante las innovaciones tecnológicas que presentaron los profesores de la UJAT. Para ello, se probaron diferentes cantidades de *clústeres*: dos, tres y cinco. La elección de estas cantidades se basó en diversos factores, tales como la complejidad de los datos y la necesidad de encontrar patrones claros en el análisis.

Se consideró que dos *clústeres* podrían proporcionar una visión general de los datos y una clara separación entre dos grupos, mientras que tres *clústeres* podrían ser más útiles para identificar patrones intermedios o relaciones más complejas. Por otro lado, se eligió probar con cinco *clústeres* para explorar aún más la complejidad de los datos y encontrar patrones más específicos, por ello se concluyó que estas tres cantidades de *clústeres* proporcionarían información valiosa para el análisis.

Una de las pruebas realizadas con *Kmeans* en *Python* fue la selección del número óptimo de *clústeres*. Para ello, se utilizó la técnica del "codo" o "*elbow*", la cual implicó la ejecución del algoritmo *Kmeans* varias veces con diferentes números de *clústeres* y se graficó la suma de las distancias cuadradas de cada punto a su *clúster* más cercano en función del número de *clústeres*. El punto donde se produjo una disminución

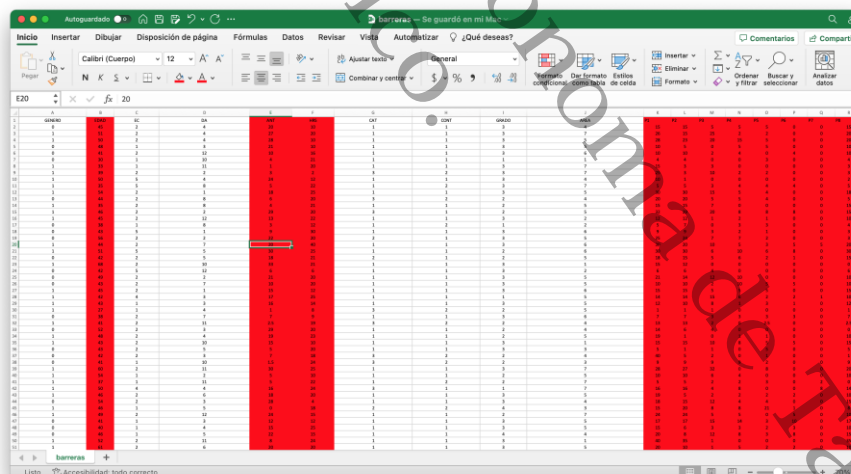
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

significativamente menor en la suma de las distancias cuadradas se consideró como el número óptimo de *clústeres*.

Antes de comenzar la programación, es necesario destacar que el modelo *Kmeans* trabaja con datos numéricos porque utiliza la distancia Euclidiana para calcular la similitud entre los puntos de datos. La distancia Euclidiana es una medida de distancia que solo tiene sentido cuando los datos son numéricos. En otras palabras, el algoritmo *Kmeans* se basa en la distancia entre los puntos de datos para agruparlos en *clústeres*. Por lo tanto, se realizó la normalización de los datos a valores numéricos a todo el *dataset*, resultando de la siguiente manera, así como las diversas modificaciones que sufrió el *dataset* en esta herramienta de minería de datos (ver figura 84).

Figura 84.

Modificaciones al dataset y normalización



Nota: Elaboración propia.

A continuación, se enlistan los cambios del *dataset* original con los 826 registros, para su lectura en la herramienta *Python*:

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Género, contando con dos opciones de hombre y mujer como respuestas, se modificó a los valores de 0 y 1 respectivamente.
- Estado civil, que contenía seis posibles respuestas: soltero, casado, viudo, divorciado, unión libre y separado, los valores asignados para estos datos fueron 0, 1, 2, 3, 4, y 5 respectivamente.
- División académica, conformada por doce posibles respuestas: Ciencias Agropecuarias, Ciencias Económico-Administrativas, Educación y Artes, Ciencias y Tecnologías de la Información, Ciencias de la Salud, Ciencias Sociales y Humanidades, Multidisciplinaria de los Ríos, Multidisciplinaria de Jalpa de Méndez, Multidisciplinaria de Comalcalco, Ciencias Básicas, Ingeniería y Arquitectura y Ciencias Biológicas, los valores asignados partieron de 0 a 11, respectivamente.
- Categorías de contratación en UJAT, pregunta que contó con las opciones de Profesor de tiempo completo, Profesor de medio tiempo y Profesor de asignatura, modificados por 0, 1 y 2.
- Tipo de contratación en UJAT, con las opciones de Base e Interino, fue modificado por 0 y 1, respectivamente.
- Grado académico, conformado por las opciones de Licenciatura, Maestría, Especialidad y Doctorado, suplantados por los valores 0, 1, 2 y 3.
- Área de formación de su último grado académico, compuesto de las opciones de: I. Física, Matemáticas y Ciencias de la Tierra, II. Biología y Química, III. Medicina y Salud, IV. Humanidades y de la Conducta, V. Ciencias Sociales y Económicas, VI.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Biocnología y Agropecuarias y VII. Ingeniería e Industria, cambiados por los valores de 0 a 6.

- Preguntas 1 a 20, fueron modificadas en nombre por "PX", es decir, P en abreviación a pregunta y X representa al número consecuente de cada una.
- Las preguntas con respuestas de opción múltiple con los valores de Totalmente en desacuerdo, En desacuerdo, Ni de acuerdo ni en desacuerdo, De acuerdo y Totalmente de acuerdo, también fueron modificados por los números 0, 1, 2, 3 y 4 respectivamente.

Para esta investigación se utilizó *Python* en su forma nativa, es decir, código. Se inició con la importación de las librerías *numpy*, la cual es una biblioteca para trabajar con matrices y vectores multidimensionales, lo que la hace útil para la manipulación de datos científicos y matemáticos; *pandas* que es eficaz para la manipulación y análisis de datos; *matplotlib.pyplot* para la visualización de datos capaz de crear gráficos de alta calidad, como gráficos de líneas, de barras, de dispersión, histogramas, entre otros.

Por su parte *sklearn.cluster.Kmeans* es utilizada para proporcionar una implementación eficiente y escalable del algoritmo de *Kmeans* que es adecuada para grandes conjuntos de datos y *sklearn.preprocessing* que proporciona una amplia variedad de funciones para la normalización de datos. Estas librerías se muestran en la figura 85, donde se observa la implementación de ellas para comenzar la programación del modelo de clasificación *Kmeans* o agrupamiento (ver figura 85).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 85.

Importación de librerías en Python

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn import preprocessing

%matplotlib inline

profesores= pd.read_excel('barriers.xlsx')
```

Nota: Elaboración propia.

A continuación, se requirió comprobar que *Python* lograra leer el archivo sin errores y detectar valores nulos que pudiesen pasar desapercibidos anteriormente, para esto se ingresó la etiqueta *info*, con el nombre asignado al conjunto de datos dentro de la herramienta (ver figura 86).

Figura 86.

Conjunto de datos leídos en Python

```
In [3]: profesores.info()
0  GENERO  826 non-null  int64
1  EDAD    826 non-null  int64
2  EC      826 non-null  int64
3  DA      826 non-null  int64
4  ANT     826 non-null  float64
5  HRS     826 non-null  int64
6  CAT     826 non-null  int64
7  CONT    826 non-null  int64
8  GRADO   826 non-null  int64
9  AREA    826 non-null  int64
10 P1      826 non-null  float64
11 P2      826 non-null  float64
12 P3      826 non-null  float64
13 P4      826 non-null  float64
14 P5      826 non-null  float64
15 P6      826 non-null  float64
16 P7      826 non-null  int64
17 P8      826 non-null  float64
18 P9      826 non-null  int64
19 P10     826 non-null  int64
```

Nota: Elaboración propia.

Se le solicitó a *Python* que mostrara los encabezados de los datos, debido a que fueron procesados correctamente (ver figura 87).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 87.

Encabezado de dataset en Python

```
In [4]: profesores.head()
```

```
Out[4]:
```

	GENERO	EDAD	EC	DA	ANT	HRS	CAT	CONT	GRADO	AREA	...	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20
0	0	45	2	4	20.0	10	1	1	3	4	...	5	5	5	5	5	5	5	5	2	1
1	1	51	2	4	27.0	20	1	1	3	7	...	4	4	2	2	3	2	3	3	3	3
2	1	50	2	4	28.0	10	1	1	3	5	...	4	4	2	3	5	2	4	4	3	4
3	0	48	1	3	21.0	10	1	1	3	5	...	5	5	5	5	5	5	5	5	5	5
4	0	41	2	12	10.0	16	1	1	3	6	...	4	5	2	3	4	4	5	5	1	1

5 rows x 30 columns

Nota: Elaboración propia.

A continuación, se seleccionaron las columnas a implementar, donde se descartaron las que no eran necesarias para el análisis de los datos con la función *drop*, junto al valor de *axis 1*, indicando que se trataba de una columna (ver figura 88).

Figura 88.

Columnas implementadas en Python

```
In [5]: profesores_variables=profesores.drop(['EDAD', 'ANT', 'HRS', 'P1', 'P2', 'P3', 'P4', 'P5', 'P6', 'P7', 'P8'], axis=1)
```

```
In [6]: profesores_variables.describe()
```

```
Out[6]:
```

	DA	CAT	CONT	GRADO	AREA	P9	P10	P11	P12	P13	P14	P15	P16	P
count	100	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.0000
mean	183	1.802663	1.301453	2.423729	4.222760	3.539952	3.446731	3.949153	3.960048	2.981840	3.043584	3.352300	3.272397	3.6295
std	179	0.941201	0.459167	0.711545	1.663517	1.203622	1.179705	1.174406	1.142926	1.259499	1.226194	1.187041	1.171226	1.0757
min	100	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.0000
max	100	1.000000	1.000000	2.000000	3.000000	3.000000	3.000000	3.000000	4.000000	2.000000	2.000000	2.000000	2.000000	3.0000
min	100	1.000000	1.000000	2.000000	4.000000	4.000000	4.000000	4.000000	4.000000	3.000000	3.000000	4.000000	3.000000	4.0000
max	100	3.000000	2.000000	3.000000	5.000000	4.000000	4.000000	5.000000	5.000000	4.000000	4.000000	4.000000	4.000000	4.0000
min	100	3.000000	2.000000	4.000000	7.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.0000

Nota: Elaboración propia.

Esta información se almacenó en un objeto llamado *profesores_variables*, donde se mostraron los valores mínimo, máximo, promedio, desviación estándar y los cuantiles

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

de cada columna, datos que variaron de manera considerable entre ellos como en puede observar en la siguiente figura (ver figura 89).

Figura 89.

Valores de ítems en Python

```
In [6]: profesores_variables.describe()
```

DA	CAT	CONT	GRADO	AREA	P9	P10	P11	P12	P13	P14	P15	P16	P
100	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.0000
183	1.802663	1.301453	2.423729	4.222760	3.539952	3.446731	3.949153	3.960048	2.981840	3.043584	3.352300	3.272397	3.6295
179	0.941201	0.459167	0.711545	1.663517	1.203622	1.179705	1.174406	1.142926	1.259499	1.226194	1.187041	1.171226	1.0757
100	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.0000
100	1.000000	1.000000	2.000000	3.000000	3.000000	3.000000	3.000000	4.000000	2.000000	2.000000	2.000000	2.000000	3.0000
100	1.000000	1.000000	2.000000	4.000000	4.000000	4.000000	4.000000	4.000000	3.000000	3.000000	4.000000	3.000000	4.0000
100	3.000000	2.000000	3.000000	5.000000	4.000000	4.000000	5.000000	5.000000	4.000000	4.000000	4.000000	4.000000	4.0000
100	3.000000	2.000000	4.000000	7.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.0000

Nota: Elaboración propia.

Posteriormente, se realizó la normalización de los valores utilizando *Python*, tal como se muestra en la figura 90. Este proceso se basó en la resta de cada valor de la celda por el valor mínimo encontrado en su respectiva columna, dividido por la resta del máximo valor entre el mínimo valor de la columna. De esta manera, todos los valores fueron transformados a un rango de 0 a 1, asegurando que el valor mínimo sea igual a 0 y el valor máximo sea igual a 1. Los datos normalizados fueron almacenados en la variable "*profesores_norm*" (ver figura 90). Esta técnica de normalización es útil para poner los datos en una escala común y facilitar su comparación.

Al tener todos los valores en el rango de 0 a 1, se eliminan las diferencias de escala entre las variables, lo que permite un análisis más preciso y coherente de los datos. La

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

normalización también ayuda a evitar que variables con rangos numéricos diferentes dominen el análisis y afecten negativamente los resultados.

Figura 90.

Normalización de datos en Python

```
In [7]: profesores_norm=(profesores_variables-profesores_variables.min())/(profesores_variables.max()-profesores_variables.m
profesores_norm
```

```
Out [7]:
```

	GENERO	EC	DA	CAT	CONT	GRADO	AREA	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	
0	0.0	0.2	0.272727	0.0	0.0	0.666667	0.500000	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.25	0.00
1	1.0	0.2	0.272727	0.0	0.0	0.666667	1.000000	0.25	0.25	0.75	0.75	0.25	0.25	0.50	0.25	0.50	0.50	0.50	0.50	0.50
2	1.0	0.2	0.272727	0.0	0.0	0.666667	0.666667	0.75	0.50	0.75	0.75	0.25	0.50	1.00	0.25	0.75	0.75	0.50	0.75	0.75
3	0.0	0.0	0.181818	0.0	0.0	0.666667	0.666667	0.75	0.75	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	0.0	0.2	1.000000	0.0	0.0	0.666667	0.833333	0.25	0.75	0.75	1.00	0.25	0.50	0.75	0.75	1.00	1.00	0.00	0.00	0.00
...
821	0.0	0.2	0.909091	0.5	0.0	0.333333	1.000000	1.00	1.00	1.00	1.00	1.00	0.75	1.00	1.00	1.00	1.00	1.00	0.25	0.25
822	0.0	1.0	0.363636	1.0	1.0	0.333333	0.500000	0.00	0.00	0.50	1.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.00
823	1.0	0.2	0.454545	0.0	0.0	0.666667	0.500000	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.75	0.25
824	1.0	0.2	0.181818	0.0	0.0	0.666667	0.666667	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	1.00
825	1.0	0.2	0.181818	1.0	0.0	0.333333	1.000000	0.50	0.50	0.75	0.50	0.75	0.75	0.50	0.25	0.50	0.50	0.25	0.25	0.25

Nota: Elaboración propia.

Al aplicar la función describe, se mostraron los valores estadísticos ya normalizados, la función describe de pandas regresa estadísticas descriptivas incluyendo: media (valor de la media de los datos), mediana (valor de la diana de los datos), máx (valor máximo de los datos), mín (valor mínimo de los datos), std (estandarización de los datos) y conteos para una columna en particular de los datos, es necesario destacar que la función describe únicamente regresa los valores de estas estadísticas para las columnas numéricas. (ver figura 91).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 91.

Función describe en Python

```
In [8]: profesores_norm.describe()
```

Out[8]:

	GENERO	EC	DA	CAT	CONT	GRADO	AREA	P9	P10	P11	P12	P13
count	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000	826.000000
mean	0.532688	0.243099	0.512217	0.401332	0.301453	0.474576	0.537127	0.634988	0.611683	0.737288	0.740012	0.495460
std	0.499233	0.249474	0.309489	0.470601	0.459167	0.237182	0.277253	0.300905	0.294926	0.293602	0.285732	0.314875
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	0.000000	0.200000	0.181818	0.000000	0.000000	0.333333	0.333333	0.500000	0.500000	0.500000	0.750000	0.250000
50%	1.000000	0.200000	0.545455	0.000000	0.000000	0.333333	0.500000	0.750000	0.750000	0.750000	0.750000	0.500000
75%	1.000000	0.200000	0.727273	1.000000	1.000000	0.666667	0.666667	0.750000	0.750000	1.000000	1.000000	0.750000
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Nota: Elaboración propia.

Antes de crear los *clústeres* con el método *Kmeans*, se supuso que no se tenía un criterio para determinar la cantidad de *clústeres* a crear y que tampoco se sabía cuál sería la cantidad óptima de *clústeres* a formar, para lo que se implementó un método llamado codo de yambú, el cual consiste en ir creando diferentes cantidades de *clústeres* y calcular que tan similares son los individuos dentro de los mismos e ir plasmando esta información dentro de una gráfica; cabe destacar que mientras más similares sean estos, más distantes se encontrarán los *clústeres* que se formen.

Para medir esta distancia se utilizó la suma de los cuadrados dentro de cada grupo, también llamada *WCSS (Within-Cluster Sum of Squares)*, la cual es una métrica utilizada en técnicas de agrupamiento (*clustering*) para evaluar la calidad de los *clústeres* obtenidos. Representa la suma de los cuadrados de las distancias entre cada punto y el centroide de su respectivo *clúster*. En otras palabras, mide la dispersión interna de los puntos dentro de cada *clúster* (ver figura 92).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 92.

Cálculo WCSS en Python

```
In [9]: wcss = []  
  
for i in range(1,11):  
    kmeans = KMeans(n_clusters = i, max_iter = 300)  
    kmeans.fit(profesores_norm) #Aplico K-Means al dataset  
    wcss.append(kmeans.inertia_)  
  
In [10]: plt.plot(range(1, 11), wcss)  
plt.title("Codo de Jambu")  
plt.xlabel('Número de clústers')  
plt.ylabel('WCSS') #Es un indicador de qué tan similares son los individuos dentro de los clústers  
plt.show
```

Nota: Elaboración propia.

Lo primero que se realizó fue una lista vacía, a la que se le llamó WCSS, para ir almacenar los valores de WCSS que se calcularon; después se realizó un *loop* con el comando *for* para crear diferentes cantidades de agrupaciones, asignando los valores a la variable *i*, desde un rango que era desde 1 hasta 11, aunque solo se querían diez *clústers*, en el rango del *for* se colocó el número al que se quería llegar más 1, es decir, 11. Después se le solicitó que creará un modelo con el método de *Kmeans*, así como la cantidad de *clústeres* que cambiarían conforme al valor de *i*.

A continuación, se aplicó el modelo *Kmeans* al objeto en donde se guardaron las características normalizadas de los profesores, con la función *fit*. Una vez aplicado, se obtuvo el valor de WCSS el cual se encontró en el valor de *inertia_* dentro del modelo.

En la siguiente figura se puede observar que el valor de WCSS fue disminuyendo conforme iba aumentando el valor de *clústeres*, de manera que, al llegar a los 10 *clústeres*, el valor de WCSS ya es muy pequeño en comparación al que se tenía con un solo *clúster*, lo que era bueno, sin embargo, se realizarían demasiadas particiones, por

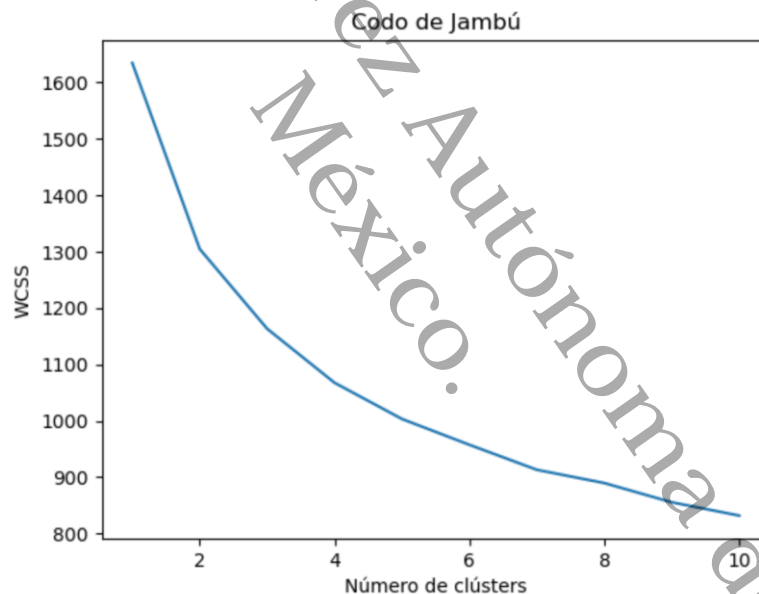
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

lo que se buscó un punto en el que el valor de WCSS dejara de disminuir de manera drástica, que en primera instancia fue el número 2 (ver figura 93).

Cuando ya se tuvieron los valores de WCSS, se utilizó la función *plot* para crear una gráfica bidimensional, cuyo eje de las x, tomó los valores del 1 al 10 y el eje de las y, tomó los valores contenidos en WCSS. El título del gráfico fue especificado con la función *title*, para cada eje se utilizó la función *xlabel* y *ylabel*, respectivamente.

Figura 93.

Gráfico de codo de Jambú



Nota: Elaboración propia.

Una vez conocida la primera cantidad de *clústeres* a formar, se aplicó la función de *Kmeans*, con un máximo de interacciones de 300, *Python* finalmente mostro los datos en una tabla definiendo los *clústeres* para cada registro (ver figura 94).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 94.

Tabla de clústeres en Python

```
In [18]: clustering = KMeans(n_clusters = 2, max_iter = 300) #Crea el modelo
clustering.fit(df_norm) #Aplica el modelo al dataset

Out[18]:
KMeans
KMeans(n_clusters=2)

In [13]: df['KMeans_Clusters'] = clustering.labels_ #Los resultados del clustering se guardan en labels dentro del modelo
df.head()

Out[13]:
```

	Género	P20	P22	P27	P29	P30	KMeans_Clusters
0	Mujer	5	5	5	2	5	1
1	Hombre	2	4	3	3	5	0
2	Hombre	3	4	4	3	3	1
3	Mujer	4	5	5	5	2	1
4	Mujer	4	5	5	1	2	1

Nota: Elaboración propia.

Para visualizar de manera gráfica los *clústeres*, se empleó la técnica de Análisis de Componentes Principales (PCA, por sus siglas en inglés) debido a la alta dimensionalidad de los datos que comprendían 20 preguntas. El PCA es una técnica que reduce la cantidad de variables originales creando un conjunto más pequeño de nuevas variables, conocidas como componentes principales, que capturan la mayor variabilidad de los datos.

En este caso, se utilizó la función *pca* del paquete *sklearn* en *Python*, indicando que se querían obtener 2 componentes principales. Estos componentes se agregaron al modelo previamente creado, lo que permitió representar los *clústeres* de manera visual en un gráfico bidimensional (ver figura 95). Esta representación reducida de los datos facilitó la interpretación y la identificación de patrones o relaciones entre los *clústeres*.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 95.

Técnica PCA en Python

```
In [15]: from sklearn.decomposition import PCA
pca = PCA(n_components=2)
pca_df = pca.fit_transform(df_norm)
pca_df_df = pd.DataFrame(data = pca_df, columns = ['C1','C2'])
pca_respuestas_df = pd.concat([pca_df_df, df[['KMeans_Clusters']]], axis=1)
pca_respuestas_df
```

Out [15]:

	C1	C2	KMeans_Clusters
0	-0.609393	0.294970	1
1	0.261197	0.504032	0
2	0.016535	0.131278	1
3	-0.393927	0.269649	1
4	-0.394048	-0.419576	1
...
821	-0.609393	0.294970	1
822	0.605272	-0.546258	0
823	0.290397	0.332085	0
824	0.290367	0.159779	0
825	0.310007	-0.192522	0

826 rows x 3 columns

Nota: Elaboración propia.

Para visualizar los resultados de la primera prueba, se utilizó la biblioteca *Matplotlib* en *Python*. Se creó una figura con una cuadrícula de 7x7 utilizando la función *figure*. Dentro de esta figura, se generó un único gráfico ubicado en la posición (1,1,1) de la cuadrícula. Se asignó el título "Resultados encontrados" al gráfico y se etiquetaron los ejes "x" e "y". Los *clústeres* se representaron utilizando los colores rosa y azul, respectivamente.

Esta visualización permitió observar la distribución y separación de los datos en los dos *clústeres* (ver figura 96).

Figura 96.**Código de gráfico Kmeans en Python**

```
In [19]: # Configuración de los parámetros del gráfico
fig = plt.figure(figsize = (6,6))
ax = fig.add_subplot(1,1,1)
ax.set_xlabel('y', fontsize = 15)
ax.set_ylabel('x', fontsize = 15)
ax.set_title('Resultados encontrados', fontsize = 15)

# Definición de los colores para cada cluster
color_theme = np.array(["pink", "blue"])

# Ejecución del algoritmo KMeans con 2 clusters
kmeans = KMeans(n_clusters=2)
kmeans.fit(pca_respuestas_df.iloc[:, 0:2])

# Asignación de cada punto a su cluster correspondiente
pca_respuestas_df["KMeans_Clusters"] = kmeans.labels_

# Graficación de los puntos con diferentes colores según el cluster
ax.scatter(x = pca_respuestas_df.C1, y = pca_respuestas_df.C2,
          c=color_theme[pca_respuestas_df.KMeans_Clusters], s = 50)
plt.show()
```

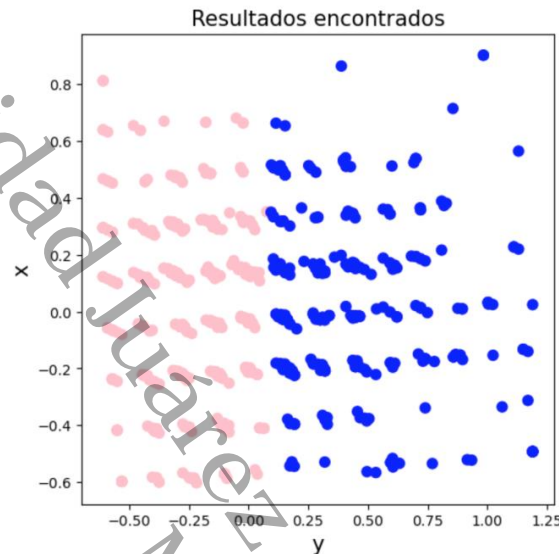
Nota: Elaboración propia.

Utilizando *Python*, se pudo generar una visualización gráfica de los *clústeres* obtenidos. Se empleó un gráfico de dispersión bidimensional que mostraba los grupos identificados mediante sus etiquetas de *clúster* asignadas. Esta representación permitió una fácil interpretación de las relaciones entre los puntos y facilitó la detección de patrones y tendencias en los datos.

Como resultado, se obtuvo un gráfico que mostraba los dos *clústeres* generados durante la primera prueba de *Kmeans* en *Python* (ver figura 97).

Figura 97.

Gráfico Kmeans obtenido con dos clústeres



Nota: Elaboración propia.

Sin embargo, es necesario comprobar la calidad de los *clústeres* obtenidos, esto se determinó mediante una matriz de confusión. La matriz de confusión compara los valores reales de las etiquetas de cada muestra con las etiquetas asignadas por el algoritmo de *clustering* y muestra la cantidad de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. La precisión del *clustering* se puede evaluar a partir de estas medidas, esto realizó mediante el código de *confusion_matrix* en *Python*, a partir del PCA calculado con anterioridad.

La matriz de confusión que se observa en la figura 98, era de un formato de 2x2 y mostró los resultados del modelo para las dos clases diferentes o *clústeres*. Cada fila representó las instancias en una clase real, mientras que cada columna representó las

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

instancias en una clase predicha. En este caso, la matriz de confusión mostró que el *clustering* agrupó correctamente 339 instancias en el primer grupo y 487 instancias en el segundo grupo. Como no hay valores en la diagonal secundaria de la matriz de confusión, esto significó que no hubo falsos positivos o falsos negativos, es decir, todas las instancias fueron clasificadas correctamente (ver figura 98).

Figura 98.

Matriz de confusión para los dos clústeres

```
In [25]: from sklearn.metrics import confusion_matrix

# Ejecución del algoritmo KMeans con 2 clusters
kmeans = KMeans(n_clusters=2)
kmeans.fit(pca_respuestas_df.iloc[:, 0:2])

# Asignación de cada punto a su cluster correspondiente
pca_respuestas_df["KMeans_Clusters"] = kmeans.labels_

# Creación de la matriz de confusión
cm = confusion_matrix(pca_respuestas_df["KMeans_Clusters"], pca_respuestas_df["KMeans_Clusters"])
print(cm)

[[ 339  0]
 [  0 487]]
```

Nota: Elaboración propia.

Para la segunda prueba se utilizó el valor de 3 como el número de *clústeres* a mostrar, donde los datos se comportaron de manera similar a la prueba anterior, sin embargo, se observó que, a diferencia de aquellos, estos se encontraron más cercanos.

Inicialmente se creó una figura de 7x7 con la función `figure` de *matplotlib*. Se le indicó que se crearía un solo gráfico dentro de la figura con los parámetros (1,1,1), a continuación, se especificó el título como 'Resultados encontrados', etiquetas (ejes 'x' y 'y') y colores de los *clústeres* que fueron rojo, azul y verde, respectivamente (ver figura 99).

Figura 99.**Código de gráfico Kmeans en Python**

```
In [26]: # Configuración de los parámetros del gráfico
fig = plt.figure(figsize = (6,6))
ax = fig.add_subplot(1,1,1)
ax.set_xlabel('y', fontsize = 15)
ax.set_ylabel('x', fontsize = 15)
ax.set_title('Resultados encontrados', fontsize = 15)

# Definición de los colores para cada cluster
color_theme = np.array(["red", "blue", "green"])

# Ejecución del algoritmo KMeans con 3 clusters
kmeans = KMeans(n_clusters=3)
kmeans.fit(pca_respuestas_df.iloc[:, 0:2])

# Asignación de cada punto a su cluster correspondiente
pca_respuestas_df["KMeans_Clusters"] = kmeans.labels_

# Graficación de los puntos con diferentes colores según el cluster
ax.scatter(x = pca_respuestas_df.C1, y = pca_respuestas_df.C2,
           c=color_theme[pca_respuestas_df.KMeans_Clusters], s = 50)
plt.show()
```

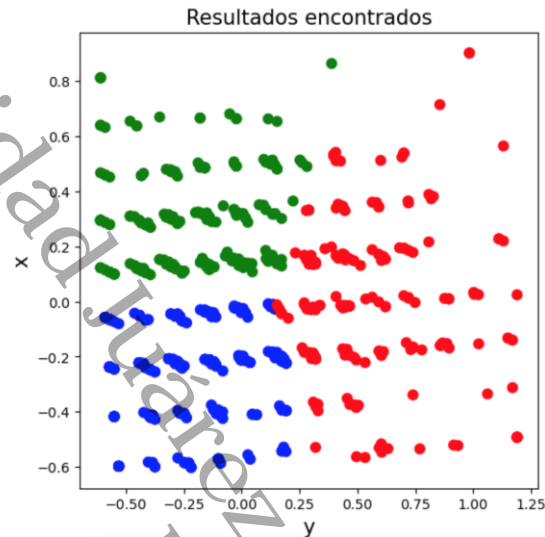
Nota: Elaboración propia.

Para obtener una visualización de los resultados del *clustering* con tres *clústeres*, como se puede observar en la figura 100, se utilizó un gráfico de dispersión en dos dimensiones, donde cada punto representaba una muestra y su color indicaba a qué *clúster* pertenecía. Esto permitió visualizar la separación entre los *clústeres* y la distribución de las muestras dentro de cada uno.

Finalmente, se obtuvo el gráfico de los tres *clústeres* utilizando la misma técnica que se utilizó para los dos *clústeres*. Es importante destacar que la elección del número de *clústeres* adecuado dependió de la naturaleza de los datos y de los objetivos de la investigación, por lo que se realizaron estas tres pruebas con diferentes valores de *k* para encontrar el número óptimo de *clústeres*, explicados con anterioridad. Finalmente, se obtuvo el gráfico de los tres *clústeres* utilizados para la primera prueba de *Kmeans* en *Python* (ver figura 100).

Figura 100.

Gráfico de Kmeans obtenido con tres clústeres



Nota: Elaboración propia.

Era necesario evaluar la calidad de los *clústeres* obtenidos, es por lo que también se implementó la matriz de confusión para esta segunda prueba. La matriz de confusión compara los valores reales de las etiquetas de cada muestra con las etiquetas asignadas por el algoritmo de *clustering* y muestra la cantidad de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. La precisión del *clustering* se puede evaluar a partir de estas medidas.

La matriz de confusión obtenida, como se puede observar en la figura 101, mostró que el *clustering* dividió los datos en tres *clústeres* diferentes. El primer *clúster* con 232 muestras, el segundo *clúster* contenía 219 muestras y el tercer *clúster* con 288 muestras.

La diagonal principal de la matriz de confusión mostró los verdaderos positivos para cada *clúster*, es decir, la cantidad de muestras que fueron asignadas correctamente

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

al *clúster* correspondiente. Los valores fuera de la diagonal principal indicaron los falsos positivos o negativos, es decir, la cantidad de muestras que fueron asignadas incorrectamente a un *clúster* diferente.

En este caso, el primer *clúster* contenía 232 muestras y todas ellas fueron clasificadas correctamente. El segundo *clúster* contuvo 219 muestras y 39 de ellas fueron asignadas incorrectamente al primer *clúster* y 48 fueron asignadas incorrectamente al tercer *clúster*. Finalmente, el tercer *clúster* contiene 288 muestras y todas ellas fueron clasificadas correctamente (ver figura 101).

Figura 101.

Matriz de confusión para los tres clústeres

```
In [19]: from sklearn.metrics import confusion_matrix

# Ejecución del algoritmo KMeans con 3 clusters
kmeans = KMeans(n_clusters=3)
kmeans.fit(pca_respuestas_df.iloc[:, 0:2])

# Asignación de cada punto a su cluster correspondiente
pca_respuestas_df["KMeans_Clusters"] = kmeans.labels

# Creación de la matriz de confusión
cm = confusion_matrix(df["KMeans_Clusters"], pca_respuestas_df["KMeans_Clusters"])
print(cm)

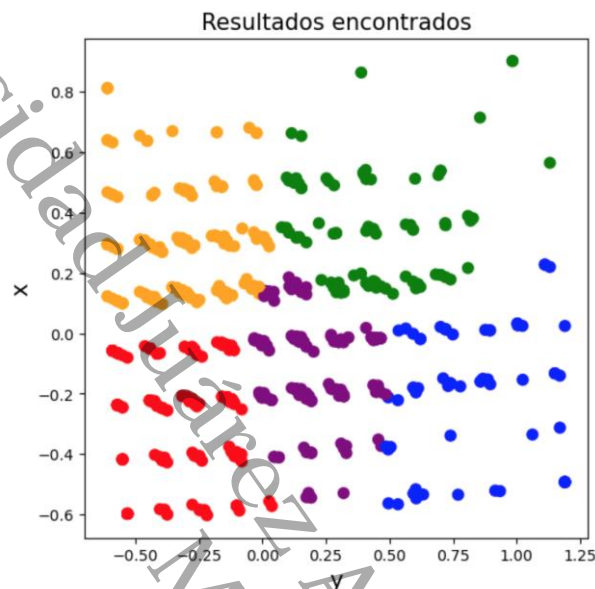
[[232  0 288]
 [ 39 219 48]
 [  0  0  0]]
```

Nota: Elaboración propia.

Para la tercera prueba se utilizó el valor de 5 para el número de *clústeres* a mostrar, donde los datos se comportaron de manera similar al gráfico anterior, sin embargo, se observó que, a diferencia de aquellos, estos se encontraban aún más cercanos (ver figura 102).

Figura 102.

Gráfico *Kmeans* obtenido con cinco clústeres



Nota: Elaboración propia.

La evaluación de la calidad del *clustering* es una tarea fundamental en el análisis de datos. Una de las técnicas más comunes para evaluar la precisión de los algoritmos de *clustering* es la matriz de confusión. Esta matriz compara los valores reales de las etiquetas de cada muestra con las etiquetas asignadas por el algoritmo de *clustering* y muestra la cantidad de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos.

En este contexto, en la presente investigación se realizó una tercera prueba de *clustering* utilizando el algoritmo *Kmeans* en *Python*, y se obtuvo una matriz de confusión que será analizada detalladamente. La matriz de confusión que se puede observar en la figura 103, es una tabla de cuatro filas y cuatro columnas. Cada fila representa las clases

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

reales de las muestras, mientras que cada columna representa las etiquetas de *clúster* asignadas por el algoritmo de *clustering*. Los números dentro de la matriz indican la cantidad de muestras asignadas a cada combinación de clase y etiqueta de *clúster*.

En la primera fila de la matriz, se puede observar que todas las muestras de la clase 1 (primer grupo) fueron asignadas al tercer *clúster*, excepto por una muestra que fue asignada al cuarto *clúster*. En la segunda fila, se puede observar que 100 muestras de la clase 2 (segundo grupo) fueron asignadas al primer *clúster*, 108 muestras al tercer *clúster* y 98 muestras al quinto *clúster*, mientras que ninguna muestra de esta clase fue asignada al segundo, cuarto o quinto *clúster*. La tercera y cuarta fila de la matriz están completamente compuestas de ceros, lo que indica que no se asignaron muestras a los *clústeres* correspondientes (ver figura 103).

Figura 103.

Matriz de confusión para los cinco *clústeres*

```
# Creación de la matriz de confusión
cm = confusion_matrix(df["KMeans_Clusters"], pca_respuestas_df["KMeans_Clusters"])
print(cm)

[[ 0 209 85 225 1]
 [100 0 108 0 98]
 [ 0 0 0 0 0]
 [ 0 0 0 0 0]
 [ 0 0 0 0 0]]
```

Nota: Elaboración propia.

En general, la matriz de confusión muestra que el algoritmo de *clustering* logró agrupar algunas muestras correctamente, pero también hubo muchas muestras que fueron asignadas a *clústeres* incorrectos. Por lo tanto, se puede inferir que la precisión del *clustering* no fue muy alta en este caso de 5.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Reglas de asociación

Las reglas de asociación son una técnica popular en minería de datos que permiten descubrir patrones interesantes en grandes conjuntos de datos. En particular, las reglas de asociación buscan encontrar relaciones frecuentes entre los diferentes elementos de un conjunto de datos, lo que puede proporcionar información valiosa para la toma de decisiones en diversas aplicaciones, como el marketing, la recomendación de productos y servicios, así como la optimización de procesos empresariales.

En este contexto, *Python* ofrece varias herramientas y bibliotecas que facilitan la implementación y el análisis de las reglas de asociación. En este sentido, en esta sección se presentarán tres pruebas realizadas al conjunto de datos de esta investigación, mediante la herramienta *Python*.

Python nativamente trabaja con código, por ello en primera instancia el código que se observa en la figura 104 fue empleado para cargar el archivo CSV que contenía los datos. El objetivo de este código era utilizar el algoritmo *Apriori* para encontrar patrones frecuentes de asociación entre los elementos del *dataset*. Para lograr esto, se utilizaron tres librerías de *Python*: *pandas*, *mlxtend.preprocessing* y *mlxtend.frequent_patterns*.

La librería llamada *pandas* es una herramienta para el análisis de datos que permite cargar y manipular archivos de diferentes formatos, incluyendo CSV. La librería *mlxtend.preprocessing* contiene funciones para preprocesar los datos, como la función *TransactionEncoder*, que se utiliza para convertir los datos en una matriz binaria donde cada fila representa un dato y cada columna representa un elemento de los datos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Se utilizó el nombre de transacciones para hacer las reglas de asociación debido a que las transacciones representan un conjunto de *ítems* que se adquieren juntos con cierta frecuencia, por lo que, al analizar las transacciones, es posible identificar patrones y asociaciones entre diferentes datos.

Además, el uso de transacciones y *TransactionEncoder* en *Python*, permiten identificar las reglas de asociación más relevantes y útiles para la toma de decisiones en el negocio. Por estas razones, se consideró que el nombre de transacciones o *transactions* en el código era apropiado para el análisis de las reglas de asociación de esta investigación.

El código de la figura 104 cargó los datos de un archivo CSV utilizando la librería *pandas*, convirtió los datos en una matriz binaria utilizando la función *TransactionEncoder* de la librería *mlxtend.preprocessing* y utilizó la función *apriori* de la librería *mlxtend.frequent_patterns* para encontrar los *ítems* frecuentes. Finalmente, se utilizó la función *association_rules* para generar las reglas de asociación a partir de los *ítems* frecuentes encontrados (ver figura 104).

Figura 104.

Importación de librerías y carga de datos en Python

```
In [1]: import pandas as pd
        from mlxtend.preprocessing import TransactionEncoder
        from mlxtend.frequent_patterns import apriori, association_rules

        # Cargar datos del archivo CSV
        df = pd.read_csv('datosaletras.csv', header=None)
```

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

A continuación, el siguiente código transformó los datos contenidos en un objeto *Dataframe* de *pandas* en una lista de listas. Este proceso se llevó a cabo utilizando dos bucles *for* anidados para iterar a través de cada elemento del *Dataframe*. Primero, se inicializó una lista vacía denominada "*transactions*" que almacenaría las reglas resultantes. Luego, se utilizó un bucle *for* para iterar a través de cada fila del *Dataframe*, donde "i" es el índice de la fila actual, como se observa en la figura 105.

Dentro del primer bucle, se inicializó una lista vacía denominada "*transaction*" que almacenaría los elementos de la regla actual. Luego, se utilizó un segundo bucle *for* para iterar a través de cada columna del *Dataframe*, donde "j" es el índice de la columna actual. Dentro del segundo bucle, se utilizó la función *pd.isna* para verificar si el valor en la posición actual es un valor faltante (*NaN*). Si el valor no era un valor faltante, se agregó a la lista "*transaction*". Al final del segundo bucle, se agregó la lista "*transaction*" a la lista "*transactions*" (ver figura 105).

Figura 105.

Transformación de los datos

```
In [2]: # Transformar datos a una lista de listas
transactions = []
for i in range(len(df)):
    transaction = []
    for j in range(len(df.columns)):
        if not pd.isna(df.values[i,j]):
            transaction.append(df.values[i,j])
    transactions.append(transaction)
```

Nota: Elaboración propia.

El siguiente código, que se puede observar en la figura 106, codificó los *items* de las transacciones en una matriz binaria utilizando la función *TransactionEncoder* de la

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

librería *mlxtend.preprocessing*. Este proceso se realizó en tres pasos. Se creó un objeto *Dataframe* de *pandas* utilizando la matriz binaria "te_ary". La función *Dataframe* convirtió la matriz binaria en una tabla legible por humanos y la etiquetó utilizando los nombres de los *ítems* originales. Los nombres de los *ítems* se obtuvieron utilizando la propiedad "columns_" del objeto *TransactionEncoder*. El resultado se almacenó en la variable "df".

Seguido de esto se utilizó la función *apriori* de la librería *mlxtend.frequent_patterns* para encontrar conjuntos de *ítems* frecuentes a partir de la tabla de datos "df". La función *apriori* tomó dos argumentos principales: "min_support", que establecieron el soporte mínimo que debía tener un conjunto de *ítems* para ser considerado frecuente, y "use_colnames", que indicó si se debían utilizar los nombres de los *ítems* originales en lugar de las columnas numeradas por la matriz binaria. El resultado se almacenó en la variable "frequent_itemsets".

Figura 106.

Configuración de soporte mínimo

```
In [3]: # Codificar ítems en binario
te = TransactionEncoder()
te_ary = te.fit_transform(transactions)
df = pd.DataFrame(te_ary, columns=te.columns_)

In [4]: # Encontrar conjuntos de ítems con soporte mínimo del 10%
frequent_itemsets = apriori(df, min_support=0.1, use_colnames=True)
```

Nota: Elaboración propia.

Una vez realizado lo anterior, se generaron las reglas de asociación a partir de los conjuntos de *ítems* frecuentes encontrados anteriormente mediante la función *apriori*. Este proceso se realizó utilizando la función *association_rules* de la librería *mlxtend.frequent_patterns*. El resultado se almacenó en la variable "rules".

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

La función `association_rules` se utilizó con dos argumentos principales. El primer argumento fue `"frequent_itemsets"`, que correspondía al objeto devuelto por la función `apriori` y contenía los conjuntos de `ítems` frecuentes encontrados en los datos. El segundo argumento fue `"metric"`, que se utilizó para especificar la medida utilizada para evaluar la calidad de las reglas de asociación. En este caso, se eligió la métrica `"confidence"`, que representa la probabilidad condicional de que la regla de asociación sea verdadera dado que el antecedente de la regla se cumple.

Además, se estableció un mínimo de confianza del 55% y un soporte mínimo del 10% utilizando el argumento `"min_threshold"`. Estos valores se utilizaron como criterios para seleccionar las reglas de asociación más relevantes y significativas. Al aplicar la función `association_rules` con estos parámetros, se obtuvieron las reglas de asociación que cumplían con los umbrales establecidos.

Figura 107.

Reglas de asociación generadas

```

In [5]: # Generar reglas de asociación con confianza del 55%
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.55)

In [6]: print(rules)

```

	antecedents \		
0	(CINCO)		
1	(CINCO)		
2	(CINCO)		
3	(CINCO)		
4	(CINCO)		
...			
9168	(PQUINCE De acuerdo, PDIECISIETE De acuerdo, B...		
9169	(PQUINCE De acuerdo, CERO, PDIECIOCHO De acuerdo)		
9170	(PQUINCE De acuerdo, Profesor de Tiempo Comple...		
9171	(PQUINCE De acuerdo, Profesor de Tiempo Comple...		
9172	(PQUINCE De acuerdo, Base, PDIECIOCHO De acuerdo)		
...			
	consequents	antecedent support \	
0	(DIEZ)	0.445755	
1	(Base)	0.445755	
2	(CER0)	0.445755	
3	(Casado)	0.445755	
4	(Hombre)	0.445755	
...			
9168	(Profesor de Tiempo Completo, CERO, PDIECIOCHO...	0.176887	
9169	(PDIECISIETE De acuerdo, Base, Profesor de Tie...	0.180425	
9170	(PDIECISIETE De acuerdo, PDIECIOCHO De acuerdo...	0.180425	
9171	(PDIECISIETE De acuerdo, CERO, Base)	0.143868	
9172	(PDIECISIETE De acuerdo, CERO, Profesor de Tie...	0.182783	

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Era necesario visualizar las reglas de asociación por completo, debido a que la pantalla de *Python* no las mostraba todas, por ello se decidió exportarlas en archivos PDF. El código mostrado en la figura 108 utilizó la librería FPDF para crear un documento PDF y agregar contenido al mismo, por ello se creó un objeto FPDF utilizando la clase FPDF y se almacenó en la variable "*pdf*". El constructor de la clase FPDF aceptaba varios argumentos opcionales para establecer la orientación, la unidad y el formato del documento PDF. En este caso, se utilizó la orientación horizontal (*landscape*), la unidad pulgada (*in*) y el formato de papel carta (*letter*).

Por último, se agregó una página al documento PDF utilizando el método *add_page()* del objeto "*pdf*". Luego, se estableció la fuente y el tamaño de letra para el título utilizando el método *set_font()*. El método *cell()* se utilizó para agregar texto a la página. Se agregó el texto "Reglas de asociación generadas" centrado en la página y se estableció un salto de línea con el argumento "*ln=1*". Luego se agregó una línea separadora de 50 guiones mediante el método *cell()* (ver figura 108).

Figura 108.

Generación del archivo PDF

```
In [7]: from fpdf import FPDF
        # Crear objeto FPDF
        pdf = FPDF()

In [8]: pdf = FPDF(orientation='L', unit='in', format='letter')
        # Agregar página y establecer fuente y tamaño de letra para el título
        pdf.add_page()
        pdf.set_font("Arial", size=16)
```

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

En esta sección se presenta la figura 109 que muestra las reglas de asociación obtenidas a partir de un análisis de datos utilizando *Python*, exportadas en un archivo PDF. En total, se obtuvieron 25,764 reglas de asociación, sin embargo, no todas son importantes o necesarias, por lo cual se prosiguió con distintas pruebas modificando los valores de confianza y soporte mínimos; si bien puede parecer que obtener una gran cantidad de reglas de asociación proporciona una visión más completa de los datos, esto también puede llevar a una sobrecarga de información y dificultad para identificar los verdaderos patrones importantes (ver figura 109).

Figura 109.

Reglas de asociación con 55% confianza y 10% SM

Regla 284: PDIECISIETE Totalmente de acuerdo, nan, CERO -> PDIECIOCHO Totalmente de acuerdo, PQUINCE Totalmente de acuerdo
Soporte: 0.11, Confianza: 0.57, Lift: 4.87
Regla 285: PDIECISIETE Totalmente de acuerdo, CERO -> PDIECIOCHO Totalmente de acuerdo, nan, PQUINCE Totalmente de acuerdo
Soporte: 0.11, Confianza: 0.57, Lift: 4.87
Regla 286: PDIECIOCHO Totalmente de acuerdo, nan, PQUINCE Totalmente de acuerdo -> PDIECISIETE Totalmente de acuerdo, CERO
Soporte: 0.11, Confianza: 0.92, Lift: 4.87
Regla 287: PDIECISIETE Totalmente de acuerdo, CERO -> PDIECIOCHO Totalmente de acuerdo, PQUINCE Totalmente de acuerdo
Soporte: 0.11, Confianza: 0.57, Lift: 4.87
Regla 288: PDIECIOCHO Totalmente de acuerdo, PQUINCE Totalmente de acuerdo -> PDIECISIETE Totalmente de acuerdo, nan, CERO
Soporte: 0.11, Confianza: 0.92, Lift: 4.87
Regla 289: PDOCE Totalmente de acuerdo, nan, PDIECISIETE Totalmente de acuerdo -> PDIECIOCHO Totalmente de acuerdo, Profesor de Tiempo Completo, CERO

Nota: Elaboración propia.

A continuación, se presenta la segunda prueba, con código que se presenta en la figura 110 el cual fue utilizado para cargar el archivo CSV que contenía los datos en forma de transacciones, donde cada transacción representó una lista de elementos. El objetivo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

de este código era utilizar el algoritmo *Apriori* para encontrar patrones frecuentes de asociación entre los elementos de las transacciones, es decir, los datos de los profesores.

Para lograr esto, se utilizan tres librerías de *Python*: *pandas*, *mlxtend.preprocessing* y *mlxtend.frequent_patterns*.

El código cargó los datos de un archivo CSV utilizando la librería *pandas*, convirtió los datos en una matriz binaria utilizando la función *TransactionEncoder* de la librería *mlxtend.preprocessing* y utilizó la función *apriori* de la librería *mlxtend.frequent_patterns* para encontrar los *ítems* frecuentes. Finalmente, se utilizó la función *association_rules* para generar las reglas de asociación a partir de los *ítems* frecuentes encontrados (ver figura 110).

Figura 110.

Importación de librerías y carga de datos en Python

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules
from fpdf import FPDF

# Cargar datos desde el archivo CSV y convertirlos en una lista de transacciones
data = pd.read_csv('datosaletras.csv', header=None)
transactions = []
for i in range(data.shape[0]):
    transactions.append([str(data.values[i,j]) for j in range(data.shape[1])])
```

Nota: Elaboración propia.

El código que se utilizó a continuación se puede observar en la figura 111, este transformó los datos contenidos en un objeto *Dataframe* de *pandas* en una lista de listas, donde cada lista representaba una transacción. Este proceso se llevó a cabo utilizando dos bucles *for* anidados para iterar a través de cada elemento del *Dataframe*.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Dentro del primer bucle, se inicializó una lista vacía denominada *"transaction"* que almacenaría los elementos de la transacción actual. Luego, se utilizó un segundo bucle *for* para iterar a través de cada columna del *Dataframe*, donde *"j"* era el índice de la columna actual. Dentro del segundo bucle, se utilizó la función *pd.isna* para verificar si el valor en la posición actual es un valor faltante (*NaN*). Si el valor no era un valor faltante, se agregaría a la lista *"transaction"*. Al final del segundo bucle, se agregó la lista *"transaction"* a la lista *"transactions"*, así mismo se declaró que se utilizaría un 70% de confianza mínima con un 15% de soporte mínimo (ver figura 111).

Figura 111.

Configuración de algoritmo *apriori* y soporte mínimo

```
# Aplicar algoritmo Apriori para generar conjuntos de ítems frecuentes
frequent_itemsets = apriori(data, min_support=0.15, use_colnames=True)

# Generar reglas de asociación con una confianza del x%
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.70)

# Filtrar reglas de asociación por lift mayor a 1 y ordenarlas de manera descendente
rules_lift = rules[rules['lift'] > 1].sort_values(by='lift', ascending=False)
```

Nota: Elaboración propia.

Para poder visualizar todas las reglas de asociación obtenidas en *Python*, se decidió exportarlas en archivos PDF ya que no se podían mostrar todas en la pantalla. Para lograr esto, se utilizó la librería *FPDF* que permitió crear un documento PDF y agregar contenido al mismo. Se creó un objeto *FPDF* a partir de la clase *FPDF*, el cual se almacenó en la variable *"pdf"*. El constructor de la clase *FPDF* permitió establecer varios argumentos opcionales como la orientación, la unidad y el formato del documento

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

PDF. En este caso, se seleccionó la orientación horizontal (*landscape*), la unidad pulgada (*in*) y el formato de papel carta (*letter*).

Posteriormente, se agregó una página al documento PDF utilizando el método `add_page()` del objeto `"pdf"`. Se estableció la fuente y el tamaño de letra para el título mediante el método `set_font()`. Luego, se utilizó el método `cell()` para agregar texto a la página, el cual mostró el título "Reglas de asociación generadas" centrado en la página y se estableció un salto de línea con el argumento `"ln=1"`. Finalmente, se agregó una línea separadora de 50 guiones mediante el método `cell()` (ver figura 112).

Figura 112.

Generación del archivo PDF

```
# Crear objeto FPDF
pdf = FPDF()

# Agregar página y establecer fuente y tamaño de letra para el título
pdf.add_page(orientation='L')
pdf.set_font("Arial", size=12)

# Agregar título y línea separadora
pdf.cell(270, 10, txt="Reglas de asociación generadas", ln=1, align="C")
pdf.cell(270, 10, txt="--*20", ln=1, align="C")
```

Nota: Elaboración propia.

A continuación, se muestra el resultado del análisis de datos utilizando *Python* (ver figura 113). En este análisis, se generaron un total de 2,319 reglas de asociación. Sin embargo, para esta investigación en particular, no todas las reglas eran relevantes, por lo que se realizó un proceso de filtrado para seleccionar aquellas reglas que eran

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

verdaderamente significativas y pertinentes para el análisis específico que se llevaría a cabo posteriormente.

Estas reglas de asociación filtradas fueron exportadas en un documento PDF para facilitar su visualización y estudio detallado. La exportación en formato PDF permitió presentar las reglas de manera clara y ordenada, lo que facilitó su análisis e interpretación en el contexto de la investigación en curso.

Figura 113.

Reglas de asociación con 70% confianza y 15% SM



Nota: Elaboración propia.

En la figura 114 se muestra el código utilizado en la tercera prueba para cargar un archivo CSV con datos de transacciones que representan una lista de elementos. El objetivo era encontrar patrones frecuentes de asociación entre los elementos de las transacciones utilizando el algoritmo *Apriori*. Para lograr esto, se utilizaron tres librerías de *Python*: *pandas*, *mlxtend.preprocessing* y *mlxtend.frequent_patterns*. El código cargó

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

los datos del archivo CSV utilizando la librería *pandas*, luego los convirtió en una matriz binaria mediante la función *TransactionEncoder* de la librería *mlxtend.preprocessing*.

Posteriormente, se empleó la función *apriori* de la librería *mlxtend.frequent_patterns* para encontrar los *ítems* frecuentes. Por último, se utilizó la función *association_rules* para generar las reglas de asociación a partir de los *ítems* frecuentes encontrados.

Figura 114.

Importación de librerías y carga de datos en Python

```
import pandas as pd
from mlxtend.preprocessing import TransactionEncoder
from mlxtend.frequent_patterns import apriori, association_rules
from fpdf import FPDF

# Cargar datos desde el archivo CSV y convertirlos en una lista de transacciones
data = pd.read_csv('datosaletras.csv', header=None)
transactions = []
for i in range(data.shape[0]):
    transactions.append([str(data.values[i,j]) for j in range(data.shape[1])])

# Convertir datos nominales en valores booleanos
te = TransactionEncoder()
te_ary = te.fit(transactions).transform(transactions)
data = pd.DataFrame(te_ary, columns=te.columns_)
```

Nota: Elaboración propia.

El siguiente código, que se presenta en la figura 115, tuvo como objetivo convertir los datos almacenados en un objeto *Dataframe* de *pandas* en una lista de listas, donde cada lista representaba una transacción. Para lograrlo, se utilizaron dos bucles *for* anidados que recorrieron cada elemento del *Dataframe*. En el primer bucle, se creó una lista vacía llamada "*transaction*" que almacenaría los elementos de la transacción actual.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Posteriormente, en el segundo bucle, se iteró a través de cada columna del *Dataframe*, utilizando "j" como el índice de la columna actual.

Dentro del segundo bucle, se utilizó la función *pd.isna* para verificar si el valor en la posición actual era un valor faltante (NaN). Si el valor no era un valor faltante, se agregó a la lista "transaction". Al final del segundo bucle, la lista "transaction" se agregó a la lista "transactions" y se estableció un 90% de confianza mínima con un 20% de soporte mínimo (ver figura 115).

Figura 115.

Configuración de algoritmo apriori y soporte mínimo

```
# Aplicar algoritmo Apriori para generar conjuntos de ítems frecuentes
frequent_itemsets = apriori(data, min_support=0.20, use_colnames=True)

# Generar reglas de asociación con una confianza del x%
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.90)

# Filtrar reglas de asociación por lift mayor a 1 y ordenarlas de manera descendente
rules_lift = rules[rules['lift'] > 1].sort_values(by='lift', ascending=False)

# Crear objeto FPDF
pdf = FPDF()
```

Nota: Elaboración propia.

Se decidió exportar todas las reglas de asociación obtenidas en *Python* en archivos PDF para poder visualizarlas en su totalidad. Para realizar esta tarea, se utilizó la librería FPDF que permitió la creación de un documento PDF y agregar contenido al mismo. Se inició creando un objeto FPDF a partir de la clase FPDF y se almacenó en la variable "pdf". El constructor de la clase FPDF permitió establecer varios argumentos opcionales, incluyendo la orientación, la unidad y el formato del documento PDF. En este

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

caso, se seleccionó la orientación horizontal (*landscape*), la unidad pulgada (*in*) y el formato de papel carta (*letter*).

Posteriormente, se agregó una página al documento PDF utilizando el método `add_page()` del objeto `"pdf"`. Se estableció la fuente y el tamaño de letra para el título utilizando el método `set_font()`. Luego, se utilizó el método `cell()` para agregar texto a la página, el cual mostró el título "Reglas de asociación generadas" centrado en la página y se estableció un salto de línea con el argumento `"ln=1"`. Finalmente, se agregó una línea separadora de 50 guiones utilizando el método `cell()` para mejorar la legibilidad del documento (ver figura 116).

Figura 116.

Generación del archivo PDF

```
# Agregar página y establecer fuente y tamaño de letra para el título
pdf.add_page(orientation='L')
pdf.set_font("Arial", size=12)

# Agregar título y línea separadora
pdf.cell(270, 10, txt="Reglas de asociación generadas", ln=1, align="C")
pdf.cell(270, 10, txt="-"*20, ln=1, align="C")
```

Nota: Elaboración propia.

A continuación, la figura 117 muestra el resultado de nuestro análisis de datos utilizando *Python*, donde se presentan las reglas de asociación obtenidas y exportadas en un documento PDF para su mejor visualización. En este caso, se generaron únicamente 54 reglas de asociación (ver figura 117).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 117.

Reglas de asociación con 90% confianza y 20% SM



Nota: Elaboración propia.

La reducción considerable de 25,764 reglas de asociación a solo 54 es muy significativa en términos de simplificación y enfoque en la información relevante. Si bien puede parecer que obtener una gran cantidad de reglas de asociación proporciona una visión más completa de los datos, esto también puede llevar a una sobrecarga de información y dificultad para identificar patrones importantes. Por otro lado, al reducir la cantidad de reglas de asociación a un número más manejable y significativo, se puede enfocar la atención en las relaciones más importantes entre las variables, lo que facilita la interpretación de los resultados y la toma de decisiones informadas.

Además, la reducción de reglas de asociación también puede ahorrar tiempo y recursos al eliminar la necesidad de examinar y analizar una gran cantidad de datos innecesarios o irrelevantes.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

4.2 Resultados obtenidos en las herramientas de minería de datos

En la sección anterior se expusieron las pruebas llevadas a cabo, las cuales fueron útiles para seleccionar los modelos que ofrecen evaluaciones de calidad positivas y obtener información relevante para discutir las conclusiones. La interpretación de los resultados se organiza según las técnicas de minería de datos utilizadas: clasificación, agrupación y reglas de asociación. Se busca comparar los resultados obtenidos en cada herramienta según los algoritmos empleados.

4.2.1 Árboles de decisión

De acuerdo con las pruebas realizadas con tres variables distintas para el modelo de árbol de decisión, se determinó que la variable que generaba información relevante y métricas de evaluación aceptables fue con la variable P12. La accesibilidad a la red en los espacios universitarios es inestable, por lo tanto, en este apartado se presentan los modelos del árbol de decisión obtenidos en *RapidMiner*, *SAS Enterprise Miner* y *Python* con la variable P12 como pregunta raíz.

RapidMiner

De acuerdo con el modelo de árbol de decisión obtenido en *RapidMiner*, se lograron métricas de evaluación superiores al 90%, destacando una precisión de clasificación del 98.16%. Esto significa que el modelo fue capaz de clasificar correctamente el 98.16% de los datos analizados.

Para evaluar el rendimiento del modelo en términos de clasificación correcta, se utilizó una matriz de confusión. En la figura 118 se pueden observar los porcentajes de datos correctamente clasificados para cada una de las categorías. Los porcentajes de

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

clasificación correcta fueron los siguientes: 91.47% para "Totalmente en desacuerdo", 94.78% para "En desacuerdo", 92.14% para "Ni de acuerdo ni en desacuerdo", 99.44% para "De acuerdo", y 92.50% para "Totalmente de acuerdo". Estos resultados demuestran que el modelo fue altamente preciso en la clasificación de las respuestas en cada una de las categorías.

En comparación con la variable P10, que obtuvo una exactitud del 96.19%, la variable P12 demostró un rendimiento aún mejor, alcanzando una precisión de clasificación del 98.16%. Esto indica que el modelo fue especialmente eficaz al predecir y clasificar las respuestas relacionadas con la variable P12 en comparación con otras variables.

Figura 118.

Matriz de confusión de árbol de decisión pregunta raíz P12

accuracy: 98.16% +/- 3.75% (micro average: 98.16%)

	true Totalmente de ...	true De acuerdo	true Ni de acuerdo n...	true En desacuerdo	true Totalmente en ...	class precision
pred. Totalmente de...	277	55	4	3	1	91.47%
pred. De acuerdo	44	195	42	19	1	94.78%
pred. Ni de acuerdo...	2	24	18	12	0	92.14%
pred. En desacuerdo	3	20	15	44	7	99.44%
pred. Totalmente en...	1	3	3	4	29	92.50%
class recall	94.71%	95.66%	91.95%	93.66%	96.32%	

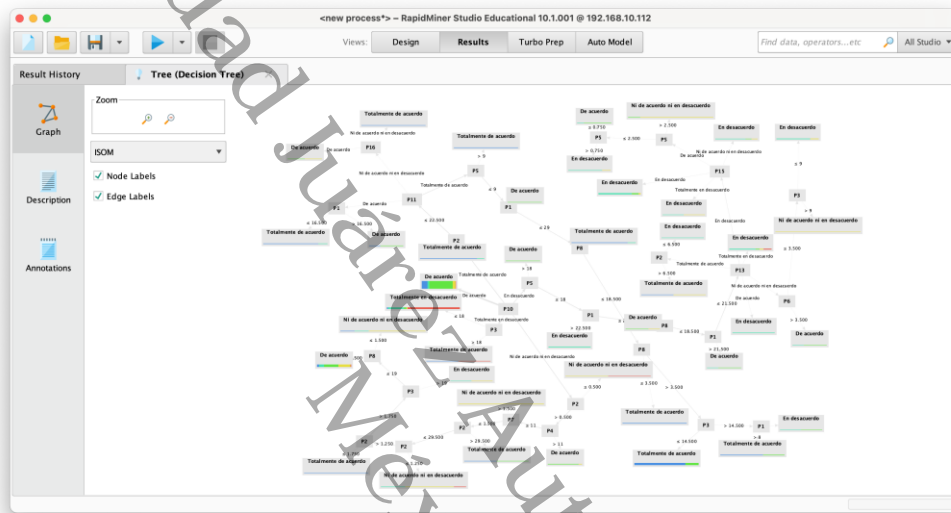
Nota: Elaboración propia.

Después de realizar el análisis de los datos, se procedió a visualizar el árbol de decisiones resultante. La figura 119 muestra el árbol de decisiones completo, el cual consta de 57 niveles y 208 nodos (ver figura 119). Esta representación gráfica

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos proporciona una visión clara de la estructura del árbol y la secuencia de decisiones tomadas en función de las características de los datos.

Figura 119.

Prueba de árbol de decisión de la pregunta raíz P12



Nota: Elaboración propia.

El análisis del árbol de decisión reveló un patrón de clasificaciones dispersas pero interconectadas, lo cual implicó un análisis exhaustivo. Para facilitar la comprensión de los resultados más relevantes y su interpretación, se proporciona la tabla 9 (ver tabla 9). Esta tabla ofrece una visión más clara y detallada de los aspectos destacados, permitiendo una mejor comprensión de los hallazgos obtenidos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 9.

Prueba de árbol de decisión de la pregunta raíz P12

No.	Clase	Variables	Decisión
1	De acuerdo	$P2 > 5$; $GRAD = 3$; {Totalmente de acuerdo=0, De acuerdo=3, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=0}	P20 = Totalmente en desacuerdo
2	De acuerdo	$P3 > 5$; {Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=4, En desacuerdo=0, Totalmente en desacuerdo=1}	P16 = De acuerdo
3	De acuerdo	$P10 =$ De acuerdo; {Totalmente de acuerdo=0, De acuerdo=2, Ni de acuerdo ni en desacuerdo=1, En desacuerdo=0, Totalmente en desacuerdo=0}	P13 = De acuerdo
4	De acuerdo	$P18 =$ De acuerdo; {Totalmente de acuerdo=0, De acuerdo=3, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=2, Totalmente en desacuerdo=0}	P10 = De acuerdo
5	De acuerdo	{Totalmente de acuerdo=1, De acuerdo=0, Ni de acuerdo ni en desacuerdo=4, En desacuerdo=0, Totalmente en desacuerdo=1}	P13 = Ni de acuerdo ni en desacuerdo
6	De acuerdo	$P1 > 1.2$; $P3 > 5$; {Totalmente de acuerdo=2, De acuerdo=7, Ni de acuerdo ni en desacuerdo=1, En desacuerdo=1, Totalmente en desacuerdo=0}	P19 = De acuerdo
7	De acuerdo	$P1 > 5$; $P3 > 10$; {Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=2, En desacuerdo=0, Totalmente en desacuerdo=0}	P16 = De acuerdo
8	De acuerdo	$P3 < 10$; $P2 > 10$; {Totalmente de acuerdo=2, De acuerdo=1, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=0}	P14 = Totalmente de acuerdo
9	De acuerdo	$CAT = 1$; {Totalmente de acuerdo=23, De acuerdo=82, Ni de acuerdo ni en desacuerdo=1, En desacuerdo=3, Totalmente en desacuerdo=0}	P15 = De acuerdo
10	De acuerdo	$GRAD = 3$; {Totalmente de acuerdo=0, De acuerdo=7, Ni de acuerdo ni en desacuerdo=2, En desacuerdo=0, Totalmente en desacuerdo=0}	P19 = Totalmente de acuerdo
11	De acuerdo	$P8 > 2$; {Totalmente de acuerdo=1, De acuerdo=1, Ni de acuerdo ni en desacuerdo=1, En desacuerdo=0, Totalmente en desacuerdo=9}	P17 = Totalmente de acuerdo
12	De acuerdo	$P10 =$ En desacuerdo; $P2 > 22.5$; {Totalmente de acuerdo=0, De acuerdo=4, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=0}	P17 = De acuerdo
13	De acuerdo	$P9 =$ De acuerdo; {Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=4, En desacuerdo=0, Totalmente en desacuerdo=0}	P19 = Ni de acuerdo ni en desacuerdo
14	En desacuerdo	$P4 > 5$; {Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=5, Totalmente en desacuerdo=1}	P14 = De acuerdo
15	En desacuerdo	$P4 < 13.5$; $P1 > 26$; {Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=8, En desacuerdo=1, Totalmente en desacuerdo=0}	P16= Totalmente de acuerdo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

No.	Clase	Variables	Decisión
16	En desacuerdo	$P3 > 4$; {Totalmente de acuerdo=3, De acuerdo=0, Ni de acuerdo ni en desacuerdo=1, En desacuerdo=1, Totalmente en desacuerdo=0}	$P19 =$ Totalmente de acuerdo
17	En desacuerdo	{Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=10, Totalmente en desacuerdo=0}	$P9 =$ De acuerdo
18	En desacuerdo	{Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=3, Totalmente en desacuerdo=0}	$P13 =$ En desacuerdo
19	Ni de acuerdo ni en desacuerdo	{Totalmente de acuerdo=4, De acuerdo=27, Ni de acuerdo ni en desacuerdo=23, En desacuerdo=15, Totalmente en desacuerdo=3}	$P11 =$ De acuerdo
20	Ni de acuerdo ni en desacuerdo	{Totalmente de acuerdo=0, De acuerdo=1, Ni de acuerdo ni en desacuerdo=9, En desacuerdo=0, Totalmente en desacuerdo=0}	$P16 =$ Ni de acuerdo ni en desacuerdo
21	Ni de acuerdo ni en desacuerdo	$P9 =$ De acuerdo; {Totalmente de acuerdo=5, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=1, Totalmente en desacuerdo=0}	$P15 =$ Totalmente de acuerdo
22	Ni de acuerdo ni en desacuerdo	$P5 < 9.5$; {Totalmente de acuerdo=1, De acuerdo=2, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=0}	$P17 =$ De acuerdo
23	Ni de acuerdo ni en desacuerdo	{Totalmente de acuerdo=5, De acuerdo=1, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=0}	$P13 =$ Totalmente de acuerdo
24	Ni de acuerdo ni en desacuerdo	$CAT = 2$; {Totalmente de acuerdo=20, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=0}	$P18 =$ Totalmente de acuerdo
25	Ni de acuerdo ni en desacuerdo	{Totalmente de acuerdo=1, De acuerdo=2, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=0}	$P14 =$ De acuerdo
26	Totalmente en desacuerdo	$P3 > 5$; {Totalmente de acuerdo=1, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=2}	$P11 =$ Totalmente en desacuerdo
27	Totalmente en desacuerdo	$P1 > 10$; {Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=3, Totalmente en desacuerdo=0}	$P9 =$ En desacuerdo
28	Totalmente en desacuerdo	$P8 > 5$; {Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=6}	$P19 =$ Totalmente de acuerdo
29	Totalmente en desacuerdo	$HRS > 10$; {Totalmente de acuerdo=0, De acuerdo=1, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=0, Totalmente en desacuerdo=1}	$P9 =$ De acuerdo
30	Totalmente en desacuerdo	$P19 =$ Totalmente de acuerdo; {Totalmente de acuerdo=0, De acuerdo=0, Ni de acuerdo ni en desacuerdo=0, En desacuerdo=2, Totalmente en desacuerdo=19}	$P10 =$ Totalmente en desacuerdo

Nota: Elaboración propia.

 Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

SAS Enterprise Miner

Con las métricas obtenidas en *SAS Enterprise Miner*, en la figura 120 se presenta a continuación la matriz de confusión para determinar el porcentaje de clasificación correcta por cada variable, donde los porcentajes de datos correctamente clasificados continúan siendo superiores al 90%, con porcentajes de error de 2.36%, 2.56%, 7.92%, 3.40% y 1.52% para los *ítems* "Totalmente en desacuerdo" (97.64%), "En desacuerdo" (97.44), "Ni de acuerdo ni en desacuerdo" (92.08%), "De acuerdo" (96.6%) y "Totalmente de acuerdo" (98.48%), respectivamente. La precisión de clasificación obtenida con la variable P12 fue de 96.44% (ver figura 120).

Figura 120.

Matriz de confusión árbol de decisión pregunta raíz P12

Procedimiento HPSPLIT						
Matriz de confusión basada en el modelo						
Real	Predichos					Tasa de error
	1	2	3	4	5	
1	187	36	13	8	0	0.2336
2	32	237	29	35	0	0.2883
3	5	21	60	14	0	0.4000
4	1	18	20	92	0	0.2977
5	2	2	0	14	0	1.0000

Nota: Elaboración propia.

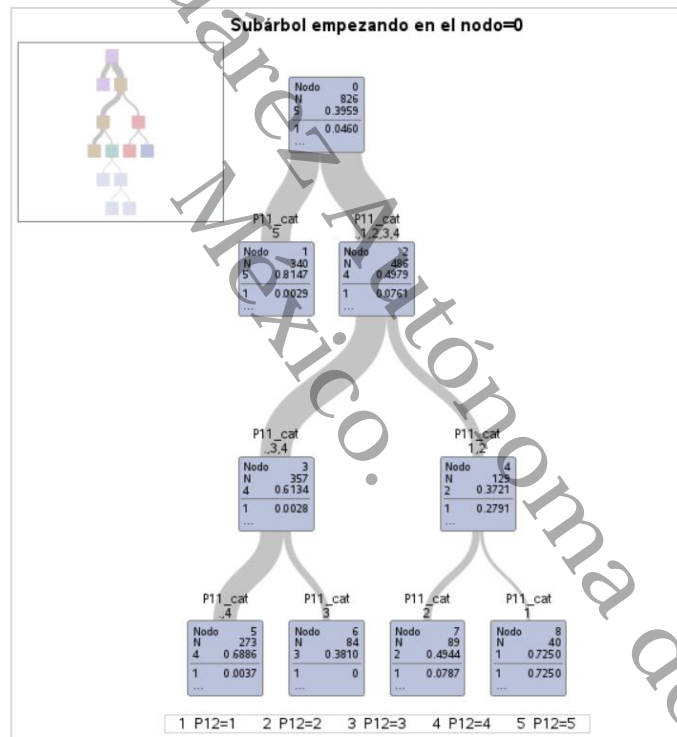
Con esta información, se logró interpretar que el uso de la variable P12 como pregunta raíz también generó un modelo de clasificación correcto y que no se presentaron cambios significativos en comparación con la prueba anterior.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tras el análisis de los datos, se visualizó el árbol de decisiones resultante, el cual tuvo una profundidad máxima de 10 niveles y 194 hojas, sin embargo, después de la poda se obtuvo una profundidad de árbol de 5 y un número de hojas de 7 después de la poda. Esto indica que el nivel de detalle en este árbol fue similar que la primera prueba, como se puede apreciar en la figura 121 (ver figura 121).

Figura 121.

Prueba de árbol de decisión de la pregunta raíz P12



Nota: Elaboración propia.

De la misma manera que en la prueba anterior de la pregunta raíz P10, no se ocultó ningún nivel del árbol generado, debido a que SAS Enterprise Miner muestra la versión podada del diagrama.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

De acuerdo con los resultados obtenidos en el diagrama, se crea la tabla 10 con las decisiones más representativas para una mayor optimización de los resultados y su interpretación (ver tabla 10).

Tabla 10.

Decisiones representativas de la pregunta raíz P12

No.	Clase	Variables	Decisión
1	De acuerdo	$P2 < 22.5; P1 > 28; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=2, \text{Totalmente en desacuerdo}=0 \}$	P14 = De acuerdo
2	De acuerdo	$P1 < 28; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=2, \text{Totalmente en desacuerdo}=0 \}$	P13 = En desacuerdo
3	De acuerdo	$P4 < 13.5; \{ \text{Totalmente de acuerdo}=4, \text{De acuerdo}=19, \text{Ni de acuerdo ni en desacuerdo}=2, \text{En desacuerdo}=9, \text{Totalmente en desacuerdo}=0 \}$	P15 = De acuerdo
4	De acuerdo	$P3 < 2.5; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=2, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P16 = Ni de acuerdo ni en desacuerdo
5	De acuerdo	$P1 < 18.5; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=9, \text{Ni de acuerdo ni en desacuerdo}=2, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P17 = De acuerdo
6	De acuerdo	$P10 = \text{Ni de acuerdo ni en desacuerdo}; \{ \text{Totalmente de acuerdo}=11, \text{De acuerdo}=42, \text{Ni de acuerdo ni en desacuerdo}=9, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P18 = De acuerdo
7	En desacuerdo	$\{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=2 \}$	P13 = Totalmente de acuerdo
8	En desacuerdo	$P1 < 2.5; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=3, \text{Ni de acuerdo ni en desacuerdo}=1, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P11 = De acuerdo
9	En desacuerdo	$P3 > 5; \{ \text{Totalmente de acuerdo}=1, \text{De acuerdo}=2, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=11, \text{Totalmente en desacuerdo}=0 \}$	P20 = En desacuerdo
10	Ni de acuerdo ni en desacuerdo	$P4 < 11; P3 > 0.5; P1 > 2.5; \{ \text{Totalmente de acuerdo}=73, \text{De acuerdo}=14, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=1 \}$	P9 = Totalmente de acuerdo
11	Ni de acuerdo ni en desacuerdo	$P9 = \text{En desacuerdo}; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=8, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P16 = De acuerdo
12	Ni de acuerdo ni en desacuerdo	$P3 > 7; P6 < 0.5; \{ \text{Totalmente de acuerdo}=4, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P13 = Totalmente de acuerdo
13	Ni de acuerdo ni en desacuerdo	$P2 < 2; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=2, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P19 = De acuerdo
14	Totalmente en desacuerdo	$GRAD = 3; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=9, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P20 = De acuerdo
15	Totalmente en desacuerdo	$\{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=2, \text{Totalmente en desacuerdo}=9 \}$	P10 = Totalmente en desacuerdo

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Python

El último modelo de clasificación a presentar es *Python*, donde se evaluó la precisión del modelo de árbol de decisión en la clasificación conforme a la segunda pregunta raíz que es P12: 12. La accesibilidad a la red en los espacios universitarios es inestable, donde se inicializó *Python* con la importación de las librerías *pandas* necesarias así como *sklearn* y *graphviz*, así mismo se calculó la precisión del modelo comparando las predicciones realizadas en el conjunto de prueba con los valores reales de la variable objetivo. La precisión del modelo indicó qué tan bien se desempeñó el modelo en la tarea de clasificación.

En estos casos de árboles de decisión, la precisión es importante porque permite saber cuánto se puede confiar en las predicciones realizadas por el modelo. Una precisión de 94.51%, como se puede observar en la figura 122, significó que el modelo fue bastante preciso en sus predicciones. Esto sugirió que el modelo logró aprender patrones importantes en los datos de entrenamiento y generalizó bien en los datos de prueba. En otras palabras, el modelo fue capaz de predecir con alta precisión la respuesta de los profesores a la pregunta P12 basado en sus características (ver figura 122).

Figura 122.

Cálculo de la precisión del modelo – P12

```
In [4]: # Realizar predicciones en el conjunto de prueba
y_pred = model.predict(X_test)

# Calcular la precisión del modelo
accuracy = metrics.accuracy_score(y_test, y_pred)
print("Precisión:", accuracy)

Precisión: 0.9451612903225806
```

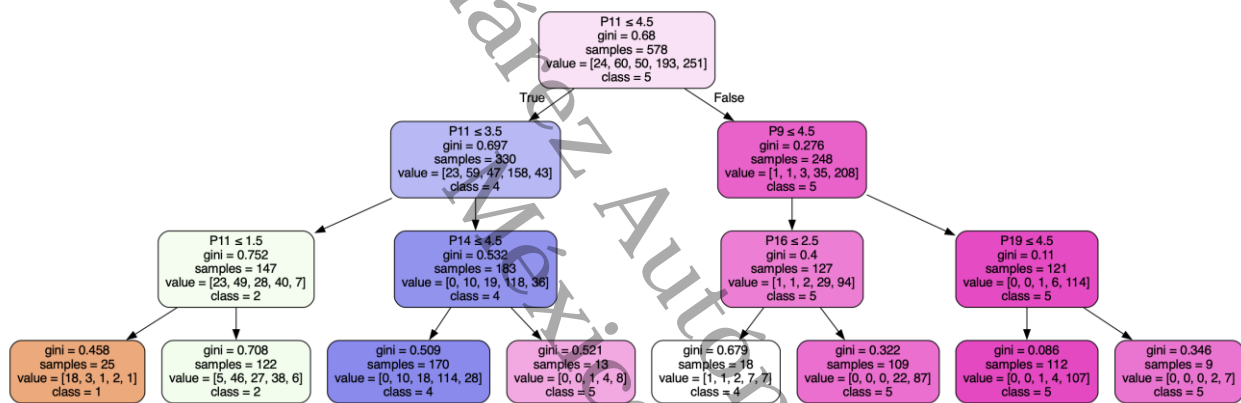
Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Por lo tanto, se concluyó que el modelo resultó útil y altamente preciso para predecir las respuestas a la pregunta P12 en la población de encuestados de la que se obtuvieron los datos; a continuación, se presenta el árbol de decisiones obtenido a partir de esta segunda prueba, correspondiente a la variable o pregunta raíz P12 (ver figura 123).

Figura 123.

Árbol de decisión generado para P12



Nota: Elaboración propia.

Con base en los resultados del diagrama, se tomaron las siguientes decisiones clave, las cuales se presentan en la tabla 11, para una mejor interpretación de los resultados (ver tabla 11). Estas decisiones son fundamentales para comprender las ramificaciones del árbol de decisiones y las acciones recomendadas en función de las características y valores de entrada. Al examinar detenidamente estas decisiones, se puede obtener una comprensión más profunda de cómo se clasifican y se toman decisiones en el modelo de árbol de decisiones.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 11.

Decisiones representativas de la pregunta raíz P12

No.	Clase	Variables	Decisión
1	De acuerdo	$P2 < 22.5; P1 > 28; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=2, \text{Totalmente en desacuerdo}=0 \}$	P14 = De acuerdo
2	De acuerdo	$P1 < 28; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=2, \text{Totalmente en desacuerdo}=0 \}$	P13 = En desacuerdo
3	De acuerdo	$P4 < 13.5; \{ \text{Totalmente de acuerdo}=4, \text{De acuerdo}=19, \text{Ni de acuerdo ni en desacuerdo}=2, \text{En desacuerdo}=9, \text{Totalmente en desacuerdo}=0 \}$	P15 = De acuerdo
4	De acuerdo	$P3 < 2.5; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=2, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P16 = Ni de acuerdo ni en desacuerdo
5	De acuerdo	$P1 < 18.5; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=9, \text{Ni de acuerdo ni en desacuerdo}=2, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P17 = De acuerdo
6	De acuerdo	$P1 > 1.2; P3 > 5; \{ \text{Totalmente de acuerdo}=2, \text{De acuerdo}=7, \text{Ni de acuerdo ni en desacuerdo}=1, \text{En desacuerdo}=1, \text{Totalmente en desacuerdo}=0 \}$	P19 = De acuerdo
7	De acuerdo	$P1 > 5; P3 > 10; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=2, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P16 = De acuerdo
8	De acuerdo	$P3 < 10; P2 > 10; \{ \text{Totalmente de acuerdo}=2, \text{De acuerdo}=1, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P14 = Totalmente de acuerdo
9	De acuerdo	$CAT = 1; \{ \text{Totalmente de acuerdo}=23, \text{De acuerdo}=82, \text{Ni de acuerdo ni en desacuerdo}=1, \text{En desacuerdo}=3, \text{Totalmente en desacuerdo}=0 \}$	P15 = De acuerdo
10	De acuerdo	$GRAD = 3; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=7, \text{Ni de acuerdo ni en desacuerdo}=2, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P19 = Totalmente de acuerdo
11	En desacuerdo	$P4 > 5; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=5, \text{Totalmente en desacuerdo}=1 \}$	P14 = De acuerdo
12	En desacuerdo	$P4 < 13.5; P1 > 26; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=8, \text{En desacuerdo}=1, \text{Totalmente en desacuerdo}=0 \}$	P16 = Totalmente de acuerdo
13	En desacuerdo	$P3 > 4; \{ \text{Totalmente de acuerdo}=3, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=1, \text{En desacuerdo}=1, \text{Totalmente en desacuerdo}=0 \}$	P19 = Totalmente de acuerdo
14	En desacuerdo	$\{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=10, \text{Totalmente en desacuerdo}=0 \}$	P9 = De acuerdo
15	En desacuerdo	$\{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=3, \text{Totalmente en desacuerdo}=0 \}$	P13 = En desacuerdo
16	Ni de acuerdo ni en desacuerdo	$P4 < 11; P3 > 0.5; P1 > 2.5; \{ \text{Totalmente de acuerdo}=73, \text{De acuerdo}=14, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=1 \}$	P9 = Totalmente de acuerdo
17	Ni de acuerdo ni en desacuerdo	$P9 = \text{En desacuerdo}; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=8, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P16 = De acuerdo
18	Ni de acuerdo ni en desacuerdo	$P3 > 7; P6 < 0.5; \{ \text{Totalmente de acuerdo}=4, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=0 \}$	P13 = Totalmente de acuerdo
19	Totalmente en desacuerdo	$P3 > 5; \{ \text{Totalmente de acuerdo}=1, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=2 \}$	P11 = Totalmente en desacuerdo
20	Totalmente en desacuerdo	$P1 > 10; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=3, \text{Totalmente en desacuerdo}=0 \}$	P9 = En desacuerdo
21	Totalmente en desacuerdo	$P8 > 5; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=0, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=6 \}$	P19 = Totalmente de acuerdo
22	Totalmente en desacuerdo	$HRS > 10; \{ \text{Totalmente de acuerdo}=0, \text{De acuerdo}=1, \text{Ni de acuerdo ni en desacuerdo}=0, \text{En desacuerdo}=0, \text{Totalmente en desacuerdo}=1 \}$	P9 = De acuerdo

Nota: Elaboración propia.

4.2.2 Agrupación

En esta investigación se buscó identificar las barreras que impidieron el uso de innovaciones tecnológicas por parte de los profesores de la UJAT en el contexto de la pandemia COVID-19. Para lograr este objetivo, se utilizó el algoritmo *Kmeans* para generar agrupaciones de datos y se evaluaron diferentes opciones de agrupamiento, considerando tanto el número de *clústeres* como las variables a analizar.

Además de las pruebas de agrupamiento, se utilizó la validación del número óptimo de *clústeres*. Los *softwares* de minería de datos, *RapidMiner* y *SAS Enterprise Miner*, así como el lenguaje de programación *Python*, ofrecen diferentes opciones para evaluar la calidad de los *clústeres* generados por algoritmos como el *Kmeans*. En particular, estos programas utilizan la matriz de confusión como una métrica de evaluación externa para determinar si el modelo de *clustering* funciona correctamente. La matriz de confusión es una tabla que muestra la precisión de las predicciones del modelo al comparar los resultados reales de un conjunto de datos con los resultados generados por el modelo.

A través de esta herramienta, se pueden analizar diferentes aspectos de la calidad de los *clústeres*, como la precisión, la tasa de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos. Por lo tanto, la matriz de confusión se convierte en una herramienta esencial para evaluar la calidad de los *clústeres* generados por los algoritmos de *clustering*, permitiendo tomar decisiones informadas y mejorar los resultados obtenidos. Los resultados mostraron que el modelo de dos agrupaciones fue el más adecuado para esta investigación y permitió identificar las barreras que

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

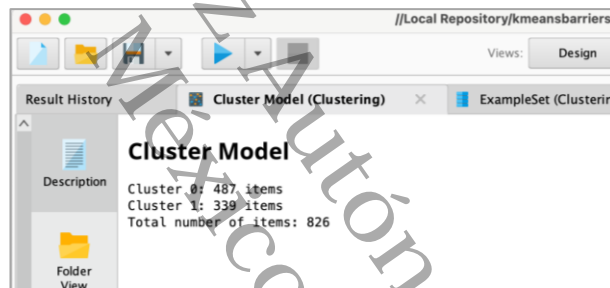
enfrentaron los profesores en la adaptación de tecnologías durante la pandemia COVID-19.

RapidMiner

Durante las pruebas de minería de datos utilizando la herramienta *RapidMiner*, se empleó el modelo de agrupación para determinar el número óptimo de *clústeres* que mejor representara los datos. Al realizar la prueba con dos *clústeres*, se obtuvo que el *Clúster 0* se conformó por 487 *ítems*, mientras que el *Clúster 1* se integró por 339 *ítems*.

Figura 124.

Ítems que conforman los dos clústeres



Nota: Elaboración propia.

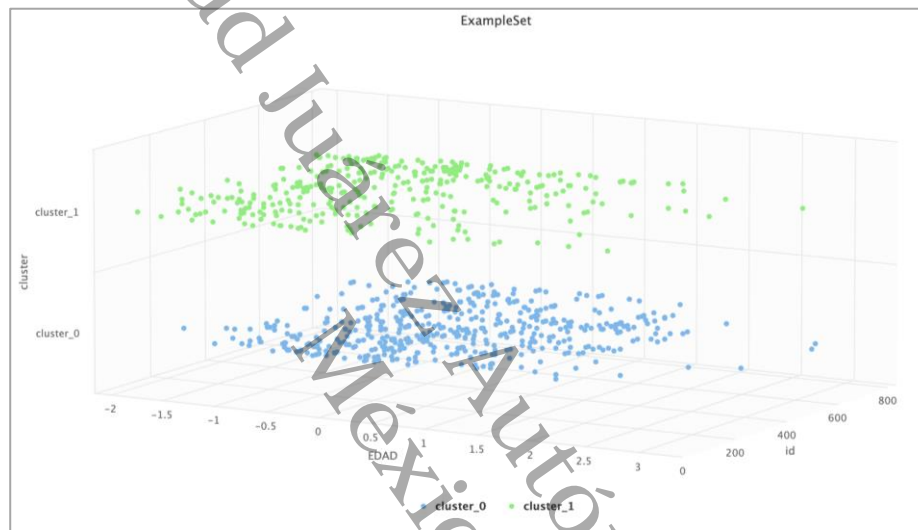
En la figura 125 se analizaron las variables de edad y *clúster*, donde se identificó que el *clúster* azul agrupó a los profesores con edades comprendidas entre 24 y 52 años, mientras que el *clúster* verde se conformó por docentes con edades entre más de 52 y hasta 80 años. Sin embargo, se observó que en el *clúster* verde existen algunos puntos que se encuentran alejados del centroide, en particular, los profesores de 24 a 28 años.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

De manera similar, en el *clúster* azul se encontraron puntos distantes en el rango de edad de 75 a 80 años. Estos resultados sugieren que, con dos agrupaciones, los datos no se clasifican de forma completa y precisa (ver figura 125).

Figura 125.

Kmeans prueba con dos agrupaciones



Nota: Elaboración propia.

En el caso de *RapidMiner*, aunque la herramienta puede calcular la medida de la silueta para evaluar la calidad del modelo de *Kmeans*, esta medida no proporciona una evaluación completa de la calidad del modelo debido a la subjetividad inherente a la técnica. Además, interpretar los resultados de la medida de la silueta requiere un conocimiento profundo del conjunto de datos y del problema que se está abordando.

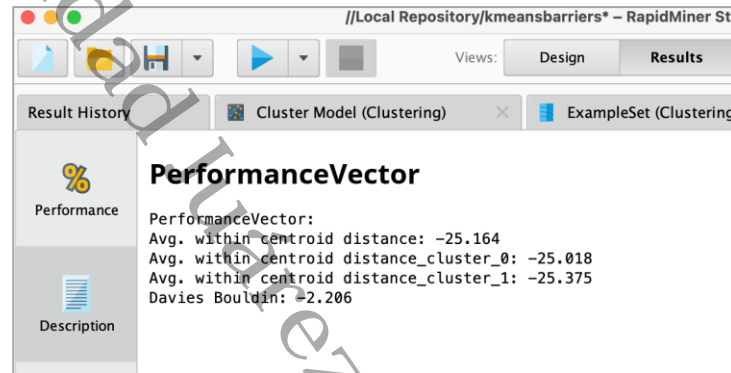
A partir de la versión 9.2 de *RapidMiner*, la medida de silueta fue eliminada y fue reemplazada por la medida de promedio, que proporciona una evaluación más simple y

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

fácil de interpretar. En la prueba de dos *clústeres*, se obtuvieron los siguientes resultados utilizando la medida de promedio (ver figura 126).

Figura 126.

Promedio de distancias centroides con dos clústeres



Nota: Elaboración propia.

Los resultados del análisis de *clustering* indicaron que se produjeron grupos altamente cohesivos y bien separados. Esto se evidencia por el valor negativo de la distancia media dentro de los centroides (-25.164), lo que significa que los puntos dentro de cada *clúster* están muy cerca de sus centroides correspondientes. Además, se pudo observar que la distancia media dentro de cada uno de los *clústeres* es similar, ya que los valores para "Avg. within centroid distance_cluster_0" y "Avg. within centroid distance_cluster_1" son casi idénticos, lo que sugiere que los dos *clústeres* tuvieron la misma cohesión.

La medida de Davies Bouldin, con un valor de -2.206, mostró una excelente separación entre los *clústeres*, ya que esta medida cuantifica la relación entre la dispersión dentro de los *clústeres* y la separación entre los *clústeres*. Un valor negativo

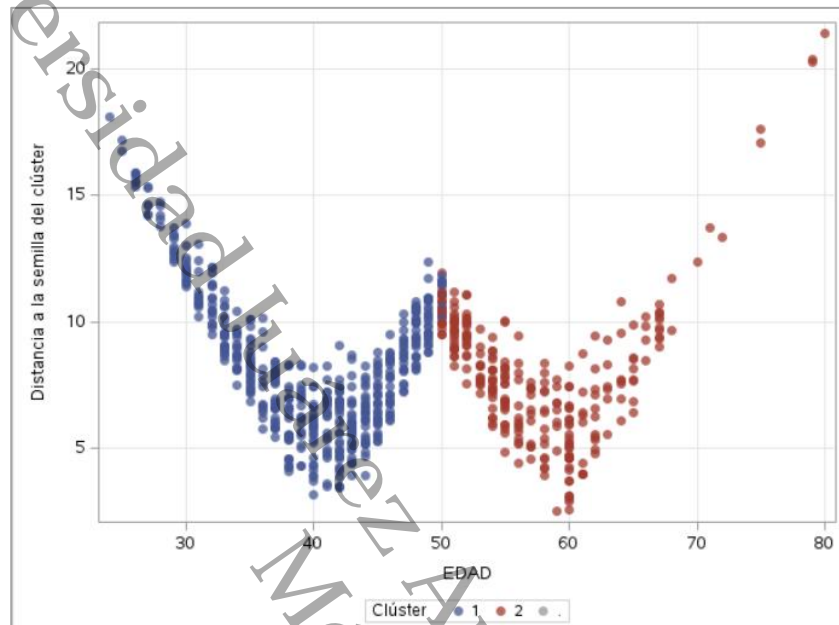
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

de la medida Davies Bouldin indica una excelente separación entre los *clústeres*, por lo tanto, un valor de -2.206 en esta medida sugiere que la separación entre los *clústeres* es adecuada. En resumen, estos resultados sugieren que el análisis de *clustering* produjo grupos altamente cohesivos y bien separados, con una separación global entre los *clústeres* muy adecuada.

SAS Enterprise Miner

En las pruebas de minería de datos realizadas se aplicó el modelo de agrupación para determinar el número óptimo de *clústeres* que mejor representara los datos. Al utilizar el algoritmo *Kmeans* con dos *clústeres*, se obtuvo que el *Clúster 1* estaba conformado por 487 *ítems*, mientras que el *Clúster 2* se integró por un total de 339 *ítems*. En la figura 127, se analizó la relación entre las variables de edad y *clúster*. Se puede observar que el *clúster* azul representa a los profesores con edades comprendidas entre 24 y 50 años, mientras que el *clúster* rojo representa a docentes con edades entre 52 y 80 años.

Sin embargo, en el *clúster* azul, algunos puntos se encuentran lejos del centroide, especialmente en el caso de profesores de 24 a 25 años. Del mismo modo, en el *clúster* rojo, hay puntos distantes en el rango de edad de 70 a 80 años. Estos resultados sugieren que, con dos agrupaciones, los datos se clasifican correctamente en su totalidad (ver figura 127).

Figura 127.*Kmeans prueba con dos agrupaciones*

Nota: Elaboración propia.

La herramienta *SAS Enterprise Miner* ofrece un resumen completo de los *clústeres* generados, como se puede apreciar en la figura 128. Este resumen proporciona información valiosa, como la frecuencia de los *ítems* en cada *clúster*, la desviación estándar RMS obtenida, la distancia máxima desde la semilla hasta la observación, el *clúster* más cercano a cada dato y la distancia entre los centroides de los *clústeres*.

Estos detalles permiten obtener una visión más detallada y precisa de los resultados obtenidos en la formación de los *clústeres*, lo que facilita el análisis y la interpretación de los patrones y características presentes en los datos (ver figura 128).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 128.

Resumen de clústeres

Resumen de clúster						
Clúster	Frecuencia	Desviación estándar RMS	Distancia máxima desde semilla hasta la observación	Radio sobrepasado	Clúster más cercano	Distancia entre Centroides del clúster
1	487	1.8982	18.0812		2	17.5218
2	339	1.8306	21.4122		1	17.5218

Nota: Elaboración propia.

Los resultados sugieren que el algoritmo de *clustering* ha generado grupos (*clústeres*) altamente cohesivos y bien definidos, como lo indica el valor de la distancia entre los centroides del *clúster* (17.52). Este valor significa que los puntos dentro de cada grupo están muy cercanos a sus respectivos centroides. Un valor elevado en la distancia entre centroides indica una buena separación entre los *clústeres*, es decir, que los puntos dentro de cada grupo están muy próximos entre sí y alejados de los puntos de otros *clústeres*. Esto demuestra que el algoritmo ha logrado agrupar objetos similares juntos y separarlos adecuadamente para identificar grupos homogéneos y bien definidos.

Python

Para llevar a cabo esta investigación, se utilizó *Python* en su forma nativa, es decir, mediante código. En primer lugar, se importaron las librerías *numpy*, que es una biblioteca para trabajar con matrices y vectores multidimensionales útil para la manipulación de datos científicos y matemáticos; *pandas*, eficaz para la manipulación y análisis de datos; y *matplotlib.pyplot*, capaz de crear gráficos de alta calidad, como gráficos de líneas, de barras, de dispersión, histogramas, entre otros.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

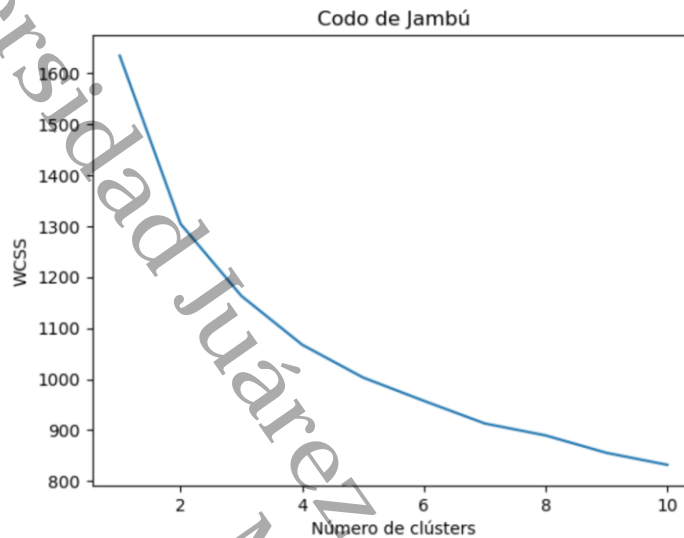
Además, se utilizó *sklearn.cluster.Kmeans*, que proporciona una implementación eficiente y escalable del algoritmo de *Kmeans* adecuada para grandes conjuntos de datos, y *sklearn.preprocessing*, que proporciona una amplia variedad de funciones para la normalización, estandarización, codificación de variables categóricas, entre otras. Antes de crear los *clústeres* con el método *Kmeans*, se supuso que no se tenía un criterio para determinar la cantidad de *clústeres* a crear y que tampoco se sabía cuál sería la cantidad óptima de *clústeres* a formar. Para resolver este problema, se implementó el método del codo de Jambú.

Este método consiste en crear diferentes cantidades de *clústeres* y calcular qué tan similares son los individuos dentro de los mismos y plasmar esta información dentro de una gráfica. Para medir esta similitud se utilizó la suma de los cuadrados dentro de cada grupo, también conocida como WCSS (*within clusters squares*). En la figura 129 se puede observar que el valor de WCSS fue disminuyendo a medida que aumentaba el número de *clústeres*.

Al llegar a 10 *clústeres*, el valor de WCSS ya era muy pequeño en comparación con el que se tenía con un solo *clúster*, lo cual era bueno. Sin embargo, se realizarían demasiadas particiones, por lo que se buscó un punto en el que el valor de WCSS dejara de disminuir de manera drástica, que en primera instancia fue el número 2, tal como se puede apreciar en la figura correspondiente (ver figura 129).

Figura 129.

Gráfico de codo de Jambú



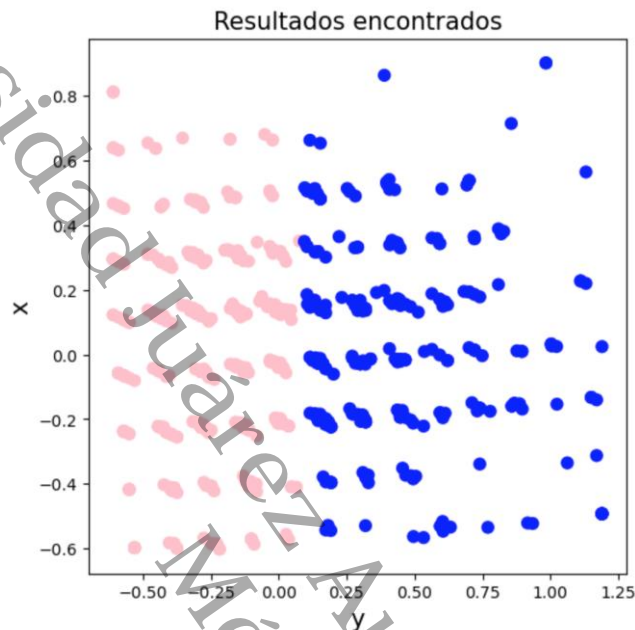
Nota: Elaboración propia.

Después de determinar el número inicial de *clústeres* a formar, se aplicó la función *Kmeans*. Además, utilizando *Python*, se pudo visualizar gráficamente los *clústeres* obtenidos. Esto se logró a través de un gráfico de dispersión en dos dimensiones, donde se mostraron los grupos resultantes identificados por sus etiquetas de *clúster* asignadas.

Como resultado, se obtuvo un gráfico que representaba los dos *clústeres* utilizados en la primera prueba de *Kmeans* en *Python* (ver figura 130). Esta representación visual permitió una fácil identificación y comprensión de los patrones y agrupamientos presentes en los datos analizados.

Figura 130.

Gráfico Kmeans obtenido primera prueba



Nota: Elaboración propia.

Sin embargo, es necesario comprobar la calidad de los *clústeres* obtenidos, esto se determinó mediante una matriz de confusión. La matriz de confusión que se observa en la figura 131, era de un formato de 2x2 y mostró los resultados del modelo para las dos clases diferentes o *clústeres*. Cada fila representó las instancias en una clase real, mientras que cada columna representó las instancias en una clase predicha.

En este caso, la matriz de confusión mostró que el *clustering* agrupó correctamente 339 instancias en el primer grupo y 487 instancias en el segundo grupo, es necesario destacar que se obtuvo un mismo número de *ítems* en los *clústeres*, como en las herramientas *RapidMiner* y *SAS Enterprise Miner*, sin embargo, en esta ocasión los

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

clústeres salieron de forma contraria. Como no hay valores en la diagonal secundaria de la matriz de confusión, esto significó que no hubo falsos positivos o falsos negativos, es decir, todas las instancias fueron clasificadas correctamente (ver figura 131).

Figura 131.

Matriz de confusión de dos clústeres

```
In [25]: from sklearn.metrics import confusion_matrix
# Ejecución del algoritmo KMeans con 2 clusters
kmeans = KMeans(n_clusters=2)
kmeans.fit(pca_respuestas_df.iloc[:, 0:2])
# Asignación de cada punto a su cluster correspondiente
pca_respuestas_df["KMeans_Clusters"] = kmeans.labels_
# Creación de la matriz de confusión
cm = confusion_matrix(pca_respuestas_df["KMeans_Clusters"], pca_respuestas_df["KMeans_Clusters"])
print(cm)

[[ 339  0]
 [  0 487]]
```

Nota: Elaboración propia.

4.2.3 Reglas de asociación

En esta sección se exhiben las reglas de asociación producidas por las herramientas, seleccionando las métricas que lograron mejores resultados. Se eligieron un total de 10 reglas mediante la ejecución de los algoritmos de asociación y las evaluaciones realizadas en cada herramienta.

RapidMiner

RapidMiner es una herramienta de minería de datos que ofrece una amplia gama de algoritmos para el análisis y la exploración de datos. En este caso, se utilizó *RapidMiner* para generar reglas de asociación con un alto nivel de confianza (90%) y un soporte mínimo del 20%. Este enfoque permitió descubrir patrones interesantes en los

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

datos y encontrar relaciones significativas entre diferentes variables del *dataset* (ver figura 132).

Figura 132.

Reglas de asociación obtenidas con 90% de confianza

No.	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s
1438	EDAD, HRS, P1, P5, P7	P6	0.136	0.911	0.988	-0.162	0.136
1439	EDAD, HRS, P2, P5, P7	P6	0.136	0.911	0.988	-0.162	0.136
1440	EDAD, P1, P2, P5, P7	P6	0.136	0.911	0.988	-0.162	0.136
1441	HRS, P1, P2, P5, P7	P6	0.136	0.911	0.988	-0.162	0.136
1442	EDAD, HRS, P1, P5, P7	P2, P6	0.136	0.911	0.988	-0.162	0.136
1443	EDAD, HRS, P2, P5, P7	P1, P6	0.136	0.911	0.988	-0.162	0.136
1444	EDAD, P1, P2, P5, P7	HRS, P6	0.136	0.911	0.988	-0.162	0.136
1445	HRS, P1, P2, P5, P7	EDAD, P6	0.136	0.911	0.988	-0.162	0.136
1446	EDAD, HRS, P1, P2, P5, P7	P6	0.136	0.911	0.988	-0.162	0.136
1447	EDAD, HRS, P1, P5, P7	P8, P6	0.136	0.911	0.988	-0.162	0.136
1448	EDAD, HRS, P1, P5, P7	P3, P6	0.136	0.911	0.988	-0.162	0.136
1449	EDAD, HRS, P1, P5, P7	P4, P6	0.136	0.911	0.988	-0.162	0.136
1450	EDAD, HRS, P2, P5, P7	P8, P6	0.136	0.911	0.988	-0.162	0.136
1451	EDAD, HRS, P2, P5, P7	P3, P6	0.136	0.911	0.988	-0.162	0.136
1452	EDAD, HRS, P2, P5, P7	P4, P6	0.136	0.911	0.988	-0.162	0.136
1453	EDAD, P1, P2, P5, P7	P8, P6	0.136	0.911	0.988	-0.162	0.136
1454	EDAD, P1, P2, P5, P7	P3, P6	0.136	0.911	0.988	-0.162	0.136

Nota: Elaboración propia.

Se lograron obtener 31 reglas con un *lift* mayor a uno, las cuales resultan ser útiles para su interpretación. Estos resultados ejemplifican la relevancia de contar con un número moderado de reglas para obtener información eficiente.

SAS Enterprise Miner

SAS Enterprise Miner es una herramienta de minería de datos avanzada que utiliza técnicas de modelado estadístico y de aprendizaje automático para descubrir patrones y relaciones en conjuntos de datos. En esta investigación, se utilizó *SAS Enterprise Miner* para generar reglas de asociación con un nivel de confianza del 90% y un soporte mínimo del 20%. Estas reglas proporcionan información valiosa sobre la relación entre diferentes

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

variables. A continuación, se presentan los resultados obtenidos y su análisis correspondiente, como se puede observar en la figura 133 se establecieron los parámetros especificados con anterioridad (ver figura 133).

Figura 133.

Reglas de asociación obtenidas con 90% de confianza

Association Report														
Relations	Expected Confidence (%)	Confidence (%)	Support (%)	Lift	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	Rule Item 2	Rule Item 3	Rule Item 4	Rule Item 5	Rule Index
2	61.54	76.32	55.77	1.24	29.00	UNO ==> TRES	UNO	TRES	UNO	====>	TRES			1
2	73.08	90.63	55.77	1.24	29.00	TRES ==> UNO	TRES	UNO	TRES	====>	UNO			2
2	73.08	83.78	59.62	1.15	31.00	CINCO ==> UNO	CINCO	UNO	CINCO	====>	UNO			3
2	71.15	81.58	59.62	1.15	31.00	UNO ==> CINCO	UNO	CINCO	UNO	====>	CINCO			4
2	80.77	90.63	55.77	1.12	29.00	TRES ==> CERO	TRES	CERO	TRES	====>	CERO			5
2	61.54	69.05	55.77	1.12	29.00	CERO ==> TRES	CERO	TRES	CERO	====>	TRES			6
2	80.77	86.49	61.54	1.07	32.00	CINCO ==> CERO	CINCO	CERO	CINCO	====>	CERO			7
2	71.15	76.19	61.54	1.07	32.00	CERO ==> CINCO	CERO	CINCO	CERO	====>	CINCO			8
2	73.08	73.81	59.62	1.01	31.00	CERO ==> UNO	CERO	UNO	CERO	====>	UNO			9
2	80.77	81.58	59.62	1.01	31.00	UNO ==> CERO	UNO	CERO	UNO	====>	CERO			10

Nota: Elaboración propia.

A partir de estos ajustes, se encontraron 25 reglas con un *lift* mayor a uno que pueden resultar útiles para la interpretación de los datos. Estas pruebas demuestran que es importante reducir el número de reglas para obtener información eficiente.

Python

A continuación, la figura 134 muestra el resultado de nuestro análisis de datos utilizando *Python*, donde se presentan las reglas de asociación obtenidas y exportadas en un documento PDF para su mejor visualización, esto con un 90% de confianza y el 20% de soporte mínimo. En este caso, se generaron únicamente 54 reglas de asociación (ver figura 134).

Figura 134.

Reglas de asociación con 90% confianza y 20% SM



Nota: Elaboración propia.

Un total de 54 reglas de asociación resultaron ser altamente significativas, lo que muestra la importancia de simplificar y enfocarse en la información relevante. Aunque puede parecer que una mayor cantidad de reglas de asociación proporciona una visión más completa de los datos, esto puede generar sobrecarga de información y dificultad para identificar patrones importantes. Por otro lado, la reducción del número de reglas de asociación a un número manejable y significativo permite enfocarse en las relaciones más importantes entre las variables, facilitando la interpretación de los resultados y la toma de decisiones informadas. Además, la reducción de reglas de asociación también puede ahorrar tiempo y recursos al eliminar la necesidad de examinar y analizar una gran cantidad de datos innecesarios o irrelevantes.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

4.3 Patrones descubiertos en árboles de decisión

En esta sección se presentan los patrones descubiertos a través de la aplicación de técnicas de minería de datos en las tres herramientas seleccionadas: *RapidMiner*, *SAS Enterprise Miner* y *Python*. La minería de datos, como bien se ha descrito con anterioridad, es una herramienta valiosa para analizar grandes conjuntos de datos y descubrir patrones y tendencias que de otra manera podrían pasar desapercibidos.

Cada plataforma de minería de datos tiene su propio conjunto de herramientas y algoritmos que pueden ayudar a encontrar patrones y conocimientos útiles en los datos. A continuación, se describirán los patrones descubiertos en cada plataforma y cómo pueden ser útiles para esta investigación.

4.3.1 *RapidMiner*

Las decisiones detectadas en los resultados conforme a la clasificación otorgada permiten generar la siguiente categorización de los patrones identificados en *RapidMiner*:

Clase De acuerdo

En esta clasificación se detectaron un total de 295 profesores que se identificaron con los patrones enlistados con los números del 1 al 13 de acuerdo con la tabla 9 previamente expuesta, donde se describe que:

1. Profesores que hace aproximadamente más de 5 años comenzaron a usar recursos de Internet para apoyar sus actividades docentes y cuentan con el grado académico de Doctorado, están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable, así mismo están totalmente en

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- desacuerdo de que han invertido pocos recursos económicos en prepararse para usar la tecnología en la práctica docente.
2. Profesores que hace aproximadamente más de 5 años utilizaron por primera vez un Ambiente Virtual de Aprendizaje están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable también están de acuerdo con que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.
 3. Profesores que están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable indicaron que también están de acuerdo con que el equipo tecnológico de apoyo a las actividades docentes es escaso así mismo coincidieron en que cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.
 4. Profesores que están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable están de acuerdo con que la institución carece de un sistema que evalúe las competencias tecnológicas de su personal y están de acuerdo con que el equipo tecnológico de apoyo a las actividades docentes es escaso.
 5. Profesores que están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable pero no están ni de acuerdo ni en desacuerdo de que cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

6. Profesores que hace aproximadamente más de 1.2 años comenzaron a usar computadora para apoyar sus actividades docentes y que hace aproximadamente 5 o más años usaron por primera vez un Ambiente Virtual de Aprendizaje están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable pero también están de acuerdo en que han invertido poco tiempo en prepararse para usar la tecnología en su práctica docente.
7. Profesores hace aproximadamente más de 5 años comenzaron a usar computadora para apoyar sus actividades docentes y que hace aproximadamente más de 10 años comenzó a usar recursos de Internet para apoyar sus actividades docentes, están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable y están de acuerdo en que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.
8. Profesores con menos de 10 años en que usaron por primera vez un Ambiente Virtual de Aprendizaje, pero con más de 10 años que utilizan recursos de Internet para apoyar sus actividades docentes indicaron estar de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable y están totalmente de acuerdo en que los ayuden a resolver problemas de *software* en la UJAT resulta complicado.
9. Profesores que están de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable indicaron pertenecer a la categoría de

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Base, así mismo están de acuerdo con que existe una política institucional poco clara que promueva la adopción de tecnología en el aula.

10. Profesores con grado académico de Doctorado están que están de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable indicaron también que están totalmente de acuerdo en que han invertido poco tiempo en prepararse para usar la tecnología en su práctica docente.
11. Profesores que están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable indicaron que hace más de 2 años utilizan el correo electrónico para trabajar con sus estudiantes y que están totalmente de acuerdo con que la institución carece de un sistema que diagnostique las competencias tecnológicas de su personal.
12. Profesores que respondieron estar en desacuerdo con que el equipo tecnológico de apoyo a las actividades docentes es escaso, están de acuerdo que la accesibilidad a la red en los espacios universitarios es inestable y que están de acuerdo con que la institución carece de un sistema que diagnostique las competencias tecnológicas de su personal.
13. Profesores que están de acuerdo en que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT y que la accesibilidad a la red en los espacios universitarios es inestable, indicaron no estar ni de acuerdo ni en desacuerdo con que han invertido poco tiempo en prepararse para usar la tecnología en la práctica docente.

Clase En desacuerdo

En esta clasificación se detectaron un total de 187 profesores que se identificaron con los patrones enlistados con los números del 14 al 18 de acuerdo con la tabla 9 previamente expuesta, donde se describe que:

14. Profesores que desde hace aproximadamente más de 5 años usan las redes sociales para trabajar con sus estudiantes indicaron que estuvieron en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable sin embargo mencionaron que también están de acuerdo con que les ayuden a resolver problemas de *software* en la UJAT resulta complicado.
15. Profesores que desde hace aproximadamente menos de 13.5 años usan las redes sociales para trabajar con sus estudiantes y que desde hace aproximadamente 26 años comenzaron a usar computadora para apoyar sus actividades docentes estuvieron en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable sin embargo están totalmente de acuerdo en que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.
16. Profesores que indicaron estar en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable hace aproximadamente menos de 4 años usaron por primera vez un Ambiente Virtual de Aprendizaje, sin embargo, están totalmente de acuerdo con que han invertido poco tiempo en prepararse para usar la tecnología en su práctica docente.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

17. Profesores que indicaron estar en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable están de acuerdo con que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.

18. Profesores que están en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable también están en desacuerdo con que cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.

Clase Ni de acuerdo ni en desacuerdo

En esta clasificación se detectaron un total de 132 profesores que se identificaron con los patrones enlistados con los números del 19 al 25 de acuerdo con la tabla 9 previamente expuesta, donde se describe que:

19. Profesores que no estuvieron ni de acuerdo ni en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable estuvieron de acuerdo con que la conectividad de la red representa un problema recurrente.

20. Profesores que no estuvieron ni de acuerdo ni en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable indicaron que no estuvieron ni de acuerdo ni en desacuerdo con que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.

21. Profesores que estuvieron de acuerdo con que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT, indicaron que no estuvieron ni de acuerdo ni en desacuerdo con que la

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- accesibilidad a la red en los espacios universitarios es inestable, sin embargo, estuvieron totalmente de acuerdo con que existe una política institucional poco clara que promueva la adopción de tecnología en el aula.
22. Profesores que hace menos de 9.5 años utilizan WhatsApp para trabajar con sus estudiantes, indicaron que no estuvieron ni de acuerdo ni en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, estuvieron de acuerdo con que la institución carece de un sistema que diagnostique las competencias tecnológicas de su personal.
23. Profesores que indicaron no estar ni de acuerdo ni en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable indicaron estar totalmente de acuerdo en que cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.
24. Profesores de categoría Interino indicaron que no estuvieron ni de acuerdo ni en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable y que están totalmente de acuerdo con que la institución carece de un sistema que evalúe las competencias tecnológicas de su personal.
25. Profesores que no estuvieron ni de acuerdo ni en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable señalaron que estuvieron de acuerdo con que les ayuden a resolver problemas de *software* en la UJAT resulta complicado.

Clase Totalmente en desacuerdo

En esta clasificación se detectaron un total de 212 profesores que se identificaron con los patrones enlistados con los números del 26 al 30 de acuerdo con la tabla 9 previamente expuesta, donde se describe que:

26. Profesores que hace aproximadamente más de 5 años utilizaron por primera vez un Ambiente Virtual de Aprendizaje están totalmente en desacuerdo de que la accesibilidad a la red en los espacios universitarios es inestable, por lo tanto, para ellos la conectividad de la red no representa un problema recurrente en sus actividades docentes.
27. Profesores que hace aproximadamente más de 10 años comenzaron a usar computadora para apoyar sus actividades docentes están totalmente en desacuerdo que la accesibilidad a la red en los espacios universitarios es inestable, así mismo están en desacuerdo de que existen carencias en la infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.
28. Profesores que hace aproximadamente más de 5 cinco años usan el correo electrónico para trabajar con sus estudiantes indicaron que están totalmente en desacuerdo que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, estuvieron totalmente de acuerdo en que han invertido poco tiempo en prepararse para usar la tecnología en la práctica docente.
29. Profesores que tienen un número de horas superior a 10 frente a grupos durante la semana indicaron que están totalmente en desacuerdo que la

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

accesibilidad a la red en los espacios universitarios es inestable, sin embargo, indicaron que están totalmente de acuerdo en que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.

30. Profesores que están totalmente de acuerdo en que han invertido poco tiempo en prepararse para usar la tecnología en la práctica docente, mencionan que están totalmente en desacuerdo que la accesibilidad a la red en los espacios universitarios es inestable así mismo mencionan que están totalmente en desacuerdo con que el equipo tecnológico de apoyo a las actividades docentes es escaso en la UJAT.

4.3.2 SAS Enterprise Miner

Clase De acuerdo

En esta clasificación se detectaron un total de 301 profesores que se identificaron con los patrones enlistados con los números del 1 al 6 de acuerdo con la tabla 10 previamente expuesta, donde se describe que:

1. Profesores que estuvieron de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable hace menos de 22.5 años comenzaron a usar recursos de Internet para apoyar sus actividades docentes y hace aproximadamente más de 28 años comenzaron a usar computadora para apoyar sus actividades docentes, están de acuerdo en que les ayuden a resolver problemas de *software* en la UJAT resulta complicado.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

2. Profesores que indicaron estar de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable señalaron hace aproximadamente menos de 28 años comenzaron a usar computadora para apoyar sus actividades docentes, sin embargo, están en desacuerdo con que cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.
3. Profesores que indicaron que hace aproximadamente menos de 13.5 años usan las redes sociales para trabajar con sus estudiantes, estuvieron de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, estuvieron de acuerdo con que existe una política institucional poco clara que promueva la adopción de tecnología en el aula.
4. Profesores que hace aproximadamente menos de 13.5 años usaron por primera vez un Ambiente Virtual de Aprendizaje indicaron estar de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, pero señalaron no estar ni de acuerdo ni en desacuerdo con que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.
5. Profesores que hace aproximadamente menos de 18.5 años comenzaron a usar computadora para apoyar sus actividades docentes estuvieron de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, pero señalaron estar de acuerdo con que la institución carece de un sistema que diagnostique las competencias tecnológicas de su personal.
6. Profesores que estuvieron de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, indicaron que no estuvieron ni de acuerdo ni

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

en desacuerdo con que el equipo tecnológico de apoyo a las actividades docentes es escaso, sin embargo, indicaron estar de acuerdo con que la institución carece de un sistema que evalúe las competencias tecnológicas de su personal.

Clase En desacuerdo

En esta clasificación se detectaron un total de 119 profesores que se identificaron con los patrones enlistados con los números del 7 al 9 de acuerdo con la tabla 10 previamente expuesta, donde se describe que:

7. Profesores que indicaron estar en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, señalaron estar totalmente de acuerdo cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.
8. Profesores que hace aproximadamente menos de 2.5 años comenzaron a usar computadora para apoyar sus actividades docentes señalaron estar en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, indicaron que estuvieron de acuerdo con que la conectividad de la red representa un problema recurrente.
9. Profesores que señalaron estar en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable y con más de 5 años en que usaron por primera vez un Ambiente Virtual de Aprendizaje, estuvieron en desacuerdo con que han invertido pocos recursos económicos en prepararse para usar la tecnología en su práctica docente.

Clase Ni de acuerdo ni en desacuerdo

En esta clasificación se detectaron un total de 193 profesores que se identificaron con los patrones enlistados con los números del 10 al 13 de acuerdo con la tabla 10 previamente expuesta, donde se describe que:

10. Profesores que tienen menos de 11 años usando las redes sociales para trabajar con sus estudiantes, que también hace más de 0.5 años usaron por primera vez un Ambiente Virtual de Aprendizaje y que hace aproximadamente más de 2.5 años comenzaron a usar computadora para apoyar sus actividades docentes señalaron estar ni de acuerdo ni en desacuerdo en que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo indicaron estar totalmente de acuerdo con que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.
11. Profesores que indicaron estar en desacuerdo con que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT, señalaron no estar ni de acuerdo ni en desacuerdo en que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, respondieron estar totalmente de acuerdo con que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.
12. Profesores indicaron que hace aproximadamente más de 7 años usaron por primera vez un Ambiente Virtual de Aprendizaje y que hace 0.5 años comenzaron a utilizar Facebook para trabajar con sus estudiantes, señalaron estar ni de acuerdo ni en desacuerdo en que la accesibilidad a la red en los espacios

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

universitarios es inestable, sin embargo, respondieron estar totalmente de acuerdo cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.

13. Profesores que indicaron estar ni de acuerdo ni en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, hace aproximadamente menos de 2 años comenzaron a utilizar recursos de Internet para apoyar sus actividades docentes, sin embargo, están de acuerdo con que han invertido poco tiempo en prepararse para usar la tecnología en su práctica docente.

Clase Totalmente en desacuerdo

En esta clasificación se detectaron un total de 213 profesores que se identificaron con los patrones enlistados con los números del 14 al 15 de acuerdo con la tabla 10 previamente expuesta, donde se describe que:

14. Profesores del grado académico de Doctorado indicaron estar totalmente en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, señalaron estar de acuerdo en que han invertido pocos recursos económicos en prepararse para usar la tecnología en su práctica docente.

15. Profesores que indicaron estar totalmente de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, indicaron estar totalmente en desacuerdo con que el equipo tecnológico de apoyo a las actividades docentes es escaso.

4.3.3 Python

Clase De acuerdo

En esta clasificación se detectaron un total de 298 profesores que se identificaron con los patrones enlistados con los números del 1 al 13 de acuerdo con la tabla 11 previamente expuesta, donde se describe que:

1. Profesores que estuvieron de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable hace menos de 22.5 años comenzaron a usar recursos de Internet para apoyar sus actividades docentes y hace aproximadamente más de 28 años comenzaron a usar computadora para apoyar sus actividades docentes, están de acuerdo en que les ayuden a resolver problemas de *software* en la UJAT resulta complicado.
2. Profesores que indicaron estar de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable señalaron hace aproximadamente menos de 28 años comenzaron a usar computadora para apoyar sus actividades docentes, sin embargo, están en desacuerdo con que cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.
3. Profesores que indicaron que hace aproximadamente menos de 13.5 años usan las redes sociales para trabajar con sus estudiantes, estuvieron de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, estuvieron de acuerdo con que existe una política institucional poco clara que promueva la adopción de tecnología en el aula.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

4. Profesores que hace aproximadamente menos de 13.5 años usaron por primera vez un Ambiente Virtual de Aprendizaje indicaron estar de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, pero señalaron no estar ni de acuerdo ni en desacuerdo con que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.
5. Profesores que hace aproximadamente menos de 18.5 años comenzaron a usar computadora para apoyar sus actividades docentes estuvieron de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable, pero señalaron estar de acuerdo con que la institución carece de un sistema que diagnostique las competencias tecnológicas de su personal.
6. Profesores que hace aproximadamente más de 1.2 años comenzaron a usar computadora para apoyar sus actividades docentes y que hace aproximadamente 5 o más años usaron por primera vez un Ambiente Virtual de Aprendizaje están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable pero también están de acuerdo en que han invertido poco tiempo en prepararse para usar la tecnología en su práctica docente.
7. Profesores hace aproximadamente más de 5 años comenzaron a usar computadora para apoyar sus actividades docentes y que hace aproximadamente más de 10 años comenzó a usar recursos de Internet para apoyar sus actividades docentes, están de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable y están de acuerdo en que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

8. Profesores con menos de 10 años en que usaron por primera vez un Ambiente Virtual de Aprendizaje, pero con más de 10 años que utilizan recursos de Internet para apoyar sus actividades docentes indicaron estar de acuerdo en que la accesibilidad a la red en los espacios universitarios es inestable y están totalmente de acuerdo en que los ayuden a resolver problemas de *software* en la UJAT resulta complicado.
9. Profesores que están de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable indicaron pertenecer a la categoría de Base, así mismo están de acuerdo con que existe una política institucional poco clara que promueva la adopción de tecnología en el aula.
10. Profesores con grado académico de Doctorado están que están de acuerdo con que la accesibilidad a la red en los espacios universitarios es inestable indicaron también que están totalmente de acuerdo en que han invertido poco tiempo en prepararse para usar la tecnología en su práctica docente.

Clase En desacuerdo

En esta clasificación se detectaron un total de 156 profesores que se identificaron con los patrones enlistados con los números del 11 al 15 de acuerdo con la tabla 11 previamente expuesta, donde se describe que:

11. Profesores que desde hace aproximadamente más de 5 años usan las redes sociales para trabajar con sus estudiantes indicaron que estuvieron en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable sin

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

embargo mencionaron que también están de acuerdo con que les ayuden a resolver problemas de *software* en la UJAT resulta complicado.

12. Profesores que desde hace aproximadamente menos de 13.5 años usan las redes sociales para trabajar con sus estudiantes y que desde hace aproximadamente 26 años comenzaron a usar computadora para apoyar sus actividades docentes estuvieron en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable sin embargo están totalmente de acuerdo en que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.
13. Profesores que indicaron estar en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable hace aproximadamente menos de 4 años usaron por primera vez un Ambiente Virtual de Aprendizaje, sin embargo, están totalmente de acuerdo con que han invertido poco tiempo en prepararse para usar la tecnología en su práctica docente.
14. Profesores que indicaron estar en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable están de acuerdo con que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.
15. Profesores que están en desacuerdo con que la accesibilidad a la red en los espacios universitarios es inestable también están en desacuerdo con que cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.

Clase Ni de acuerdo ni en desacuerdo

En esta clasificación se detectaron un total de 165 profesores que se identificaron con los patrones enlistados con los números del 16 al 18 de acuerdo con la tabla 11 previamente expuesta, donde se describe que:

16. Profesores que tienen menos de 11 años usando las redes sociales para trabajar con sus estudiantes, que también hace más de 0.5 años usaron por primera vez un Ambiente Virtual de Aprendizaje y que hace aproximadamente más de 2.5 años comenzaron a usar computadora para apoyar sus actividades docentes señalaron estar ni de acuerdo ni en desacuerdo en que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo indicaron estar totalmente de acuerdo con que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.
17. Profesores que indicaron estar en desacuerdo con que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT, señalaron no estar ni de acuerdo ni en desacuerdo en que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, respondieron estar totalmente de acuerdo con que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente.
18. Profesores indicaron que hace aproximadamente más de 7 años usaron por primera vez un Ambiente Virtual de Aprendizaje y que hace 0.5 años comenzaron a utilizar Facebook para trabajar con sus estudiantes, señalaron estar ni de acuerdo ni en desacuerdo en que la accesibilidad a la red en los espacios

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

universitarios es inestable, sin embargo, respondieron estar totalmente de acuerdo cuando tienen un problema con algún equipo es difícil conseguir apoyo técnico.

Clase Totalmente en desacuerdo

En esta clasificación se detectaron un total de 207 profesores que se identificaron con los patrones enlistados con los números del 19 al 22 de acuerdo con la tabla 11 previamente expuesta, donde se describe que:

19. Profesores que hace aproximadamente más de 5 años utilizaron por primera vez un Ambiente Virtual de Aprendizaje están totalmente en desacuerdo de que la accesibilidad a la red en los espacios universitarios es inestable, por lo tanto, para ellos la conectividad de la red no representa un problema recurrente en sus actividades docentes.
20. Profesores que hace aproximadamente más de 10 años comenzaron a usar computadora para apoyar sus actividades docentes están totalmente en desacuerdo que la accesibilidad a la red en los espacios universitarios es inestable, así mismo están en desacuerdo de que existen carencias en la infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.
21. Profesores que hace aproximadamente más de 5 cinco años usan el correo electrónico para trabajar con sus estudiantes indicaron que están totalmente en desacuerdo que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, estuvieron totalmente de acuerdo en que han invertido poco tiempo en prepararse para usar la tecnología en la práctica docente.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

22. Profesores que tienen un número de horas superior a 10 frente a grupos durante la semana indicaron que están totalmente en desacuerdo que la accesibilidad a la red en los espacios universitarios es inestable, sin embargo, indicaron que están totalmente de acuerdo en que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.

4.4 Patrones descubiertos en agrupación

En este análisis se presentarán los patrones descubiertos utilizando las herramientas *RapidMiner*, *SAS Enterprise Miner* y *Python*, utilizando el algoritmo *Kmeans*. Se destaca que en las tres herramientas se obtuvieron dos *clústeres* con tamaños de 487 y 339 *ítems* respectivamente, lo que indica que los resultados son consistentes y comparables entre las diferentes herramientas. Aunque es importante mencionar que el orden de los *clústeres* en *Python* puede ser diferente.

Además, se encontraron patrones similares en los tres conjuntos de datos analizados, lo cual demuestra la eficacia del algoritmo *Kmeans* en la identificación de patrones en conjuntos de datos de gran tamaño. Este hallazgo resalta la validez y la consistencia de los resultados obtenidos a través de diferentes herramientas, lo que fortalece la confianza en las conclusiones y en la interpretación de los patrones identificados

4.4.1 Información de los *clústeres* obtenidos

En la figura 135 se analizaron las variables de edad y *clúster*, donde se identificó que el *clúster* azul agrupó a los profesores con edades comprendidas entre 24 y 52 años, mientras que el *clúster* verde se conformó por docentes con edades entre 52 y 80 años.

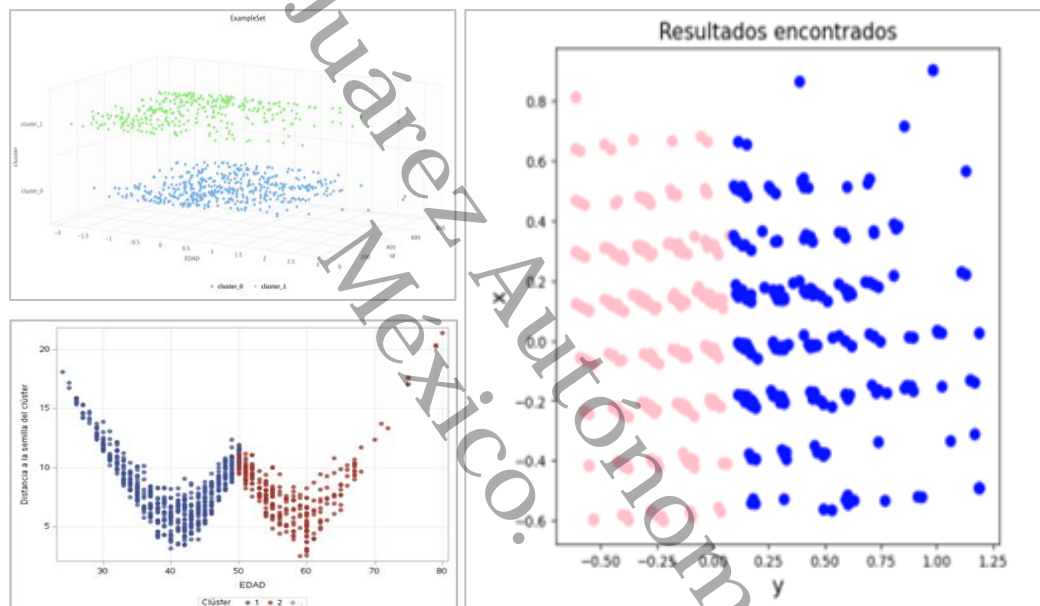
Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Sin embargo, se observó que en el *clúster* verde existen algunos puntos que se encuentran alejados del centroide, en particular, los profesores de 24 a 28 años.

De manera similar, en el *clúster* azul se encontraron puntos distantes en el rango de edad de 75 a 80 años (ver figura 135).

Figura 135.

Kmeans prueba con dos agrupaciones



Nota: Elaboración propia.

Al realizar la prueba con dos *clústeres*, se obtuvo que el primer se conformó por 487 *ítems*, mientras que el segundo *clúster* se integró por 339 *ítems*. Es necesario destacar que los 826 *ítems* estaban compuestos por 12 preguntas adicionales, la cuáles se emplearon para el conglomerado de respuestas, dando el total de 9,912 datos totales en el *dataset*.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Es posible que la apariencia visual de los gráficos de los *clústeres* sea diferente en cada herramienta debido a las diferencias en el diseño y la interfaz de usuario. Cada herramienta puede tener diferentes opciones de visualización y estilos de gráficos que pueden afectar la forma en que se representan los datos. Además, la representación gráfica de los *clústeres* puede ser influenciada por la elección de las variables y parámetros utilizados en la herramienta de análisis.

Por lo tanto, es importante destacar que, aunque los gráficos de los *clústeres* puedan verse diferentes en cada herramienta, los datos subyacentes y los patrones descubiertos son los mismos. La consistencia en los resultados es más relevante que la apariencia visual de los gráficos.

Basándonos en los diagramas y los informes de *clústeres* generados en las tres herramientas (*RapidMiner*, *SAS Enterprise Miner* y *Python*), se encontró que el número total de profesores en el *dataset* con edades entre 24 años y 52 años o menos fue de 618. Por otro lado, se identificaron 208 *ítems* de profesores con edades entre más de 52 y 80 años. El *clúster* 1 y el *clúster* 0 estuvieron compuestos por 339 *ítems* cada uno.

El *clúster* 0 se caracterizó por incluir profesores con edades entre 52 y 80 años, lo cual representó el 41% de los datos clasificados. Con base en estos hallazgos, se realizó la interpretación de los agrupamientos que permiten identificar los patrones generales en el *dataset*, considerando las edades de los profesores y sus respuestas, conforme al orden de las preguntas en el *dataset* (ver tabla 12).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 12.

Carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	53
En desacuerdo (2)	151
Ni de acuerdo ni en desacuerdo (3)	104
De acuerdo (4)	333
Totalmente de acuerdo (5)	185

Nota: Elaboración propia.

En primera instancia se obtuvo que el 7% de los profesores estaban totalmente en desacuerdo con esto, sin embargo, una mayor cantidad del 18% estuvo en desacuerdo. Por otra parte, se contempló que el 13% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 40% estuvo de acuerdo con esta cuestión, mientras que el 22% respondieron que estaban totalmente de acuerdo.

Estos resultados sugieren que existe una percepción generalizada entre los profesores de la UJAT de que hay carencias en la infraestructura tecnológica para el soporte de las actividades docentes. Aunque una minoría significativa de los profesores no está de acuerdo con esta afirmación, la mayoría de los encuestados parece estar de acuerdo o totalmente de acuerdo.

Por lo tanto, se puede inferir que es importante que la UJAT preste atención a las necesidades tecnológicas de los profesores para garantizar que puedan llevar a cabo sus actividades docentes de manera efectiva y eficiente. Es importante tener en cuenta que el uso de tecnologías educativas y la infraestructura tecnológica son elementos clave

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

para el desarrollo de la educación en línea y el aprendizaje digital, por lo que resulta esencial contar con una infraestructura sólida y adecuada que permita el desempeño efectivo de los docentes y estudiantes.

Sin embargo, los resultados de la encuesta sugieren que, al menos en la percepción de los profesores encuestados, la infraestructura tecnológica en la UJAT no es una limitante significativa para el desempeño docente. No obstante, es importante continuar evaluando y mejorando la infraestructura tecnológica en la institución para garantizar el éxito en la implementación de tecnologías educativas y la educación en línea. Por otra parte, para *El equipo tecnológico de apoyo a las actividades docentes es escaso*, se obtuvo lo siguiente:

Tabla 13.

El equipo tecnológico de apoyo a las actividades docentes es escaso

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	58
En desacuerdo (2)	148
Ni de acuerdo ni en desacuerdo (3)	137
De acuerdo (4)	333
Totalmente de acuerdo (5)	150

Nota: Elaboración propia.

Se obtuvo que el 7% de los profesores estaban totalmente en desacuerdo con esto, sin embargo, una cantidad del 18% estuvo en desacuerdo. Por otra parte, se contempló que el 17% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 40% estuvo de acuerdo con esta cuestión, mientras que el 17% respondieron que estaban totalmente de acuerdo.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

De acuerdo con los resultados obtenidos, parece haber una división bastante equitativa entre los encuestados que están de acuerdo y los que no lo están con la afirmación de que el equipo tecnológico de apoyo a las actividades docentes es escaso en la UJAT. Sin embargo, un porcentaje significativo de encuestados (40%) estuvo de acuerdo con esta afirmación, lo que sugiere que todavía hay espacio para mejorar la infraestructura tecnológica en la universidad. Además, el hecho de que un 17% de los encuestados estén totalmente de acuerdo podría indicar que hay algunas áreas donde la falta de tecnología está afectando seriamente las actividades docentes.

Por otro lado, es positivo que un 7% de los encuestados estén totalmente en desacuerdo, lo que sugiere que la universidad está haciendo esfuerzos para proporcionar suficiente equipo tecnológico para sus profesores. En general, estos resultados indican que es importante seguir trabajando en mejorar la infraestructura tecnológica en la UJAT para apoyar adecuadamente las actividades docentes. No obstante, se contempló que la conectividad de la red representó un problema muy importante en las actividades de los docentes, como se puede ver a continuación en la tabla 14:

Tabla 14.

La conectividad de la red representa un problema recurrente

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	40
En desacuerdo (2)	89
Ni de acuerdo ni en desacuerdo (3)	84
De acuerdo (4)	273
Totalmente de acuerdo (5)	340

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

En estos resultados se reflejó que el 5% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 11% estuvo en desacuerdo. Por otra parte, se contempló que el 10% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 33% estuvo de acuerdo con esta cuestión, mientras que el 41% respondieron que estaban totalmente de acuerdo.

De estos resultados se puede concluir que la mayoría de los encuestados, un 74% en total, están de acuerdo o totalmente de acuerdo en que la conectividad de la red representa un problema recurrente en la UJAT. Esto sugiere que hay una necesidad importante de mejorar la infraestructura tecnológica relacionada con la conectividad de la red en la institución. Además, es importante tener en cuenta que, aunque una minoría (16%) no percibe la conectividad como un problema, todavía es importante considerar las necesidades de estos grupos al abordar este problema en particular.

Continuando con la red de la universidad, se consideró el tema de su accesibilidad, donde los profesores de la institución caso de estudio respondieron lo siguiente:

Tabla 15.

La accesibilidad a la red en los espacios universitarios es inestable

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	38
En desacuerdo (2)	82
Ni de acuerdo ni en desacuerdo (3)	82
De acuerdo (4)	297
Totalmente de acuerdo (5)	327

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

En estos resultados se encontró que el 5% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 10% estuvo en desacuerdo. Por otra parte, se contempló que el 10% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 36% estuvo de acuerdo con esta cuestión, mientras que el 39% respondieron que estaban totalmente de acuerdo.

Los resultados de la encuesta indican que existe una percepción generalizada de que la accesibilidad a la red en los espacios universitarios es inestable. A pesar de que solo el 5% de los profesores estuvieron en total desacuerdo con esta afirmación, un porcentaje mayor (10%) también estuvo en desacuerdo. Además, más de un tercio de los encuestados (36%) estuvieron de acuerdo con esta afirmación y otro 39% estuvo totalmente de acuerdo. Esto sugiere que la inestabilidad de la accesibilidad a la red en los espacios universitarios es un problema significativo para una proporción considerable de profesores en la UJAT. Si bien la red, su accesibilidad e inestabilidad fueron problemas muy visibles en los datos anteriores, era necesario realizar el conteo de las opiniones de los profesores respecto al apoyo técnico de la UJAT:

Tabla 16.

Cuando tengo un problema con algún equipo es difícil conseguir apoyo técnico

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	106
En desacuerdo (2)	233
Ni de acuerdo ni en desacuerdo (3)	168
De acuerdo (4)	208
Totalmente de acuerdo (5)	111

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

En estos resultados se encontró que el 13% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 28% estuvo en desacuerdo. Por otra parte, se contempló que el 20% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 25% estuvo de acuerdo con esta cuestión, mientras que el 14% respondieron que estaban totalmente de acuerdo.

De acuerdo con los resultados obtenidos, la mayoría de los profesores no están de acuerdo con que sea difícil conseguir apoyo técnico cuando tienen un problema con algún equipo. Sin embargo, hay una cantidad significativa de encuestados (39%) que están de acuerdo o totalmente de acuerdo con esta afirmación.

Esto sugiere que puede haber casos aislados en los que los profesores encuentren dificultades para recibir soporte técnico, aunque en general la situación no es grave. Por lo tanto, sería importante revisar y mejorar el servicio de soporte técnico para garantizar que todos los profesores tengan acceso a la ayuda que necesitan cuando se enfrentan a problemas con los equipos.

Tabla 17.

Que te ayuden a resolver problemas de software en la UJAT resulta complicado

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	87
En desacuerdo (2)	220
Ni de acuerdo ni en desacuerdo (3)	211
De acuerdo (4)	186
Totalmente de acuerdo (5)	122

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Por otra parte, en estos resultados se encontró que el 10% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 27% estuvo en desacuerdo. Por otra parte, se contempló que el 26% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 22% estuvo de acuerdo con esta cuestión, mientras que el 15% respondieron que estaban totalmente de acuerdo.

Estos resultados sugieren que hay una percepción generalizada entre los encuestados de que es difícil obtener ayuda para resolver problemas de *software* en la UJAT. Aunque un porcentaje significativo de profesores (37%) no están de acuerdo con esta afirmación, el hecho de que el 37% esté de acuerdo o totalmente de acuerdo sugiere que hay margen de mejora en la asistencia técnica y el soporte de *software* en la universidad. También es importante destacar que un cuarto de los encuestados no tiene una opinión clara sobre esta cuestión, lo que podría indicar que hay una falta de información o transparencia en cuanto a cómo obtener ayuda técnica en la UJAT. Se podría considerar que es necesario mejorar la accesibilidad y calidad del soporte técnico para resolver problemas de *software* en la universidad.

Tabla 18.

Existe una política institucional poco clara que promueva la adopción de tecnología en el aula

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	63
En desacuerdo (2)	154
Ni de acuerdo ni en desacuerdo (3)	182
De acuerdo (4)	283
Totalmente de acuerdo (5)	144

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

En la tabla 18, previamente expuesta, se señalan los resultados obtenidos respecto a *Existe una política institucional poco clara que promueva la adopción de tecnología en el aula*, donde se encontró que el 8% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 19% estuvo en desacuerdo.

Por otra parte, se contempló que el 22% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 34% estuvo de acuerdo con esta cuestión, mientras que el 17% respondieron que estaban totalmente de acuerdo.

Estos resultados sugieren que existe una percepción generalizada entre los encuestados de que la UJAT carece de una política clara que fomente la adopción de tecnología en el aula. Aunque un porcentaje significativo de profesores (27%) no estuvo de acuerdo o estuvo en desacuerdo con esta afirmación, la mayoría de los encuestados (51%) estuvo de acuerdo o totalmente de acuerdo.

Es posible que esta falta de una política clara pueda estar limitando la integración efectiva de la tecnología en las actividades docentes, lo que puede estar afectando negativamente la calidad del aprendizaje de los estudiantes. Por lo tanto, sería importante que la institución implemente políticas claras y efectivas para fomentar la adopción y uso adecuado de la tecnología en el aula.

Por otra parte, se planteó la interrogante sobre la adecuación de su estructura administrativa actual para establecer un sistema eficiente de utilización de tecnología. En este contexto, se presentan a continuación los resultados obtenidos a través del cuestionario que respondieron los profesores de la UJAT:

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 19.

La estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	60
En desacuerdo (2)	177
Ni de acuerdo ni en desacuerdo (3)	194
De acuerdo (4)	268
Totalmente de acuerdo (5)	127

Nota: Elaboración propia.

Se identificó que el 7% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 21% estuvo en desacuerdo. Por otra parte, se contempló que el 24% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 33% estuvo de acuerdo con esta cuestión, mientras que el 15% respondieron que estaban totalmente de acuerdo.

Estos resultados sugieren que un porcentaje significativo de encuestados perciben que la estructura administrativa de la UJAT no es adecuada para establecer un sistema de utilización de tecnología eficiente. Aunque el 7% de los profesores estaban totalmente en desacuerdo y el 21% estuvieron en desacuerdo, más de la mitad de los encuestados (48%) estuvo de acuerdo o totalmente de acuerdo con esta afirmación. Esto sugiere que hay un desafío en la estructura administrativa de la UJAT para facilitar una adopción efectiva de la tecnología en el aula, sin embargo, es necesario observar las respuestas de los docentes ante la carencia de un sistema para diagnosticar las competencias tecnológicas del personal de la UJAT (ver tabla 20).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 20.

La institución carece de un sistema que diagnostique las competencias tecnológicas de su personal

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	36
En desacuerdo (2)	93
Ni de acuerdo ni en desacuerdo (3)	192
De acuerdo (4)	325
Totalmente de acuerdo (5)	180

Nota: Elaboración propia.

En estos resultados se encontró que el 5% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 11% estuvo en desacuerdo. Por otra parte, se contempló que el 23% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 39% estuvo de acuerdo con esta cuestión, mientras que el 22% respondieron que estaban totalmente de acuerdo.

Estos resultados sugieren que una gran parte de los encuestados cree que la institución carece de un sistema que diagnostique las competencias tecnológicas de su personal. Aunque un pequeño porcentaje de los profesores está en desacuerdo con esta afirmación, una cantidad significativa de ellos está de acuerdo o totalmente de acuerdo, lo que indica que la implementación de un sistema de diagnóstico de competencias tecnológicas podría ser beneficioso para la universidad.

La evaluación de competencias tecnológicas del personal es un aspecto fundamental para garantizar un adecuado aprovechamiento de la tecnología en cualquier institución educativa, siendo esta una de las cuestiones expuestas a los docentes:

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 21.

La institución carece de un sistema que evalúe las competencias tecnológicas de su personal

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	33
En desacuerdo (2)	88
Ni de acuerdo ni en desacuerdo (3)	188
De acuerdo (4)	337
Totalmente de acuerdo (5)	180

Nota: Elaboración propia.

Para esta pregunta, se encontró que el 4% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 10% estuvo en desacuerdo. Por otra parte, se contempló que el 23% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 41% estuvo de acuerdo con esta cuestión, mientras que el 22% respondieron que estaban totalmente de acuerdo.

De acuerdo con los resultados, una mayoría significativa de los encuestados (63%) está de acuerdo en que la institución carece de un sistema que evalúe las competencias tecnológicas de su personal, mientras que solo el 14% de los encuestados está en desacuerdo o totalmente en desacuerdo con esta afirmación. Además, un 23% de los encuestados no mostraron una postura clara al respecto. Esto sugiere que existe una necesidad percibida de implementar un sistema de evaluación de competencias tecnológicas en la institución, y que esta es una cuestión importante para una gran cantidad de profesores; por otro lado, se realizó una pregunta acerca del tiempo que han invertido los profesores en su preparación tecnológica para sus actividades.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 22.

He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	106
En desacuerdo (2)	233
Ni de acuerdo ni en desacuerdo (3)	168
De acuerdo (4)	208
Totalmente de acuerdo (5)	111

Nota: Elaboración propia.

En estos resultados se encontró que el 30% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 40% estuvo en desacuerdo. Por otra parte, se contempló que el 12% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 16% estuvo de acuerdo con esta cuestión, mientras que el 2% respondieron que estaban totalmente de acuerdo.

Estos resultados indican que la mayoría de los profesores encuestados han invertido tiempo en prepararse para usar la tecnología en su práctica docente. Un total del 70% de los profesores estuvieron en desacuerdo o totalmente en desacuerdo con la afirmación, lo que sugiere que consideran haber dedicado tiempo a prepararse en este aspecto. Por otro lado, un 16% de los encuestados estuvo de acuerdo con la afirmación, lo que indica que hay un grupo minoritario de profesores que siente que ha invertido poco tiempo en la preparación para el uso de la tecnología en la enseñanza.

Finalmente se añadió una pregunta respecto a la cantidad de recursos económicos que han invertido los profesores en su preparación tecnológica para la práctica docente:

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tabla 23.

He invertido pocos recursos económicos en prepararme para usar la tecnología en mi práctica docente

Respuestas	Total de respuestas
Totalmente en desacuerdo (1)	277
En desacuerdo (2)	314
Ni de acuerdo ni en desacuerdo (3)	122
De acuerdo (4)	143
Totalmente de acuerdo (5)	20

Nota: Elaboración propia.

En estos resultados se encontró que el 28% de los profesores estaban totalmente en desacuerdo con esto, así mismo, una cantidad del 38% estuvo en desacuerdo. Por otra parte, se contempló que el 15% de los encuestados no estaban ni de acuerdo ni en desacuerdo, no obstante, una mayor cifra del 17% estuvo de acuerdo con esta cuestión, mientras que el 2% respondieron que estaban totalmente de acuerdo.

Estos resultados indican que la mayoría de los profesores encuestados no están de acuerdo con la afirmación de que han invertido pocos recursos económicos en prepararse para usar la tecnología en su práctica docente. Un 66% de los encuestados estuvo en desacuerdo o totalmente en desacuerdo, mientras que solo el 19% estuvo de acuerdo o totalmente de acuerdo. Además, un 15% de los encuestados no estaban ni de acuerdo ni en desacuerdo con esta afirmación. Esto sugiere que la mayoría de los profesores encuestados han dedicado recursos económicos para mejorar su capacidad tecnológica en su trabajo docente.

4.5 Patrones descubiertos en reglas de asociación

En este apartado se mostrarán las reglas de asociación generadas por las diferentes herramientas, enfocándonos en aquellas que obtuvieron los mejores resultados según las evaluaciones realizadas en cada una de ellas. Se seleccionaron un total de 10 reglas resultantes de la ejecución de los algoritmos de asociación.

A partir de los resultados obtenidos, se construyó un diagrama de Venn para identificar las reglas que coincidieron en las tres herramientas: *RapidMiner*, *SAS Enterprise Miner* y *Python*. Esto nos permitió detectar patrones comunes y consistentes entre las diferentes plataformas.

A continuación, se presentarán los patrones identificados utilizando esta técnica, destacando las reglas de asociación más relevantes y significativas para el análisis de los datos.

4.5.1 *RapidMiner*

En este apartado se presentan los *itemsets* generados en *RapidMiner*, para ello se empleó un soporte mínimo del 20% y una confianza mínima del 90%, con estas métricas de calidad se obtuvo un *lift* mayor a uno lo que indica que las reglas mostradas se repiten dentro del conjunto de datos. Esto quiere decir que se detectaron patrones reales y no son reglas construidas por una cuestión aleatoria o de azar.

En la figura 136 se puede apreciar que se estableció un soporte mínimo del 20% y un mínimo de confianza del 90%, además de requerir un mínimo de cinco *items* en los datos antecesores (ver figura 136).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Figura 136.

Reglas de asociación en RapidMiner

Nº	Premises	Conclusion	Support	Confidence	LaPlace	Gain	p-s
1	EDAD, P1, P2, ANT, P8	HRS	0.944	1	1	-0.944	0.944
2	HRS, P1, P2, ANT, P8	EDAD	0.944	1	1	-0.944	0.944
3	EDAD, HRS, P1, ANT, P3	P2	0.891	1	1	-0.891	0.891
4	EDAD, P1, P2, ANT, P3	HRS	0.891	1	1	-0.891	0.891
5	HRS, P1, P2, ANT, P3	EDAD	0.891	1	1	-0.891	0.891
6	EDAD, HRS, P2, ANT, P5	P1	0.815	1	1	-0.815	0.815
7	EDAD, P1, P2, ANT, P5	HRS	0.815	1	1	-0.815	0.815
8	HRS, P1, P2, ANT, P5	EDAD	0.815	1	1	-0.815	0.815
9	EDAD, HRS, P2, ANT, P4	P1	0.806	1	1	-0.806	0.806
10	EDAD, P1, P2, ANT, P4	HRS	0.806	1	1	-0.806	0.806
11	HRS, P1, P2, ANT, P4	EDAD	0.806	1	1	-0.806	0.806
12	EDAD, HRS, P1, ANT, P6	P2	0.452	1	1	-0.452	0.452
13	EDAD, HRS, P2, ANT, P6	P1	0.452	1	1	-0.452	0.452
14	EDAD, P1, P2, ANT, P6	HRS	0.452	1	1	-0.452	0.452
15	HRS, P1, P2, ANT, P6	EDAD	0.452	1	1	-0.452	0.452
16	EDAD, HRS, P1, ANT, P7	P2	0.154	1	1	-0.154	0.154
17	EDAD, HRS, P2, ANT, P7	P1	0.154	1	1	-0.154	0.154

Nota: Elaboración propia.

Estos datos representan las reglas de asociación encontradas en el *dataset* en la aplicación de *RapidMiner*. En cada regla se especifican los *ítems* involucrados, el porcentaje de veces que ocurren juntos con relación al total de transacciones, el *lift* y el número de veces que se presentan juntos. Cada regla se identifica por un índice numérico, el cual representa las respuestas de los profesores en escala Likert.

Tabla 24.

Respuestas en escala Likert

Respuestas	Número asignado
Totalmente en desacuerdo	1
En desacuerdo	2
Ni de acuerdo ni en desacuerdo	3
De acuerdo	4
Totalmente de acuerdo	5

Nota: Elaboración propia.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Por ejemplo, la regla número 1 indica que el ítem UNO se asocia con el ítem TRES con un porcentaje del 61.54%, lo que significa que en el 61.54% de las transacciones en las que aparece UNO, también aparece TRES. El *lift* de 1.24 indica que la presencia de UNO aumenta en un 24% la probabilidad de que también esté presente TRES. La regla aparece dos veces en el conjunto de datos y se expresa en forma de implicación: "UNO => TRES", lo que significa que la presencia de UNO implica la presencia de TRES.

Como resultado, en total se obtuvieron 10 reglas con un *lift* mayor a uno que pueden ser útiles para la interpretación. Se generaron 31 reglas con una configuración que incluye un mínimo de dos *ítems* en los datos antecedentes. Es importante destacar que todas las reglas resultantes tienen como consecuencia el ítem "Totalmente en desacuerdo" en la variable P19.

Esto sugiere que la falta de tiempo dedicado a la preparación para el uso de la tecnología en la práctica docente no es un obstáculo común que puede afectar negativamente la actitud de los profesores hacia su implementación. Sin embargo, del total de 31 reglas generadas, se emplearon 25 para la interpretación de reglas obtenidas, siendo estas las de mayor relevancia. Estos patrones identificados son un factor clave para comprender cómo superar esta barrera y lograr una adopción efectiva de la tecnología en la educación.

4.5.2 SAS Enterprise Miner

Se generaron 10 reglas utilizando esta herramienta con un soporte mínimo del 20% y una confianza del 90%. El tamaño de las reglas varía de un ítem como mínimo a

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

cinco *ítems* como máximo. La mayoría de las reglas generadas tienen dos o tres *ítems*, como se puede observar en la siguiente figura (ver figura 137).

Figura 137.

Reglas de asociación SAS Enterprise Miner

Association Report														
Relations	Expected Confidence (%)	Confidence (%)	Support (%)	Lift	Transaction Count	Rule	Left Hand of Rule	Right Hand of Rule	Rule Item 1	Rule Item 2	Rule Item 3	Rule Item 4	Rule Item 5	Rule Index
2	61.54	76.32	55.77	1.24	29.00	UNO ==> TRES	UNO	TRES	UNO	====>	TRES			1
2	73.08	90.63	55.77	1.24	29.00	TRES ==> UNO	TRES	UNO	TRES	====>	UNO			2
2	73.08	83.78	59.62	1.15	31.00	CINCO ==> UNO	CINCO	UNO	CINCO	====>	UNO			3
2	71.15	81.58	59.62	1.15	31.00	UNO ==> CINCO	UNO	CINCO	UNO	====>	CINCO			4
2	80.77	90.63	55.77	1.12	29.00	TRES ==> CERO	TRES	CERO	TRES	====>	CERO			5
2	61.54	69.05	55.77	1.12	29.00	CERO ==> TRES	CERO	TRES	CERO	====>	TRES			6
2	80.77	86.49	61.54	1.07	32.00	CINCO ==> CERO	CINCO	CERO	CINCO	====>	CERO			7
2	71.15	76.19	61.54	1.07	32.00	CERO ==> CINCO	CERO	CINCO	CERO	====>	CINCO			8
2	73.08	73.81	59.62	1.01	31.00	CERO ==> UNO	CERO	UNO	CERO	====>	UNO			9
2	80.77	81.58	59.62	1.01	31.00	UNO ==> CERO	UNO	CERO	UNO	====>	CERO			10

Nota: Elaboración propia.

Se lograron identificar 25 reglas con un *lift* superior a uno luego de aplicar los ajustes necesarios, lo cual resulta relevante para la interpretación de los datos. Estos resultados evidencian la importancia de disminuir la cantidad de reglas para obtener información más eficiente, debido a que en las anteriores pruebas se alcanzaron 75 reglas de asociación, de las cuáles no todas tenían una verdadera relevancia.

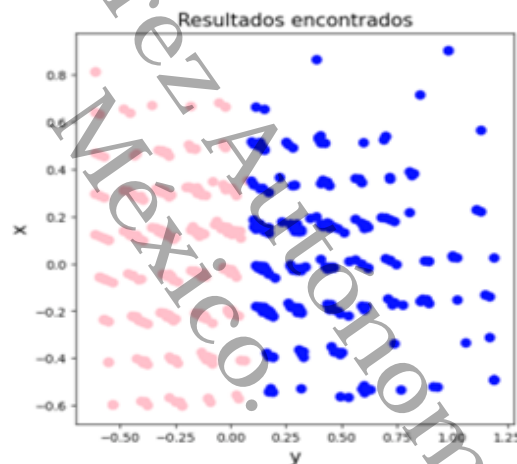
En esta segunda evaluación de las métricas de calidad utilizando *SAS Enterprise Miner*, se encontró una coincidencia en los resultados con los parámetros obtenidos en *RapidMiner*. Además, se detectó que los patrones en las reglas resultantes también coinciden con la variable P19. He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente, se destacó otra regla de asociación relacionada con la variable P20.

4.5.3 Python

En *Python* se generaron 54 reglas con una confianza mínima del 90% y un soporte mínimo de 0.20. La mayoría de las reglas están compuestas por cuatro *ítems*, mientras que las reglas de tres *ítems* son menos comunes. En la figura 138 se muestran las diez reglas con las mejores métricas y mayor frecuencia según la herramienta, sin importar el tamaño de los *ítemsets* (ver figura 138).

Figura 138.

Reglas de asociación con 90% confianza y 20% SM



Nota: Elaboración propia.

A partir de los resultados obtenidos con esta última herramienta, se ha obtenido una mayor cantidad de información que puede ser interpretada. Además, se encontraron coincidencias con los resultados generados por *RapidMiner* y *SAS Enterprise Miner*. En particular, se observó que en la columna del ítem consecuente predominan los resultados que indican que los encuestados estuvieron "totalmente en desacuerdo", por lo tanto, del

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

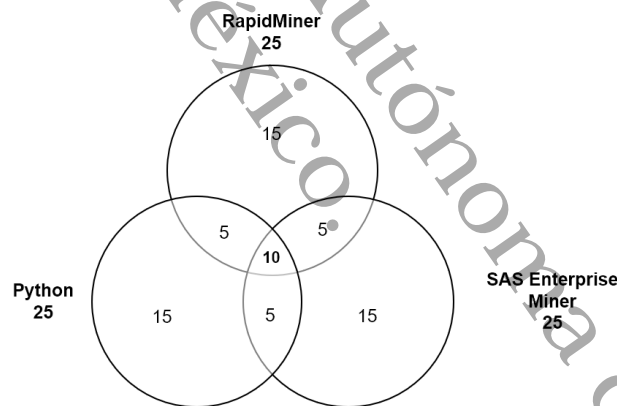
total de 54 reglas generadas, se emplearon 25 para la interpretación de reglas obtenidas, siendo estas las de mayor relevancia.

Interpretación de las reglas obtenidas

Se realizó un análisis comparativo de las reglas generadas en las tres herramientas para identificar similitudes y diferencias que permitieran una interpretación más eficiente de la información. En la figura 139 se presenta un diagrama de Venn que muestra el número de reglas que coincidieron en las tres herramientas: *RapidMiner*, *SAS Enterprise Miner* y *Python* (ver figura 139).

Figura 139.

Diagrama de Venn de las reglas de asociación



Nota: Elaboración propia.

Como se puede apreciar en la figura 139 previamente mostrada, se encontraron un total de 10 reglas que coinciden en las tres herramientas: *RapidMiner*, *SAS Enterprise Miner* y *Python*, sin embargo, se encontraron los siguientes resultados conforme al diagrama de Venn:

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- En el círculo que representa a *RapidMiner*: 15 reglas que no coinciden con las otras dos herramientas.
- En el círculo que representa a *SAS Enterprise Miner*: 15 reglas que no coinciden con las otras dos herramientas.
- En el círculo que representa a *Python*: 15 reglas que no coinciden con las otras dos herramientas.
- En la intersección entre *RapidMiner* y *SAS*: 5 reglas en común.
- En la intersección entre *RapidMiner* y *Python*: 5 reglas en común.
- En la intersección entre *SAS Enterprise Miner* y *Python*: 5 reglas en común.

Después de obtener los resultados, se eligieron las 10 reglas que coincidieron en las tres herramientas para su análisis e interpretación. Además, se tomó en cuenta el valor informativo de las reglas para su selección. Por otra parte, se omitió la interpretación de la variable consecuente de las reglas, ya que en todos los casos esta variable es P9. Totalmente en desacuerdo, lo que sugiere que el profesor ha invertido mucho tiempo en prepararse para usar la tecnología en su práctica docente.

Reglas generadas en RapidMiner, SAS Enterprise Miner y Python

1. {P10 = Totalmente de acuerdo, P12 = Totalmente de acuerdo, P11 = Totalmente de acuerdo} = P9 Totalmente de acuerdo.

Fueron 253 profesores (31% de los encuestados) que cumplieron esta regla, afirmando que estuvieron totalmente de acuerdo con que:

- El equipo tecnológico de apoyo a las actividades docentes es escaso,

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- La accesibilidad a la red en los espacios universitarios es inestable,
- La conectividad de la red representa un problema recurrente,

Coincidiendo en que existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.

2. {CONT = Base, P18 = De acuerdo, P10 = De acuerdo, TW = 0} = CAT Profesor de Tiempo Completo.

Un total de 305 encuestados (37% de los profesores) con contratación de Base, estuvieron de acuerdo con que:

- La institución carece de un sistema que evalúe las competencias tecnológicas de su personal,
- El equipo tecnológico de apoyo a las actividades docentes es escaso,

Sin embargo, nunca han utilizado Twitter para trabajar con sus estudiantes y cuentan con una categoría Profesor de Tiempo Completo.

3. {P9 = Totalmente de acuerdo, P12 = Totalmente de acuerdo, P11 = Totalmente de acuerdo} = P10 Totalmente de acuerdo.

Un total de 150 docentes (18% de los encuestados) coincidieron con esta regla, en donde estuvieron totalmente de acuerdo:

- Existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT,
- La accesibilidad a la red en los espacios universitarios es inestable,
- La conectividad de la red representa un problema recurrente,

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Finalmente coincidieron con que el equipo tecnológico de apoyo a las actividades docentes es escaso.

4. {P16 = De acuerdo, P15 = De acuerdo, CAT = Profesor de Tiempo Completo} = CONT Base, FAC = 0.

Un total de 516 profesores coincidieron en estar de acuerdo con:

- La estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente,
- Existe una política institucional poco clara que promueva la adopción de tecnología en el aula,
- Ser profesores pertenecientes a la categoría de Profesor de Tiempo Completo,

Resultando en ser docentes con contratación de Base en la UJAT pero que nunca han utilizado la red social Facebook para trabajar con sus estudiantes.

5. {P14 = De acuerdo, P13 = De acuerdo, P17 = De acuerdo} = CAT Profesor de Tiempo Completo.

En este caso 305 docentes con categoría de Tiempo Completo, coincidieron en estar de acuerdo con que:

- Que te ayuden a resolver problemas de *software* en la UJAT resulta complicado,
- Cuando tengo un problema con algún equipo es difícil conseguir apoyo técnico,

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- La institución carece de un sistema que diagnostique las competencias tecnológicas de su personal.

6. {GRAD = Doctorado, P12 = Totalmente de acuerdo, P11 = Totalmente de acuerdo, P18 = Totalmente de acuerdo} = P1 DIEZ.

Se encontró que 467 profesores de la UJAT con grado académico de Doctorado estuvieron totalmente de acuerdo con que:

- La accesibilidad a la red en los espacios universitarios es inestable,
- La conectividad de la red representa un problema recurrente,
- La institución carece de un sistema que evalúe las competencias tecnológicas de su personal,

Además, indicaron que hace aproximadamente 10 años comenzaron a usar computadora para apoyar sus actividades docentes.

7. {P18 = Totalmente de acuerdo, P11 = Totalmente de acuerdo, CONT = Base} = P12 Totalmente de acuerdo, P17 Totalmente de acuerdo.

En esta regla 391 docentes respondieron estar totalmente de acuerdo con:

- La institución carece de un sistema que evalúe las competencias tecnológicas de su personal,
- La conectividad de la red representa un problema recurrente,
- La accesibilidad a la red en los espacios universitarios es inestable,

Además, todos ellos formaban parte del tipo de contratación de Base en la UJAT y por otra parte estuvieron de acuerdo con que la institución carece de un sistema que diagnostique las competencias tecnológicas de su personal.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

8. {P8 = DIEZ, CAT = Profesor de Tiempo Completo, P17 = De acuerdo} = P18 De acuerdo.

Son 431 profesores que desde hace aproximadamente 10 años utilizan el correo electrónico para trabajar con sus estudiantes y pertenecen a la categoría de Profesor de Tiempo Completo, así mismo están de acuerdo con que:

- La institución carece de un sistema que diagnostique las competencias tecnológicas de su personal,
- La institución carece de un sistema que evalúe las competencias tecnológicas de su personal.

9. {P16= De acuerdo, P9 = De acuerdo, AVA = 10, CONT = Base} = P10 De acuerdo.

En esta regla un total de 407 profesores con contratación de Base en la UJAT y que hace aproximadamente 10 años utilizaron por primera vez un Ambiente Virtual de Aprendizaje, coincidieron en que estaban de acuerdo en:

- La estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente,
- Existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT.

10. {P15 = Ni de acuerdo ni en desacuerdo, P18 = Ni de acuerdo ni en desacuerdo, P13 = Ni de acuerdo ni en desacuerdo} = P14 Ni de acuerdo ni en desacuerdo.

Se encontró que un total de 221 docentes respondieron que no estuvieron ni de acuerdo ni en desacuerdo con que:

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Existe una política institucional poco clara que promueva la adopción de tecnología en el aula,
- La institución carece de un sistema que evalúe las competencias tecnológicas de su personal,
- Cuando tengo un problema con algún equipo es difícil conseguir apoyo técnico,

Así mismo, estuvieron no estuvieron ni de acuerdo ni en desacuerdo con que te ayuden a resolver problemas de *software* en la UJAT resulta complicado.

Para el caso de las reglas de asociación 1, 3, 4, 5, 9 y 10 logran identificar que la dimensión de recursos disponibles y sistema educativo se encuentra atenuada dentro de los patrones hallados, esto se debe a que domina el hecho de que los profesores si han invertido tiempo en prepararse para usar la tecnología en su práctica docente, sin embargo, en base a los datos proporcionados a partir de las reglas mencionadas parece haber una preocupación compartida entre los profesores encuestados sobre las carencias en infraestructura tecnológica en la UJAT.

En particular, se destaca que un porcentaje significativo de 47% de profesores con contrato de base coincide en que la institución carece de un sistema para evaluar las competencias tecnológicas de su personal, tal como se observó en las reglas de asociación 2 y 7, la cuales tienen que ver con las dimensiones sociodemográficas, uso de recursos tecnológicos y recursos disponibles y sistema educativo; lo que podría indicar una necesidad de desarrollo profesional en esta área, así como de mejorar la formación y capacitación en tecnología para el personal docente.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Por otro lado, la falta de acceso a la red y la escasez de equipo tecnológico de apoyo, como se menciona en las reglas 7 y 10, asociadas con la dimensión de recursos disponibles, pueden afectar negativamente la calidad de la enseñanza y el aprendizaje en la institución, esto debido a que pueden impedir el uso efectivo de tecnología en el aula y en las actividades docentes en general. Esto puede limitar la capacidad de los profesores para utilizar herramientas y recursos tecnológicos en su enseñanza y dificultar la comunicación y colaboración con los estudiantes.

Además, puede limitar la capacidad de los estudiantes para acceder a materiales y recursos en línea, así como para participar en actividades en línea y colaborar con sus compañeros. Es por esto que la percepción de los profesores con respecto a la infraestructura tecnológica y la estructura administrativa, como se mencionó en las reglas de asociación 6 y 8 (relacionadas con la dimensión de uso de recursos tecnológicos), puede ser un indicador de la necesidad de mejorar la gestión y el soporte técnico para el uso efectivo de la tecnología en la UJAT, puesto que los profesores manifestaron que necesitaban tener acceso a una infraestructura tecnológica adecuada y un soporte técnico efectivo para poder utilizar la tecnología de manera eficaz en sus actividades docentes.

Si la infraestructura y la estructura administrativa son inadecuadas, puede resultar difícil para los profesores adoptar nuevas tecnologías y utilizarlas para mejorar la calidad de la enseñanza y el aprendizaje. Mejorar la gestión y el soporte técnico puede ayudar a superar estas barreras y permitir a los profesores aprovechar al máximo las tecnologías disponibles para mejorar la experiencia educativa de los estudiantes.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Así mismo, como se observó en la regla de asociación 6, la cual pertenece a la dimensión de sociodemográficos y recursos disponibles y sistema educativo, los profesores con grado académico de Doctorado destacaron que la inestabilidad de la accesibilidad a la red y la carencia de un sistema que evalúe las competencias tecnológicas del personal de la institución fueron unas de las principales barreras que afrontaron ante el uso de las innovaciones tecnológicas; estos resultados también indicaron que los docentes con Doctorado suelen tener una mayor formación en tecnología y una mayor experiencia en su uso en la enseñanza y el aprendizaje.

Por lo tanto, son más conscientes de la importancia de contar con una infraestructura tecnológica adecuada y de la necesidad de evaluar las competencias tecnológicas del personal de la institución para garantizar un uso efectivo de la tecnología en la educación.

Es interesante observar que este grupo mencionó también haber comenzado a utilizar la computadora para apoyar sus actividades docentes hace aproximadamente 10 años y que de la misma manera habían trabajado desde hace aproximadamente 10 años con sus estudiantes a través del correo electrónico, lo que sugirió entonces una brecha generacional en cuanto al uso de la tecnología, esta se debió a que los profesores con grado académico de Doctorado pertenecían a una generación que no creció en un entorno tecnológico, y por lo tanto, tenían menos habilidades tecnológicas y menos experiencia en el uso de herramientas digitales en comparación con generaciones más jóvenes que habían estado expuestas a la tecnología desde una edad temprana.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Además, es posible que estos profesores hayan recibido una formación académica en una época en la que la tecnología no era una herramienta común en la educación. Es necesario destacar que las redes sociales pueden jugar un papel muy importante en la educación, si se saben emplear de la manera correcta; sin embargo para el caso de los docentes de la UJAT la falta de uso redes sociales como Facebook y Twitter (reglas de asociación 2 y 4) para trabajar con los estudiantes indicó que existía una falta de adopción de herramientas digitales en la práctica docente, así como la eficacia de las actividades, el uso de tecnologías como Facebook y Twitter pueden ser útiles para apoyar las actividades docentes en la educación superior, ofreciendo una serie de beneficios como la posibilidad de comunicación constante y el acceso a recursos y materiales educativos en línea.

Además, es necesario destacar la importancia de que los profesores reciban capacitación y formación en el uso de estas herramientas para poder aprovechar su potencial de manera efectiva.

Por otro lado, en la regla de asociación 9 relacionada con la dimensión de recursos disponibles, se encontró que los profesores con contratación de Base en la UJAT que han utilizado por primera vez un Ambiente Virtual de Aprendizaje hace aproximadamente 10 años coinciden en que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente y existen carencias en infraestructura tecnológica para el soporte de las actividades docentes.

Por último, la falta de claridad o información insuficiente sobre ciertas cuestiones relacionadas con la adopción de tecnología en la UJAT puede ser una oportunidad para

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

proporcionar información y recursos adicionales que ayuden a los profesores a comprender mejor la situación y tomar decisiones más informadas en cuanto a la adopción de tecnologías en las aulas físicas, no solo de manera virtual.

Finalmente, en la regla de asociación 10 asociada con la dimensión de recursos disponibles, se encontró un grupo de 221 profesores que no están ni de acuerdo ni en desacuerdo con algunas afirmaciones relacionadas con la adopción de tecnología en la UJAT, lo que puede indicar que existe una falta de claridad o información insuficiente sobre estas cuestiones por parte de los docentes. De acuerdo con la identificación de los patrones con la técnica de reglas de asociación y su relación con las dimensiones: sociodemográficas, uso de recursos tecnológicos y recursos disponibles y sistemas educativos, queda expuesto que existen reglas donde se demuestra la afectación de estas barreras en los profesores de la UJAT.

Cabe mencionar que esta regla de asociación resultó ser diferente a las demás, ya que en lugar de mostrar una correlación entre variables, indicó una falta de correlación o ambigüedad en la percepción de los docentes sobre el tema de la tecnología en la educación, mientras las otras 9 reglas se relacionaron con la importancia del acceso a recursos tecnológicos, principalmente de la accesibilidad a la red en los espacios universitarios y con que la conectividad de la red fue un problema recurrente para ellos.

4.6 Análisis de usabilidad

En esta sección se presentan en la tabla 25 los resultados obtenidos del análisis de usabilidad de las herramientas de minería de datos *RapidMiner*, *SAS Enterprise Miner* y *Python*. La usabilidad de estas herramientas fue evaluada de acuerdo con la escala de

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

usabilidad SUS (*System Usability Scale*), que se enfoca en medir la facilidad de uso percibida por los usuarios. Esta escala consta de 5 ítems (facilidad, eficiencia, flexibilidad, satisfacción del usuario y capacidad de integración) y se utiliza para obtener una puntuación numérica que indica el nivel de usabilidad de una herramienta (ver tabla 25).

Tabla 25.

Escala de usabilidad SUS utilizada

Escala	Descripción	Valor asignado
Excelente	La herramienta tuvo un funcionamiento óptimo, rápido y responsivo, facilitando al usuario su uso, así como navegación.	3
Bueno	La herramienta tuvo un funcionamiento aceptable, sin embargo, presentó ralentizaciones, lo que en algunos momentos retrasó su uso, así como navegación.	2
Malo	La herramienta no tuvo un funcionamiento adecuado, presentó diversos fallos durante su uso y la navegación en ella resultó tediosa.	1

Nota: Elaboración propia.

En la tabla 26, se presenta la matriz de calificación de las herramientas, para tal caso los cinco atributos considerados para evaluación contienen tres características que sirvieron de referencia. En ese sentido, cabe señalar que el total de características fueron quince, considerando que el puntaje más alto de acuerdo con la escala es 3, entonces la puntuación máxima que pueden obtener las herramientas es 45. La tabla 26 proporciona una visión general de cómo se calificaron las herramientas en cada uno de los atributos

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

evaluados, lo que permite comparar y determinar cuál herramienta obtuvo la puntuación más alta en general (ver tabla 26).

Tabla 26.

Puntuación de las herramientas en escala de usabilidad SUS

Parámetros	Características	RapidMiner	SAS Enterprise Miner	Python
Facilidad	• Claridad y estética de la interfaz	2	3	1
	• Accesibilidad de la documentación	3	3	3
	• Tutoriales proporcionados por la herramienta	2	3	3
Eficiencia	• Simplicidad	2	3	3
	• Accesibilidad de las opciones	2	2	2
	• Menús en la herramienta	3	3	1
Flexibilidad	• Capacidad de personalización de la herramienta para adaptarse a las necesidades del usuario	2	3	1
	• Facilidad para agregar o quitar funcionalidades según las necesidades del usuario	2	3	1
	• Posibilidad de automatizar tareas para una mayor eficiencia en el uso de la herramienta	1	3	1
Satisfacción del usuario	• Experiencia general del uso	2	3	2
	• Retroalimentación	3	3	2
	• Capacidad de la herramienta para cumplir con las expectativas del usuario	2	3	2
Capacidad de integración	• Capacidad de la herramienta para integrarse con otras herramientas o sistemas	2	2	3
	• Facilidad para importar y exportar datos desde y hacia otras herramientas o sistemas	3	3	3
	• Capacidad de la herramienta para trabajar con múltiples fuentes de datos y formatos de archivos	3	3	3

Nota: Elaboración propia con base en Musyaffa *et al.* (2023).

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Basándonos en las características evaluadas en la matriz previa, se presentan en la tabla 27 los resultados clasificados por parámetros, los cuales fueron obtenidos tras el análisis de las tres herramientas de minería de datos. En dicha evaluación, se concluyó que *SAS Enterprise Miner* demostró una experiencia superior en términos de usabilidad en comparación con las otras dos herramientas, obteniendo una puntuación de 43.

Tabla 27.

Resultados obtenidos de la evaluación por herramienta

Parámetros	Herramientas		
	<i>RapidMiner</i>	<i>SAS Enterprise Miner</i>	<i>Python</i>
Facilidad de aprendizaje	7	9	7
Eficiencia	7	8	6
Flexibilidad	5	9	3
Satisfacción del usuario	7	9	6
Capacidad de integración	8	8	9
Total	34	43	31

Nota: Elaboración propia.

En este capítulo se llevaron a cabo pruebas de usabilidad con las herramientas de minería de datos *RapidMiner*, *SAS Enterprise Miner* y *Python*. Estas pruebas permitieron realizar un análisis para emitir una recomendación en el capítulo siguiente. Las pruebas de evaluación de usabilidad también contribuyeron a la selección de los modelos que se utilizaron para analizar los resultados obtenidos en las herramientas de minería de datos *RapidMiner*, *SAS Enterprise Miner* y *Python*. Durante la exposición de resultados, se pudo observar que cada herramienta de minería de datos utilizada (*RapidMiner*, *SAS Enterprise Miner* y *Python*) proporcionó métricas de calidad y patrones identificados a través de diferentes técnicas, como clasificación, agrupación y reglas de asociación.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

En cuanto a las métricas de calidad, se evaluaron aspectos como la precisión, la exactitud y la confianza, entre otros, para determinar qué tan bien se ajustaban los modelos a los datos y qué tan confiables eran las predicciones realizadas. Cada herramienta proporcionó sus propias métricas y se pudieron comparar para obtener una visión más completa y precisa de los resultados.

En relación con los patrones identificados, se utilizaron técnicas como la clasificación para asignar categorías a los datos, la agrupación para encontrar similitudes entre los elementos y la generación de reglas de asociación para descubrir relaciones entre los diferentes elementos del conjunto de datos. Estos patrones ayudaron a obtener *insights* y conocimientos importantes sobre los datos analizados.

Durante el análisis de los resultados, se encontraron tanto coincidencias como diferencias entre las herramientas. Las coincidencias en los patrones identificados y las métricas de calidad obtenidas en las tres herramientas proporcionaron mayor confianza en los resultados, ya que se validaron a través de diferentes enfoques. Con base en los hallazgos identificados en el análisis, se concluye que *SAS Enterprise Miner* ha demostrado ser la herramienta con mejor experiencia en usabilidad, obteniendo altas puntuaciones en los atributos evaluados. Esto sugiere que *SAS Enterprise Miner* ofrece una interfaz intuitiva y fácil de usar, lo que facilita la exploración y el análisis de datos.

Capítulo 5. Conclusiones, recomendaciones y trabajos futuros

En este capítulo, se presentarán las conclusiones obtenidas a partir del análisis y desarrollo de esta investigación. Además, se proporcionarán recomendaciones basadas en los resultados obtenidos y se discutirán posibles direcciones para futuros trabajos e investigaciones en el área.

Las conclusiones extraídas de esta investigación son fundamentales para evaluar el logro de los objetivos establecidos inicialmente. Se analizarán los resultados obtenidos en relación con los criterios de éxito y se discutirá en qué medida se han cumplido. Además, se destacarán los principales hallazgos y contribuciones del proyecto/estudio, resaltando su relevancia e impacto en el campo de estudio.

En base a los resultados y conclusiones obtenidos, se presentarán recomendaciones prácticas que puedan guiar a los profesionales, investigadores o interesados en el tema a tomar decisiones informadas. Estas recomendaciones estarán respaldadas por los hallazgos del estudio y se enfocarán en acciones concretas que podrían implementarse para mejorar la situación, resolver desafíos identificados o aprovechar oportunidades detectadas. Finalmente se propondrán temas o aspectos específicos que podrían ser explorados en futuros trabajos, teniendo en cuenta las áreas que requieren mayor profundización, los enfoques alternativos que podrían considerarse y las posibles extensiones o aplicaciones del estudio actual.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

5.1 Conclusiones

1 Los profesores de la UJAT han compartido sus opiniones sobre los obstáculos que enfrentan para adoptar tecnologías innovadoras, lo cual respalda la idea presentada en este documento de que toda investigación científica y académica debe estar delimitada. En este trabajo se identificaron tres dimensiones importantes: características sociodemográficas, uso de recursos tecnológicos y recursos disponibles y sistema educativo. Cabe mencionar que llevar a cabo una investigación implica un proceso para generar nuevo conocimiento, ya sea a partir de estudios previos o de una idea original.

A lo largo de las diferentes etapas de este estudio, que incluyen las generalidades, el marco de la investigación, el desarrollo metodológico y los resultados presentados en cada capítulo, se han establecido conclusiones basadas en tres elementos principales: el logro del objetivo general y de los objetivos específicos, así como la exposición de los hallazgos obtenidos en los resultados que responden a la pregunta de investigación planteada.

1 El objetivo general de la investigación fue identificar patrones de conducta en profesores de la UJAT ante el uso de innovaciones tecnológicas mediante técnicas de minería de datos. Para lograr esto, se establecieron cinco objetivos específicos que incluyeron la identificación de la herramienta adecuada de minería de datos, la aplicación de técnicas de árbol de decisión, agrupamiento y reglas de asociación para representar y analizar los datos, y la interpretación de los resultados obtenidos.

Según lo expuesto anteriormente, se ha cumplido con los objetivos específicos al haber utilizado la técnica de árboles de decisión para la clasificación, la técnica de

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Kmeans para la agrupación en dos grupos utilizando cada herramienta y las reglas de asociación para encontrar coincidencias entre las reglas obtenidas de las tres herramientas seleccionadas. Cada técnica fue aplicada en *RapidMiner*, *SAS Enterprise Miner* y *Python*. Los resultados obtenidos fueron analizados e interpretados para identificar patrones de conducta. Como resultado, se puede afirmar que se logró completar el objetivo general.

Conforme a los resultados presentados se destaca que en *RapidMiner*, *SAS Enterprise Miner* y *Python*, al implementar los árboles de decisión coincidieron con el tipo de clasificación designado, donde se presentaron similitudes y relaciones que se pueden identificar entre los patrones. Por ejemplo, se constató que el 40% de los profesores (327) coincidían en que la accesibilidad a la red en los espacios universitarios era inestable, lo que sugería un problema generalizado que afectaba a gran parte del personal docente.

Asimismo, se encontró que la institución carecía de un sistema que evaluara o diagnosticara las competencias tecnológicas de su personal, lo que reflejaba una falta de atención o recursos destinados a la capacitación tecnológica del personal docente.

También se observó que, aunque algunos profesores tenían experiencia previa en el uso de tecnología en sus actividades docentes, esto no necesariamente se traducía en habilidades tecnológicas desarrolladas. En efecto, los patrones indicaban que algunos profesores que llevaban más tiempo usando tecnología en sus actividades docentes estaban más de acuerdo en que habían invertido poco tiempo en prepararse para usar la tecnología. Otro aspecto que se destacó fue la dificultad para obtener apoyo técnico o resolver problemas de *software* en la universidad, lo que sugería que había una falta de

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

recursos o personal dedicado al soporte técnico. Esto podía ser un obstáculo para la adopción de la tecnología en el aula.

1 Además, se encontró una relación entre la experiencia académica y las opiniones de los profesores sobre la tecnología en la universidad. Por ejemplo, los profesores con grado de Doctorado tendían a estar más de acuerdo en que habían invertido poco tiempo en prepararse para usar la tecnología. Por otro lado, se destacó que los profesores que utilizan redes sociales y ambientes virtuales de aprendizaje para trabajar con sus estudiantes tienden a estar más conscientes de las carencias en infraestructura tecnológica y en la inadecuada estructura administrativa para el uso de tecnología en la universidad.

También se constató que 34% de los profesores (283) coincidían en que la política institucional no estaba clara o no promovía adecuadamente la adopción de la tecnología en el aula, lo que sugiere que había un problema más profundo en la cultura institucional en relación con la tecnología. Estos patrones pusieron de relieve la necesidad de una política más clara y una mayor atención a la capacitación y el soporte técnico para el personal docente. Aunque en este aspecto *RapidMiner*, *SAS Enterprise Miner* y *Python* coincidieron, cada herramienta generó información distinta que se logra complementar, consiguiendo generar una mejor interpretación. Por ejemplo, en el caso de *RapidMiner* parece haber una falta de consenso entre los profesores sobre la infraestructura tecnológica y la accesibilidad a la red en los espacios universitarios. Algunos profesores indicaron que existen carencias en la infraestructura tecnológica, mientras que otros no están de acuerdo con esta afirmación. Además, algunos profesores indicaron que la

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

accesibilidad a la red es inestable, mientras que otros no están de acuerdo ni en desacuerdo.

Sin embargo, parece haber un consenso general en que los profesores no han invertido suficiente tiempo en prepararse para utilizar la tecnología en su práctica docente (30% o 244). Esto sugiere que podría ser útil implementar más programas de formación y capacitación en tecnología para los profesores, lo que podría mejorar su eficacia en el uso de herramientas digitales y mejorar la calidad de la educación. También se destaca la necesidad de proporcionar un apoyo técnico adecuado para los problemas técnicos que los profesores puedan encontrar al utilizar tecnología en sus clases, en el caso de *SAS Enterprise Miner* en particular, los profesores que utilizan redes sociales y ambientes virtuales de aprendizaje desde hace más de 13.5 años parecen ser los más críticos en cuanto a la accesibilidad a la red y la política institucional.

Por otro lado, se observa que los profesores que comenzaron a usar computadoras más temprano en su carrera docente están más conscientes de la falta de un sistema para diagnosticar las competencias tecnológicas de los profesores en la institución. En general, estos resultados sugieren que los profesores más experimentados en el uso de tecnología en su práctica docente tienen una perspectiva más crítica de la infraestructura tecnológica y las políticas institucionales en la universidad. Por último, en el caso de *Python*, complementa lo previamente mencionado puesto que se menciona nuevamente que la inestabilidad y accesibilidad a la red era un problema recurrente para los docentes.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Por otro lado, de acuerdo con el *dataset*, se obtuvo que el 53% de los docentes encuestados pertenecían al género Masculino, mientras que el 47% restante eran del género Femenino, así mismo un 59% de encuestados o 618 docentes tenían edades que oscilaban entre los 24 y 52 años, mientras que 208 profesores o el 41% restante, tenían edades comprendidas entre más de 52 y hasta 80 años, estos datos fueron empleados para el algoritmo *Kmeans*, donde se identificaron patrones clasificados por uso de recursos tecnológicos mismos que fueron obtenidos a través de la técnica de agrupación de minería de datos, donde se logró identificar información coincidente en las tres herramientas.

Por ejemplo, se reveló que el 6% de los profesores en la UJAT no están de acuerdo con la afirmación de que hay carencias en la infraestructura tecnológica para el soporte de las actividades docentes, mientras que el 18% está en desacuerdo. Por otro lado, el 13% no tiene una posición clara al respecto, pero el 40% está de acuerdo y el 22% está totalmente de acuerdo. En general, se puede inferir que existe una percepción generalizada entre los profesores de la UJAT de que hay carencias en la infraestructura tecnológica para el soporte de las actividades docentes. Aunque una minoría significativa de los profesores no está de acuerdo con esta afirmación, la mayoría de los encuestados parece estar de acuerdo o totalmente de acuerdo. Por lo tanto, se puede decir que es importante que la UJAT preste atención a las necesidades tecnológicas de los profesores para garantizar que puedan llevar a cabo sus actividades docentes de manera efectiva y eficiente. No obstante, los resultados del análisis de los *clústeres* sugieren que, al menos

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

en la percepción de los profesores encuestados, la infraestructura tecnológica en la UJAT no es una limitante significativa para el desempeño docente.

Por otra parte, se encontró que el 74% de los encuestados están de acuerdo o totalmente de acuerdo en que la conectividad de la red representa un problema recurrente en la institución. Aunque una minoría significativa de 120 profesores (16%) no percibe este problema, los resultados sugieren que es importante mejorar la infraestructura tecnológica relacionada con la conectividad de la red para satisfacer las necesidades de la mayoría de los profesores.

En detalle, se encontró que el 13% de los profesores (106 docentes) estaban totalmente en desacuerdo con la afirmación de que era difícil conseguir apoyo técnico cuando tenían un problema con algún equipo. Esto indica que consideraban que era fácil obtener el apoyo necesario en caso de necesitar asistencia técnica. Por otro lado, un 28% (233 docentes) estuvo en desacuerdo con la afirmación, lo que sugiere que experimentaron dificultades o limitaciones al tratar de obtener apoyo técnico en situaciones relacionadas con problemas de equipos.

En cuanto a los profesores que manifestaron una posición neutral, se observó que el 20% (168 docentes) no estaban ni de acuerdo ni en desacuerdo con la afirmación. Esto indica que tenían una opinión neutral o no tenían una postura clara sobre la facilidad o dificultad de conseguir apoyo técnico en caso de problemas con equipos. Sin embargo, una cifra mayor, el 25% (208 docentes), estuvo de acuerdo con la afirmación, lo que indica que consideraron que era difícil obtener apoyo técnico en caso de tener problemas con equipos.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Un 14% de los profesores (111 docentes) respondieron que estaban totalmente de acuerdo, lo que sugiere que experimentaron una dificultad significativa para conseguir apoyo técnico en situaciones relacionadas con problemas de equipos. Estos resultados reflejan una variedad de opiniones entre los profesores en relación con la disponibilidad y facilidad de obtener apoyo técnico cuando enfrentan problemas con equipos. Estos hallazgos pueden ser útiles para identificar áreas de mejora y brindar un mejor soporte técnico a los profesores en la institución.

De los resultados se puede decir que, aunque la mayoría de los profesores no perciben dificultades para conseguir apoyo técnico cuando tienen un problema con algún equipo, todavía hay un porcentaje significativo de encuestados (41%) que están de acuerdo o totalmente de acuerdo en que esto puede ser un problema. Por lo tanto, sería importante revisar y mejorar el servicio de soporte técnico para garantizar que todos los profesores tengan acceso a la ayuda que necesitan cuando se enfrentan a problemas con los equipos, incluso en casos aislados.

En los resultados obtenidos, se pudo observar que una proporción significativa de los profesores tenía opiniones diferentes con respecto a la afirmación "Es complicado que te ayuden a resolver problemas de *software* en la UJAT". En detalle, se identificó que el 11% de los profesores (87 docentes) estaban totalmente en desacuerdo con esta afirmación, lo que indica que no consideraban difícil recibir ayuda para resolver problemas de *software* en la UJAT. Por otro lado, un 27% (220 docentes) estuvo en desacuerdo, lo que sugiere que tenían ciertas dificultades o limitaciones al recibir ayuda en este aspecto.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

En cuanto a los profesores que manifestaron una posición neutral, se encontró que el 26% (211 docentes) no estaban ni de acuerdo ni en desacuerdo con la afirmación. Esto indica que tenían una opinión neutral o no tenían una postura clara sobre la facilidad o dificultad de recibir ayuda en temas de *software*. Sin embargo, una cifra mayor, el 23% (186 profesores), estuvo de acuerdo con la afirmación, lo que indica que experimentaron dificultades para recibir ayuda en la resolución de problemas de *software* en la UJAT. Finalmente, un 15% de los profesores (122 docentes) respondieron que estaban totalmente de acuerdo, lo que sugiere que consideraban muy complicado recibir asistencia en temas de *software* en la institución.

De acuerdo con los resultados obtenidos, la mayoría de los profesores no consideran que sea complicado obtener ayuda para resolver problemas de *software* en la UJAT. Sin embargo, hay una proporción significativa de 307 encuestados (38%) que están de acuerdo o totalmente de acuerdo con esta afirmación. Esto sugiere que puede haber casos aislados en los que los profesores encuentren dificultades para recibir ayuda en el área de *software*, aunque en general la situación no es grave. Por lo tanto, sería importante seguir mejorando el servicio de soporte técnico para garantizar que todos los profesores tengan acceso a la ayuda que necesitan cuando se enfrentan a problemas con el *software*.

Los resultados del análisis de *clústeres* indicaron que un porcentaje significativo de los encuestados perciben que la estructura administrativa de la UJAT no es adecuada para establecer un sistema de utilización de tecnología eficiente. Aunque 60 profesores (7%) estaban totalmente en desacuerdo y otros 177 docentes (21%) estuvieron en

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

desacuerdo, más de la mitad, 395 profesores o el 48% de los encuestados estuvo de acuerdo o totalmente de acuerdo con esta afirmación. Esto sugiere que hay un desafío en la estructura administrativa de la UJAT para facilitar una adopción efectiva de la tecnología en el aula.

Por otro lado, la mayoría de los encuestados (63%) están de acuerdo en que la institución carece de un sistema que evalúe las competencias tecnológicas de su personal, o bien 517 profesores. Solo el 14% (121 docentes) está en desacuerdo o totalmente en desacuerdo con esta afirmación, mientras que un 23% (188) no mostró una postura clara al respecto. Estos resultados sugieren que hay una necesidad percibida de implementar un sistema de evaluación de competencias tecnológicas en la institución, lo que indica que es una cuestión importante para una gran cantidad de profesores.

Finalmente, los resultados muestran que la mayoría de los profesores encuestados han invertido tiempo en prepararse para usar la tecnología en su práctica docente, con un 70% de ellos estando en desacuerdo o totalmente en desacuerdo con la afirmación de que han invertido poco tiempo. Solo un 16% estuvo de acuerdo con la afirmación, lo que sugiere que hay un grupo minoritario de 149 profesores que siente que ha dedicado poco tiempo a la preparación para el uso de la tecnología en la enseñanza.

Se puede concluir que la mayoría de los profesores encuestados han invertido tiempo en prepararse para el uso de la tecnología en su práctica docente. Esto es positivo, ya que sugiere que existe un esfuerzo por parte de los profesores para adaptarse a las nuevas herramientas tecnológicas y mejorar la calidad de su enseñanza. Sin embargo, también es importante destacar que un grupo minoritario de profesores siente

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

que ha invertido poco tiempo en la preparación para el uso de la tecnología, lo que indica que aún hay espacio para mejorar y para fomentar la capacitación en tecnología educativa. Estos resultados sugieren que hay un compromiso por parte de los profesores en mejorar su práctica docente a través del uso de la tecnología, pero aún hay desafíos por enfrentar en cuanto a la preparación y capacitación en este ámbito.

En el análisis previo mediante árboles de decisión y la técnica de agrupación, se destacaron las barreras que impidieron a los profesores adoptar innovaciones tecnológicas en su enseñanza, en gran parte debido a problemas de accesibilidad y conectividad en los espacios universitarios, lo que se relaciona con la falta de recursos disponibles y deficiencias en el sistema educativo. Estos obstáculos afectan los aspectos pedagógicos y el ambiente académico, generando consecuencias en las tres dimensiones estudiadas. Sin embargo, al aplicar las reglas de asociación, también se obtuvo información relevante enfocada en los propios docentes.

En el caso de las reglas de asociación, se empleó el algoritmo A priori en *RapidMiner*, *SAS Enterprise Miner* y *Python*, donde 10 de las reglas resultantes fueron coincidentes en las herramientas empleadas. Destacando tres aspectos, primero el caso de que existe un total de 516 profesores en la UJAT que están de acuerdo en que la estructura administrativa y la política institucional son inadecuadas para promover la adopción de tecnología en el aula. Además, se trata de profesores de tiempo completo y con contrato base en la UJAT, lo que sugiere que son docentes con cierta estabilidad laboral y experiencia en la institución.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Sin embargo, a pesar de su experiencia y posición en la institución, este grupo de profesores nunca ha utilizado Facebook para trabajar con sus estudiantes, lo que podría indicar una falta de interés o conocimiento sobre el uso de esta herramienta en la enseñanza. Estos resultados sugieren que existen barreras institucionales y pedagógicas que impiden una adopción más amplia y efectiva de la tecnología en la enseñanza, y que hay oportunidades de mejora en la formación y capacitación de los docentes en este aspecto.

Para el segundo caso, es que un grupo significativo de 467 profesores con grado académico de Doctorado en la UJAT percibe la inestabilidad de la accesibilidad a la red y la conectividad como problemas recurrentes que afectan su práctica docente, así como la falta de un sistema de evaluación de competencias tecnológicas en la institución. Además, se puede inferir que estos profesores han estado utilizando la computadora como herramienta para apoyar sus actividades docentes desde hace al menos 10 años.

Esta información puede ser relevante para la institución para enfocar los esfuerzos en mejorar la conectividad y la accesibilidad a la red, así como en la implementación de un sistema de evaluación de competencias tecnológicas del personal docente.

En tercer lugar, se encuentra un escenario donde un total de 431 profesores de tiempo completo que han utilizado el correo electrónico para trabajar con sus estudiantes durante aproximadamente 10 años están conscientes de que la institución no tiene un sistema para evaluar y diagnosticar sus competencias tecnológicas.

Esto sugiere que los profesores están preocupados por su propia capacitación y desarrollo en el uso de la tecnología en su práctica docente, y están en busca de una

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

evaluación y diagnóstico que les permita mejorar y avanzar en este aspecto. Además, esto también puede indicar que la institución no ha proporcionado suficiente apoyo y capacitación en tecnología para los profesores, lo que lleva a una preocupación por parte de los docentes y la necesidad de un sistema para evaluar sus habilidades y competencias.

Es importante destacar que las tres situaciones anteriores se encuentran reflejadas en diversos patrones, primordialmente en los relacionados con los árboles de decisión en donde, así como en los *clústeres* relacionados con los docentes en donde el 59% de encuestados o 618 docentes tenían edades que oscilaban entre los 24 y 52 años, mientras que 208 profesores o el 41% restante, tenían edades comprendidas entre 52 y 80 años. En cuanto a las reglas de asociación también se detectó que en todas las reglas resultantes se obtuvieron relaciones con las diferentes barreras, coincidiendo en todos los casos con que la accesibilidad a la red en los espacios universitarios es inestable, la conectividad en la red representa un problema recurrente y que la estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente. A pesar de que los docentes han dedicado tiempo y recursos para prepararse en el uso de tecnología en su práctica docente, es probable que hayan tenido diversos obstáculos que afrontar ante la implementación de innovaciones tecnológicas.

1 A partir de los resultados obtenidos en esta investigación, se puede concluir que las barreras que enfrentan los docentes de la UJAT en el uso de innovaciones tecnológicas están vinculadas con diversas dimensiones, como lo son la sociodemográfica, el uso de recursos tecnológicos y los recursos disponibles en el

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos sistema educativo. Estas dimensiones tienen un impacto significativo en el uso de innovaciones tecnológicas por parte de los profesores. A través de las técnicas de minería de datos empleadas, como los árboles de decisión, agrupación y reglas de asociación, se han identificado los siguientes patrones relacionados con estas barreras (ver tabla 28).

Tabla 28.

Patrones identificados en la investigación

Patrones	Barreras
Se necesita implementar o utilizar más innovaciones tecnológicas para trabajar con los estudiantes	Uso de recursos tecnológicos
Existen carencias en la infraestructura tecnológica en la UJAT	Recursos disponibles y sistema educativo
Problemas constantes de conectividad	Recursos disponibles y sistema educativo
La gestión administrativa no es apropiada para implementar un sistema eficiente de uso de tecnologías en la universidad	Recursos disponibles y sistema educativo
La UJAT no cuenta con un sistema para evaluar las habilidades tecnológicas de su personal.	Recursos disponibles y sistema educativo
La política institucional existente en la UJAT para promover la adopción de tecnología en el aula es poco clara y no brinda las directrices necesarias para fomentar el uso de la tecnología por parte de los docentes	Recursos disponibles y sistema educativo
Se dispone de un número insuficiente de dispositivos tecnológicos para apoyar las actividades docentes	Recursos disponibles y sistema educativo

Nota: Elaboración propia.

Las barreras mencionadas anteriormente no dependen únicamente de los docentes porque están relacionadas con aspectos más amplios y complejos del sistema educativo y de la institución en la que trabajan. Por ejemplo, la accesibilidad a la red y la conectividad pueden estar influenciadas por factores externos como la infraestructura de telecomunicaciones en la región donde se ubica la universidad, la política gubernamental en materia de conectividad, entre otros.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

De igual manera, la estructura administrativa y la política institucional son temas que involucran a toda la institución y no solo a los docentes.

Además, estas barreras seguirán existiendo siempre porque el uso de tecnología en la educación no es un proceso lineal y homogéneo, sino que está influenciado por múltiples factores que van cambiando constantemente. Por ejemplo, la aparición de nuevas tecnologías y su integración en la educación implica la necesidad de actualización constante de los docentes y de los recursos tecnológicos disponibles. Además, la adopción de tecnología en la educación también está influenciada por las tendencias pedagógicas y las expectativas de los estudiantes y de la sociedad en general, lo que hace que la dinámica sea cambiante y evolutiva. En consecuencia, es difícil eliminar por completo estas barreras, aunque sí se pueden identificar y buscar estrategias para reducir su impacto en el proceso de enseñanza-aprendizaje.

Por lo anterior, se concluye que la utilización de técnicas de minería de datos en la investigación permitió identificar patrones de conducta significativos en los profesores de la UJAT. Estos resultados proporcionarán información valiosa para la institución, ya que podrán entender mejor las barreras que enfrentan los docentes en la adopción de las innovaciones tecnológicas, y por lo tanto podrán diseñar políticas y programas que aborden estas barreras de manera más efectiva.

Además, esta investigación puede ser de beneficio para la UJAT al permitir que la institución pueda identificar los recursos y las áreas en las que necesitan mejorar para poder proporcionar un entorno tecnológico adecuado para sus docentes. Al comprender mejor las barreras específicas que enfrentan los profesores, la UJAT podrá desarrollar

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

estrategias específicas para abordarlas y, por lo tanto, mejorar la calidad de la educación que ofrecen. La utilización de técnicas de minería de datos en esta investigación permitió obtener información relevante que puede ser de gran beneficio para la UJAT, ya que les permitirá mejorar su entorno tecnológico, abordar las barreras específicas que enfrentan los docentes ante el uso de las innovaciones tecnológicas y, en última instancia, mejorar la calidad de la educación que ofrecen.

5.2 Discusión

En el contexto actual de la educación superior, la adopción de innovaciones tecnológicas se ha convertido en un tema relevante para garantizar un aprendizaje efectivo y adaptado a las necesidades de los estudiantes. Sin embargo, los profesores de educación superior se enfrentan a diversas barreras que dificultan la adopción y el aprovechamiento pleno de estas innovaciones, incluyendo la minería de datos. La pandemia por COVID-19 ha acelerado la necesidad de utilizar herramientas tecnológicas en la enseñanza, lo que ha llevado a numerosos estudios que exploran las barreras que los docentes enfrentan en este contexto.

En este sentido, se han realizado varios estudios que analizan las barreras y desafíos que enfrentan los profesores de educación superior en la adopción de innovaciones tecnológicas durante la pandemia. Entre estos estudios, destacan cinco investigaciones relevantes que arrojan luz sobre las barreras en la adopción de tecnologías en diferentes contextos y regiones. El estudio realizado por Perdomo (2022), titulado "Docentes y barreras enfrentadas en la enseñanza remota de emergencia por COVID-19", reveló la existencia de obstáculos significativos en el uso de innovaciones

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

tecnológicas, incluyendo la minería de datos, para el proceso de enseñanza-aprendizaje. El trabajo de Pinto *et al.* (2022) titulado "Brecha digital en la formación inicial docente: desafíos en los ambientes de aprendizaje durante la pandemia COVID-19 en La Guajira (Colombia)" destacó la falta de acceso a recursos tecnológicos y habilidades digitales adecuadas como obstáculos para la utilización efectiva de innovaciones tecnológicas.

Por su parte, el estudio de García y Silva (2022) titulado "Percepción académica sobre las barreras en la adopción de innovaciones tecnológicas durante la pandemia por la COVID-19" reveló obstáculos como la falta de capacitación adecuada, la resistencia al cambio y las preocupaciones sobre la privacidad y seguridad de los datos en relación con el uso de la minería de datos y otras innovaciones tecnológicas.

El estudio de Kakepoto *et al.* (2021) "Pedagogical Shift: Faculty Insights about E-Teaching Barrieris during COVID Pandemic" se centró en las barreras enfrentadas por los profesores universitarios en Pakistán durante la pandemia de COVID-19 al hacer la transición a la enseñanza en línea. Los resultados destacaron la resistencia a adquirir nuevas habilidades tecnológicas, la falta de programas de desarrollo docente en TI/Computación y la necesidad de que los profesores adopten un papel activo en la implementación exitosa de la enseñanza en línea.

Finalmente, el estudio de García *et al.* (2021) titulado "The Transformation of Higher Education After the COVID Disruption: Emerging Challenges in an Online Learning Scenario" identificó diversas barreras y desafíos desde la perspectiva de estudiantes, profesores e instituciones universitarias. Entre los desafíos mencionados se encuentran el problema técnico y la brecha digital para los estudiantes, la falta de capacitación en

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

técnicas en línea para los profesores y la necesidad de mejorar la infraestructura tecnológica y promover una cultura participativa a nivel institucional.

Los cinco estudios mencionados previamente proporcionan una visión amplia de las barreras y desafíos que los profesores de educación superior enfrentan en la adopción y el uso efectivo de innovaciones tecnológicas, incluida la minería de datos, en el contexto de la pandemia de COVID-19. Aunque los estudios se realizaron en diferentes países, comparten algunas similitudes en cuanto a las barreras identificadas.

En primer lugar, la falta de acceso a recursos tecnológicos y la brecha digital aparecen como obstáculos comunes en varios estudios. Tanto el estudio de Perdomo (2022) como el de Pinto *et al.* (2022) destacan la importancia de contar con los recursos adecuados, incluyendo dispositivos y conectividad, para utilizar eficazmente las innovaciones tecnológicas en los entornos de aprendizaje. Esta falta de acceso limita la capacidad de los profesores para aprovechar las herramientas tecnológicas, incluida la minería de datos, en su enseñanza.

En segundo lugar, la falta de habilidades digitales y de capacitación adecuada también se menciona en varios estudios. Tanto García y Silva (2022) como Kakepoto *et al.* (2021) señalan que muchos profesores carecen de las habilidades necesarias en tecnología de la información y computación para utilizar plenamente las innovaciones tecnológicas en su práctica docente. Esto destaca la importancia de programas de desarrollo docente que se centren en mejorar estas habilidades y brindar capacitación específica en tecnología. En tercer lugar, la resistencia al cambio y la falta de apoyo institucional son barreras recurrentes en los estudios. García y Silva (2022) destacan la

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

resistencia de algunos profesores a aceptar la creciente influencia de la tecnología en la academia en general, lo que puede dificultar la adopción de innovaciones tecnológicas, como la minería de datos. Asimismo, la falta de apoyo administrativo y gerencial por parte de las instituciones universitarias también se menciona en varios estudios como un obstáculo para la implementación efectiva de la enseñanza en línea y el uso de tecnologías emergentes.

Además de estas barreras comunes, cada estudio también aporta información única sobre los desafíos específicos en cada contexto. El estudio de Perdomo (2022) destaca las preocupaciones sobre la privacidad y seguridad de los datos como una barrera para el uso de la minería de datos en la enseñanza remota de emergencia. El estudio de Kakepoto *et al.* (2021) resalta la falta de interés y resistencia entre los profesores pakistaníes para adquirir nuevas habilidades en tecnología de la información y computación.

En general, los estudios sugieren que superar estas barreras requiere un enfoque integral que abarque tanto aspectos técnicos como pedagógicos. Es necesario mejorar la infraestructura tecnológica y garantizar el acceso equitativo a los recursos tecnológicos. Además, se deben proporcionar programas de desarrollo docente que incluyan capacitación específica en tecnología, fomentar una cultura de cambio y colaboración, y brindar un mayor apoyo administrativo y gerencial. La colaboración entre profesores, directores y administradores también es crucial para abordar estos desafíos y promover una educación de calidad.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

La discusión de los resultados de los cinco estudios mencionados y los patrones identificados en la investigación de la UJAT finalmente proporcionan una visión más amplia de las barreras y desafíos que enfrentan los docentes de educación superior en la adopción de innovaciones tecnológicas, incluyendo la minería de datos. A continuación, se presentará una comparación y análisis de estos resultados:

1. Barreras relacionadas con los recursos tecnológicos: En los estudios mencionados, se encontró que la falta de acceso a recursos tecnológicos y la carencia de habilidades digitales adecuadas son obstáculos importantes para los profesores en la adopción de innovaciones tecnológicas. Esto incluye tanto la infraestructura tecnológica insuficiente como los dispositivos y la conectividad limitada. Estas barreras también se identificaron en la investigación de la UJAT, donde se destacó la falta de implementación o uso suficiente de innovaciones tecnológicas debido a la carencia de recursos tecnológicos y problemas constantes de conectividad.
2. Barreras institucionales y de apoyo: Los estudios mencionados resaltan la importancia del apoyo institucional y administrativo para promover la adopción de innovaciones tecnológicas. La falta de políticas claras y directrices, así como la falta de programas de desarrollo docente que incluyan capacitación en tecnología, fueron barreras identificadas. En el caso de la UJAT, se encontró que la gestión administrativa no era apropiada para implementar un sistema eficiente de uso de tecnologías, y que la política institucional existente para promover la adopción de tecnología en el aula era poco clara.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

3. Barreras de habilidades y resistencia al cambio: El estudio de Pakistán destaca que los profesores poseen habilidades tecnológicas básicas, pero muestran resistencia a adoptar plenamente el aprendizaje en línea como alternativa a las clases presenciales. Esto también se relaciona con la falta de interés en adquirir nuevas habilidades tecnológicas. En los demás estudios, se encontró que la falta de capacitación adecuada, la resistencia al cambio y la falta de confianza tecnológica son barreras significativas. En la investigación de la UJAT, se identificó que la universidad no cuenta con un sistema para evaluar las habilidades tecnológicas del personal docente.
4. Barreras relacionadas con la brecha digital: Tanto en los estudios mencionados como en la investigación de la UJAT, se destaca la brecha digital como un desafío importante. Esto incluye la falta de acceso equitativo a recursos tecnológicos y la ampliación de la brecha debido a problemas socioeconómicos y geográficos. En el estudio de La Guajira (Colombia), se resaltó la brecha digital en la formación inicial docente como un obstáculo específico durante la pandemia.
5. Barreras de capacitación y desarrollo profesional: Varios de los estudios mencionados y la investigación de la UJAT destacan la importancia de la capacitación y el desarrollo profesional para superar las barreras en la adopción de innovaciones tecnológicas. La falta de programas de desarrollo docente y la necesidad de capacitación en habilidades tecnológicas específicas fueron identificadas como barreras en múltiples.

5.3 Recomendaciones

En esta sección, se presentan las buenas prácticas que se han identificado durante el desarrollo de la investigación para que los usuarios de los productos generados y otros investigadores que deseen replicar los métodos, tecnologías y datos utilizados, puedan observarlas adecuadamente para así obtener resultados confiables y precisos.

Se presentan recomendaciones en dos tipos: en primer lugar, se exponen recomendaciones al nivel técnico sobre las herramientas implementadas en el estudio encuentran recomendaciones a partir de la información descubierta con los datos analizados y en segundo lugar están las recomendaciones para mejorar la adopción de tecnología en la práctica docente en la UJAT.

Tras el análisis de usabilidad de varias herramientas, se concluyó que *SAS Enterprise Miner* obtuvo el mejor puntaje. Por esta razón, se recomienda esta herramienta por ser fácil de utilizar, incluso para aquellos usuarios que no tienen experiencia en minería de datos, además de requerir poca documentación para su uso.

SAS Enterprise Miner es una herramienta de *software* especializada en análisis y minería de datos, diseñada para manejar grandes cantidades de información de manera eficiente. Además de su facilidad de uso, esta herramienta es altamente personalizable, lo que permite adaptarla a las necesidades específicas de cada usuario o proyecto.

Por otro lado, cuenta con una amplia gama de algoritmos y modelos de análisis predefinidos, lo que facilita la exploración y el análisis de los datos.

Además, *SAS Enterprise Miner* cuenta con una amplia comunidad de usuarios, lo que se traduce en una gran cantidad de recursos disponibles para aprender y mejorar el

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

uso de esta herramienta. Todo esto hace que *SAS Enterprise Miner* sea una opción altamente recomendable para el análisis y minería de datos, especialmente para aquellos usuarios que no tienen experiencia previa en este tipo de tareas y buscan una solución fácil de usar y con resultados precisos.

Estas características la convierten en una opción adecuada para esta investigación. *SAS Enterprise Miner* es un *software* muy versátil y puede ser utilizado en diferentes escenarios. De acuerdo con lo anterior, se presentan los siguientes escenarios para la herramienta de minería de datos *SAS Enterprise Miner*:

- Análisis de riesgo crediticio: Las instituciones financieras pueden utilizar *SAS Enterprise Miner* para analizar y predecir el riesgo crediticio de sus clientes, lo que les permitirá tomar decisiones informadas en cuanto a la aprobación de créditos y préstamos.
- Marketing y publicidad: Las empresas pueden utilizar *SAS Enterprise Miner* para analizar los patrones de compra y comportamiento de los clientes, lo que les permitirá crear campañas publicitarias más efectivas y dirigidas a su público objetivo.
- Medicina y salud: Los profesionales de la salud pueden utilizar *SAS Enterprise Miner* para analizar grandes cantidades de datos médicos y crear modelos predictivos para predecir enfermedades o identificar factores de riesgo en pacientes.
- Análisis de datos de redes sociales: Las empresas y organizaciones pueden utilizar *SAS Enterprise Miner* para analizar grandes cantidades de datos de redes

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

sociales y comprender mejor el comportamiento y opiniones de sus clientes o usuarios.

- Control de calidad: Las empresas pueden utilizar *SAS Enterprise Miner* para analizar datos de producción y detectar patrones que indiquen problemas de calidad o ineficiencias en el proceso productivo, lo que les permitirá tomar medidas correctivas y mejorar su proceso de producción

Es necesario destacar que implementar herramientas de minería de datos como *RapidMiner* y *Python* puede resultar complicado para aquellos usuarios que se están iniciando en el mundo de la minería de datos, debido a que estas herramientas requieren un mayor conocimiento técnico y habilidades en programación. A diferencia de *SAS Enterprise Miner*, que ha sido diseñado para usuarios no especializados, con poca documentación para su uso y una curva de aprendizaje no tan amplia.

Por esta razón, se recomienda a los usuarios novatos en la minería de datos utilizar herramientas más amigables y fáciles de utilizar como *SAS Enterprise Miner*, que les permitan realizar análisis y descubrimientos de patrones sin la necesidad de tener una amplia formación técnica. En cambio, *RapidMiner* y *Python* son más adecuados para usuarios que ya tienen un nivel avanzado de conocimiento en la minería de datos y habilidades de programación, ya que ofrecen una mayor flexibilidad y personalización en los análisis de datos.

Por otro lado, es recomendable que las tres técnicas de árboles de decisión, agrupación y reglas de asociación sean empleadas en más de una herramienta a nivel tecnológico y en minería de datos por varias razones:

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- En primer lugar, cada técnica tiene sus propias fortalezas y debilidades, y la combinación de ellas puede llevar a mejores resultados y un análisis más completo. Por ejemplo, los árboles de decisión son útiles para la clasificación de datos y la identificación de patrones, mientras que la agrupación es útil para identificar grupos similares de datos y la regresión de estos. Por otro lado, las reglas de asociación son útiles para identificar patrones de co-ocurrencia en grandes conjuntos de datos.
- En segundo lugar, cada herramienta tiene diferentes características y funciones que pueden complementarse entre sí. Al utilizar más de una herramienta, los usuarios pueden acceder a una variedad de funcionalidades y personalizar el análisis según sus necesidades específicas.

Además, al utilizar varias herramientas, los usuarios también pueden reducir la dependencia de una sola herramienta y diversificar su enfoque en la minería de datos. Esto puede resultar útil en caso de fallos técnicos o actualizaciones que afecten a una herramienta específica. Con respecto a las recomendaciones para mejorar la adopción de tecnología en la práctica docente en la UJAT, se definen los siguientes puntos:

- Promover la formación continua de los docentes en el uso de tecnología educativa: Se sugiere que la institución brinde capacitación constante a los docentes en el uso de tecnología educativa, con el objetivo de que se sientan más seguros y familiarizados con su uso en el aula. De esta forma, se podrán reducir algunas de las barreras identificadas en la investigación, como la falta de habilidades y competencias tecnológicas.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Mejorar la infraestructura tecnológica en la universidad: Se recomienda a la institución invertir en la adquisición y mantenimiento de recursos tecnológicos para los docentes, como computadoras, tabletas, proyectores y pizarras digitales. De esta manera, se podrán reducir las barreras relacionadas con la falta de equipo tecnológico y se fomentará el uso de tecnología en la práctica docente.
- Implementar programas de formación y actualización tecnológica para los docentes, con el fin de mejorar su competencia digital y habilidades en el uso de las tecnologías de la información y comunicación.
- Establecer un sistema de seguimiento y evaluación para medir el impacto del uso de tecnologías en el proceso de enseñanza y aprendizaje, con el fin de identificar las áreas que requieren mejoras y ajustes.
- Crear un equipo de soporte técnico especializado que brinde asistencia a los docentes en el uso de tecnologías y solución de problemas técnicos, con el fin de reducir la frustración y la desmotivación que pueden surgir al enfrentarse a problemas tecnológicos.

En conclusión, es importante mencionar que las recomendaciones expuestas en esta investigación no buscan eliminar por completo las barreras que impiden la adopción de innovaciones tecnológicas por parte de los docentes en la UJAT. Sin embargo, estas recomendaciones pueden formar parte del inicio de un proyecto aún más grande, que permita abordar de manera más eficaz y con un mejor alcance la resolución de estas barreras.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

La implementación de estas recomendaciones puede ser considerada como un primer paso hacia una cultura institucional que promueva la adopción de tecnología en el aula y el desarrollo de habilidades tecnológicas en el personal docente. La combinación de técnicas como árboles de decisión, agrupación y reglas de asociación en más de una herramienta de minería de datos a nivel tecnológico puede contribuir en gran medida a la identificación de patrones y conductas de los estudiantes, lo que permitirá la toma de decisiones más informadas y eficaces en la institución. Estas recomendaciones pueden ser el comienzo de un camino hacia una mejora continua en el uso de tecnología en el ámbito educativo, y en la formación de profesionales más capacitados en el uso de herramientas tecnológicas.

5.4 Trabajos futuros

En este apartado se exponen posibles trabajos futuros que se derivan de la investigación realizada. Estos trabajos pueden abordar cuestiones que no se pudieron tratar en la investigación actual o bien aplicar la misma metodología a otras problemáticas o disciplinas, ya sea en el ámbito educativo o tecnológico. A continuación, se presentan algunas sugerencias para posibles investigaciones futuras relacionadas con los resultados de esta investigación.

- La tecnología avanza rápidamente con el tiempo, lo que abre la posibilidad de utilizar nuevas herramientas en la minería de datos. Por lo tanto, se pueden realizar futuras investigaciones utilizando herramientas diferentes a las empleadas en este estudio, tales como las descartadas en el presente.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Se puede desarrollar un sistema de seguimiento de competencias tecnológicas de los docentes en la UJAT, que permita medir el impacto de los programas de capacitación y formación en tecnología.
- Esta investigación deja abierta un área de oportunidad para crear un estudio sobre las diferencias en el uso de las innovaciones tecnológicas en el aula entre las distintas Divisiones Académicas en la UJAT, con el fin de identificar factores específicos que influyen en su adopción.
- Existe la posibilidad de crear un análisis del impacto de la política institucional sobre la adopción de tecnología en el aula, identificando los factores que dificultan su implementación y sugiriendo mejoras en la política para fomentar su adopción dentro de la UJAT.
- Crear una investigación que se centre en la implementación de técnicas de minería de datos adicionales, donde se empleen nuevas, tales como redes neuronales y SVM, para ampliar el análisis de patrones de conducta y su relación con el uso de tecnología en el aula, bien sea en docentes y/o alumnos.
- Realizar un estudio que permita evaluar los cambios en la adopción de tecnología en el aula a lo largo del tiempo, comparando los resultados con los obtenidos en esta investigación y detectando posibles mejoras en la adopción de tecnología en la UJAT, bien sea de manera general o para las Divisiones Académicas que mayor relación tengan con las áreas tecnológicas.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Realizar una investigación similar de barreras ante el uso de innovaciones tecnológicas en profesores de la UJAT, encuestando a los docentes durante el período postpandemia, debido a que las respuestas y realidades probablemente serán muy distintas a las encontradas en la presente.

En conclusión, los trabajos futuros propuestos en esta investigación no deben ser descartados ya que pueden dar continuidad y relevancia a este estudio sobre las barreras ante el uso de innovaciones tecnológicas en profesores de la UJAT. Estos trabajos pueden ampliar y profundizar el conocimiento sobre la problemática y permitir la aplicación de técnicas y herramientas más avanzadas en la minería de datos para un mejor análisis de los resultados. La realización de nuevos estudios aplicando herramientas distintas también podría permitir una comparación y validación de los resultados obtenidos en esta investigación, lo que podría llevar a una mayor comprensión y solución de las barreras identificadas. Estos trabajos futuros podrían contribuir significativamente a la mejora de la calidad educativa y tecnológica en la UJAT.

Nombre del trabajo recepcional para obtener el grado. Arial 10, justificado a la izquierda.

Referencias

- Aguilera, I., García, Á., Luengo, J., Damas, S., Melero, F., Valle, J. & Herrera, F. (2023). TSFEDL: A Python library for time series spatio-temporal feature extraction and prediction using deep learning. *Neurocomputing*, 517(2), 223–228. <https://doi.org/10.1016/j.neucom.2022.10.062>
- Álvarez, E. (2021). Uso crítico y seguro de tecnologías digitales de profesores universitarios. *Formación universitaria*, 14(1), 33-44. <https://cutt.ly/SDDh6Cr>
- Amarda, J., Suarna, N., & Nurdiawan, O. (2022). KLASIFIKASI INDUSTRI KERAJINAN BERDASARKAN PROVINSI MENGGUNAKAN METODE DECISION TREE UNTUK MEMBANDINGKAN PENGUJIAN ALGORITMA ID3 ALGORITMA C45 DAN ALGORITMA CART. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 6(2), 790–796. <https://doi.org/10.36040/jati.v6i2.5736>
- Andry, J., Tannady, H., Dwinoor, G. & Freggy, D. (2023). Analysis of the Omicron virus cases using data mining methods in RapidMiner applications. *Microbes and Infectious Diseases*, 1(1), 10-13. <https://doi.org/10.21608/mid.2023.194619.1469>
- Arancibia, M., Cabero, J. y Marín, V. (2020). Creencias sobre la enseñanza y uso de las tecnologías de la información y la comunicación (TIC) en docentes de educación superior. *Formación Universitaria*, 13(3), 89-100. <https://cutt.ly/IDDjij0>
- Aung, S., Funabiki, N., Aung, L., Htet, H., Kyaw, H. & Sugawara, S. (2022). An implementation of java programming learning assistant system platform using node.js. *2022 10th International Conference on Information and Education Technology (ICIET)*, 47–52. <https://doi.org/10.1109/ICIET55102.2022.9779047>

1

19

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Báez, I., Sosa, S., Jaimes, M., Yáñez, A. y Díaz, R. (2022). Aplicación del proceso de descubrimiento del conocimiento para la detección de diabetes. *South Florida Journal of Development*, 3(2), 2345-2357. <https://doi.org/10.46932/sfjdv3n2-060>
- Bahit, E. (2018). *Introducción al lenguaje Python* (1ª ed.). Escuela de Informática Eugenia Bahit. <https://cutt.ly/JDDIKTk>
- Bai, L., Liao, C., Wang, C., Zhang, M., Meng, F., Fan, M. & Zhang, B. (2022). Study on height prediction of water flowing fractured zone in deep mines based on Weka platform. *Sustainability*, 15(1), 737. <https://doi.org/10.3390/su15010737>
- Belinchón, Y. (2019). Minería de datos. *Repositorio de la Universidad Carlos III de Madrid*, 1, 1-8. <https://cutt.ly/KUVgKVm>
- Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. (2017). Julia: A fresh approach to numerical computing. *SIAM review. Society for Industrial and Applied Mathematics*, 59(1), 65–98. <https://doi.org/10.1137/141000671>
- Boubker, O., Naoui, K., Ouajdouni, A. & Arroud, M. (2022). The effect of action-based entrepreneurship education on intention to become an entrepreneur. *MethodsX*, 9, 101657. <https://doi.org/10.1016/j.mex.2022.101657>
- Briceño, J. y Castellanos, M. (2021). Percepciones de docentes universitarios frente al cambio de modalidad presencial a remota por la COVID-19: comparación entre profesores presenciales y virtuales. *Revista Educación Superior y Sociedad (ESS)*, 33(2), 351-376. <https://doi.org/10.54674/ess.v33i2.455>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Brusco, M. (2021). Logistic regression via Excel spreadsheets: Mechanics, model selection, and relative predictor importance. *INFORMS Transactions on Education*, 23(1), 1–11. <https://doi.org/10.1287/ited.2021.0263>
- Cañete, L., Monroy, R. & Medina, M. (2021). A review and experimental comparison of multivariate decision trees. *IEEE access: practical innovations, open solutions*, 9, 110451–110479. <https://doi.org/10.1109/access.2021.3102239>
- Castrillón, O., Castillo, L. y Castaño, C. (2022). Minería de datos aplicada a la detección de cáncer gástrico. *CIT Informacion Tecnologica*, 33(4), 151–160. <http://bit.ly/3kX9xv4>
- Chiu, C., Qin, J., Zhang, Y., Yu, J. & Wu, Y. (2022). Self-supervised learning with random-projection quantizer for speech recognition. *arXiv*, 3915–3924. <http://bit.ly/3Ru0bTK>
- Comisión Económica para América Latina y el Caribe [CEPAL] y la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura [UNESCO]. (2020). *La educación en tiempos de la pandemia de COVID-19*. <https://cutt.ly/dBSrI7H>
- Custode, L. & Iacca, G. (2023). Evolutionary learning of interpretable decision trees. *IEEE access: practical innovations, open solutions*, 11, 6169–6184. <https://doi.org/10.1109/access.2023.3236260>
- Diario Oficial de la Federación [DOF]. (2005). *Reglamento de la ley Federal de Derechos de Autor*. Razón Pública. <https://cutt.ly/0DFqLKs>
- Diario Oficial de la Federación [DOF]. (2010). *Reglamento de la Ley Federal de Protección de Datos Personales*. Razón Pública. <https://acortar.link/LR3ock>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Díaz, J. A., Ruiz, M. & Martín, M. (2023). A survey on the use of association rules mining techniques in textual social media. *Artificial Intelligence Review*, 56(2), 1–26. <https://doi.org/10.1007/s10462-022-10196-3>
- Díaz, S. (2018). Software Comercial. *Con-Ciencia Boletín Científico de la Escuela Preparatoria No. 3*, 5(9). <https://cutt.ly/6DDlqL4>
- Fernandes, A., Hartono, H. & Chakir, A. (2022). Data set analysis using rapid miner to predict cost insurance forecast with data mining methods. *Journal of Hunan University Natural Sciences*, 49(6). <http://bit.ly/3WSeP8q>
- Ganim, A., Al, H., Mohammed, S. & Ahmed, S. (2022). A descriptive statistical analysis of overweight and obesity using big data. *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–6. <https://doi.org/10.1109/HORA55278.2022.9800098>.
- García, M., Reyes, J. y Godínez, G. (2017). Las TIC en la educación superior, innovaciones y retos. *Revista Iberoamericana de las Ciencias Sociales y Humanísticas [RICSH]*, 6(12), 299-316. <https://cutt.ly/KDDk1w5>
- García, V. y Silva, M. (2021). Barreras para la adopción de innovaciones tecnológicas. <https://cutt.ly/ZDD4t97>
- García, V. y Silva, M. (2021). Innovación tecnológica disruptiva durante la pandemia. Factores intrínsecos que dificultan su adopción entre las y los docentes. En G. Medina y S. Aquino (Coords.), *La tecnología educativa en tiempos de pandemia* (pp. 31-44). Gradus Editora.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

García, V. y Silva, M. (2022). Percepción académica sobre las barreras en la adopción de innovaciones tecnológicas durante la pandemia por la covid-19. *Apertura*, 14(1), 96-113. <https://doi.org/10.32870/ap.v14n1.2150>

12 27
García, V., Garrido, A. & Martín, R. (2021). The transformation of higher education after the COVID disruption: Emerging challenges in an online learning scenario. *Frontiers in Psychology*, 12, 616059. <https://doi.org/10.3389/fpsyg.2021.616059>

Giorgi, F., Ceraolo, C. & Mercatelli, D. (2022). The R language: An engine for bioinformatics and data science. *Life (Basel, Switzerland)*, 12(5), 648. <https://doi.org/10.3390/life12050648>

GNU. (2021). *Licencias*. GNU. <https://cutt.ly/sDDkVQP>

Godoy, Y. y Avelino, R. (2023). Sistema de información para la gestión del proceso de contratación de la Universidad del Magdalena. [Tesis de licenciatura, Universidad de Magdalena]. Repositorio Digital Institucional. <https://acortar.link/L6w3pQ>

González, M. y González, J. (2013). Aplicación del estándar ISO/IEC 9126-3 en el modelo de datos conceptual entidad-relación. *SCIELO*, 22(35). <https://bit.ly/3HvtUXU>

Google Cloud. (2023). *Lista de precios*. Google Cloud. <https://cutt.ly/eDDkXt4>

Google. (2023). *Google Cloud Datalab*. Google. <https://cutt.ly/tDDkKKp>

Gordan, M., Sabbaghi, S., Ismail, Z., Ghaedi, K., Carroll, P., McCrum, D. & Samali, B. (2022). State-of-the-art review on advancements of *data mining* in structural health monitoring. *Measurement: Journal of the International Measurement Confederation*, 193, 110939. <https://bit.ly/3jlyG1T>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Grolemund, G. & Wickham, H. (2017). *R for Data Science*. O'Reilly.
<https://acortar.link/AlozNZ>

Grupo de Trabajo de Tecnología e Innovación Educativa. (2021). El uso de la tecnología para innovar la práctica docente: Retos y lecciones aprendidas en la pandemia. *Diálogo Interamericano*, 7(1). <https://cutt.ly/KBSeBnQ>

Gümüş, M., Kukul, V. & Düzenli, Ö. (2023). Mapping the research on Educational Technology: An overview through text mining. *Kastamonu eğitim dergisi*, 31(2), 202–216. <https://doi.org/10.24106/kefdergi.1271474>

Gunandi, A., Awang, H., Alhawad, E. & Shabaan, L. (2023). Customer value and Data Mining in segmentation analysis. *International Journal of Information Technology and Computer Science Applications*, 1(1), 20–34. <https://doi.org/10.58776/ijitcsa.v1i1.16>

Gurazada, S., Gao, S., Burstein, F. & Buntine, P. (2022). Predicting patient length of Stay in Australian Emergency Departments using Data Mining. *Sensors (Basel, Switzerland)*, 22(13), 4968. <https://doi.org/10.3390/s22134968>

Haji, S., Jacksi, K. & Salah, R. M. (2023). A semantics-based clustering approach for online laboratories using K-Means and HAC algorithms. *Mathematics*, 11(3), 548. <https://doi.org/10.3390/math11030548>

Hernández, A. (2018). *Monitorización de la actividad de sujetos en una casa inteligente empleando técnicas de data analytics*. UC3M.es. <https://cutt.ly/h2HIATc>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Hidalgo, B. & Gisbert, M. (2022). Factores determinantes que permiten establecer tipologías de profesorado en el contexto de la innovación tecnológica educativa. *RED*, 22(69). <https://doi.org/10.6018/red.499171>

21 Hudaefi, F., Hassan, M., Abduh, M. & Beik, I. (2023). Knowledge discovery of zakat administration worldwide from YouTube and Zoom via text mining. *Journal of Islamic Accounting and Business Research*, 14(1), 159–180. <https://doi.org/10.1108/jiabr-03-2022-0067>

Ibañez, J., Echeverri, T. y Castrillón, O. (2022). Predicción de la calidad de vida universitaria a través de minería de datos. *Revista Ingenierías Universidad de Medellín*, 27(40), 1–14. <https://doi.org/10.22395/rium.v21n40a1>

IBM. (2021). Reglas de asociación. *IBM Documentation*. IBM Web Site. <https://acortar.link/tyY9Lf>

IBM. (2022a). *IBM docs: Reglas de asociación*. Createspace Independent Publishing Platform. <https://cutt.ly/6DOfTrv>

IBM. (2022b). *Software IBM SPSS*. IBM. <https://cutt.ly/CDDkUoY>

IBM. (2022c). *Precios SPSS Statics*. IBM. <https://cutt.ly/vDDkO5l>

1 Ikotun, A., Ezugwu, A., Abualigah, L., Abuhaija, B. & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, 622, 178–210. <https://doi.org/10.1016/j.ins.2022.11.139>

Instituto Tecnológico y de Estudios Superiores de Monterrey [ITESM] y el Banco Interamericano de Desarrollo [BID]. (2021). *Estudio: Uso de las tecnologías*

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

digitales para el proceso de enseñanza-aprendizaje en Latinoamérica. Inter-American Development Bank. <https://cutt.ly/8TLKZ5Z>

Internacional de la Educación América Latina [IEAL]. (2021). *Experiencias didáctico-pedagógicas y laborales de la práctica docente en tiempos de la COVID-19, en América Latina* (1a ed.). Editorial Internacional de la Educación América Latina. <https://cutt.ly/JBStv4L>

Internacional de la Educación para América Latina [IEAL]. (2021). *Tendencias-Situación laboral y educativa de América Latina en el contexto de la pandemia COVID-19*. Estudio Casarli. <https://cutt.ly/1BSwRIY>

Isnaini, K., Sulistiyani, D. & Putri, Z. (2021). Pelatihan Desain Menggunakan Aplikasi Canva. *Selaparang: Jurnal Pengabdian Masyarakat Berkemajuan*, 5(1), 291-295.

Jara, F., Cháves, J., Villa, I. y Novillo, J. (2021). Rol del docente para la educación virtual en tiempos de pandemia: Retos y oportunidades. *Polo del conocimiento*, 63(6), 30-45. <https://cutt.ly/rDOfIH8>

Java. (2023). *¿Qué es la tecnología Java y para que la necesito?* Java. <https://cutt.ly/DDOsSvC>

Jiménez, J. (2019). *Introducción a R y Rstudio* (1^{ra} ed.). <https://cutt.ly/VDDMLFD>

Joyanes, L. (2019). *Inteligencia De Negocios Y Analítica De Datos* (1a ed.). Alfa Omega.

Juárez, B. y Torres, I. (2022). Evaluación del nivel de usabilidad basado en la NTP-ISO/IEC 9126 de un producto software generado por una herramienta para el modelado de procesos de negocio [Tesis de ingeniería, Universidad Señor de Sipán]. Repositorio Institucional. <https://acortar.link/eQkXYk>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Kakepoto, I., Talpur, Q., Memon, I., Halepoto, I. & Bux Jalbani, K. (2021). Pedagogical Shift: Faculty Insights about E-Teaching Barriers during COVID Pandemic. *International Journal of Innovation, Creativity and Change*, 15(6), 1147-1160. <https://acortar.link/7BOYaM>

Klenzi, R., Malberti, M. y Beguerí, G. (2018). Visualización en un entorno de minería de datos desde una perspectiva interacción humano computador. *Computación y Sistemas*, 22(1), 279-290. <https://cutt.ly/GDDkWBQ>

Klochko, O., Gurevych, R., Nagayev, V., Dudorova, L. & Zuziak, T. (2022). Data mining of the healthcare system based on the machine learning model developed in the Microsoft azure machine learning studio. *Journal of Physics: Conference Series*, 2288, 012006. <https://doi.org/10.1088/1742-6596/2288/1/012006>

Laboratorio Latinoamericano y Caribeño de Investigación e Innovación Educativa, SUMMA y el Centro de Estudios Para la Transformación de la Educación Social. (2021). *Experiencias de desarrollo profesional docente en América Latina en contextos COVID-19 y su vinculación con tecnologías digitales*. 20. FIT-ED. <https://cutt.ly/9BSqVkw>

Laureano, D. (2022). Las tecnologías de información y comunicación en la comprensión y producción de textos. *Revista Latinoamericana Ogmios*, 2(3), 64–84. <https://doi.org/10.53595/rlo.v2.i3.024>

Leo, R., Salgado, C., Sánchez, A. y Peralta, M. (2023). Un modelo de calidad de software con la sostenibilidad como característica transversal. *XXVIII Congreso Argentino de Ciencias de la Computación (CACIC)*, 332-341. <https://acortar.link/ICHZDR>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Lim, J. (2023). A predictive model of ischemic heart disease in middle-aged and older women using *data mining* technique. *Journal of Personalized Medicine*, 13(4), 663. <https://doi.org/10.3390/jpm13040663>

31 López, M. (2022). Educación superior en comunicación audiovisual: Desafíos de la virtualidad en tiempos de COVID-19. *ADResearch ESIC International Journal of Communication Research*, 27(27). <https://doi.org/10.7263/adresic-27-197>

Lugo, A. (2021). Análisis del desempeño docente en la modalidad a distancia en línea en el contexto de la pandemia de COVID-19. *Tecnología y contingencias*, 65-71. Editorial Transdigital. <https://cutt.ly/PDD3vEc>

Luo, X., Qu, H., Wang, Y., Yi, Z., Zhang, J. & Zhang, M. (2022). Supervised learning in multilayer spiking neural networks with spike temporal error backpropagation. *IEEE Transactions on Neural Networks and Learning Systems*, 1–13. <https://doi.org/10.1109/TNNLS.2022.3164930>

17 Maraza, B., Valderrama, E., Cari, H., Apaza, M. & Sanchez, J. (2022). A Predictive Model Implemented in KNIME Based on Learning Analytics for Timely Decision Making in Virtual Learning Environments. *International Journal of Information and Education Technology*, 12(2). <https://doi.org/10.18178/ijiet.2022.12.2.1591>

Maturana, F., Fávero, L., Lima, F., Braga, R. & Dos Santos, M. (2022). B3 financial assets data collection on Twitter: A Python text mining algorithm. *Contabilometria*, 11(1). <https://acortar.link/Z2GZwR>

Microsoft Azure. (2023). *Precios de Machine Learning Studio (clásico)*. Azure. <https://cutt.ly/tDOfDDs>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Microsoft. (2023a). *¿Qué es Azure Machine Learning?* Microsoft. <https://cutt.ly/6DDkb4t>
- Microsoft. (2023b). *Excel*. Microsoft. <https://cutt.ly/mDOFGaK>
- Montavon, G., Kauffmann, J., Samek, W. & Müller, K. (2022). Explaining the Predictions of Unsupervised Learning Models. *xxAI - Beyond Explainable AI*, 117–138. <https://acortar.link/JO6Vru>
- Nanfack, G., Temple, P. & Frénay, B. (2022). Constraint enforcement on decision trees: A survey. *ACM Computing Surveys*, 54(10), 1–36. <https://bit.ly/3WUjkiz>
- Nath, D. & Anand, A. (2021). Trade heterogeneity in the EU: Insights from the emergence of COVID-19 using time series clustering. *Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie*, 993(3), 9–26. <http://bit.ly/3DrLESA>
- Nazar, M. (2022). Tecnologías, información y derechos. *Cuadernos del INAP (CUINAP)*, 97. <http://bit.ly/3HN3qSU>
- Ningsih, W. S. & Haryanto, E. (2022). Application of RapidMiner for clustering aids cases by province using data mining k-means clustering. *Journal of Intelligent Decision Support System (IDSS)*, 5(3), 89–98. <https://doi.org/10.35335/idss.v5i3.101>
- Noorman, M., Khan, A. & Cong, D. (2022). Quality attributes and performance satisfaction of open-source library system: A survey amongst librarians in Pakistan. *Malaysian Journal of Library & Information Science*, 27(2), 19–36. <https://bit.ly/311smNq>
- Oracle. (2022a). *Oracle Business Intelligence Server*. Oracle. <https://cutt.ly/0DOFJYb>
- Oracle. (2022b). *Oracle Business Intelligence Server Pricing*. Oracle. <https://cutt.ly/6DODCZB>

25

33

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Oyola-García, A. (2021). La variable. *Revista del Cuerpo Médico Hospital Nacional Almanzor Aguinaga Asenjo*, 14(1), 90-93. <https://cutt.ly/wDOfLAJ>
- Parlindungan, R., Fernandes, B. & Tannady, H. (2022). Big data analytics in library to classification book publishers. *Journal of Positive School Psychology*, 6(2), 4303–4310. <https://www.journalppw.com/index.php/jpsp/article/view/2780>
- Perdomo, B. (2022). Docentes y barreras enfrentadas en la enseñanza remota de emergencia por COVID-19: revisión integrativa. *Foro educacional*, 38, 7–37. <https://doi.org/10.29344/07180772.38.2999>
- Pérez, A., Salgado, B., Hasperué, W., Cela., K. y Santamaría, J. (2023). Evolución de las técnicas de minería de datos para extraer datos provenientes de twitter aplicadas a la educación superior: una revisión sistemática. *South Florida Journal of Development*, 4(1), 33–55. <https://doi.org/10.46932/sfjdv4n1-002>
- Pérez, J., Senent, J. & Jimeno, P. (2022). The application of spreadsheets for teaching hydrological modeling and climate change impacts on streamflow. *Computer Applications in Engineering Education*. Zenodo. <http://bit.ly/3HO4EgS>
- Pérez, M., Falla, V., Guzmán, M., Arbulu, M. y Garcia, L. (2022). Factores determinantes en la gestión de recursos humanos de la empresa Electronorte S.A. Chiclayo en la mejora de los procesos de incorporación de innovación tecnológica. *Revista Científica Emprendimiento Científico Tecnológico*, 3, 1–46. <http://bit.ly/3kWHewE>
- Pinto-Santos, A. R., George-Reyes, C. E., & Cortés-Peña, O. F. (2022). Brecha digital en la formación inicial docente: desafíos en los ambientes de aprendizaje durante la

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

pandemia COVID-19 en La Guajira (Colombia). *Formación Universitaria*, 15(5), 49–60. <https://doi.org/10.4067/s0718-50062022000500049>

Python. (2022). *El tutorial de Python*. Python Docs. <https://cutt.ly/4DOsMdq>

Quintero, S., Orjuela, W. y Escobar, J. (2022). Medición y evaluación de las capacidades de innovación tecnológica: una revisión crítica de la literatura. *Rev. CEA*, 8(18). <https://doi.org/10.22430/24223182.2499>

Quiñonez, S., Chan, G. y Reyes, W. (2021). Desarrollo de la competencia digital en profesores universitarios. *Revista científica electrónica de Educación y Comunicación en la Sociedad del Conocimiento*, 21(1), 83-114. <https://cutt.ly/iDOfVw2>

Ramadhan, F. & Wibowo, A. (2022). Text *Clustering* analysis of public sentiment on twitter about tokopedia BTS using orange *data mining*. *Jurnal Darma Agung*, 30(2), 784–792. <https://doi.org/10.46930/ojsuda.v30i2.2366>

Real Academia Española. (2021). Innovación. En *Diccionario de la lengua española* (24^a edición digital). <https://cutt.ly/UDD0zk8>

Real Academia Española. (2022). Tecnología. En *Diccionario de la lengua española* (24^a edición digital). <https://cutt.ly/xDDkzuF>

Riascos, J. y Molina, J. (2015). Breves consideraciones acerca de la importancia de los árboles de decisión en el análisis de carteras. *Revista de la Facultad de Ciencias Económicas y Administrativas*, 17(1), 11-33. <https://cutt.ly/jDD0IOK>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Rincón, I. (2022). Educación, innovación, emprendimiento, crecimiento y desarrollo en América Latina. *Revista de ciencias sociales - Universidad del Zulia. Facultad de Ciencias Económicas y Sociales*, 28(3), 110–128. <https://acortar.link/rupS5J>
- Ruiz-Rabasco, Y., Hernandez-Nodarse, M., Puya-Lino, A, y González-Reyes, S. (2021). Educación Superior virtual: Validación de un instrumento en la Universidad Estatal Península de Santa Elena. *Atenas*, 1(57), 64-79. <https://cutt.ly/wDDkggv>
- Sabpril, I. & Alfin, A. (2022). Analysis of information system effectiveness using ISO/IEC 9126 and AHP (analytical Hierarchy Process) methods at PT ZYS. *JTECS : Jurnal Sistem Telekomunikasi Elektronika Sistem Kontrol Power Sistem Dan Komputer*, 2(2), 189. <https://doi.org/10.32503/jtecs.v2i2.2716>
- Saleh, Y., Muhammed, Y., Al, A., Faisal, M., Mohamad, A. & Birhan, M. (2022). Discovery of knowledge in the incidence of a type of lung cancer for patients through data mining models. *Computational Intelligence and Neuroscience*. 22. <http://bit.ly/3HMOVdOI>
- Sánchez, F. (2019). Fundamentos epistémicos de la investigación cualitativa y cuantitativa: consensos y disensos. *Revista Digital de Investigación en Docencia Universitaria*, 13(1), 102-122. <https://cutt.ly/nDDkpiJ>
- Sánchez, M., Fabián, L. y Melgoza, D. (2021). Competencias digitales docentes: una experiencia en el nivel universitario. *Hamut´ay*, 8(1), 59-66. <https://cutt.ly/xDDjMNg>
- Sánchez, M., García, J., Steffens, E. y Hernández, H. (2019). Estrategias Pedagógicas en Procesos de Enseñanza y Aprendizaje en la Educación Superior incluyendo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Tecnologías de la Información y las Comunicaciones. *Información Tecnológica*, 30(3), 277-286. <https://cutt.ly/ODDjFRz>

Santander. (2022). *Python: qué es y por qué deberías aprender a utilizarlo*. Becas-santander.com. <https://www.becas-santander.com/es/blog/Python-que-es.html>

Santiago, R. (2021). Educación a distancia en el Posgrado en México. *Atenas*, 1(57), 145-161. <https://cutt.ly/oDDj7qh>

SAS Enterprise Miner. (2022). *Reveal valuable insights with powerful data mining software*. SAS Enterprise Miner. <https://cutt.ly/cDDkwd6>

SAS Institute Inc. (2022). *¿Qué es Hadoop?* SAS. <https://cutt.ly/KDDkrdB>

Shahbazi, Z. & Byun, Y. (2022). Knowledge discovery on cryptocurrency exchange rate prediction using machine learning pipelines. *Sensors (Basel, Switzerland)*, 22(5), 1740. <https://doi.org/10.3390/s22051740>

Shaheen, M. & Khan, S. (2023). Wisdom Mining: Future of *data mining*. *Recent Patents on Engineering*, 17(1), 2–11. <https://acortar.link/Ryvr7x>

Shen, Q., Qiao, L., Guo, J., Li, P., Li, X., Li, B., Feng, W., Gan, W., Wu, W. & Ouyang, W. (2022). Unsupervised learning of accurate Siamese tracking. *arXiv*, 8101–8110. <http://bit.ly/3WSfJ4O>

Shu, X. & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817. <https://acortar.link/HZ3bm8>

Singhal, N. & Himanshu. (2022). A review on knowledge discovery from databases. *Lecture Notes in Electrical Engineering*, 457–464. <https://acortar.link/iWL6nQ>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- 12 Smedberg, H. & Bandaru, S. (2023). Interactive knowledge discovery and knowledge visualization for decision support in multi-objective optimization. *European Journal of Operational Research*, 306(3), 1311–1329. <https://acortar.link/uD0njK>
- Soriano, J. y Jiménez, D. (2022). Importancia de la innovación docente como proceso y gestión en el ámbito de Ciencias de la Salud: una revisión sistemática. *Gestionar: revista de empresa y gobierno*, 2(4), 73–85. <http://bit.ly/3HryAh4>
- Stanley, O., Chimezie, O., Daniel, O., Stanley, N. & Okere, M. (2023). Adoption of machine learning and data mining tools in the identification and prediction of diabetes disease in patients using classification mining algorithm. *IRE Journals*, 6(8), 195-204. <https://www.irejournals.com/paper-details/1704102>
- Sulistiani, I. & Nursaeni, N. (2022). Implementasi Asesmen Instrumen Kualitas Menggunakan Model Kualitas ISO/IEC 25010 dan ISO/IEC 9126 Pada Sistem Informasi Akademik SIPAKATAU Berbasis Pengalaman Pengguna. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 12(1), 96. <https://doi.org/10.35585/inspir.v12i1.2663>
- Venegas, L., Luzardo, H. y Pereira, A. (2020). Conocimiento, formación y uso de herramientas TIC aplicadas a la Educación Superior por el profesorado de la Universidad Miguel de Cervantes. *EduTec. Revista Electrónica De Tecnología Educativa*, (71), 35-52. <https://cutt.ly/FDDjCTk>
- Vijay, S., Movva, K., Reddy, V. & Chakraborty, G. (2014). Internet gambling behavioral markers: Using the power of SAS® enterprise miner™ 12.1 to predict high-risk Internet gamblers. *Oklahoma State University*. <https://acortar.link/lb8Q6r>

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Wu, M., Syu, J. & Chen, C. (2022). Kelly-based options trading strategies on settlement date via supervised learning algorithms. *Computational Economics*, 59(4), 1627–1644. <https://doi.org/10.1007/s10614-021-10226-2>
- Xu, L. (2022). Happiness Scores Analysis Report. *2022 6th International Seminar on Education, Management and Social Sciences (ISEMSS 2022)*, 2892–2901. https://doi.org/10.2991/978-2-494069-31-2_340
- Yuliarina, A. & Hendry, H. (2022). Comparison of prediction analysis of GoFood service users using the knn & Naive Bayes algorithm with RapidMiner software. *Jurnal Teknik Informatika (Jutif)*, 3(4), 847–856. <https://bit.ly/3jjwEzh>
- Zhang, M., Fan, J., Sharma, A. & Kukkar, A. (2022). Data mining applications in university information management system development. *Journal of Intelligent Systems*, 31(1), 207–220. <https://doi.org/10.1515/jisys-2022-0006>
- Zhao, B., Ji, S., Xu, J., Tian, Y., Wei, Q., Wang, Q., Lyu, C., Zhang, X., Lin, C., Wu, J. & Beyah, R. (2022). A large-scale empirical analysis of the vulnerabilities introduced by third-party components in IoT firmware. *Proceedings of the 31st ACM SIGSOFT International Symposium on Software Testing and Analysis*, 442-454. <https://doi.org/10.1145/3533767.3534366>

Glosario

A

AGPL: Affero General Public License

API: Application Programming Interfaces

B

BID: Banco Interamericano de Desarrollo

C

CEPAL: Comisión Económica para América Latina y el Caribe

D

DOF: Diario Oficial de la Federación

F

FURPS: Functionality, Usability, Reliability, Performance and Supportability

G

GNU: GNU No es Unix

GPL: General Public License

I

IBM: International Business Machines Corporation

IEAL: Internacional de la Educación para América Latina

ITESM: Instituto Tecnológico y de Estudios Superiores de Monterrey

K

KDD: Knowledge Discovery in Databases

M

MEFI: Modelo Integrado de Formación Educativa

O

OCDE: Organización para la Cooperación y el Desarrollo Económicos

UNESCO: United Nations Educational, Scientific, and Cultural Organization

S

SPSS: Statistical Package for Social Sciences

SQL: Structured Query Language

Y

YALE: Yet Another Learning Environment

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Anexo A. Cuestionario aplicado a los profesores para la recolección de datos

Barreras para la adopción de innovaciones tecnológicas

Instrucciones: Estimado Profesor de la UJAT. Le pedimos que lea con cuidado los enunciados relativos a las barreras para la adopción de innovaciones tecnológicas y exprese de la manera más honesta su sentir respecto a la escala presentada. Los resultados del estudio nos permitirán identificar factores que deben modificarse para acelerar y mejorar las iniciativas. Le tomará 5 minutos responder. Muchas gracias.

* Todas las preguntas deben ser contestadas de manera obligatoria.

Género:

- Mujer
 Hombre

Edad: _____

Estado civil:

- Soltero
 Casado
 Viudo
 Divorciado
 Unión libre
 Separado

División Académica:

- Ciencias Agropecuarias
 Ciencias Económico-Administrativas
 Educación y Artes
 Ciencias y Tecnologías de la Información
 Ciencias de la Salud
 Ciencias Sociales y Humanidades

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Multidisciplinaria de los Ríos
- Multidisciplinaria de Jalpa de Méndez
- Multidisciplinaria de Comalcalco
- Ciencias Básicas
- Ingeniería y Arquitectura
- Ciencias Biológicas

3 Antigüedad en la UJAT (años): _____

Número de horas frente a grupo a la semana: _____

Categorías de contratación en UJAT:

- Profesor de tiempo completo
- Profesor de medio tiempo
- Profesor de asignatura

Tipo de contratación en UJAT:

- Base
- Interino

Grado académico:

- Licenciatura
- Maestría
- Especialidad

3 **Área de formación de su último grado académico:**

- I. Física, Matemáticas y Ciencias de la Tierra
- II. Biología y Química
- III. Medicina y Salud
- IV. Humanidades y de la Conducta
- V. Ciencias Sociales y Económicas
- VI. Biotecnología y Agropecuarias
- VII. Ingeniería e Industria

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Sección 2 de 4

Uso de recursos tecnológicos

Favor de expresar en AÑOS, el tiempo que ha empleado estos recursos. Puede colocar cero si nunca los ha empleado.

1. ¿Aproximadamente hace cuántos años comenzó a usar computadora para apoyar sus actividades docentes? _____
2. ¿Aproximadamente hace cuántos años comenzó a usar recursos de Internet para apoyar sus actividades docentes? _____
3. ¿Aproximadamente hace cuántos años usó por primera vez un Ambiente Virtual de Aprendizaje? _____
4. ¿Aproximadamente desde hace cuántos años usa las redes sociales para trabajar con sus estudiantes? _____
5. ¿Hace cuántos años que usa WhatsApp para trabajar con sus estudiantes?

6. ¿Hace cuántos años que usa Facebook para trabajar con sus estudiantes?

7. ¿Hace cuántos años que usa Twitter para trabajar con sus estudiantes? _____
8. ¿Hace cuántos años que usa el correo electrónico para trabajar con sus estudiantes? _____
9. ¿Qué otras redes sociales o recursos emplean para trabajar con sus estudiantes?

Sección 3 de 4

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Recursos disponibles y sistema educativo

Por favor exprese su grado de acuerdo o desacuerdo con los enunciados que se expresan.

10. Existen carencias en infraestructura tecnológica para el soporte de las actividades docentes en la UJAT

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

11. El equipo tecnológico de apoyo a las actividades docentes es escaso

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

12. La conectividad de la red representa un problema recurrente

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

13. La accesibilidad a la red en los espacios universitarios es inestable

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

14. Cuando tengo un problema con algún equipo es difícil conseguir apoyo técnico

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

15. Que te ayuden a resolver problemas de *software* en la UJAT resulta complicado

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

16. Existe una política institucional poco clara que promueva la adopción de tecnología en el aula

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

17. La estructura administrativa es inadecuada para establecer un sistema de utilización de tecnología eficiente

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

18. La institución carece de un sistema que diagnostique las competencias tecnológicas de su personal

- Totalmente en desacuerdo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

19. La institución carece de un sistema que evalúe las competencias tecnológicas de su personal

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

20. Hace falta premiar a los docentes que adoptan las innovaciones

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

21. Hace falta premiar a los docentes que promueven las innovaciones

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

22. Es poco el tiempo que le puedo dedicar a planear actividades académicas con tecnología

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Totalmente de acuerdo

23. Trabajar con nuevos recursos tecnológicos altera demasiado mis rutinas

- Totalmente en desacuerdo
 En desacuerdo
 Ni de acuerdo ni en desacuerdo
 De acuerdo
 Totalmente de acuerdo

24. Quisiera dedicarle más tiempo a aprender a trabajar en ambientes virtuales

- Totalmente en desacuerdo
 En desacuerdo
 Ni de acuerdo ni en desacuerdo
 De acuerdo
 Totalmente de acuerdo

25. La tecnología tiene cierto riesgo para el proceso de enseñanza-aprendizaje

- Totalmente en desacuerdo
 En desacuerdo
 Ni de acuerdo ni en desacuerdo
 De acuerdo
 Totalmente de acuerdo

26. Siempre será mejor la enseñanza tradicional que la virtual

- Totalmente en desacuerdo
 En desacuerdo
 Ni de acuerdo ni en desacuerdo
 De acuerdo
 Totalmente de acuerdo

27. La tecnología genera muchos vicios académicos en el alumnado

- Totalmente en desacuerdo
 En desacuerdo
 Ni de acuerdo ni en desacuerdo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- De acuerdo
- Totalmente de acuerdo

28. La tecnología induce malas prácticas académicas en el alumnado

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

29. Es imposible llegar a manejar los recursos tecnológicos dada la velocidad de los avances

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

30. Mi disciplina no ayuda a dominar la tecnología para la enseñanza

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

31. Saber utilizar la tecnología requiere de una inversión alta y constante de todo tipo de recursos

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

32. Es mayor la inversión (tiempo, dinero) que hay que hacer en tecnología que los beneficios que aporta

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

33. Evito usar el Internet para trámites de la vida cotidiana (pagos, bancos y otros)

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

34. He invertido poco tiempo en prepararme para usar la tecnología en mi práctica docente

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

35. He invertido pocos recursos económicos en prepararme para usar la tecnología en mí práctica docente

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

36. Considero poco importante invertir mis propios recursos en equipo tecnológico para usar en mis clases

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

37. Andar buscando en Internet recursos educativos para el aprendizaje es una pérdida de tiempo

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

38. Trabajar con tecnología a distancia me produce estrés

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

39. Llevar cursos de capacitación en línea me provoca ansiedad o desesperación

- Totalmente en desacuerdo
- En desacuerdo
- Ni de acuerdo ni en desacuerdo
- De acuerdo
- Totalmente de acuerdo

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Sección 4 de 4

Uso de aulas virtuales

Expresé por favor su sentir respecto a la escala presentada

40. Si usted es usuario de las aulas virtuales que implementó la UJAT para compensar las clases presenciales ¿Cuál fue la razón principal por la que decidió hacerlo?

- Porque fue una instrucción
- Porque deseo aprender la modalidad en línea
- Porque deseo concluir el programa con mis alumnos
- Porque es mi obligación
- Porque me gusta
- Ninguna de las anteriores
- Otra...

41. Si usted no es usuario de las aulas virtuales que implementó la UJAT para compensar las clases presenciales ¿Cuál fue la razón principal por la que decidió no hacerlo?

- Porque no me considero competente en esa modalidad
- Porque no me avisaron con anticipación
- Porque trabajar en línea me produce estrés
- Porque invierto demasiado tiempo trabajando virtualmente
- Ninguna de las anteriores
- Otra...

42. Si usted se registró en el Diplomado en línea de Microsoft Teams que se ofreció en la UJAT, y LO CONCLUYÓ, elija de los siguientes enunciados, TRES que representen algunas razones por las que lo cursó

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- Porque estoy acostumbrado a trabajar en línea
- Porque ese tipo de capacitación es una obligación para el docente universitario
- Porque es una gran oportunidad para aprender algo nuevo
- Porque no quiero que mis alumnos piensen que soy antitecnológico
- Porque el curso era muy dinámico e interesante
- Porque la Universidad hizo un esfuerzo que hay que retribuir
- Porque lo necesitaba para usar las aulas virtuales
- Otra...

43. Si usted no se registró en el Diplomado en línea de Microsoft Teams que se ofreció en la UJAT, elija de los siguientes enunciados, TRES que representen algunas razones por las que no lo cursó.

- Porque no me gusta ese tipo de modalidad
- Porque mis conocimientos para usar tecnología son limitados
- Porque no quería estresarme
- Porque nunca seré un profesor virtual
- Porque no me avisaron con tiempo
- Porque no me interesa aprender en esta etapa de mi vida
- Otra...

2

Anexo B. Ley Federal de Derechos de Autor

Capítulo I: Disposiciones Generales

Artículo 1.- La presente Ley es de orden público y de observancia general en toda la República y tiene por objeto la protección de los datos personales en posesión de los particulares, con la finalidad de regular su tratamiento legítimo, controlado e informado, a efecto de garantizar la privacidad y el derecho a la autodeterminación informativa de las personas.

Artículo 2.- Son sujetos regulados por esta Ley, los particulares sean personas físicas o morales de carácter privado que lleven a cabo el tratamiento de datos personales, con excepción de:

- I. Las sociedades de información crediticia en los supuestos de la Ley para Regular las Sociedades de Información Crediticia y demás disposiciones aplicables, y
- II. Las personas que lleven a cabo la recolección y almacenamiento de datos personales, que sea para uso exclusivamente personal, y sin fines de divulgación o utilización comercial.

Artículo 3.- Para los efectos de esta Ley, se entenderá por:

- I. Aviso de Privacidad: Documento físico, electrónico o en cualquier otro formato generado por el responsable que es puesto a disposición del titular, previo al tratamiento de sus datos personales, de conformidad con el artículo 15 de la presente Ley.
- II. Bases de datos: El conjunto ordenado de datos personales referentes a una persona identificada o identificable.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- 2
- III. **Bloqueo:** La identificación y conservación de datos personales una vez cumplida la finalidad para la cual fueron recabados, con el único propósito de determinar posibles responsabilidades en relación con su tratamiento, hasta el plazo de prescripción legal o contractual de éstas. Durante dicho periodo, los datos personales no podrán ser objeto de tratamiento y transcurrido éste, se procederá a su cancelación en la base de datos que corresponde.
- IV. **Consentimiento:** Manifestación de la voluntad del titular de los datos mediante la cual se efectúa el tratamiento de estos.
- V. **Datos personales:** Cualquier información concerniente a una persona física identificada o identificable.
- VI. **Datos personales sensibles:** Aquellos datos personales que afecten a la esfera más íntima de su titular, o cuya utilización indebida pueda dar origen a discriminación o conlleve un riesgo grave para éste. En particular, se consideran sensibles aquellos que puedan revelar aspectos como origen racial o étnico, estado de salud presente y futuro, información genética, creencias religiosas, filosóficas y morales, afiliación sindical, opiniones políticas, preferencia sexual.
- 2
- VII. **Días:** Días hábiles.
- VIII. **Disociación:** El procedimiento mediante el cual los datos personales no pueden asociarse al titular ni permitir, por su estructura, contenido o grado de desagregación, la identificación del mismo.
- IX. **Encargado:** La persona física o jurídica que sola o conjuntamente con otras trate datos personales por cuenta del responsable.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- 2
- X. Fuente de acceso público: Aquellas bases de datos cuya consulta puede ser realizada por cualquier persona, sin más requisito que, en su caso, el pago de una contraprestación, de conformidad con lo señalado por el Reglamento de esta Ley.
- XI. Instituto: Instituto Federal de Acceso a la Información y Protección de Datos, a que hace referencia la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental.
- XII. Ley: Ley Federal de Protección de Datos Personales en Posesión de los Particulares.
- XIII. Reglamento: El Reglamento de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares.
- XIV. Responsable: Persona física o moral de carácter privado que decide sobre el tratamiento de datos personales.
- XV. Secretaría: Secretaría de Economía.
- 8
- XVI. Tercero: La persona física o moral, nacional o extranjera, distinta del titular o del responsable de los datos.
- XVII. Titular: La persona física a quien corresponden los datos personales.
- XVIII. Tratamiento: La obtención, uso, divulgación o almacenamiento de datos personales, por cualquier medio. El uso abarca cualquier acción de acceso, manejo, aprovechamiento, transferencia o disposición de datos personales.
- XIX. Transferencia: Toda comunicación de datos realizada a persona distinta del responsable o encargado del tratamiento.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

Artículo 4.- Los principios y derechos previstos en esta Ley, tendrán como límite en cuanto a su observancia y ejercicio, la protección de la seguridad nacional, el orden, la seguridad y la salud públicos, así como los derechos de terceros.

Artículo 5.- A falta de disposición expresa en esta Ley, se aplicarán de manera supletoria las disposiciones del Código Federal de Procedimientos Civiles y de la Ley Federal de Procedimiento Administrativo. Para la substanciación de los procedimientos de protección de derechos, de verificación e imposición de sanciones se observarán las disposiciones contenidas en la Ley Federal de Procedimiento Administrativo.

2

Universidad Juárez Autónoma de Tabasco.
México.

Anexo C. Ley Federal de Protección de Datos Personales

Capítulo I: Disposiciones Generales

Artículo 1.- La presente Ley es de orden público y de observancia general en toda la República y tiene por objeto la protección de los datos personales en posesión de los particulares, con la finalidad de regular su tratamiento legítimo, controlado e informado, a efecto de garantizar la privacidad y el derecho a la autodeterminación informativa de las personas.

Artículo 2.- Son sujetos regulados por esta Ley, los particulares sean personas físicas o morales de carácter privado que lleven a cabo el tratamiento de datos personales, con excepción de:

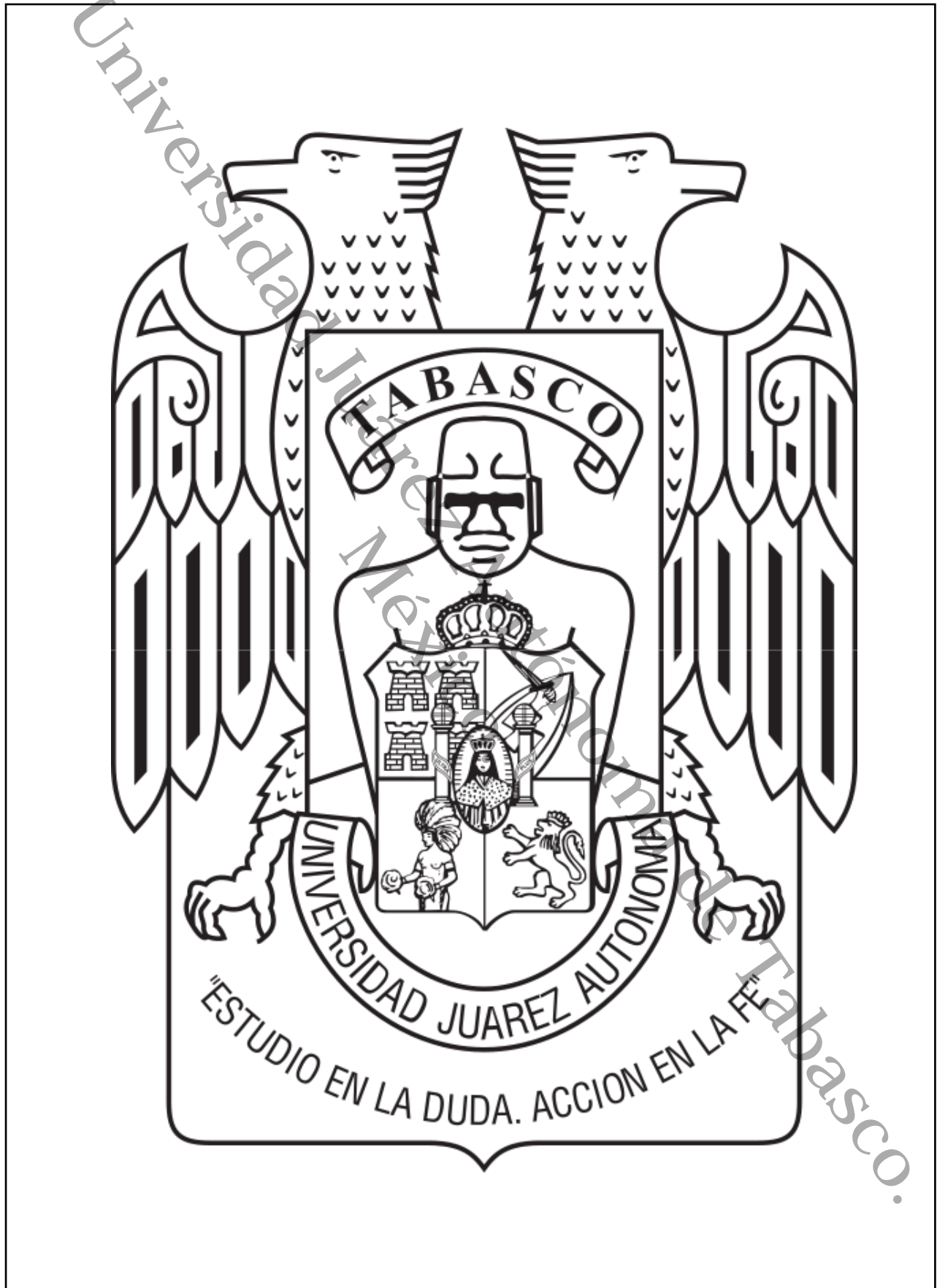
- I. Las sociedades de información crediticia en los supuestos de la Ley para Regular las Sociedades de Información Crediticia y demás disposiciones aplicables, y
- II. Las personas que lleven a cabo la recolección y almacenamiento de datos personales, que sea para uso exclusivamente personal, y sin fines de divulgación o utilización comercial.

Artículo 3.- Para los efectos de esta Ley, se entenderá por:

- I. Aviso de Privacidad: Documento físico, electrónico o en cualquier otro formato generado por el responsable que es puesto a disposición del titular, previo al tratamiento de sus datos personales, de conformidad con el artículo 15 de la presente Ley.
- II. Bases de datos: El conjunto ordenado de datos personales referentes a una persona identificada o identificable.

Barreras para el uso de innovaciones tecnológicas en profesores de educación superior desde la minería de datos

- 2
- III. Bloqueo: La identificación y conservación de datos personales una vez cumplida la finalidad para la cual fueron recabados, con el único propósito de determinar posibles responsabilidades en relación con su tratamiento, hasta el plazo de prescripción legal o contractual de éstas. Durante dicho periodo, los datos personales no podrán ser objeto de tratamiento y transcurrido éste, se procederá a su cancelación en la base de datos que corresponde.
- IV. Consentimiento: Manifestación de la voluntad del titular de los datos mediante la cual se efectúa el tratamiento de estos.
- V. Datos personales: Cualquier información concerniente a una persona física identificada o identificable.
- 18
- VI. Datos personales sensibles: Aquellos datos personales que afecten a la esfera más íntima de su titular, o cuya utilización indebida pueda dar origen a discriminación o conlleve un riesgo grave para éste. En particular, se consideran sensibles aquellos que puedan revelar aspectos como origen racial o étnico, estado de salud presente y futuro, información genética, creencias religiosas, filosóficas y morales, afiliación sindical, opiniones políticas, preferencia sexual.



Marcela de los Angeles Yanes Pérez.pdf

 Universidad Juárez Autónoma de Tabasco

Detalles del documento

Identificador de la entrega

trn:oid:::3117:582715705

Fecha de entrega

24 abr 2026, 2:00 p.m. GMT-6

Fecha de descarga

24 abr 2026, 2:46 p.m. GMT-6

Nombre del archivo

Marcela de los Angeles Yanes Pérez.pdf

Tamaño del archivo

11.2 MB

357 páginas

64.212 palabras

445.987 caracteres




10% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe


- ▶ Bibliografía
- ▶ Texto citado
- ▶ Coincidencias menores (menos de 20 palabras)
- ▶ Abstract

Fuentes principales

- 9%  Fuentes de Internet
- 2%  Publicaciones
- 0%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad




N.º de alerta de integridad para revisión

-  **Caracteres reemplazados**
440 caracteres sospechosos en N.º de páginas
Las letras son intercambiadas por caracteres similares de otro alfabeto.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.

Fuentes principales

- 9%  Fuentes de Internet
- 2%  Publicaciones
- 0%  Trabajos entregados (trabajos del estudiante)

Fuentes principales

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	Internet	dilemascontemporaneoseduccionpoliticayvalores.com	5%
2	Internet	doczz.es	<1%
3	Internet	docs.google.com	<1%
4	Internet	www.scielo.org.mx	<1%
5	Internet	repositorio.uwiener.edu.pe	<1%
6	Internet	www.apaga.es	<1%
7	Internet	www.coursehero.com	<1%
8	Internet	www.grupouno.net	<1%
9	Internet	docplayer.es	<1%
10	Internet	dspace.uazuay.edu.ec	<1%
11	Publicación	Mendez Sosa, Ellen Andrea. "Evaluating Contamination Levels in Fresh Foods Mar..."	<1%

12	Internet	www.dykinson.com	<1%
13	Internet	crea.ujaen.es	<1%
14	Internet	hdl.handle.net	<1%
15	Internet	repositorio.ucv.edu.pe	<1%
16	Internet	repositorio.autonoma.edu.co	<1%
17	Internet	ve.scielo.org	<1%
18	Internet	docta.ucm.es	<1%
19	Internet	jsret.knpub.com	<1%
20	Internet	cdn.goconqr.com	<1%
21	Publicación	Tai-Ying Chiang. "Analysis of Responsible Tourism Practices in the Wake of COVID..."	<1%
22	Internet	repository.unipiloto.edu.co	<1%
23	Internet	ri.ujat.mx	<1%
24	Internet	repositorio.ucsg.edu.ec	<1%
25	Internet	idss.iocspublisher.org	<1%

26	Internet	udgvirtual.udg.mx	<1%
27	Internet	mdh.diva-portal.org	<1%
28	Internet	cienciadigital.org	<1%
29	Internet	renati.sunedu.gob.pe	<1%
30	Internet	view.genial.ly	<1%
31	Publicación	Andrea-Jalla Castañeda-Maizel, Haydee-Katherine Tito-Estevez, Julio-César Mateu...	<1%
32	Internet	repositorio.utn.edu.ec	<1%
33	Publicación	Abid Hussain, Muhammad Rafiq. "Perceptions of Academic Librarians Toward Op...	<1%
34	Publicación	Liyang Bai, Changlong Liao, Changxiang Wang, Meng Zhang, Fanbao Meng, Mingj...	<1%
35	Publicación	Siyuan Li, Yongpan Wang, Chaopeng Dong, Shouguo Yang et al. "LibAM: An Area ...	<1%
36	Internet	dialnet.unirioja.es	<1%
37	Internet	pt.slideshare.net	<1%
38	Internet	bi-ppr.com.mx	<1%
39	Internet	es.altapps.net	<1%

40	Internet	iesalc.unesco.org	<1%
41	Internet	learn.microsoft.com	<1%
42	Internet	mediateca.ctera.org.ar	<1%
43	Internet	ouci.dntb.gov.ua	<1%
44	Publicación	Flores Rodriguez, Magda Rivana. "El plan de vigilancia comunitaria del Parque Na..."	<1%
45	Publicación	Hassounah, Jamil. "Participative Decision-Making and Employee Attitudes: A Stud..."	<1%
46	Publicación	Mario Alas Solís, Asly Mayté Berríos Palacios. "Adaptación y Escalamiento de Enfo..."	<1%
47	Publicación	Mosime, Afnan Yousef. "The Effect of Entrepreneurship Education on Entreprene..."	<1%
48	Internet	revistas.unap.edu.pe	<1%
49	Internet	rstudio-pubs-static.s3.amazonaws.com	<1%
50	Internet	www.slideshare.net	<1%