



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



**IDENTIFICACIÓN DE PATRONES DE CONDUCTA, ASOCIADAS
CON LAS BARRERAS PARA EL USO DE INNOVACIONES
TECNOLÓGICAS EN ESTUDIANTES DE EDUCACIÓN SUPERIOR,
APLICANDO TÉCNICAS DE 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:

Wendy Guadalupe Azuara García

Director(es) de Trabajo Recepcional:

**Dra. Martha Patricia Silva Payró
Dr. Guillermo De los Santos Torres**

Cuerpo Académico: Gestión de Tecnologías de la Información

Línea de Generación y Aplicación del Conocimiento de la Maestría:

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

Cunduacán, Tabasco

Febrero, 2023



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



**IDENTIFICACIÓN DE PATRONES DE CONDUCTA, ASOCIADAS
CON LAS BARRERAS PARA EL USO DE INNOVACIONES
TECNOLÓGICAS EN ESTUDIANTES DE EDUCACIÓN SUPERIOR,
APLICANDO TÉCNICAS DE 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:

Wendy Guadalupe Azuara García

Director(es) de Trabajo Recepcional:

Dra. Martha Patricia Silva Payró
Dr. Guillermo De los Santos Torres

Cuerpo Académico: Gestión de Tecnologías de la Información

Jurado Revisor:

Dr. Pablo Payró Campos
Dr. Gerardo Arceo Moheno
Dra. Verónica García Martínez
Mtro. Rafael Mena De la Rosa

Línea de Generación y Aplicación del Conocimiento de la Maestría:
**Administración, diseño e implementación de integración de
soluciones de TI**

Cunduacán, Tabasco

Febrero, 2023

Cunduacán, Tabasco a 23 de enero de 2023

Asunto: Oficio de Autorización

A quien corresponda:

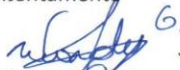
El que suscribe, autoriza por medio del presente escrito a la Universidad Juárez Autónoma de Tabasco para que utilice tanto física como digitalmente la tesis de maestría denominada, **"IDENTIFICACIÓN DE PATRONES DE CONDUCTA, ASOCIADAS CON LAS BARRERAS PARA EL USO DE INNOVACIONES TECNOLÓGICAS EN ESTUDIANTES DE EDUCACIÓN SUPERIOR, APLICANDO TÉCNICAS DE 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.

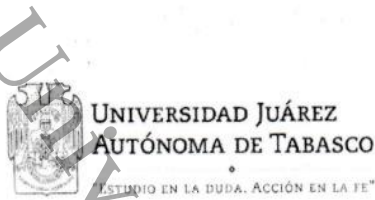
Por lo antes manifestado, libero a la Universidad Juárez Autónoma de Tabasco de cualquier reclamación legal que pudiera ejercer respecto al uso y manipulación de la tesis mencionada y para los fines estipulados en este documento.

Sin otro particular, agradezco de antemano la atención prestada a la presente y me despido enviándole un cordial saludo.

Atentamente


Wendy Guadalupe Azuara García
Matrícula: 202H19007

C.c.p. Dr. Eddy Arquímedes García Alcocer.- Encargado del despacho de la coordinación de posgrado.
Estudiante.



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. 0661/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 **"Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos"**, a realizar por el **C. Wendy Guadalupe Azuara García**, 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. Óscar Alberto González González
Director



C.e.p. Dr. Eddy Arquímedes García Alcocer. 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



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. 0662/DACYTI/CP/2022

Asunto: Asignación de Director de Tesis

Dr. Guillermo De los Santos Torres
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 **"Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos"**, a realizar por el **C. Wendy Guadalupe Azuara García**, 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. Óscar Alberto González González
Director



C.c.p. Dr. Eddy Arquímedes García Alcoer. Encargado del Despacho de la Coordinación de Posgrado
Alumno
Archivo

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



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



F6: Respuesta de jurado

Cunduacán, Tabasco, a 06 de diciembre de 2022.

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 *"Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos"*, realizada por el C. Wendy Guadalupe Azuara García, 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

Dr. Pablo Payro Campos

Dra. Veronica Garcia Martinez

Dr. Gerardo Arceo Moreno

Mtro. Rafael Mena de la Rosa



c.c.p. Dr. Eddy Arquimedes Garcia Alcocer. Encargada del despacho de la Coordinación de Posgrado Estudiante.

Universidad Juárez Autónoma de Tabasco



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



Cunduacán, Tabasco, a 23 de enero de 2023.

Asunto: Cesión de derechos

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

El que suscribe la presente, declara que el trabajo de tesis titulado, "IDENTIFICACIÓN DE PATRONES DE CONDUCTA, ASOCIADAS CON LAS BARRERAS PARA EL USO DE INNOVACIONES TECNOLÓGICAS EN ESTUDIANTES DE EDUCACIÓN SUPERIOR, APLICANDO TÉCNICAS DE MINERÍA DE DATOS" es de mi autoría intelectual y por lo tanto cedo 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 derecho de autor ante las autoridades competentes

Atentamente

[Firma manuscrita]
Wendy Guadalupe Azuara García

[Firma manuscrita]
Dra. Martha Patricia Silva Payró

[Firma manuscrita]
Dr. Guillermo De los Santos Torres

c.c.p. Dr. Eddy Arquímedes García Aicocer. Encargado del despacho de la Coordinación de Posgrado. Estudiante.





DIVISIÓN ACADÉMICA DE CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



Cunduacán, Tabasco a 20 de enero de 2023
Oficio No. 0160/DACYTI/CP/2023

Asunto: Autorización de impresión de Tesis

C. Wendy Guadalupe Azuara García
Matricula: 202H19007

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 **"Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos"**, para presentar examen y 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. 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

Este trabajo fue realizado como parte del programa de posgrado de la División Académica de Ciencias y Tecnologías de la Información (DACyTI) de la Universidad Juárez Autónoma de Tabasco como becaria por el Consejo Nacional de Ciencia y Tecnología (CONACYT) en México.

De forma particular agradezco a la Dra. Martha Patricia Silva Payró y el Dr. Guillermo De los Santos Torres, profesores investigadores de la Universidad Juárez Autónoma de Tabasco quienes fueron directores de este proyecto de tesis y mostraron siempre disposición y apoyo para esta servidora.

Universidad Juárez Autónoma de Tabasco.
México.

Dedicatorias

La familia es el primer grupo social al que pertenecemos, es el pilar que nos sostiene para seguir adelante cuando nos sentimos desorientados y procura lo mejor para nosotros de modo que podamos crecer en las diversas áreas de la vida.

En ese mismo sentir, quiero dedicar este trabajo a mi esposo con quien he decidido construir mi familia y que desde el principio estuvo alentando mi crecimiento en esta etapa que se ha concluido.

Mis dedicatorias se extienden a mi familia de origen, mis padres. Quienes se esforzaron por brindarme el beneficio de una educación, misma que me permitió emprender este proyecto.

Universidad Juárez Autónoma de Tabasco.
México.

Resumen

El uso de innovaciones tecnológicas en estudiantes es un fenómeno social que se ha visto perjudicado por una serie de factores denominados barreras y que interfieren en el rendimiento académico de la población estudiantil en cualquier nivel educativo. En la actualidad este tema ha incrementado su relevancia debido a la dinámica social provocada por la pandemia del coronavirus, ocasionando que las actividades académicas se realicen a través de plataformas virtuales a causa del distanciamiento social. Sin embargo, ha sido objeto de estudio para diferentes autores desde hace años, los cuales han tratado de conocer y analizar los factores que han influido en la apropiación tecnológica en estudiantes llevando clases con la modalidad a distancia. Por tal motivo, el presente trabajo a través de un enfoque cuantitativo tuvo como propósito la aplicación de las técnicas de minería de datos tales como clasificación, agrupamiento y reglas de asociación, para identificar patrones de conducta relacionados con las barreras relacionadas al uso de innovaciones tecnológicas en estudiantes a partir de los datos previamente recolectados de estudiantes de licenciatura en una universidad pública del sureste mexicano. Para el desarrollo de este proyecto se seleccionaron las herramientas Orange, WEKA y Lenguaje R con las cuales se realizaron diferentes pruebas y evaluaciones que permitieron realizar un análisis de usabilidad. Asimismo, las herramientas empleadas fueron de utilidad durante las diferentes etapas del proceso metodológico KDD (Descubrimiento de Conocimiento de Base de Datos). De acuerdo con las pruebas realizadas y la interpretación de resultados, se concluye que los principales patrones de conducta en los estudiantes se relacionan principalmente con el acceso a internet, falta de atención de los profesores y uso compartido de recursos tecnológicos.

Introducción

El presente trabajo contiene cinco capítulos que presentan las generalidades de la investigación, autores teóricos y referentes, el desarrollo metodológico, resultados y conclusiones. Con el fin de presentar la identificación de patrones de conducta asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos. Por lo tanto, en este apartado se presenta un resumen de cada capítulo desarrollado en este proyecto.

El capítulo uno se encuentra denominado generalidades, cuyo contenido presenta el planteamiento del problema y sus antecedentes, dando pauta a la construcción de la pregunta de investigación, alcances y limitaciones así como sus objetivos. Asimismo, este capítulo contiene la justificación del proyecto, el enfoque metodológico y los procesos a emplear.

En el capítulo dos se encuentra el marco general de la investigación, en este se presenta el marco teórico que expone a diferentes autores teóricos con respecto a las barreras económica, acceso, pedagógica y ambiente académico, mismo que se encuentra estructurado en tres subtemas. Por otro lado este capítulo contiene un marco referencial que presenta trabajos nacionales e internacionales aplicando diferentes técnicas para la resolución del problema de investigación. Por último, se encuentran el marco conceptual, tecnológico y legal que contienen aspectos técnicos de minería de datos como términos aplicados al proyecto, herramientas y tipos de licencias de software.

El desarrollo metodológico se encuentra en el capítulo tres de este trabajo, consta de tres apartados principales: información general acerca del instrumento de recolección de datos, análisis y selección de herramientas de minería de datos a emplear, terminando con el proceso KDD constituido de ocho tareas o actividades que fueron desarrolladas

con las herramientas Orange, WEKA y Lenguaje R aplicando las técnicas de clasificación, agrupación y reglas de asociación.

En el capítulo de cuatro, se presentan las pruebas y resultados aplicando las tres técnicas de minería de datos y las herramientas seleccionadas. El propósito de esto fue mostrar la construcción de los diferentes modelos para seleccionar los que generaban mejores resultados y poder realizar la interpretación de los patrones identificados. También, con las pruebas y resultados obtenidos con cada herramienta, se elaboró un análisis de usabilidad con el fin de presentar la herramienta mejor evaluada de acuerdo con el usuario.

El último capítulo consta de tres apartados, el primero es el de las conclusiones permitiendo contrastar el cumplimiento de los objetivos y la respuesta a la pregunta de investigación. El segundo presenta las recomendaciones clasificadas como tecnológicas y recomendaciones resultantes de los patrones de conducta identificados. Finalmente, se presentan los trabajos futuros sugeridos en el área pedagógica y el área de tecnologías de acuerdo con el trabajo de investigación presentado.

Índice general

Índice de tablas	xix
Índice de figuras	xx
Capítulo 1. Generalidades	1
1.1 Barreras para el uso de innovaciones tecnológicas.....	1
1.2 Planteamiento del problema	4
1.2.1 Definición del problema.....	4
1.2.2 Delimitación de la investigación	6
1.2.3 Pregunta de investigación.....	7
1.3 Objetivos.....	8
1.3.1 Objetivo general.....	8
1.3.2 Objetivos específicos.....	8
1.4 Justificación	8
1.5 Metodología utilizada.....	9
1.5.1 Enfoque de la investigación	9
1.5.2 Fuentes de investigación	10
1.5.3 Técnicas de recolección de datos.....	10
1.5.4 Metodología para aplicar.....	11
Capítulo 2. Marco de la investigación	13
2.1 Marco teórico	13
2.1.1 Acceso y costo para el uso de tecnologías en la formación superior.....	13
2.1.2 Interacción con las tecnologías desde el enfoque pedagógico	15
2.1.3 La influencia del ambiente académico en estudiantes universitarios	17

2.2	Marco referencial	20
2.2.1	El uso de tecnologías en estudiantes mexicanos.....	20
2.2.2	El uso de innovaciones tecnológicas en estudiantes de educación superior desde diferentes enfoques en otros países.....	24
2.2.3	Aspectos relevantes.....	28
2.3	Marco conceptual.....	30
2.3.1	Barrera económica.....	30
2.3.2	Barrera del acceso.....	30
2.3.3	Barrera pedagógica.....	31
2.3.4	Barrera del ambiente académico	31
2.3.5	Patrones de conducta.....	31
2.3.6	Minería de datos	32
2.3.7	Proceso KDD	33
2.3.8	Algoritmos en la minería de datos.....	33
2.3.9	Técnicas de minería de datos	34
2.3.10	Modelo FURPS	35
2.4	Marco tecnológico.....	36
2.4.1	Herramientas de minería de datos.....	36
2.4.2	Herramientas complementarias	41
2.5	Marco legal	42
2.5.1	Licencia de software GPL	42
2.5.2	Licencia de software propietario.....	42

2.5.3 Ley Federal de Protección de Datos Personales en Posesión de Particulares 46

Capítulo 3. Aplicación de la metodología y desarrollo 48

3.1 Información general 48

3.1.1 Descubrimiento y entendimiento de la información 48

3.1.2 Diseño y descripción del instrumento 50

3.1.3 Visualización y exploración de los datos 54

3.2 Análisis y selección de herramientas de minería de datos 58

3.2.1 Procedimiento y análisis de indicadores 60

3.2.2 Características de las herramientas seleccionadas 66

3.3 Proceso KDD 70

3.3.1 Aprendizaje y dominio de la aplicación 72

3.3.2 Crear un conjunto de datos destino 76

3.3.3 Limpieza y procesamiento de datos 78

3.3.4 Reducción y proyección de los datos 82

3.3.5 Selección de funciones y algoritmos de minería de datos 85

3.3.6 Minería de datos 88

3.3.7 Interpretación 114

3.3.8 Uso del conocimiento descubierto 115

Capítulo 4. Pruebas y Resultados 117

4.1 Pruebas realizadas por herramienta 117

4.1.1 Orange 118

4.1.2 WEKA 139

4.1.3	Lenguaje R.....	158
4.1.4	Análisis de usabilidad de las herramientas	177
4.2	Interpretación de resultados	184
4.2.1	Clasificación.....	184
4.2.2	Agrupamiento.....	198
4.2.3	Reglas de asociación.....	213
4.2.4	Resultados del análisis de usabilidad	227
Capítulo 5.	Conclusiones, recomendaciones y trabajos futuros	231
5.1	Conclusiones	231
5.2	Recomendaciones	242
5.3	Trabajos futuros.....	245
Referencias	248
Glosario	269

Índice de tablas

Tabla 1. Dimensiones de la barrera pedagógica.	17
Tabla 2. Distribución de estudiantes por división académica y género.	49
Tabla 3. Ítems contenidos por sección de la encuesta.	51
Tabla 4. Variables de investigación.	53
Tabla 5. Factores y criterios del modelo FURPS.	59
Tabla 6. Análisis de la selección de las herramientas para minería de datos aplicando la metodología FURPS.	61
Tabla 7. Ponderación de las herramientas de acuerdo con la metodología FURPS.	63
Tabla 8. Herramientas de minería de datos clasificadas por tipo de licencia.	64
Tabla 9. Clasificación por etapas de la metodología KDD.	72
Tabla 10. Conjunto de datos.	77
Tabla 11. Cambio de nomenclatura en las variables.	84
Tabla 12. Algoritmos de minería de datos.	87
Tabla 13. Decisiones del árbol generado en Orange	186
Tabla 14. Decisiones del árbol generado en WEKA.	191
Tabla 15. Decisiones del árbol generado en lenguaje R	196
Tabla 16. Escala de ponderación de las herramientas.	227
Tabla 17. Matriz de evaluación para las herramientas	228
Tabla 18. Resultados obtenidos de la evaluación por herramienta	229
Tabla 19. Principales patrones identificados.	240

Índice de figuras

Figura 1. Evolución de la disponibilidad de TI en hogares mexicanos.	2
Figura 2. Errores de caracteres especiales.	56
Figura 3. Verificación de duplicidad de datos.	57
Figura 4. Proceso KDD.	71
Figura 5. Interfaz Gráfica de Orange.	74
Figura 6. Interfaz gráfica de Explorer en WEKA.	75
Figura 7. Entorno de trabajo de Lenguaje R y RStudio.	76
Figura 8. Depuración de conflictos entre datos.	80
Figura 9. Comparación del dataset posterior a la limpieza de datos.	82
Figura 10. Proyección de datos en Orange.	83
Figura 11. Selección del dataset en Orange.	89
Figura 12. Representación gráfica de la función distributions.	91
Figura 13. Evaluación del árbol de decisión en Orange.	92
Figura 14. Ventana de configuración de K-means en Orange.	93
Figura 15. Gráfica de Silueta en Orange.	94
Figura 16. Aplicación de Reglas de asociacion en Orange.	95
Figura 17. Estudio de los datos en WEKA.	96
Figura 18. Generación del modelo árbol de decisión WEKA.	97
Figura 19. Aplicación del modelo cluster en WEKA.	99
Figura 20. Aplicación del modelo de asociación WEKA.	100
Figura 21. Importación del dataset a Lenguaje R.	101
Figura 22. Descarga de librerías del árbol de decisión en Lenguaje R.	103
Figura 23. Creación del modelo de árbol de decisión en Lenguaje R.	104
Figura 24. Modelo de evaluación del árbol de decisión en lenguaje R.	104
Figura 25. Elaboración del árbol de decisión en Lenguaje R.	105
Figura 26. Descarga de librerías de K-means en Lenguaje R.	106
Figura 27. Escalado de datos en Lenguaje R.	107
Figura 28. Creación de clústeres en Lenguaje R.	108
Figura 29. Graficar clústeres en Lenguaje R.	109
Figura 30. Métodos para encontrar el número óptimo de clústeres en Lenguaje R.	109
Figura 31. Creación de nuevos clústeres con el número óptimo.	110
Figura 32. Descarga de librerías del algoritmo Apriori en Lenguaje R.	111
Figura 33. Elaboración de Transacciones en Lenguaje R.	111
Figura 34. Exportación de transacciones generadas en Lenguaje R.	112
Figura 35. Parametros del algoritmo Apriori Lenguaje R.	112
Figura 36. Visualización de reglas de asociación en Lenguaje R.	113
Figura 37. Parámetros de inicio del árbol de decisión.	119
Figura 38. Resultados de métricas de evaluación en árbol de decisión con pregunta raíz P.1.	120
Figura 39. Matriz de árbol de decisión con pregunta raíz P.1.	121
Figura 40. Prueba de árbol de decisión con pregunta raíz P.1.	121
Figura 41. Resultados de métricas de evaluación en árbol de decisión con pregunta raíz P.3.	122

Figura 42. Matriz de árbol de decisión con pregunta raíz P.3.....	123
Figura 43. Prueba de árbol de decisión con pregunta raíz P.3.....	123
Figura 44. Resultados de métricas de evaluación en árbol de decisión con pregunta raíz P.7	124
Figura 45. Matriz de árbol de decisión con pregunta raíz P.7.....	125
Figura 46. Prueba de árbol de decisión con pregunta raíz P.7.....	126
Figura 47. K-means interactivo con dos agrupaciones.....	127
Figura 48. Diagrama de silueta con dos clústeres en Orange.....	128
Figura 49. Gráfica Plot en Orange con dos agrupaciones.....	129
Figura 50. K-means interactivo con cuatro agrupaciones.....	130
Figura 51. Diagrama de silueta con cuatro agrupaciones	131
Figura 52. Gráfica Plot en Orange con cuatro agrupaciones	132
Figura 53. K-means interactivo con cinco agrupaciones	133
Figura 54. Diagrama de silueta con cuatro agrupaciones	134
Figura 55. Gráfica Plot en Orange con cinco agrupaciones	135
Figura 56. Reglas de asociación en orange con cincuenta por ciento de confianza ..	136
Figura 57. Reglas de asociación en Orange con soporte mínimo del quince por ciento.	137
Figura 58. Reglas de asociación en orange con soporte mínimo del veinte por ciento	138
Figura 59. Métricas de evaluación en WEKA de árbol de decisión con pregunta raíz P.1	141
Figura 60. Visualización del diagrama de árbol con pregunta raíz P.1	142
Figura 61. Métricas de evaluación en WEKA de árbol de decisión con pregunta raíz P.3	143
Figura 62. Visualización del diagrama de árbol con pregunta raíz P.3.....	144
Figura 63. Métricas de evaluación en WEKA de árbol de decisión con pregunta raíz P.7	145
Figura 64. Visualización del diagrama de árbol con pregunta raíz P.7	146
Figura 65. Modelo de entrenamiento de dos clústeres en WEKA	148
Figura 66. K-means prueba de dos agrupaciones por edad.....	148
Figura 67. Modelo de entrenamiento de cuatro clústeres en WEKA	149
Figura 68. Diagrama de cuatro agrupaciones en WEKA	150
Figura 69. Entrenamiento de cinco clústeres en WEKA	150
Figura 70. K-means prueba de cinco agrupaciones por materia en WEKA.....	151
Figura 71. Diagramas de cinco agrupaciones con la variable P.22 en WEKA	152
Figura 72. Métricas del algoritmo Apriori en WEKA con cincuenta por ciento de confianza	153
Figura 73. Reglas de asociación en WEKA con cincuenta por ciento de confianza ...	154
Figura 74. Métricas del algoritmo Apriori en WEKA con setenta por ciento de confianza	155
Figura 75. Reglas de asociación en WEKA con setenta por ciento de confianza.....	155
Figura 76. Métricas del algoritmo Apriori en WEKA con noventa por ciento de confianza	156
Figura 77. Reglas de asociación en WEKA con noventa por ciento de confianza.....	157

Figura 78. Árbol de decisión Lenguaje R con pregunta raíz P.1.....	159
Figura 79. Visualización del diagrama de árbol con pregunta raíz P.1	160
Figura 80. Métricas de evaluación en Lenguaje R de árbol de decisión con pregunta raíz P.1	161
Figura 81. Árbol de decisión Lenguaje R con pregunta raíz P.3.....	162
Figura 82. Métricas de evaluación en Lenguaje R de árbol de decisión con pregunta raíz P.3	163
Figura 83. Árbol de decisión Lenguaje R con pregunta raíz P.7.....	164
Figura 84. Métricas de evaluación en Lenguaje R de árbol de decisión con pregunta raíz P.7	165
Figura 85. Métricas generales de calidad con dos agrupamientos en Lenguaje R.....	167
Figura 86. Diagramas de dos agrupaciones en Lenguaje R.....	168
Figura 87. Vectores con cuatro agrupamientos en Lenguaje R.....	168
Figura 88. Diagramas de cuatro agrupaciones en Lenguaje R.....	169
Figura 89. Vectores con cinco agrupamientos en Lenguaje R	170
Figura 90. Diagramas de cinco agrupaciones en Lenguaje R	171
Figura 91. Métricas del algoritmo Apriori en Lenguaje R con cincuenta por ciento de confianza	172
Figura 92. Reglas de asociación en Lenguaje R con cincuenta por ciento de confianza	173
Figura 93. Métricas del algoritmo Apriori en Lenguaje R con setenta por ciento de confianza	174
Figura 94. Reglas de asociación en Lenguaje R con setenta por ciento de confianza	175
Figura 95. Métricas del algoritmo Apriori en Lenguaje R con noventa por ciento de confianza	175
Figura 96. Reglas de asociación en Lenguaje R con noventa por ciento de confianza	176
Figura 97. Curva de aprendizaje de herramientas para minería de datos	178
Figura 98. Error en Orange: no se asignó el nodo raíz.....	181
Figura 99. Error de lenguaje R: no se encontró la función repart.plot	182
Figura 100. Error de lenguaje R: Error en el argumento escrito	182
Figura 101. Resultados de métricas de evaluación del modelo de árbol de decisión en Orange	185
Figura 102. Matriz de confusión del árbol de decisión en Orange.....	185
Figura 103. Representación gráfica de árbol decisión en Orange.....	186
Figura 104. Métricas de evaluación del modelo de árbol de decisión en WEKA	190
Figura 105. Representación gráfica de árbol decisión en WEKA	191
Figura 106. Métricas de evaluación del modelo de árbol de decisión en lenguaje R .	194
Figura 107. Diagrama de árbol Lenguaje R.....	195
Figura 108. Métricas de calidad del algoritmo k-Means en Orange.....	199
Figura 109. K-means Orange agrupamiento por edad	200
Figura 110. K-means Orange gráficos de dispersión de acuerdo con el clúster.....	201
Figura 111. Métricas de calidad del algoritmo SimpleKMeans en WEKA.....	204
Figura 112. SimpleKMeans WEKA gráficos de dispersión de acuerdo con el clúster	205
Figura 113. Métricas de evaluación de agrupación en R.....	209

Figura 114. KMeans Lenguaje R gráficos de agrupamiento..... 210
Figura 115. Reglas de asociación Orange 215
Figura 116. Resultados de métricas de calidad del algoritmo Apriori en WEKA 216
Figura 117. Resultados de reglas de asociación en WEKA 217
Figura 118. Resultados de métricas de calidad del algoritmo Apriori en Lenguaje R 218
Figura 119. Resultados de reglas de asociación en Lenguaje R..... 218
Figura 120. Diagrama de Venn con las reglas de asociación..... 219

Universidad Juárez Autónoma de Tabasco
México.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Capítulo 1. Generalidades

1.1 Barreras para el uso de innovaciones tecnológicas

El uso de innovaciones tecnológicas ha cobrado importancia en el ámbito educativo en todos los niveles. Para el caso de la educación superior y sus estudiantes, esto no ha sido la excepción, ya que les ha permitido obtener diversos beneficios de su uso y apropiación; sin embargo, al mismo tiempo ha generado obstáculos o barreras para que los estudiantes universitarios logren adaptarse a las nuevas tecnologías.

Es por tal motivo que las barreras tecnológicas en estudiantes universitarios han sido abordadas por diferentes autores, inclusive desde el inicio de las clases en la modalidad a distancia. De acuerdo con Sánchez & Luján (2020) se presentan dos factores: el que se debe directamente al proceso de formación es el pedagógico y el que está indirectamente relacionado con el proceso de formación es el acceso a internet y que serán explicados en lo sucesivo.

2

2

En lo referente a la barrera pedagógica contiene dos aspectos principales, la modalidad de educación, ya sea a distancia o presencial y las dificultades que presenta el docente con relación al uso de Tecnologías de la Información (TI) y su implementación en el aula mencionado por Marín *et al.*, (2017), enfatizando en la necesidad de que las universidades comiencen a considerar dentro de sus prioridades la formación de docentes, sobre el uso de las tecnologías con miras a enfrentar nuevos desafíos dentro del aula de manera ágil y oportuna con las nuevas generaciones de estudiantes que se vienen adaptando a la cultura digital.

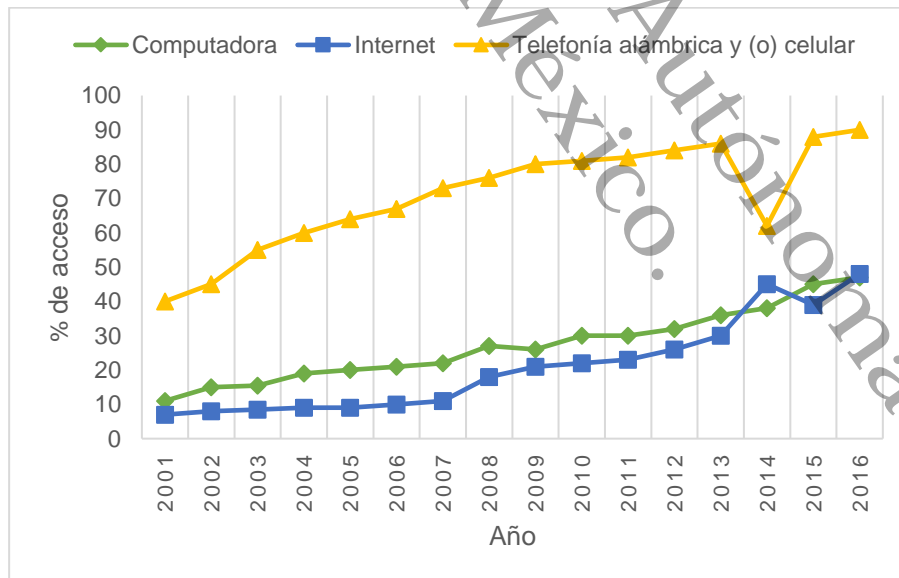
1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Redactar aspectos generales que permitan al lector tener una idea o panorama del tema que se investiga y de las organizaciones o sectores socioeconómicos a los que se dirige la investigación. Opcionalmente, puede subdividirse en temas. No confundir con la redacción del marco referencial. El primer párrafo de cada capítulo NO lleva sangría en la primera línea.

En lo concerniente al acceso a internet en la Figura 1 se muestra la evolución de la disponibilidad de las TI en los hogares mexicanos en el periodo comprendido del 2001 al 2016.

Figura 1.
Evolución de la disponibilidad de TI en hogares mexicanos.



Nota: Gómez et al., (2018).

Aunque el acceso a internet aumenta en los hogares mexicanos, no logra alcanzar el 100% de cobertura por lo que hay una población que no cuenta con el recurso. A esto

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

habría que añadir que existen estudiantes que rentan el servicio de internet o se conectan en espacios públicos o ajenos, como la casa de un amigo o vecino (Morares *et al.*, 2020).

Hasta el momento se han abordado aspectos de accesibilidad a internet, así como el factor pedagógico; sin embargo, hay otros factores que no han sido abordados y que están relacionados con la motivación de los estudiantes para mantenerse en la educación universitaria en línea, tales como la situación económica y el ambiente organizacional, los cuales fueron mencionados por Troche (2019) y que se exponen en lo subsecuente.

Sin lugar a duda, abordar el aspecto económico es de suma importancia ya que se puede presentar en el estudiante inclusive desde mucho antes de entrar a la universidad y puede ser, un factor decisivo para ingresar a la educación superior y mantenerse en ella una vez dentro, ya que la falta de recursos económicos para acceder y disponer de las tecnologías y sus innovaciones, lo coloca en franca desventaja (Martínez, 2021).

El segundo factor motivacional es el ambiente organizacional, Rodríguez y Espinoza (2017) explican que los estudiantes de nivel superior tienden a sentirse incómodos dentro de un ambiente virtual al trabajar de forma colaborativa con otros compañeros debido a la poca participación y supervisión del docente. En ese mismo sentido, Ramírez (2019) explica que esta circunstancia puede ocasionar falta de confianza hacia las autoridades escolares provocando la poca colaboración del estudiante a la resolución del problema. Así mismo dentro del ambiente organizacional

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

se puede incluir el estrés académico en los alumnos ocasionado por las demandas del entorno o ambiente escolar por la carga académica (Álvarez-Silva *et al.*, 2018).

Por lo anteriormente expuesto, se considera que las preocupaciones en los estudiantes se inclinan a factores como la accesibilidad, economía, aspectos pedagógicos y académicos que se pueden definir como las barreras para lograr concluir sus estudios, así como para hacer uso de la tecnología.

1.2 Planteamiento del problema

1.2.1 Definición del problema

La población a nivel mundial está enfrentando un gran desafío desde diferentes enfoques, como el social, la salud y la educación debido a la pandemia por COVID 19. En el aspecto educativo se realizó un cierre masivo de instituciones educativas en más de 190 países alrededor del mundo donde más de 1,200 millones de estudiantes dejaron de tener clases presenciales (Comisión Nacional de América Latina [CEPAL], 2020). A partir de esta situación, las instalaciones del hogar se han vuelto una extensión de la escuela.

Cada país de forma particular ha tomado medidas ante las circunstancias descritas y que actualmente prevalecen, tal es el caso de México que consideró implementar diferentes medios como la radio, televisión y plataformas virtuales para continuar con las clases desde casa. En ese sentido se destaca que en México un 55.7% de la población estudiantil de nivel superior de escuelas públicas y privadas, ha tenido que hacer uso de la computadora como herramienta principal seguido de teléfonos inteligentes para recibir

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

sus clases a distancia y no perder el ciclo escolar (Instituto Nacional de Estadística Geografía e Informática [INEGI], 2021).

Del mismo modo a nivel local, Tabasco no se encuentra al margen de este desafío educativo provocado por la pandemia, debido a que ya es una realidad en cualquier estudiante universitario la educación a distancia así como las clases a través de plataformas de educación virtual y la implementación de herramientas tecnológicas como sucedió en la Institución de Educación Superior (IES) caso de estudio. En ella se ha presenciado que sus estudiantes han sufrido estragos derivados del uso de tecnologías y sus innovaciones se han visto perjudicados por una serie de barreras que se destacan en la brecha digital como se verá en lo subsecuente.

La primera barrera es la económica que considera los precios para adquirir dispositivos tecnológicos y servicio de internet ya que hay muchas personas que no cuentan con recursos suficientes para adquirirlos. La segunda barrera es el acceso que hace referencia a la infraestructura para el uso de las tecnologías, ya que actualmente en muchos lugares del estado hay estudiantes que no cuentan con acceso a internet por encontrarse en zonas alejadas de la ciudad y de las redes de conexión (Arellano-Becerril y Gómez-Zermeño, 2021). La tercera barrera es la pedagógica y se refiere a la preparación del docente para la materia que imparte, así como la capacitación en tecnologías para impartir sus clases de manera ágil y oportuna (Pérez-López *et al.*, 2021).

2 Por último, la barrera del ambiente académico se refiere a la percepción que tienen los estudiantes sobre su contexto escolar actual en los entornos virtuales, la carga

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

académica y la interacción con los compañeros y profesores. De modo que se involucran aspectos motivacionales que son importantes para la formación académica del estudiante (Muñoz-Vidal y Beltrán-Véliz, 2020).

Las reacciones y comportamientos de los estudiantes ante las situaciones presentadas en cada una de las barreras se han agravado con la pandemia, volviendo más visibles las carencias de disponibilidad tecnológica que ya existían (Vicentini, 2020). Con base en lo anterior, es necesario conocer los patrones en los cuales se hace visible esta problemática.

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 estudiantes sobre el uso de Tecnologías de la Información (TI) durante la situación social que se enfrenta con la pandemia COVID-19.
- En esta investigación se presentan estudios y conceptos sobre las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, que guardan relación con el instrumento utilizado para la realización de las encuestas, con estos dos elementos se podrá tener un mayor entendimiento del fenómeno de estudio de este trabajo.
- La aplicación del instrumento se efectuó mediante Formularios de Google.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

- El número de personas encuestadas es de 1,070 estudiantes de educación superior, contenidas en la base de datos de García y Silva (2021).
- El grupo de personas encuestadas corresponden a una Institución de Educación Superior (IES) en la que sus participantes pertenecen a diferentes programas educativos de licenciatura.
- Se usaron de técnicas de minería de datos.
- Las herramientas que se contemplaron para esta investigación son: *WEKA*, lenguaje R y *Orange*.

Limitaciones

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

- El dataset se generó tomando las encuestas aplicadas de García y Silva (2021), las cuales se llevaron a cabo con alumnos de una IES en el sureste mexicano.
- El estudio se centró en las percepciones que tiene el alumno con respecto al uso de tecnologías.
- El tiempo de aplicación del instrumento fue de mayo a diciembre de 2020.
- Se aplicaron las técnicas de minería de datos para la identificación de patrones de conducta de los alumnos.

1.2.3 Pregunta de investigación

¿Qué patrones de conducta relacionados con las barreras económica, de acceso, pedagógica y ambiente académico afectan el uso de innovaciones tecnológicas en los estudiantes de educación superior?

7

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

1.3 Objetivos

1.3.1 Objetivo general

Identificar patrones de conducta, relacionados con las barreras para el uso de innovaciones tecnológicas en estudiantes de licenciatura de una Institución de Educación Superior, empleando técnicas de minería de datos.

1.3.2 Objetivos específicos

- Aplicar la técnica de árbol de decisión para una representación analítica de los patrones relacionados con las barreras para el uso de innovaciones tecnológicas.
- Emplear la técnica de agrupamiento para identificar las similitudes entre los patrones de conducta relacionados con las barreras para el uso de innovaciones tecnológicas.
- Aplicar las reglas de asociación para identificar las relaciones entre los estudiantes y las barreras para el uso de innovaciones tecnológicas.
- Analizar patrones de conducta obtenidos de las barreras para el uso de innovaciones tecnológicas en los estudiantes para la toma de decisiones futuras.

1.4 Justificación

Con la implementación de *Microsoft Teams* en la universidad del sureste de México para llevar a cabo las actividades académicas a raíz de la pandemia COVID 19, Medina (2021) esclarece que de 34 mil estudiantes pertenecientes a la institución únicamente alrededor de 17 mil 534 estudiantes pudieron ingresar a las aulas virtuales. De modo que los obstáculos para el uso de innovaciones tecnológicas en los estudiantes requieren atención para atenuar su impacto en el rendimiento académico.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Existen diferentes instrumentos metodológicos aplicados a estudios de carácter social dependiendo de la elección del autor: entrevistas, encuestas, foros de discusión, solo por mencionar algunos ejemplos. Sin embargo, a pesar de lo útiles que pueden llegar a ser, en estudios con grandes cantidades de datos se dificulta la localización y análisis de información relevante (Martínez-Abad y Hernández-Ramos, 2018).

2

Por lo tanto, este trabajo propone usar técnicas de minería de datos, debido a que su aplicación en el ámbito educativo ha sido de gran utilidad en los últimos años permitiendo deducir fenómenos sociales (Menacho 2017). A través de un análisis profundo de datos se generan patrones de conducta, obteniendo información novedosa y potencialmente útil a diferencia de un análisis estadístico convencional (Rosado y Verjel, 2017). Con los resultados generados de la aplicación de las técnicas de minería de datos, la IES del sureste de México podrá tomar decisiones informadas orientadas a implementar recursos que favorezcan el aprendizaje de los estudiantes en función de sus necesidades.

1.5 Metodología utilizada

1.5.1 Enfoque de la investigación

El presente trabajo al tener un impacto social en el ámbito educativo y considerando la aplicación de minería de datos, se requiere que el enfoque metodológico de la investigación sea del tipo cuantitativo ya que de acuerdo con Rivadeneira (2017), en las últimas décadas las investigaciones cuantitativas han sido de utilidad para el estudio de fenómenos sociales debido a que se caracteriza por atender a un problema,

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

generar objetivos en el estudio, realizar el planteamiento de una hipótesis o preguntas de investigación y de igual manera requerir de instrumentos para la recolección de información que sirvan para medir las variables propuestas en la investigación.

1.5.2 Fuentes de investigación

Al ser una investigación con enfoque cuantitativo la fuente de investigación se relaciona con los datos estadísticos que se utilizan para la aplicación de técnicas de minería de datos. Por lo tanto, la fuente de los datos estadísticos para este trabajo es la base de datos de García y Silva (2021), mismos que recolectaron previamente la información de los estudiantes de licenciatura de la IES caso de estudio, a través de una encuesta realizada en Formularios de Google. Por otro lado, también se consideró la búsqueda de literatura como: tesis, artículos de divulgación científica o conferencias que permitieron dar un soporte teórico a este trabajo. Se puede decir de acuerdo con Prieto (2018), que el tipo de fuente utilizada es primaria ya que la información original se obtuvo de los alumnos y no ha sido evaluada o interpretada por alguien más. De igual forma, los aspectos teóricos que se obtuvieron de investigaciones publicadas previamente por otros autores.

1.5.3 Técnicas de recolección de datos

La aplicación de la encuesta ocurrió en mayo a diciembre 2020. Este instrumento lleva por título: Barreras para el uso de innovaciones tecnológicas de estudiantes universitarios. Su aplicación surgió derivado de la pandemia por COVID-19 y se realizó mediante Formularios de Google. De acuerdo con el número de encuestas realizadas se

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

obtuvo un total de 1,070 estudiantes encuestados de nivel licenciatura, pertenecientes a una Universidad Pública del sureste México, los cuales fueron seleccionados de forma no probabilística. Con la participación de los estudiantes, se procuró obtener información acerca de las pautas de conducta en los estudiantes (Hernández y Carpio, 2019).

De la aplicación de encuestas se extrajo información para generar un *dataset* que es de utilidad para la identificación de patrones, con dicho conjunto de datos se aborda una metodología específica que será expuesta en el siguiente apartado.

1.5.4 Metodología para aplicar

Previo a la aplicación metodológica referente al modelo FURPS (derivado de su acrónimo en inglés *Functionality Usability Reliability Performance & Supportability*), y el proceso KDD por sus siglas en inglés *Knowledge Discovery in Databases*, es indispensable conocer las necesidades de información de la IES caso de estudio, así como del diseño del instrumento. Por lo tanto, este aspecto se consideró dentro de la metodología.

Modelo FURPS

Para la selección de la herramienta de minería de datos que se ocupó en esta investigación se aplicó el modelo FURPS, que de acuerdo con Vaca (2017) consta de diferentes criterios que sirven como métricas para la evaluación de *software* de minería de datos. Con los resultados obtenidos se seleccionó el *software* que mejor rendimiento.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Proceso KDD

Para efectos de esta investigación el análisis del comportamiento de los estudiantes se realizó con el fin de hallar patrones de conducta, para ello, se aplicaron los principios que conforman el proceso de minería de datos apegado a la metodología KDD, que consiste en una serie de pasos expuestos en un inicio por Fayyad, Piatetsky-Shapiro & Smyth (1996) y posteriormente ha sido retomado por otros autores como es el caso de Giraldo y Vargas (2017) o Calvache-Fernández *et al.*, (2018) que concuerdan con los pasos de aplicación del método y ha sido de utilidad en diferentes áreas del conocimiento como el ámbito académico.

A partir de lo expuesto previamente, los pasos metodológicos aplicados se muestran a continuación:

1. Aprender el dominio de la aplicación.
2. Crear un conjunto de datos destino.
3. Limpieza y procesamiento de datos.
4. Reducción y proyección de datos.
5. Elección de las funciones y los algoritmos de minería de datos.
6. Minería de datos (identificación de patrones).
7. Interpretación.
8. Uso del conocimiento descubierto.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Capítulo 2. Marco de la investigación

2.1 Marco teórico

No existe un consenso sobre la prioridad con que deban atenderse las barreras que influyen en los patrones de conducta asociados al uso de tecnología e innovaciones tecnológicas en estudiantes en lo general y por consiguiente en estudiantes de educación superior. Al menos es lo que logra discernirse tras la revisión de bibliografía acerca del tema. Mientras que algunos autores destacan como la barrera más importante la falta de una cultura digital, otros colocan al factor económico como primordial, aunque la existencia de uno no anula la existencia del otro; así la pobreza puede ser un factor para no adquirir una cultura digital que permita participar de las dinámicas de un mercado global digitalizado o viceversa: la falta de una cultura digital puede dificultar la consecución de una prosperidad económica. Puede decirse, por consiguiente, que el acceso a la tecnología y sus innovaciones se ve limitado por el factor económico o que el factor económico se encuentra sesgado por la renuencia al acceso.

2.1.1 Acceso y costo para el uso de tecnologías en la formación superior

Trejo (1999, como se citó en Capurro, 2002) indica la falta de políticas en Latinoamérica destinadas a la promoción de la cultura digital, ocasionando que el costo de los servicios de conexión, de los equipos de cómputo y el de la capacitación de los usuarios sean vistos como las tres limitantes más importantes para el uso de la red. Finalmente explica que, si bien la cultura digital es una cultura global, es necesario que sea accesible para la población. Por su parte Capurro (2002) se sustenta en esta

10

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

3 declaración para advertir que la población de América Latina se encuentra en una disyuntiva: en condiciones de pobreza y con recursos limitados, deben decidir si satisfacer sus necesidades más básicas e inmediatas (alimento, vivienda, vestido, salud) o adquirir tecnología; generalmente inclinándose por satisfacer sus necesidades inmediatas sacrificando la adquisición de tecnología, lo que en su opinión constituye o supone un error que acrecienta la brecha respecto a poblaciones de países más ricos, puesto que el mundo en vías de globalización se encuentra cada vez más digitalizado y unirse a estas dinámicas económicas requiere la adquisición de tecnologías de la información. Reforzando esta idea, encontramos a Cabrera *et al.*, (1992) que de acuerdo con sus investigaciones acerca del tema, propone dos modelos económicos:

- Costo-beneficio: los estudiantes perciben un mayor beneficio de otras actividades remunerativas como el trabajo que las cuestiones académicas.
- Focalización del subsidio: aquellos apoyos económicos que el alumno perciben para continuar con sus estudios.

Como se muestra previamente, la percepción del alumno referente al factor económico juega un papel importante para continuar con las actividades académicas, lo que permite inferir que de igual forma tendrá un efecto en la adquisición de nuevas tecnologías.

En un artículo posterior, retomando varias de las ideas de Trejo (1999), Capurro (2002) y Cabrera *et al.*, (1992); Covi (2007), Salinas (2002) y más reciente aún Martínez (2021), centran su atención en la barrera del acceso que considera aspectos tales como

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

la promoción de la cultura digital, así como la implementación de infraestructura necesaria en comunicados urbanos y rurales que faciliten el uso y apropiación de las Tecnologías de la Información (TI) dentro de la sociedad, considerando a este factor como un problema adicional al económico que acrecienta la brecha digital. Desde su perspectiva al eliminar esta barrera podría tener un efecto en la barrera económica que se refiere al costo de los servicios de conexión y de los equipos de cómputo, así como la capacitación y desarrollo de habilidades digitales dentro de la comunidad estudiantil en las instituciones educativas.

Como puede observarse los autores antes mencionados siguen una misma línea, en la que se encuentran estrechamente relacionadas la barrera de acceso y la barrera económica, hasta el punto de que es posible interpretar que remediando la una puede atenuarse la otra.

2.1.2 Interacción con las tecnologías desde el enfoque pedagógico

En cuanto a la esfera pedagógica, resulta complejo definir la naturaleza de los obstáculos enmarcados en este terreno. Esto se debe a que los avances tecnológicos en el ámbito educativo han traído consigo una serie de problemas, que se logran identificar como limitaciones para el aprendizaje en los estudiantes dentro de las instituciones educativas (Garrido *et al.*, 2013). En este sentido, Cabero (2005) invita a las universidades a adaptarse a las necesidades y exigencias que surgen dentro de la sociedad con el propósito de integrar los beneficios de las tecnologías en los procesos de formación estudiantil. Para lograr alcanzar dicho objetivo se requiere fomentar en el

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

alumno el desarrollo de destrezas cognitivas y estrategias de aprendizaje que sirvan para incorporarlos en la sociedad de la información y el conocimiento. Por otro lado, Crovi y López (2011) en su estudio sobre el uso educativo de las tecnologías desde el punto de vista de los estudiantes, destacan principalmente que los profesores suelen motivar a los alumnos al uso de las tecnologías, pero son pocos los profesores que usan las tecnologías para impartir sus clases. Compartiendo la misma idea se encuentran Espinoza y Ricaldi (2018), quienes consideran que los estudiantes requieren no únicamente de infraestructura, sino también que el docente esté especializado para la materia que imparte y para el uso e implementación de tecnologías dentro del aula.

Es muy claro de acuerdo con las ideas expuestas previamente, que la falta de formación en los profesores y estudiantes con respecto al uso de tecnologías es un impedimento dentro de la barrera pedagógica que requiere atención (Padilla, 2008). Para ello las instituciones deben contemplar en los planes de estudio la inclusión de un entorno más tecnológico, esto lo menciona Mosquera (2008) debido a la creación de nuevas estructuras y redes de interacción social desarrolladas dentro del internet, fomentando dentro de estos espacios un comportamiento que prioriza la velocidad de comunicación.

Estos cambios en los procesos de formación han sido abordados por diferentes autores presentado diferentes dimensiones enmarcadas en el factor pedagógico. Es por lo que, de acuerdo la literatura revisada de Aguilar (2004), Crespo y Palaguachi (2020), García *et al.*, (2010), García (2019), Garrido *et al.*, (2013), Gaytan (2015), Pierrakeas *et al.*, (2004), Simpson (2013), Swan (2001) y Zimmerman (2012), se presenta en la tabla

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

1, la siguiente clasificación que propone las dimensiones que involucran a la barrera pedagógica, no tan solo desde el punto de vista de uso de las tecnologías sino también, considerando la modalidad de clases en línea.

Tabla 1.
Dimensiones de la barrera pedagógica.

Participantes	Dimensiones
Percepción de los estudiantes	Falta de tiempo Técnicas de estudio deficientes Falta de motivación Falta de competencias digitales Trabajo en equipo Capacitación para uso de tecnologías Modalidad de evaluación
Actitud y aptitud docente	Escaso seguimiento y supervisión del docente Pruebas de evaluación poco adecuadas Carencia en el diseño del material de estudio Falta de preparación profesional Implementación de tecnologías

Nota: Elaboración propia.

Tal y como se muestra en la tabla 1, es evidente que el uso de tecnologías y sus innovaciones involucra tanto aspectos docentes como aspectos del estudiante debido a que son las partes que se encuentran directamente involucrados en el proceso de enseñanza-aprendizaje.

2.1.3 La influencia del ambiente académico en estudiantes universitarios

Como se ha mostrado hasta el momento, los obstáculos que impiden el uso de tecnologías y sus innovaciones en estudiantes de nivel superior, involucran condiciones económicas, de acceso y pedagógicas. Sin embargo, también se deben considerar

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

aspectos del ente institucional, tal y como lo expone Winocur (2006), según la autora, independientemente del aspecto cultural, económico o social de cada estudiante, la universidad constituye un espacio no solamente para el desarrollo de las exigencias académicas sino también para fomentar vínculos sociales y establecer códigos de identidad institucional.

Es por esta búsqueda de identidad entre los jóvenes que en dicho escenario se ha incorporado la necesidad y práctica de nuevas tecnologías, este tema ha sido estudiado por autores que han desarrollado modelos teóricos aplicados a la modalidad de enseñanza *e-learning*. Díaz (2008) menciona que, aunque la educación a distancia no es algo nuevo en los estudios, la retención de aprendizaje electrónico es algo relativamente nuevo y ha cobrado fuerza en la actualidad de acuerdo con los avances en tecnología y las circunstancias actuales en el plano educativo. Es por este motivo que autores como Bayrón (2012), Do Carmo *et al.*, (2018), Balmori *et al.*, (2013), Moncada (2014) y Fonseca (2016); han retomado los modelos teóricos de Tinto (1975, 1987) y Bean & Metzner (1985) referente al contexto organizacional en ambientes académicos expuestos a continuación:

Modelo de Tinto (1975, 1987). Considerado como uno de los modelos más influyentes en el tema, propone que existen tres variables que pueden influir en el desempeño de los estudiantes: características familiares y personales, capacidades y experiencias preuniversitarias. Así mismo, la integración académica y social y del estudiante se puede ver influenciado de acuerdo con su desempeño escolar, en la medida

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

que el estudiante se siente comprometido puede decidir si continua o abandona sus estudios.

Modelo Bean & Metzner (1985). Este modelo fue desarrollado por los autores para estudiar la deserción estudiantil a través del proceso de formación vía electrónica. Se basa principalmente en cuatro variables que tratan de profundizar en la integración social de los estudiantes en el ente universitario: (a) rendimiento académico con base en calificaciones, (b) aspectos psicológicos como lo es el estrés y otras variables académicas, (c) experiencia académica previa, (d) variables ambientales como puede ser el trato con las autoridades académicos y la interacción con los compañeros.

Aunque en ambos modelos se considera la deserción académica, involucran variables importantes que pueden ser tomadas en cuenta al momento de utilizar las herramientas tecnológicas en instituciones académicas.

Como se ha mostrado hasta el momento por diversos autores que han estudiado la perspectiva del alumno, así como la función de los docentes en esta nueva era digital, el aspecto pedagógico se encuentra vinculado con el ambiente académico, sin embargo se logra entender cuáles son los aspectos que deben ser considerados en estas barreras para que el alumno se sienta motivado a la adaptación y uso de nuevas tecnologías.

Al margen de la urgencia con que deban atenderse estas y otras barreras, todas son una realidad que dificulta la relación de los estudiantes con las tecnologías y sus innovaciones.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

3 En el presente trabajo, se presta particular atención a las previamente aludidas: económica, de acceso, pedagógicas, y ambiente académico, desde la perspectiva de los estudiantes. Esto no implica que las demás, que originan el problema, carezcan de importancia, sino que la elección obedece a la delimitación que requiere toda investigación de carácter científico y académico, y en qué medida influyen en el comportamiento de los estudiantes de la universidad del sureste de México sujeto de investigación y en qué medida los empujan a adoptar ciertos comportamientos que dificulten su relación con las TI empleadas para su proceso de aprendizaje.

2.2 Marco referencial

Son diversas las metodologías, los alcances y los referentes teóricos y conceptuales que se han abordado para el estudio de las percepciones en los estudiantes de educación superior y su relación con el uso de las tecnologías de la información y sus innovaciones tecnológicas, así como para la identificación de patrones de conducta en los mismos.

En este apartado se presentan diferentes estudios clasificados como nacionales e internacionales, donde se emplean distintos métodos estadísticos y de entrevista, así como se encuentran estudios donde se aplica la ciencia de datos.

2.2.1 El uso de tecnologías en estudiantes mexicanos

Inclusión de las tecnologías en el aprendizaje universitario

Torres y Moreno (2013), exponen una investigación de carácter mixto sobre cómo perciben los estudiantes de licenciatura el uso de las tecnologías desde diferentes escenarios de aprendizaje. Para los fines del estudio se consideró que el enfoque

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

cuantitativo se realizara por medio de una encuesta a 308 participantes aplicando un muestreo probabilístico, en tanto al enfoque cualitativo se realizó mediante grupos de discusión con 40 participantes en total. Como resultado, los autores refieren que la

2

Universidad Veracruzana no está orientada al uso de las TI en el proceso de enseñanza-aprendizaje y se aproxima más a un modelo de enseñanza tradicional. Esto se debe a que los alumnos perciben una ausencia en el uso de tecnologías por parte del docente y de la población estudiantil.

La percepción de los estudiantes sobre el uso de las tecnologías aplicando la ciencia de datos

En este estudio realizado por Salas-Rueda *et al.*, (2019), los autores estudian la percepción de los estudiantes de una universidad en México, sobre el uso de tecnologías muy específicas como los son redes sociales (*Facebook, Twitter, Youtube*), herramientas digitales web 2.0 y aplicaciones en internet. A diferencia del estudio presentado previamente, el objeto de esta investigación no es conocer si los jóvenes estudiantes usan estas tecnologías, sino cómo las implementan en el aspecto académico y si consideran que son de utilidad en dicho contexto, además de contemplar las tecnologías antes mencionadas, se toma en cuenta el perfil del estudiante (edad, género, licenciatura). La muestra estuvo compuesta por 61 estudiantes de diferentes licenciaturas y para la aplicación de la minería de datos se usó la herramienta RapidMiner por medio de la técnica árbol de decisión. De acuerdo con los datos obtenidos Salas-Rueda *et al.*, (2019) concluyen lo siguiente:

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

- El uso de las redes sociales, herramientas web 2.0 y aplicaciones en internet son percibidas por lo estudiantes de manera positiva para su integración en los aspectos académicos.
- El número de estudiantes objeto de análisis fue de 61 individuos por lo que se debe considerar emplear una muestra más representativa.

Actitud de estudiantes universitarios hacia los cursos de modalidad mixta

2 El estudio realizado por Sotelo *et al.*, (2016) fue publicado por la Universidad de Sonora y trata de describir las actitudes de los estudiantes hacia los cursos de modalidad mixta, así como la relación o el grado de influencia en su rendimiento académico. Para este trabajo los autores consideraron un enfoque metodológico cuantitativo no experimental con una población muestra de 210 estudiantes a los cuales se les aplicó un instrumento medido por una escala Likert de 40 reactivos distribuidos en cinco factores: profesor, plataforma tecnológica, utilización del curso, interés del alumno y el uso y actitud hacia la tecnología. Para el procesamiento de los datos se usó la aplicación SPSS versión 21.

2 Dentro de los resultados obtenidos del estudio Sotelo *et al.*, (2016) destacan que las asignaturas de modalidad mixta han ganado aceptación entre los estudiantes, sin embargo, todavía hay alumnos que se encuentran renuentes a llevar clases a través de plataformas virtuales. También destacan que la percepción de los estudiantes sobre el uso de las TI en la educación ha permitido un impacto positivo que puede mejorar la calidad en la educación.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Detección de patrones de competencias digitales manifestadas por estudiantes universitarios

El estudio se abordó desde una perspectiva cuantitativa, fue realizado por Islas y Franco (2018) con el objetivo de dar a conocer la aplicación de la minería de datos en la identificación de patrones de conducta asociado a las habilidades digitales con las que cuentan los estudiantes de la Universidad de Guadalajara. En este caso, la herramienta utilizada para la aplicación de la técnica de clasificación fue *WEKA* y la población muestra estuvo compuesta por 748 estudiantes a los cuales se les aplicó un cuestionario con preguntas cerradas como instrumento para la recolección de datos.

Con los resultados obtenidos Islas y Franco (2018), concluyen que la minería de datos es de utilidad para el descubrimiento de información que no es detectable con procedimientos estadísticos normales. Como ejemplo de esto, en su estudio se detectaron tres patrones de coincidencia relacionados con el uso de los medios digitales orientados al aprendizaje, el nivel de dependencia hacia los mismos y cómo afecta en el rendimiento académico del alumno.

Al final Islas y Franco (2018) recomiendan que en los trabajos de minería de datos educativa no se centren únicamente en el pronóstico, sino que haya una contribución en aquellas situaciones que requieran de una descripción, análisis o explicación.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Identificación de patrones de comportamiento estudiantil con el uso de dispositivos móviles

2 El uso de dispositivos móviles dentro del ambiente académico hoy en día es considerado como una herramienta tecnológica. No obstante, su uso entre los estudiantes de cualquier nivel educativo puede tener ventajas y desventajas que influyen en su rendimiento académico y eso dependerá de ciertos factores. Por tal motivo, el estudio realizado por Martínez *et al.*, (2019) toma relevancia debido a que es el objetivo principal de su trabajo, mismo que se llevó a cabo en estudiantes de licenciatura de primer semestre de la Universidad de Guadalajara a través de encuestas realizadas en dos periodos semestrales aplicando formularios de Google. Los resultados obtenidos del análisis muestran que el uso de dispositivos móviles dentro de las actividades académicas impacta de forma positiva al rendimiento académico, sin embargo, en el caso de ser usado para el ocio en gran medida, tendrá efectos negativos en el aprendizaje de los estudiantes por lo que es necesario que el estudiante sea consciente de que el uso desmedido del celular para actividades de ocio puede representar un gran distractor.

2.2.2 El uso de innovaciones tecnológicas en estudiantes de educación superior desde diferentes enfoques en otros países

Detección del alumno en riesgo de titulaciones de ciencias de la Salud: aplicación de técnicas de learning analytics

31 Es un estudio realizado por Saiz *et al.*, (2018), en la Universidad de Burgos en España. Su objeto principal de fue ponderar el orden de importancia que tienen los

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

patrones de aprendizaje que son más efectivos en los estudiantes que toman clases a través de la plataforma Moodle con el propósito de poder predecir cuándo un alumno está en riesgo de no titularse y dar soluciones que logren prevenir este tipo de situaciones. Como se puede observar en este escenario, además del problema con el uso de tecnologías en los estudiantes universitarios, también se encontró una técnica distinta para la identificación de patrones de conducta.

Para la aplicación de la técnica, Saiz *et al.*, (2018) trabajaron con una población muestra de 122 alumnos de la rama de Ciencias de la Salud, identificando los siguientes patrones de conducta en los estudiantes con el uso de la plataforma Moodle: (1) Realización de las pruebas tipo test, (2) Interacción con la plataforma, (3) Acceso al *feedback* del docente, (4) Uso de Estrategias Metacognitivas de auto Planificación EMP y (5) Acceso a la información teórica.

En la investigación de Saiz *et al.*, (2018) concluyen que el patrón conductual más efectivo dentro de la comunidad universitaria objeto de muestra, se relaciona con los estudiantes que planifican, reflexionan y sistematizan sus actividades académicas a través de la plataforma Moodle. Lo que permite la detección temprana de los estudiantes en riesgo de no titularse y realizar propuestas de mejora.

Factores que influyen en el rendimiento académico de los estudiantes

El siguiente trabajo se deriva de un proyecto de investigación que estudia los factores que influyen en el rendimiento académico de estudiantes universitarios en Medellín, Colombia. Los factores que considera Cruz (2016) en este trabajo los clasifica

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

como demográficos, pedagógicos, biológicos, socioculturales, y cognoscitivos. Los factores enunciados previamente fueron evaluados mediante un cuestionario diseñado en formularios de *Visual Basic* para ser procesado posteriormente en *Microsoft Excel*. Dentro de los hallazgos más significativos, se destaca que para el proceso enseñanza-aprendizaje la institución académica debe dar atención a los factores propuestos por Cruz (2016), ya que la autora considera que son el principal motivo de disminución en el rendimiento académico de los estudiantes llegando incluso a desertar. Por lo que la universidad debe prevenir estas fallas a través de capacitaciones en el manejo del estrés en los docentes y alumnos, así como la construcción de modelos educativos que consideren el ambiente sociocultural, nivel económico y perfil de los estudiantes.

Aplicación de técnicas de minería de datos para determinar las interacciones de los estudiantes en un entorno virtual de aprendizaje

El siguiente estudio realizado por Jaramillo y Paz-Arias (2015) está enfocado en determinar las interacciones de los estudiantes de la Universidad de Loja en España de la modalidad de estudios a distancia. Para la recolección de los datos se realizó un análisis a la base de datos donde se encontraba registrada la información de los alumnos, la herramienta para minería de datos seleccionada fue RapidMiner, la cual fue seleccionada como resultado de un análisis comparativo con otras herramientas y se aplicó la técnica árbol de decisión para el análisis de resultados.

Con los hallazgos obtenidos, los autores concluyen que la interacción de los estudiantes con la plataforma de educación a distancia se encuentra en un nivel medio.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Los factores que más influyen son las interacciones con las tareas, exámenes y recursos, lo que significa que el estudiante utiliza los recursos y materiales de la plataforma a conciencia y guarda una relación con el interés que tengan hacia las materias.

Un método de minería de datos para el comportamiento de los estudiantes

El siguiente estudio realizado por Na (2020) en la Universidad de Artes y Ciencias de Hunan, China. El objetivo principal de este trabajo fue identificar patrones de conducta relacionados con el desempeño de los estudiantes dentro del aprendizaje en línea. Para la selección de los datos se ocupó la base de datos que contiene el registro de los estudiantes de la institución educativa, por lo que se tomó en consideración para el estudio una muestra aleatoria de 300 estudiantes de 12 cursos distintos. Con la implementación de la técnica de minería de datos se empleó el método de aprendizaje multitarea que permite analizar simultáneamente múltiples tareas, siendo de utilidad cuando la cantidad de datos por tarea individual es pequeña.

Con la información obtenida de dicho análisis Na (2020), concluye que los hábitos de vida de los estudiantes influyen de forma negativa considerablemente, provocando una disminución en su rendimiento académico. El autor refiere que derivado de los resultados de la investigación, se puedan orientar los hábitos de vida y de estudio entre los universitarios, así como que el número de la población muestra debe ser ampliado debido a que la distribución de la población no fue equilibrada.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Minería de datos para predecir el éxito de estudiantes

Esta investigación realizada en Turquía por Arcinas *et al.*, (2021), resaltan la influencia que tienen las técnicas de minería de datos en el contexto educativo, proponiendo como objetivo principal descubrir las características que impactan la elección de estudiantes en las instituciones académicas, a través de la aplicación de procedimientos predictivos para pronosticar el desempeño de los estudiantes y que puede estar influenciado por variables de comportamiento y actitud.

Arcinas *et al.*, (2021), mencionan que se han realizado estudios previos para predecir el rendimiento académico en estudiantes con resultados poco satisfactorios, en ese sentido propusieron implementar varias técnicas: clasificación, *clustering*, modelos predictivos, descubrimiento con modelos y refinamiento de datos. Esto con el fin de cubrir todas las áreas en las que se involucra la minería de datos en la educación y ofrecer una propuesta más precisa.

2.2.3 Aspectos relevantes

Como se presentó anteriormente en cada uno de los estudios, a pesar de que éstos fueron abordados desde la perspectiva del uso de tecnologías en estudiantes de educación superior y los factores que pueden influir en el rendimiento académico; los estudios tomaron caminos distintos y esto depende del fin o propósito con el que se desarrolló determinado trabajo. Se puede observar que las aproximaciones en las investigaciones en algunos casos guardan relación con la barrera pedagógica sin importar si la modalidad de clases es presencial, virtual o mixta, se presenta la

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

problemática, pero la perspectiva se limita al uso de la tecnología en el ambiente académico.

Por otro lado, se visualizan trabajos que incluyen factores adicionales a la barrera pedagógica y que fueron consideradas desde la percepción del autor en estudios internacionales. Por lo que deja una brecha para el abordaje del uso de innovaciones tecnológicas incluyendo la barrera del acceso, la barrera económica y la barrera del ambiente académico. Aunado a lo anterior, es necesario conocer el impacto de estos factores en los patrones de conducta de los estudiantes de la universidad del sureste de México.

Otro aspecto importante dentro de las investigaciones expuestas es el número de población muestra para el abordaje metodológico, es necesario que el número muestra sea representativo de la población estudiantil y más aún cuando las técnicas que se aplicarán son de minería de datos.

Por último, se muestra que el uso de herramientas para realizar minería de datos es variado y depende de los criterios que considere el investigador para su elección e implementación.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

2.3 Marco conceptual

2.3.1 Barrera económica

7 Rodríguez y Sandoval (2017), consideran que la barrera económica es un factor relacionado con el contexto social, porque sigue siendo determinante para adquirir dispositivos electrónicos y los servicios como el internet que se producen por su uso. El capital económico que tienen las familias de origen del estudiante, marca las diferencias en la variedad y profundidad de la penetración tecnológica en el hogar, conforme aumenta la penetración de las TI en el sistema social, los segmentos de población con nivel socioeconómico más favorable son los que tienen mayor oportunidad de adquirir información de forma más rápida, a diferencia de la población con un nivel socioeconómico más bajo, ocasionando una desigualdad entre los segmentos poblacionales tendiente a aumentar.

7 Como consecuencia de lo planteado, se puede decir que en la medida que son los usuarios los que definen el tipo de aplicación y el curso de desarrollo de la tecnología, los rezagados o excluidos por sus circunstancias, tienen cada vez menos oportunidad de opinar al respecto.

2.3.2 Barrera del acceso

3 Es aquel problema que no se resuelve únicamente con contar o disponer de aparatos electrónicos en casa como una computadora o de tener los conocimientos sobre el uso de tecnología, sino en el aspecto social, como el aprovechamiento de

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

infraestructura tecnológica que permita el uso apropiado de aquellos dispositivos obteniendo de ellos adecuados niveles de explotación (Andrés, 2014).

2.3.3 Barrera pedagógica

Castro-Villagrán *et al.*, (2016) y Sánchez *et al.*, (2012) la describen como romper con los esquemas tradicionales de aprendizaje que no van acorde al contexto que se vive mundialmente. Para ello, los aspectos que se deben considerar son la formación docente en el manejo de tecnologías, así como la actualización de los materiales didácticos conforme a las tendencias mundiales.

2.3.4 Barrera del ambiente académico

Esta barrera guarda relación con las experiencias de los estudiantes dentro del ambiente académico. Fonseca y García (2016), reportan que en la barrera del ambiente académico se hace referencia a factores como el clima institucional que guarda relación con el trato entre iguales (compañeros), así como la relación maestro-alumno e incluso la relación que se tenga entre el alumno y el personal administrativo de la institución. Otro factor que influye son las expectativas que tiene el estudiante dentro de la institución académica, como puede ser flexibilidad en los planes de estudio, planes de becas existentes, cultura organizacional que hace referencia al grado de identidad del alumno con la institución, variables pedagógicas y metodologías aplicadas.

2.3.5 Patrones de conducta

Son comportamientos que no forman parte de los rasgos de personalidad, es la reacción de una persona predispuesta ante una situación que le resulta desafiante. En el

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

caso de los estudiantes puede estar relacionado con las evaluaciones, el dominio del conocimiento adquirido durante su desempeño académico, técnicas de enseñanza o aspectos particulares de convivencia (Ballesteros *et al.*, 2013).

2.3.6 Minería de datos

Existe una variedad de definiciones dadas por autores que han estudiado sobre el tema, algunas de estas definiciones se presentan a continuación:

- De acuerdo con Rodríguez y Díaz (2009), la minería de datos es un conjunto de técnicas empleadas para la creación de mecanismos adecuados de dirección. Dichos mecanismos pueden encontrarse dentro de la estadística, identificación de patrones, así como la clasificación de datos y predicción de modelos. La aplicación de la ciencia de datos es de utilidad cuando lo que se busca es el procesamiento y análisis de una gran cantidad de datos, así como puede ser usada en diferentes áreas como la banca, telecomunicaciones, seguros, educación, política, por mencionar algunas de tantas áreas del conocimiento.
- La minería de datos es un área de estudio científico conocida como un conjunto de técnicas y herramientas aplicadas a cierto suceso o comportamiento que representa un conocimiento, que antes se desconocía, esto puede ser posible a partir del análisis de grandes volúmenes de datos (Riquelme *et al.*, 2006).
- Villegas y Luján-Mora (2016), exponen que la minería de datos también conocida como *datamining*, es la aplicación de una colección de técnicas y herramientas

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

que sirven para el procesamiento, modelado y evaluación de datos con el propósito de brindar soluciones a determinados problemas.

2.3.7 Proceso KDD

Galván *et al.*, (2014) exponen que el proceso KDD es un método utilizado para la aplicación de técnicas de minería de datos, siendo una parte integral en el descubrimiento de conocimientos partiendo de una base de datos en la que se realizan actividades como el preprocesamiento, submuestreo y transformación de los datos. De este modo será posible la identificación de patrones en los datos minados, con el propósito de ser evaluados y factibles de interpretación y de esta manera contribuir a la generación de nuevos conocimientos con los resultados (Flores *et al.*, 2019).

2.3.8 Algoritmos en la minería de datos

Medina y Gómez (2014), explican que los algoritmos en minería de datos son aplicados cuando se quiere encontrar información que no es visible o está incompleta y su selección dependerá del área de estudio, modelo y complejidad de los datos. También presenta una clasificación de estos como se muestra a continuación:

- Exploración: Utilizado para encontrar relaciones sistemáticas entre las variables cuando hay escaso o nulo conocimiento de los resultados. Este tipo de algoritmos solo funcionan como un primer acercamiento para la predicción de un modelo.
- Estadísticos: Estos algoritmos son de utilidad en datos que tratan de encontrar una función matemática ya sea lineal, cuadrática, exponencial o logarítmica. Siendo necesario que su coeficiente de correlación esté lo más cercano posible al menos

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

(-1) o al uno positivo (+1). Es de relevancia en minería de datos para corregir datos faltantes de una muestra.

- Clasificación de datos: Es el proceso de asignación de etiquetas de valores discretos para un registro sin marcar y un clasificador, cuando se dan otros atributos, el modelo predecirá la categoría del atributo según la muestra.
- Aprendizaje automático: Es una combinación de inteligencia artificial y datos estadísticos, que proporciona un número de distintos problemas y algoritmos diversos para solucionarlos. Los algoritmos van a variar dependiendo de los cuestionamientos a los que se les quiera dar respuesta, así como el conjunto de datos que se puede utilizar para su preparación, la estrategia de aprendizaje y la representación de datos son todos diferentes.
- Serie temporal: Sirve en dos situaciones, para predecir tendencias basadas únicamente en el conjunto de datos original empleado para crear un modelo y en predicción cruzada, para crear un modelo general que se pueda aplicar a múltiples series en la previsión del tiempo de valores continuos.

2.3.9 Técnicas de minería de datos

Según Molina y García (2012), las técnicas de minería de datos son una etapa dentro de la metodología KDD (Proceso de extracción del conocimiento, por sus siglas en inglés *Knowledge Discovery in Databases*), cuyo propósito es obtener patrones o modelos a partir de los datos recopilados. Así mismo, una técnica se enfoca en la

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

extracción de la información de datos a partir de los algoritmos y en la práctica un conjunto de algoritmos conforma la técnica más apropiada para la resolución de un problema.

Camana (2016) clasifica las técnicas más representativas de la minería de datos usando la metodología KDD:

- Redes neuronales: Se usan para la detección de patrones y características de los datos, una vez adiestradas las redes se puede hacer previsiones, clasificaciones y segmentación. Su comportamiento es similar a la del cerebro humano, aprendiendo de las experiencias adquiridas y aplicando el conocimiento en la resolución de problemas.
- Regresión lineal: Utilizada para relacionar datos que representan más de dos variables, permite identificar relaciones entre variables numéricas y construir modelos de regresión.
- Árboles de decisión: Son herramientas analíticas que forman parte de la inteligencia artificial, utilizado para la predicción por medio de condiciones (reglas).
- Reglas de asociación: Encuentra hechos que ocurren en común dentro del conjunto de datos que se analiza.

2.3.10 Modelo FURPS

Álvarez *et al.*, (2016) mencionan que el modelo FURPS fue desarrollado por Hewlett-Packard y sirve para evaluar la calidad de un sistema. Su nombre está construido como un acrónimo derivado de los criterios que evalúa y que se pueden dividir en dos aspectos: los requerimientos de funcionalidad (*functionality*) que son

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

criterios que evalúan las funciones que el sistema debe ser capaz de hacer y los requerimientos no funcionales como la usabilidad (*usability*), confiabilidad (*reliability*), desempeño (*performance*) y capacidad de soporte (*supportability*). En esta última clasificación el objetivo es evaluar las características del *software* desde la perspectiva del ambiente en el sistema.

2.4 Marco tecnológico

Como se ha mostrado previamente, la minería de datos es un término que se ha estudiado durante varios años, sin embargo, recientemente la utilidad de estas técnicas ha cobrado una mayor importancia en diferentes ámbitos. Es en este sentido que también las herramientas tecnológicas que pueden ser aplicadas son variadas, debido a la relevancia que se le ha dado al desarrollo e implementación de herramientas para la minería de datos. Por tal motivo, en este apartado se presentan diferentes herramientas que han sido implementadas en la aplicación de técnicas de minería de datos.

2.4.1 Herramientas de minería de datos

Waikato Environment for Knowledge Analysis

Waikato Environment for Knowledge Analysis (WEKA) es una herramienta compuesta por una serie de paquetes de código abierto que pueden emplear diferentes técnicas de procesamiento, clasificación o agrupamiento de los datos y que pueden ser integrados a cualquier proyecto que requiera un análisis de datos (Molina y García, 2012). Sin embargo, Villegas y Luján-Mora (2016), advierten que la escala de documentación orientada al usuario es escasa por lo que esto ocasiona una disminución en la usabilidad.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

RapidMiner Studio

Según Naik & Samant (2016), RapidMiner es un entorno implementado en Java donde el usuario interactúa con procesos de aprendizaje automático y minería de datos. Dentro de sus características de compatibilidad, se pueden implementar modelos desarrollados con lenguaje R o Python facilitando la ampliación de las bibliotecas de terceros dentro del *software* (RapidMiner, s.f.).

Lenguaje R

R es un lenguaje y un entorno útil cuando se requiere implementar gráficas y computación estadística con una variedad de técnicas que pueden ir desde los modelos lineales y no lineales hasta la clasificación y agrupamiento de datos. Su distribución está bajo el licenciamiento GPL y es compatible con diferentes sistemas operativos como Windows, Linux o MacOS por mencionar algunos (Fundación R, s.f.).

Microsoft Azure Analysis Services

Es un motor de análisis de datos implementado en la plataforma de *Microsoft Azure* con servicios basados en la nube, por lo que no se requiere administrar una infraestructura ya que la plataforma brinda ese servicio. Para hacer uso de la infraestructura se requiere la implementación de una instancia aplicando la herramienta Azure Resource Manager (Microsoft Azure, s.f.).

Orange

De acuerdo con Alom & Courtney (2018) es un *software* de código abierto que brinda versatilidad para programar visualmente, además proporciona soluciones de

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

modelado predictivo cuando se tiene un conjunto de datos que requiere ser analizado con el uso de técnicas de minería de datos y aprendizaje automático. Con *Orange* se pueden realizar gráficas de barras o de proyección visual para un análisis más eficiente de los datos, así como se destacan los algoritmos de *clustering* y clasificación (Benalcázar, 2017).

Python.

El uso de Python en los últimos años para el análisis de *big data* ha logrado ocupar un lugar de preferencia entre los usuarios por su facilidad de uso (incluso para usuarios que no son informáticos) ya que su entorno de trabajo cuenta con una amplia variedad de bibliotecas que están disponibles de forma gratuita (Stančin & Jović, 2019). Además, el programador tiene opciones para elegir cómo quiere leer los datos, tipo de estructura para el almacenamiento de datos y el paquete de aprendizaje automático a usar (Ozgur, 2017).

Amazon Elastic MapReduce

Es una plataforma destinada al procesamiento de grandes volúmenes de datos mediante herramientas de código abierto como Apache Hadoop y Apache Spark. Con el uso de estas y otras herramientas (marcos) de *big data* se pueden procesar datos con el propósito de ser analizados y elaborar proyectos de inteligencia de negocios. Al ser un producto de Amazon AWS ofrece las ventajas de administración en la nube lo que permitirá utilizar, transformar y mover grandes cantidades de datos que entran y salen de una base de datos de AWS (Busu, 2017).

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Oracle Data Mining

Bernal *et al.*, (2021) mencionan que éste es un componente para el análisis de datos que cuenta con diferentes algoritmos de minería de datos (clasificación, regresión, detección de anomalías, etc.) y que no requiere exportar la base de datos hacia otras herramientas. Al estar integrados ambos componentes en un mismo entorno de trabajo se logra maximizar la eficiencia de los procesos y se producen ciertas ventajas en la seguridad, preparación, facilidad de actualización y análisis de los datos.

Google Cloud Datalab

De acuerdo con Nordman (2020), la herramienta *Cloud Datalab* contribuye al análisis, visualización y transformación de los datos que se encuentren alojados dentro de la plataforma de *Google Cloud*. La herramienta está construida sobre *Jupyter* notebooks y utiliza sus propios recursos de ingeniería de cómputo que se encuentran ubicados dentro de su propio repositorio (*Google Cloud Source Repository*). En cuanto al rendimiento de *Cloud Datalab*, puede soportar grandes cantidades de datos para su análisis si el usuario lo requiere.

IBM SPSS Modeler

IBM SPSS es un programa comercial por lo que su uso está restringido por los términos del licenciamiento, esta herramienta permite aplicar técnicas de minería de datos para la identificación de patrones en datos estructurados y no estructurados mediante una Interfaz Gráfica de Usuario (GUI). En lo referido a técnicas de minería de datos, ofrece métodos varios métodos incluyendo algoritmos de segmentación,

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

clasificación o predicción dependiendo del usuario la selección de estos. Es una herramienta que proporciona configuraciones de parámetros fáciles de entender, así como capacidades de modelado automático que permiten la ejecución de varios modelos a la vez (Diao *et al.*, 2018).

SAS Enterprise Miner

Es una de las herramientas más populares en la minería de datos, su principal atributo es el manejo de grandes bases de datos lo que permite agilizar el trabajo de la minería de datos al momento de crear modelos predictivos y descriptivos con una alta precisión a través de una interfaz gráfica. *SAS Enterprise Miner*, aunque se basa en la metodología SEMMA, al igual que otras metodologías para minería de datos, permite la identificación de patrones, realizar análisis de asociación y evaluar la precisión de los modelos estadísticos que se apliquen.

Analytic Solver para Excel

De acuerdo con *Microsoft* (2019), *Analytic Solver* es una herramienta complementaria que puede ser utilizado en libros de *Microsoft Excel*, dentro de sus características de compatibilidad, pues ser utilizado en *Excel* para *Windows*, *Mac* o *Excel online*. Esta herramienta de análisis de datos puede ser de utilidad para la detección de patrones de comportamiento, implementación de modelos predictivos o aplicación de minería de texto. De los algoritmos para minería de datos que ofrece *Analytic Solver* se destacan los siguientes: de regresión lineal y logística, vecinos cercanos, árboles de clasificación, redes neuronales y algoritmos múltiples.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

MatLab

Es un entorno de computación compatible para trabajar en *Windows*, *Mac* y *Unix*. A pesar de que es una herramienta mayormente utilizada por ingenieros y analistas de datos, uno de sus atributos principales es la lectura o visualización intuitiva de los datos lo que permite que sea una herramienta atractiva para usuarios que no necesariamente cuentan con conocimientos básicos de codificación o ciencias de la computación (MathWorks, s.f.).

Por otro lado, Ozgur (2019) menciona que, en cuanto a la lectura de resultados y gráficos, a diferencia de otras herramientas para minería de datos, MatLab no se limita a la visualización bidimensional, permitiendo al usuario una mejor interpretación de los datos que analice.

2.4.2 Herramientas complementarias

Hoja de cálculo Microsoft Excel

Microsoft Excel es una herramienta que forma parte de la paquetería de Office, es una hoja de cálculo diseñada para la manipulación de grandes cantidades de datos, compuesta por celdas agrupadas por filas y columnas. Excel funciona para la realizar desde cálculos aritméticos sencillos hasta la elaboración y seguimiento de cálculos complejos que requieran del diseño de una macro para acciones que se repiten.

Algunas de las funciones de esta herramienta van desde la sumatoria de datos, filtros, construcción de tablas y diseño de gráficos (Microsoft Support, s.f.).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

2.5 Marco legal

2.5.1 Licencia de software GPL

Las herramientas *WEKA*, *Orange*, *Python* y lenguaje *R* son *softwares* que están disponibles para su descarga de forma gratuita. Sin embargo, a pesar de ser gratuito, el tipo de licencia que tienen es de *software* libre conocida como la Licencia Pública General (GPL de GNU). Cuando se habla de *software* libre significa que los usuarios tienen acceso a ejecutar, copiar, modificar, estudiar o mejorar el *software* (GNU, 2019).

2.5.2 Licencia de software propietario

En este tipo de licencia se clasifican las herramientas *RapidMiner*, *Microsoft Azure Analysis Service*, *Amazon EMR*, *Oracle DataMining*, *Google Cloud*, *IBM SPSS Modeler*, *SAS Enterprise Miner*, *Analytic Solver* y *MatLab*. Porque el tipo de licenciamiento que tiene es comercial y se encuentra restringido ya sea para su compra en el caso de ser adquirido por una empresa, uso gratuito limitado o para fines educativos donde sus limitantes se encuentran dirigidas al periodo de vigencia para el uso del *software*.

RapidMiner Studio

El propósito de la licencia educativa de *RapidMiner Studio* es que sea usado por docentes y estudiantes ya sea de nivel licenciatura o posgrado. Para hacer la solicitud de una licencia educativa es necesario contar con el correo electrónico institucional, una vez que se autorice la licencia estará vigente durante un año. Al término del periodo de vigencia podrá hacerse una solicitud de renovación sujeta de aprobación por parte de la empresa (*RapidMiner*, s.f.).

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Microsoft Azure para estudiantes

Consiste en dos aspectos, el primero es una suscripción gratuita y válida por un año, con posibilidad de renovación siempre y cuando el estudiante continúe inscrito dentro del programa educativo al que pertenece, el segundo aspecto es que *Microsoft Azure* otorga 100 dólares de crédito para hacer uso de ello en caso de que el usuario se exceda de las limitaciones del uso gratuito en las herramientas (Microsoft Azure, s.f.).

Amazon capa gratuita

De acuerdo con Amazon (s.f.), maneja tres tipos de ofertas para hacer uso de sus productos: gratis sin fecha de vencimiento, 12 meses de uso y pruebas gratuitas de 30 días. Para efectos del proyecto interesan las herramientas que tienen licencias de uso ilimitado, así como las de 12 meses de uso, debido a que son las que incluyen productos de almacenamiento de base de datos y las herramientas de *machine learning*. En el caso de ambas modalidades de uso, aunque es atractivo el factor tiempo disponible, también se debe considerar dentro de las limitaciones que tienen un tope en el número de horas para el uso de las instancias así como en la cantidad de almacenamiento disponible para los datos.

Oracle Cloud

Cuenta con una licencia gratuita de 30 días y 300 dólares en crédito para hacer uso de los servicios incluyendo los que son de paga. Posterior al periodo de prueba se puede solicitar su ampliación o los servicios ocupados empezarán a cobrarse. También se encuentran productos gratuitos de forma permanente con algunas limitaciones ya sea

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

en el tiempo de uso o la cantidad de almacenamiento: almacén de datos autónomos, almacenamiento de objetos, almacenamiento de archivos, entre otros, (Oracle Latinoamérica, s.f.).

Google Cloud programa gratuito

Cloud Datalab no tiene costo por el servicio; sin embargo, para hacer uso de la plataforma se debe tener una cuenta en *Google Cloud Platform* que sí tiene un costo por los recursos que se ocupen como recursos de almacenamiento y *data analysis services*. Actualmente la plataforma de *Google Cloud* cuenta con un programa gratuito que ofrece 90 día de servicios gratuitos con 300 dólares en créditos para aquellos servicios que no se encuentran adicionados de forma gratuita duran el periodo de prueba. (Google Cloud, s.f.).

IBM SPSS GradPack

De acuerdo con IBM (s.f.) la licencia SPSS GradPack es una suscripción exclusivamente para estudiantes que propone la empresa debido a que las licencias que venden son de uso empresarial. Dentro de los beneficios que tiene la licencia para estudiantes es un descuento del 90% en comparación con la licencia comercial por lo que es necesario para realizar la suscripción ponerse en contacto con un agente de ventas para proporcionar mayores informes sobre el precio de la licencia. En cuanto a los recursos disponibles con SPSS GradPack se destacan: estadística descriptiva, correlación, tablas de referencias cruzadas, frecuencias, pruebas T y clúster de regresión lineal.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

MatLab

MatLab es un *software* de paga, cuenta con una prueba de 30 días con lo cual se tiene acceso a las diferentes funciones que ofrece la herramienta incluyendo las de análisis de datos. Posterior a los días de prueba la herramienta solicitará el cobro por los servicios y para ello el usuario debe ponerse en contacto con servicio al cliente. Existe una opción de licencia para estudiantes, pero se encuentra restringido a los convenios realizados con las universidades (MathWorks, s.f.).

Analytic Solver para Excel

Esta es una herramienta complementaria para Microsoft Excel, su servicio puede ser adquirido para usar de forma local en la computadora o en línea. El proveedor de este complemento ofrece 15 días de prueba gratuita, que incluye funciones de optimización, simulación y minería de datos (Frontline Systems, s.f.).

SAS VIYA

Esta herramienta cuenta con diferentes paquetes gratuitos para el análisis de datos que incluye la predicción y planificación de demanda del consumidor, toma de decisiones del negocio o programación. De modo que el usuario pueda seleccionar el que mejor se adapte a sus necesidades. Cualquiera de los paquetes gratuitos que ofrece SAS Enterprise Miner cuentan con una vigencia de 14 días para su uso (SAS, s.f.).

Microsoft 365

Al contratar la paquetería de Microsoft Office 365 se tiene acceso al producto de Microsoft Excel, el tipo de paquete sugerido es el personal que tiene un costo anual de

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

\$1,299.00 pesos mexicanos o bien se pueden realizar pagos mensuales de \$129.99 pesos mexicanos. Las aplicaciones pueden ser utilizadas en un máximo de cinco dispositivos, cuenta con acceso para trabajar en línea y al ser una licencia de uso personal su uso está limitado a un usuario (Microsoft, s.f.).

2.5.3 Ley Federal de Protección de Datos Personales en Posesión de Particulares

Al hacer uso dentro de la investigación un dataset que contiene las encuestas aplicadas a estudiantes de una universidad del sureste de México, es pertinente aclarar que se debe hacer un uso responsable de la información recopilada y que hay una ley que regula en México la protección de datos personales. Esta legislación la encontramos como la Ley Federal de Protección de Datos Personales en Posesión de los Particulares la cual fue emitida en el año 2010, no se encuentran modificaciones a la mencionada ley por lo que para efectos legales sigue vigente.

De acuerdo con la Ley Federal de Protección de Datos Personales en Posesión de los Particulares (2010) en su artículo uno, el propósito de esta ley es la protección de datos personales en posesión de terceros para controlar y garantizar la privacidad y el derecho a la autodeterminación informativa de las personas.

Así mismo se entiende que los datos recabados se deben tratar de acuerdo con los principios establecidos en esta ley y no se debe hacer de forma fraudulenta o engañosa. Eso debido a que se presume de cierto nivel de confianza por parte del propietario de la información y el tercero que hace uso de los datos. Por lo que el tratamiento de la información se debe realizar bajos los términos y condiciones

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

establecidas por las partes y que deben estar basadas en la normativa de ley. Todo esto se encuentra explicitado conforme al artículo siete (Ley Federal de Protección de Datos Personales en Posesión de los Particulares, 2010).

En caso de no cumplir con los términos y condiciones establecidos en ley, así como entre el propietario de los datos y el tercero, el artículo 62 establece que se llevará a cabo un procedimiento de imposición de sanciones mediado por la autoridad competente hacia el presunto infractor (Ley Federal de Protección de Datos Personales en Posesión de los Particulares, 2010).

Universidad Juárez Autónoma de Tabasco.
México.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Capítulo 3. Aplicación de la metodología y desarrollo

3.1 Información general

3.1.1 Descubrimiento y entendimiento de la información

Esta investigación ha sido elaborada para el caso particular de una universidad pública del sureste de México. Por lo tanto, es indispensable conocer previo a la aplicación del proceso *KDD*, la calidad de la información que será utilizada para el análisis y estudio de la minería de datos.

Los datos que fueron utilizados en este trabajo se obtuvieron de la base de datos de García y Silva (2021). Para la generación de esta base de datos se realizó la aplicación de encuestas mediante Formularios de Google, cuyo objetivo principal de aplicación fue la de identificar los obstáculos que enfrentan los estudiantes de nivel superior para la aplicación de Tecnologías de la Información (TI) en el proceso de formación académica.

2 De acuerdo con el número de encuestas realizadas, el estudio se basó en un universo compuesto por 1,070 estudiantes de la universidad caso de estudio, considerando aspectos sobre el perfil de los estudiantes como edad, género, licenciatura, así como recursos personales disponibles, tiempo de uso de los recursos, recursos tecnológicos usados por el profesor, percepciones del alumno referentes a la modalidad de clases a distancia, tareas, exámenes, relación profesor alumno y finalmente, su experiencia con respecto a las clases virtuales en época de pandemia. Con ello se buscó identificar cuáles son los obstáculos que enfrentan los alumnos en el proceso de enseñanza aprendizaje dentro de la institución educativa a la que pertenecen, de acuerdo

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

con las barreras expuestas en esta investigación: económica, de acceso, pedagógica y ambiente académico. En la tabla 2 se presenta la distribución de los estudiantes encuestados por División Académica y género.

Tabla 2.

Distribución de estudiantes por división académica y género.

División académica	Masculino	Femenino	Total
Ciencias Agropecuarias	13	15	28
Ciencias Básicas	48	63	111
Ciencias Biológicas	7	17	24
Ciencias de la Salud	34	91	125
Ciencias Económico Administrativas	79	114	193
Ciencias Sociales y Humanidades	33	41	74
Ciencias y Tecnologías de la Información	33	10	43
Educación y Artes	85	177	262
Ingeniería y Arquitectura	76	43	119
Multidisciplinaria de Comalcalco	14	19	33
Multidisciplinaria de Jalpa de Méndez	15	34	49
Multidisciplinaria de los Ríos	5	4	9
Total general	442	628	1070
Total general (%)	41%	59%	100%

Nota: Elaboración propia con base en datos de García y Silva (2021).

De acuerdo con los datos expuestos previamente en la tabla 2, los estudiantes encuestados se encontraban inscritos en la universidad del sureste de México, perteneciendo a diferentes carreras de nivel licenciatura de las Divisiones Académicas presentadas, con edades que oscilan entre los 20 y los 57 años.

Como en toda investigación, se busca generar valor con el proceso metodológico aplicado, así como con los resultados obtenidos. Este trabajo no es la excepción, debido

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

a que se busque generar un valor agregado para la institución académica derivado de la aplicación de minería de datos que se realice.

3.1.2 Diseño y descripción del instrumento

El instrumento obtenido para la generación del *dataset* en esta investigación, ha sido elaborado por García y Silva (2021), docentes investigadoras pertenecientes a la IES caso de estudio. Para acceder a la información de este instrumento, es a través de la base de datos y solo si se comparte el vínculo que permite la visualización de la encuesta. Al contar con los permisos de acceso se puede realizar una exportación del *dataset* en formato CSV, por sus siglas en inglés (*Comma Separated Values*) y que posteriormente puede ser procesado a una hoja de *Excel*.

El diseño de este instrumento fue de suma importancia en esta investigación debido a que permitió analizar los principales obstáculos que enfrentan los alumnos para el uso de tecnologías y cuál es su influencia en su rendimiento académico, de igual manera podrá servir de apoyo a la institución para hacer propuestas de solución que se adapten a sus necesidades.

Aunque se contemplan los aspectos pedagógicos como la evaluación de actividades, revisión de tareas, retroalimentación a estudiantes, envío de planeación de actividades o el uso de herramientas tecnológicas por parte del profesor, la información generada con este instrumento ha sido desde la percepción de los estudiantes.

En lo que respecta a la estructura del instrumento, es un cuestionario que consta de 43 preguntas relacionadas con el uso de innovaciones tecnológicas y adicionalmente

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

se incluyen seis cuestionamientos de aspectos sociodemográficos del alumno como la edad, género, programa educativo, año de ingreso a la universidad y número de materias inscritas, haciendo un total de 49 ítems. Todos los criterios considerados en el cuestionario fueron clasificados en siete secciones como se muestra en la tabla 3.

Tabla 3.
Ítems contenidos por sección de la encuesta.

Sección	Ítem	No. de Ítems
Datos del alumno.	Genero	6
	Edad	
	División académica	
	Programa educativo que cursa	
	Ingreso a la universidad	
	Asignaturas inscritas	
Recursos personales disponibles.	Internet en casa	7
	Eficiencia del internet	
	Acudir a un café internet	
	Computadora personal	
	Equipo de cómputo compartido	
	Celular con internet	
Tiempo de uso de recursos tecnológicos para actividades académicas.	Inicio de uso de computadora	4
	Inicio de uso de recursos en internet	
	Inicio de uso de ambientes virtuales	
	Inicio de interacción con redes sociales	
Recursos tecnológicos utilizados por el profesor. Factor extrínseco.	Total de profesores asignados	7
	Total de profesores que usaron MS Teams para las clases	
	Total de profesores que usaron otra plataforma	
	Total de profesores que usaron correo electrónico como herramienta principal para las clases	
	Total de profesores que usaron red social como herramienta principal para las clases	
	Total de profesores que no utilizaron ninguna herramienta	
	Total de profesores que se comunicaron durante la contingencia con el alumno	

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Sección	Ítem	No. de Ítems
Percepciones. Factor intrínseco.	Las asignaturas teóricas son aptas para clases a distancia	10
	Las asignaturas prácticas no son aptas para clases a distancia.	
	La enseñanza virtual es más pesada que la presencial	
	Ansiedad y estrés al trabajar con tecnología	
	Con la educación a distancia se aprende menos	
	Escaso conocimiento en plataformas virtuales	
	Conocimiento tecnológico limitado	
	Capacitación en plataformas virtuales	
	Capacidad de ser autodidacta	
	Trabajos en equipo	
Evaluación de actividades y profesores.	Profesores que envían planeación de actividades	9
	Profesores que dan retroalimentación	
	Profesores que revisan tareas	
	Profesores que evalúan con examen en línea	
	Profesores que solicitan exponer en línea	
	Profesores que han evaluado al menos una vez	
	Profesores que evalúan con trabajo final	
	No se aclara la modalidad de evaluación	
Aprendizaje obtenido en clases a distancia		
Experiencias y sugerencias.	Experiencia con el internet	6
	Experiencia con las asignaturas	
	Experiencia con tareas y actividades	
	Interacción con compañeros y profesores	
	Opiniones	
	Sugerencias	

Nota: Elaboración propia con base en datos de García y Silva (2021).

Las preguntas relacionadas con el uso de tecnologías en estudiantes se encuentran a partir de la sección dos relacionada con los recursos disponibles, hasta la sección siete de experiencias y sugerencias. Cada una de las preguntas de este instrumento fueron contestadas con diferentes escalas de opción múltiple de acuerdo con las necesidades de información, algunos cuestionamientos aplicaban la escala tipo *Likert*, en otras secciones se aplicaron escalas numéricas o respuestas de tipo binaria. Excepto en la sección siete de Experiencias y sugerencias, aunque hay preguntas con opciones

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

en esta sección, en el caso de las preguntas 42 y 43 son preguntas abiertas que fueron respondidas conforme al criterio de los encuestados puesto que se le solicita al participante agregar alguna opinión o sugerencia si así lo desea.

De acuerdo con García y Silva (2021), este instrumento fue aplicado de forma virtual, a través de invitaciones realizadas a los alumnos de forma directa o utilizando como intermediarios a los profesores de asignatura para solicitar la participación de los estudiantes y a través de oficios a los Directores de las 12 Divisiones Académicas. Al tener el consentimiento del alumno para contestar la encuesta, recibió el vínculo para acceder vía web al formulario.

Basado en el instrumento de recolección de datos proporcionado para esta investigación y considerando las investigaciones realizadas sobre el planteamiento del problema abordado en este trabajo, se presenta la tabla 4 de operacionalización.

Tabla 4.
Variables de investigación.

Variables	Dimensiones	Indicadores	Autores
Independiente Estudiantes	Perfil académico del estudiante	Género. Edad. Programa educativo. Asignaturas inscritas.	
Dependiente Uso de TIC	<i>Económico y acceso.</i> Recursos personales disponibles.	Accesibilidad y eficiencia del internet. Disponibilidad de dispositivos electrónicos (celular, tableta y computadora).	Gómez <i>et al.</i> , (2018), Martínez, (2020), Rodríguez y Sandoval (2017).
	Tiempo de uso de recursos tecnológicos.	Uso de computadora, internet, redes sociales (como medio académico y social) y plataformas virtuales.	Espinoza y Ricaldi (2018).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Variables	Dimensiones	Indicadores	Autores
Dependiente Uso de TIC	<i>Pedagógicos.</i>		Marín <i>et al.</i> , (2017).
	Extrínseco: Recursos tecnológicos usados por el profesor para impartir clases.	Plataforma implementada por la universidad u otras plataformas académicas. Correo electrónico como medio principal. Redes sociales. Profesores que no usan ningún recurso tecnológico.	
	Intrínseco: Percepciones del estudiante respecto a la modalidad de clases virtual.	Asignaturas aptas para clases a distancias. Asignaturas que deben ser únicamente presenciales. Importancia de capacitación para clases virtuales. Capacidad de aprendizaje del alumno. Trabajos en equipo, modalidad virtual.	Álvarez-Silva <i>et al.</i> , (2018).
	Evaluación y actividades	Profesores que envían planeación de actividades. Retroalimentación al alumno. Aclaraciones del profesor sobre la forma de evaluar. Tareas y exposiciones en línea. Modalidad de evaluación (parcial o proyecto final). Percepción de aprendizaje del alumno.	Arancibia <i>et al.</i> , (2020).
	<i>Ambiente académico.</i>		Ramírez (2019).
	Experiencias y sugerencias.	Experiencia con las asignaturas, eficiencia del internet, tareas, actividades, interacción con los profesores y compañeros.	Rodríguez y Espinoza (2017).

Nota: Elaboración propia.

3.1.3 Visualización y exploración de los datos

Para la manipulación de los datos, se optó por transformar el *dataset* del formato original CSV por sus siglas en inglés (Comma Separated Values), al formato estándar de *Excel* .xlsx que corresponde al utilizado en las hojas de cálculo de *Microsoft Excel*. Una vez realizada la transformación del *dataset* se procedió a insertar una tabla, con la

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

finalidad de que la exploración de los datos fuese más eficiente y de este modo hacer uso de diferentes herramientas de visualización como la aplicación de filtros y tablas dinámicas, así como herramientas de revisión, de modo que puedan ser consideradas posteriormente para su implementación en el proceso de limpieza y transformación de los datos. Cabe mencionar que el *dataset* utilizado contiene datos estructurados y son de tipo categórico.

Por consiguiente, se puede decir que el objetivo principal de la exploración de datos ha sido la de verificar la calidad de estos, a través de la aplicación de diferentes técnicas de consulta, así como de ciertos criterios que se describen en lo subsecuente y que permitieron un análisis más objetivo (Infraestructura de Datos Espaciales para el Distrito Capital. [IDECA], 2019).

Fiabilidad de los datos. Se consideró que el tipo de fuente de donde provienen los datos es confiable, así como el tipo de instrumento implementado para su recolección. Se realizó un cotejo de los datos visualizados en el *dataset* con el número de personas encuestadas que aparece en el formulario, y se considera que no hay falta de información en los datos exportados, por lo tanto, el *dataset* está completo.

2

Tipos de errores. Dentro de los errores encontrados en el *dataset* se destaca principalmente el uso de caracteres especiales como sustitutos de ciertas letras o por el uso de palabras acentuadas. Tal y como se muestra en la Figura 2.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 2. Errores de caracteres especiales.

	AS	AT	AU	AV	AW	AX	AY	AZ	BA	
1	38. Con resp	39. Con resp	40. Con resp	41. Con resp	42. Si deseas	43. Si deseas agregar una opini	3n o sugerencia para			
2	Frustrante p	Regular. Con Siento que h	Hubo alguno	Ninguno	Ninguno	No		1 S	Alamente c	
3	Frustrante p	Muy pobre. l	Las tareas y e	Fue nula. Jan	xxx	xxx	internet	5	El 50% de ell	
4	Frustrante p	Buena. La ter	Siento que h	La interacci	3 Fue buena, l	Para explicar	No, porque l	0	El 50% de ell	
5	Frustrante p	Muy pobre. l	Siento que h	Fue nula. Jan	Presencial	Ninguna	Error	1	Ninguno de e	
6	Excelente. N	Excelente. La	Las actividad	Todo fue exc	Xxxx	Xxxxxx	Lcccc	10	Todos	
7	Frustrante p	Muy pobre. l	Las tareas y e	La interacci	3 Capacitar a l	Preocuparse	No. El intern	6	El 50%	
8	Frustrante p	Regular. Con	Las tareas y e	La interacci	3 Que capacie	capacitaci	3 r no	6	M	
9	Bastante bu	Excelente. La	Las actividad	Hubo alguno	excelente	Excelente	Si cuenta cor	3	El 50%	
10	Frustrante p	Regular. Con	Siento que h	La interacci	3 Q traten de c	Q no pongan	No ya que er	6	Menos del 50	
11	Bastante bu	Buena. La ter	Las actividad	Todo fue exc	Que me gust	Que siempre	Si	7	Todos	
12	Frustrante p	Muy pobre p	Siento que h	La interacci	3 Capaciten a	3 Que explique	Si	0	Ninguno	
13	Regular. Fue	Buena. La ter	Las actividad	Hubo alguno	Incre	3-ble si:	Los profesor	Todo bien	2	El 50%
14	Bastante bu	Buena. La ter	Las actividad	Hubo alguno	Que haya m	Que no dejar	Si	6	M	
15	Bastante bu	Regular. Con	Las actividad	Hubo algunos	profesores	Ninguna	Si	3	El 50%	
16	Regular. Fue	Regular. Con	Las actividad	Hubo algunos	profesores	Seria bueno	Si porque	6	M	

Nota: Elaboración propia.

Este tipo de error se debe a la exportación de los datos del Formulario de Google a una hoja de cálculo, ciertos datos no pudieron ser adaptados de un formato a otro durante la exportación, por lo que en la etapa de limpieza de datos se corrigieron.

Otros errores que también se encontraron, se debieron a cambios en el tipo de respuestas que proporcionaba el Formulario y la adición de nuevas preguntas durante su aplicación, estas modificaciones ocasionaron una serie desfases en respuestas dadas por los alumnos que ya no coincidían con ninguna pregunta, casillas en blanco, y preguntas que contienen dos escalas tipo Likert diferentes debido a los cambios realizados en el formulario.

Datos nulos. Se hallaron campos nulos o vacíos, ocasionado por la adición de nuevas preguntas que no estaban contempladas desde el principio en el cuestionario y

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

que durante su aplicación se agregaron, en este caso hubo personas encuestadas que tienen campos vacíos porque cuando participaron en la encuesta las preguntas no existían. Por otro lado, en las preguntas abiertas también se encontraron campos vacíos debido a que las respuestas no eran obligatorias en estos cuestionamientos.

Duplicidad de datos. En cuanto a este criterio, aunque es una encuesta con escala tipo *Likert*, sí se hallaron respuestas duplicadas, debido a que la aplicación fue mediante un formulario electrónico, pudo haber errores de conexión que se prestaron a confusión entre los participantes. En la Figura 3 se muestran cuáles fueron las respuestas duplicadas.

Figura 3. Verificación de duplicidad de datos.

41. Con respecto a las clases virtuales que implementó la UJA	42. Si deseas agregar alguna situación importante no considerada en	43. Si deseas agregar una opinión o sugerencia para mejorar esta experi
176 La interacción se redujo a charlas expositivas de los profesor	Sin comentarios	Sin comentarios
181 Todo fue excelente. Se interactuó con los profesores y los co	Sin comentarios	Sin comentarios
182 Todo fue excelente. Se interactuó con los profesores y los co	Sin comentarios	Sin comentarios
292 Hubo algunos profesores que utilizaron diversas estrategias	En realidad solo es que los maestros sean más conscientes con noso	Considero que las clases a distancia pueden mejorar, realmente es una
310 Hubo algunos profesores que utilizaron diversas estrategias	En realidad solo es que los maestros sean más conscientes con noso	Considero que las clases a distancia pueden mejorar, realmente es una
381 Hubo algunos profesores que utilizaron diversas estrategias	la cuarentena. Me la pasé trabajando con el celular y tuve la	que solo mandaron PDFs y no se sabe cómo evaluarán. Deberían estar
382 Hubo algunos profesores que utilizaron diversas estrategias	la cuarentena. Me la pasé trabajando con el celular y tuve la	que solo mandaron PDFs y no se sabe cómo evaluarán. Deberían estar
394 La interacción se redujo a charlas expositivas de los profesor	Sin comentarios	Sin comentarios
395 La interacción se redujo a charlas expositivas de los profesor	Sin comentarios	Sin comentarios
468 La interacción se centró más en el envío de tareas.	La carrera de médico cirujano NO ESTÁ DISEÑADA PARA LLEVARSE EN	Evaluar que las carreras del Área de la salud, implican prácticas, entre o
469 La interacción se centró más en el envío de tareas.	La carrera de médico cirujano NO ESTÁ DISEÑADA PARA LLEVARSE EN	Evaluar que las carreras del Área de la salud, implican prácticas, entre o
470 Fue nula. Jamás interactuamos con nadie.	Es IMPOSIBLE llevar MÉDICO CIRUJANO EN LINEA	LA LICENCIATURA EN MÉDICO CIRUJANO NO ESTÁ DISEÑADA PARA LLEV
471 Fue nula. Jamás interactuamos con nadie.	Es IMPOSIBLE llevar MÉDICO CIRUJANO EN LINEA	LA LICENCIATURA EN MÉDICO CIRUJANO NO ESTÁ DISEÑADA PARA LLEV
675 Todo fue excelente. Se interactuó con los profesores y los co	Sin comentarios	Sin comentarios
676 Todo fue excelente. Se interactuó con los profesores y los co	Sin comentarios	Sin comentarios
752 La interacción se centró más en el envío de tareas.	Muchos maestros no explican las tareas y solo las dejan para cumpli	Tener un planeación exacta con presencia de los maestros y las retroali
753 La interacción se centró más en el envío de tareas.	Muchos maestros no explican las tareas y solo las dejan para cumpli	Tener un planeación exacta con presencia de los maestros y las retroali
900 Fue nula. Jamás interactuamos con nadie.	Me di de baja temporal este semestre por cuestiones de Salud	En mi opinión no todos contamos con Internet.
901 Fue nula. Jamás interactuamos con nadie.	Este semestre me di de baja temporal. No tome clases virtuales por	No todos contamos con Internet.
969 La interacción se centró más en el envío de tareas.	Las clases que me gustan presenciales son inglés y francés	Inglés y francés son las que me gustarían presenciales
972 La interacción se centró más en el envío de tareas.	Las clases que me gustan presenciales son inglés y francés	Inglés y francés son las que me gustarían presenciales

Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En general, las respuestas de los participantes son diversas, sin embargo, como se muestra en la Figura 3 durante la revisión se detectaron encuestas que repitieron las mismas respuestas en todas sus preguntas.

Los criterios descritos previamente sirvieron de referencia, siendo retomados nuevamente durante el proceso *KDD* en la limpieza de los datos para realizar las correcciones pertinentes.

3.2 Análisis y selección de herramientas de minería de datos

Callejas-Cuervo *et al.*, (2018) explican que la calidad de un sistema o *software* sea para desarrollar o para ser implementado en algún conjunto de tareas, debe ser considerado desde un inicio en el proyecto que se realice, así como en cada etapa durante el desarrollo de las actividades con la finalidad de dar un apropiado seguimiento sobre los requerimientos de funcionalidad, calidad y eficiencia del sistema. Por otro lado, Ramos (2019), exponen que para el análisis de calidad de un sistema o *software* existen diversos modelos de calidad que pueden ser de utilidad para el análisis de un *software*. Cada modelo contiene un conjunto de características también conocidos como atributos o factores que sirven para ser aplicados como métricas de evaluación y análisis. Dependerá del usuario la metodología o modelo que desee aplicar para realizar las mediciones necesarias en las diferentes herramientas de *software*.

Por lo tanto, para la selección de la herramienta de minería de datos que se empleó en esta investigación, se desarrolló el modelo *FURPS*, que se aplicó en el análisis de

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

eficiencia y funcionalidad en tres *softwares* propuestos en el marco tecnológico con el fin de seleccionar la herramienta más apropiada.

De acuerdo con Constanzo (2014) y Vaca (2017) el modelo FURPS consta de diferentes criterios y factores con los que debe contar un sistema tal y como se muestra en la tabla 5.

Tabla 5.

Factores y criterios del modelo FURPS.

Factores	Criterios
Funcionalidad	Características y capacidades del programa Generalidad de las funciones Seguridad del sistema
Usabilidad	Factores humanos Factores estéticos Consistencia de la interfaz Documentación
Confiabilidad <i>Reliability</i>	Frecuencia y severidad de fallos Exactitud de las salidas Tiempo medio de fallos Capacidad de recuperación ante fallos Capacidad de predicción
Rendimiento <i>Performance</i>	Velocidad de procesamiento Tiempo de respuesta Consumo de recursos Rendimiento efectivo total Eficacia
Soporte	Extensibilidad Adaptabilidad Capacidad de prueba Capacidad de configuración Compatibilidad Requisitos de instalación

Nota: Elaboración propia.

59

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Este modelo ofrece como ventaja durante el proceso de evaluación la asignación de prioridades y posteriormente la definición de atributos de calidad que proporcionen como resultado métricas. Con base en las mediciones obtenidas se procede a la selección del *software* que ofrezca el mejor rendimiento.

Aunque existen diversos modelos que pueden servir de base para la evaluación de productos de *software*, son pocas las investigaciones que las incluyen para la selección del *software* a implementar en los estudios. Por tal motivo, el propósito de implementar este modelo para la selección del *software* en este trabajo es dar un soporte metodológico que permita ofrecer un valor adicional, así como una mejor presentación en cuanto a la selección e implementación de las herramientas utilizadas.

3.2.1 Procedimiento y análisis de indicadores

Existen diversos análisis que han aplicado el modelo FURPS para la verificación de calidad en un *software*. Tal es el caso de Yugan *et al.*, (2019), siendo este un ejemplo de cómo los 23 indicadores de este modelo fueron aplicados para analizar el rendimiento de *frameworks* JSF, por sus siglas en inglés (*JavaServer Faces*) *primefaces* y *bootfaces*. En este mismo sentido, se encuentra el trabajo realizado por Araujo (2020), en el cual su objetivo principal fue la evaluación de un sistema ERP en desarrollo para una institución educativa. Así como estos ejemplos se encuentran análisis que evalúan la calidad de diferentes herramientas tipo *software*, aplicando los indicadores de evaluación que consideren pertinentes de acuerdo con el modelo presentado.

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Para el caso de esta investigación interesan las herramientas de minería de datos, por lo tanto, se ha considerado retomar el trabajo realizado por Guzmán (2022), que aplicó el modelo FURPS para el análisis de 16 herramientas de minería de datos, considerando los 23 criterios de evaluación y otorgando una ponderación de cero a 23, donde el cero representa que no se cumple con ninguno de los criterios del modelo y 23 que se interpreta como que la herramienta evaluada cumple con absolutamente todos los atributos considerados por el proceso de evaluación de acuerdo con el modelo FURPS.

Por lo anterior, en la tabla 6 se muestra el análisis de selección de las herramientas de minería de datos aplicando la metodología propuesta.

Tabla 6. *Análisis de la selección de las herramientas para minería de datos aplicando la metodología FURPS.*

Factores / Criterios	Herramientas utilizadas para Minería de Datos															
	RapidMiner	Orange	WEKA	Lenguaje R	Python	Matlab	Microsoft SQL Server	Microsoft Excel	Microsoft Analysis Services	Oracle Business Intelligence Server	SPSS Modeler	DAAL	Amazon EMR	Google Cloud Datalab	Microsoft Azure Analysis Services	SAS Enterprise Miner
Funcionalidad																
Características y capacidades del programa	x	x	x	x	x	x	x	X	x	x	x	x	x	x	x	x
Generalidad de las funciones	x	x	x	x			x	X	x	x	x		x	x	x	x
Seguridad del sistema	x	x	x	x	x	x	x	X	x	x	x	x	x	x	x	x
Usabilidad																
Factores humanos	x	x	x	x			x	X	x	x	x		x	x	x	x
Factores estéticos			x	x					x						x	
Consistencia de la interfaz	x	x	x	x	x	x	x	X	x	x	x		x	x	x	x
Documentación	x		x	x	x	x	x	X	x		x	x	x	x	x	x

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Factores / Criterios	Herramientas utilizadas para Minería de Datos															
	RapidMiner	Orange	WEKA	Lenguaje R	Python	Matlab	Microsoft SQL Server	Microsoft Excel	Microsoft Analysis Services	Oracle Business Intelligence Server	SPSS Modeler	DAAL	Amazon EMR	Google Cloud Datalab	Microsoft Azure Analysis Services	SAS Enterprise Miner
Confiabilidad																
Frecuencia y severidad de fallos			x	x			x				x					x
Exactitud de las salidas	x	x	x	x	x	x	x	X	x	x	x		x	x	x	x
Tiempo medio de fallos	x		x				x	X	x							x
Capacidad de recuperación ante fallos				x							x					
Capacidad de predicción		x	x	x				X	x		x			x	x	x
Rendimiento																
Velocidad de procesamiento	x	x	x	x	x	x	x	X	x	x	x	x	x	x	x	x
Tiempo de respuesta		x	x	x	x	x	x	X	x	x	x		x	x	x	x
Consumo de respuesta		x	x					X	x	x			x	x	x	x
Rendimiento efectivo total	x	x	x	x	x	x	x	X	x	x	x	x	x	x	x	x
Eficacia	x	x	x	x	x	x	x	X	x	x	x		x	x	x	x
Capacidad de Soporte																
Extensibilidad	x	x	x	x			x	X	x	x	x	x	x	x	x	x
Adaptabilidad	x	x	x	x	x	x	x	X	x	x	x	x	x		x	x
Capacidad de prueba	x	x	x	x	x	x	x	X	x	x	x		x	x	x	x
Capacidad de configuración		x	x	x		x	x		x	x	x	x	x		x	x
Compatibilidad	x	x		x	x		x	X	x				x	x	x	
Requisitos de instalación	x		x	x			x	X	x	x			x		x	x

Nota: Guzmán (2022).

Como se muestra en la tabla 6, los criterios de evaluación fueron marcados en las herramientas respectivas dependiendo si cumplían con los requerimientos de evaluación de acuerdo con el análisis realizado por Guzmán (2022), en caso contrario se muestran

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

casillas vacías que se pueden considerar como incumplimiento de ciertos criterios en las herramientas.

Posterior al análisis de criterios en las herramientas, se procedió a realizar un conteo general, para conocer cuántos de los criterios aplicaban en cada herramienta. Cada una de las métricas evaluadas fueron clasificadas de acuerdo con los factores que propone este modelo: funcionalidad, usabilidad, confiabilidad, rendimiento y capacidad de soporte. Para ilustrar los resultados finales sobre esta evaluación se muestra la tabla 7, en la que se puede observar la sumatoria de los criterios de acuerdo con el análisis previo y aunque existen herramientas con una ponderación alta no se alcanza la máxima puntuación de acuerdo con el estudio realizado.

Tabla 7. Ponderación de las herramientas de acuerdo con la metodología FURPS.

Factores	Herramientas utilizadas para Minería de Datos															
	RapidMiner	Orange	WEKA	Lenguaje R	Python	Matlab	Microsoft SQL Server	Microsoft Excel	Microsoft Analysis Services	Oracle Business Intelligence Server	SPSS Modeler	DAAL	Amazon EMR	Google Cloud Datalab	Microsoft Azure Analysis Services	SAS Enterprise Miner
Funcionalidad	3	3	3	3	2	2	3	3	3	3	3	2	3	3	3	3
Usabilidad	3	2	4	4	2	2	3	3	4	2	3	1	3	3	4	3
Confiabilidad	2	2	4	4	1	1	3	3	3	3	2	0	1	2	4	2
Rendimiento	3	4	5	4	4	4	5	5	5	5	4	2	5	5	5	5
Capacidad de Soporte	5	5	5	6	3	3	6	5	3	4	4	3	6	3	6	5
Total	16	16	21	21	12	12	20	19	18	17	16	8	18	16	22	18

Nota: Guzmán (2022).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Otro aspecto que resaltar, y siendo este el objetivo primordial a retomar en esta investigación es que de las 16 herramientas evaluadas existen herramientas con licencia comercial y de acceso gratuito.

Por lo tanto, partiendo del análisis previo, se clasificaron las 16 herramientas de acuerdo con el tipo de licencia como se muestra en la tabla 8.

Tabla 8.

Herramientas de minería de datos clasificadas por tipo de licencia.

Herramientas utilizadas para minería de datos	Factores					Ponderación
	Funcionalidad	Usabilidad	Confiablez	Rendimiento	Capacidad de soporte	
<i>Herramientas gratuitas</i>						
Interactivas						
Orange	3	2	2	4	5	16
WEKA	3	4	4	5	5	21
Lenguajes de programación						
Lenguaje R	3	4	4	4	6	21
Python	2	2	1	4	3	12
<i>Herramientas comerciales</i>						
Interactivas						
RapidMiner	3	3	2	3	5	16
Microsoft Excel	3	3	3	5	5	19
SAS Enterprise Miner	3	3	2	5	5	18
SPSS Modeler	3	3	2	4	4	16
Google Cloud Datalab	3	3	2	5	3	16
Microsoft SQL Server	3	3	3	5	6	20
Microsoft Analysis Services	3	4	3	5	3	18
Oracle Business Intelligence Server	3	2	3	5	4	17
Amazon EMR	3	3	1	5	6	18
Microsoft Azure Analysis Services	3	4	4	5	6	22
Lenguajes de programación						
DAAL	2	1	0	2	3	8
Matlab	2	2	1	4	3	12

Nota: Elaboración propia con base en datos de Guzmán (2022).

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

De acuerdo con la clasificación presentada en la tabla 8, se evaluaron herramientas que tienen una licencia comercial y también de descarga gratuita. Por otro lado, se observa que se consideraron lenguajes de programación y softwares del tipo interactivo que no requieren experiencia en programación.

Aunque este estudio contempla tres herramientas para el análisis de minería de datos, se optó por considerar las herramientas que se encuentran dentro de la clasificación de licencia gratuita. En ese sentido cabe mencionar que las cuatro herramientas contempladas en ese rubro dentro del análisis de calidad realizado por Guzmán (2022) son Lenguaje R, WEKA, Orange y Python. Cabe mencionar que, de los *softwares* antes mencionados, se van a seleccionar las de mejor ponderación. Asimismo, no se hace distinción en el caso de que la herramienta requiera el uso de un lenguaje de programación.

Después de lo previamente expuesto con relación al análisis del modelo FURPS, los puntajes obtenidos por cada herramienta, así como el tipo de licencia, se determinó que las herramientas que resultaron seleccionadas para este trabajo fueron: *WEKA*, Lenguaje R y *Orange*.

El análisis presentado en este apartado se llevó a cabo con la finalidad de presentar los parámetros de selección de las tres herramientas a emplear duran el proceso metodológico.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

3.2.2 Características de las herramientas seleccionadas

Orange

Orange es un *software* para minería de datos basado en componentes *front-end* de programación visual. Contiene componentes para el procesamiento, filtrado, modelado de datos. Cuenta con la tecnología requerida para el análisis de grandes volúmenes de datos con el fin de generar conocimiento nuevo que explique el comportamiento de los datos de acuerdo con el contexto (Benalcázar, 2017).

Esta herramienta tiene dos formas para acceder a la plataforma: (1) a través de scripts desde *Python*, (2) a través de *widgets* en dentro del panel de selección. Al ser una herramienta de programación visual tiene como ventaja de que puede ser utilizada por usuarios con conocimientos básicos o avanzados sobre la usabilidad y dominio del sistema debido a que se puede tener acceso a las funciones de *Orange* mediante *widgets* que abarcan desde la visualización, selección de subconjuntos y preprocesamiento de datos. Sin embargo, si el usuario lo desea también puede hacer uso de scripts desde *Python* (Naik & Samant, 2016).

En este sentido, Mojica (2015) menciona que *Orange* se destaca por ser un entorno que contiene una variedad de componentes (*widgets*) con los que se pueden realizar diferentes tareas de ciencias de datos. Se pueden realizar gráficas de barras o de proyección visual para un análisis más eficiente de los datos, las librerías de *Orange* se encuentran ordenadas por jerarquías considerando el tipo de elemento. Los componentes considerados como de bajo nivel pueden relacionarse con componentes

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

de niveles superiores para una mejor comprensión y análisis de los datos. Dentro de los principales componentes de *Orange* se muestran los siguientes: Gestión y procesamiento de datos, clasificación, regresión, asociación, ensambles, *clustering*, evaluación y proyección.

Los *widgets* de *Orange* al ser su principal atributo para trabajar con la herramienta permiten que ésta sea compatible con las tareas de ciencias de datos más comunes. En ese sentido se destaca que se pueden emplear múltiples árboles de clasificación como *Tree*, así como la aplicación de *clustering* (agrupamiento de datos), clasificación de Bayes, regresión logística y reglas de asociación (Fernández y Bonilla, 2020).

Por último, se destaca que *Orange* es una herramienta novedosa en esta investigación, debido a que son pocos los trabajos sobre minería de datos que han implementado este *software*.

WEKA

De acuerdo con Bhatia (2019), *Waikato Environment for Knowledge Analysis* (*WEKA*) es una herramienta para la minería de datos que fue desarrollada por el grupo de aprendizaje automático de la Universidad de Nueva Zelanda, el principal objetivo de este proyecto ha sido proporcionar una amplia variedad de algoritmos de aprendizaje automático, así como ofrecer herramientas de preprocesamiento de datos, útil para el desarrollo de investigaciones académicas y comerciales. En este sentido se puede decir que esta herramienta es conocida particularmente por ser un entorno para la experimentación y el análisis de datos útil aplicar y evaluar las técnicas más relevantes

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

para el análisis de datos. De las técnicas más usadas se destacan las siguientes: segmentación o *clustering*, reglas de asociación, predicción y clasificación.

Para el análisis de los datos el único requerimiento es que la información se contenga almacenada en un formato de archivo específico conocido como ARFF (*Attribute Relation File Format*). Así mismo, cabe mencionar que *WEKA* es un *software* ampliamente difundido e implementado para la aplicación de minería de datos, esto se debe a que es una herramienta distribuida bajo la licencia GPL, lo que significa que es de libre distribución, cuenta con la mayoría de los algoritmos de aprendizaje automático y se encuentra desarrollado en Java (Hussain *et al.*, 2017).

Esta herramienta está compuesta por una serie de paquetes de código abierto que pueden emplear diferentes técnicas y algoritmos de minería de datos que pueden integrarse en cualquier proyecto. La implementación de los algoritmos puede realizarse desde una línea de comandos o interfaz de línea de ordenes CLI, por sus siglas en inglés (*Command Line Interface*). Sin embargo, a pesar de todas la funcionalidades y facilidades que ofrece el *software*, la escala de documentación orientada al usuario es escasa y se requiere de una mayor búsqueda bibliográfica para hacer contemplar su usabilidad (Villegas y Luján-Mora, 2016).

Lenguaje R

R es un lenguaje y un entorno de programación que es de utilidad cuando se requiere implementar gráficas y computación estadística. El origen de su nombre es por

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

la primera letra del nombre de sus dos autores: Robert Gentleman y Ross Ihaka, siendo desarrollado en la Universidad de Auckland en Nueva Zelanda (Bhatia, 2019).

De acuerdo con Fundación R (s.f.), R cuenta con una variedad de técnicas que pueden ir desde los modelos lineales y no lineales hasta la clasificación y agrupamiento de datos. Su distribución está bajo el licenciamiento público general GPL y se ofrecen versiones binarias previamente compiladas es compatible para diferentes sistemas operativos como *Windows*, *Linux* o *MacOS* por mencionar algunos. Dentro de sus características principales se mencionan las siguientes:

- Gráficos con calidad de publicación bien diseñados y se pueden visualizar en pantalla o se puede imprimir.
- Manejo eficaz para la instalación y almacenamiento de datos.
- Diversidad de operadores y en su mayoría para el cálculo de matrices.
- Dentro del lenguaje de programación se incluyen bucles, funciones recursivas (definidas por el usuario) así como instalaciones de entrada y salida.
- Ofrece una amplia variedad de herramientas para la clasificación, agrupación, análisis de series temporales y modelización lineal y no lineal.

Dentro del diseño de R se consideró que los usuarios también puedan agregar funciones adicionales a las que ya se contemplan y se pueden definir como nuevas funciones (Fundación R, s.f.).

Otro de los atributos de R, es que comparte el mismo lenguaje de programación que *Python*. Sin embargo, se considera que a pesar de que R es una herramienta que no

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

necesariamente presenta los datos de una manera fácil para interpretar, si se compara con otras herramientas de minería de datos como por ejemplo *SAS Enterprise Miner* que requiere usar tres lenguajes de programación para realizar las mismas tareas que se pueden hacer en R, entonces estamos hablando de que Lenguaje R es una herramienta más versátil y sencilla de comprender (Ozgun, 2017).

Por lo anterior, se puede decir que R tiene sus atributos y también aspectos a considerar, sin embargo, no deja de ser una herramienta *software* atractiva y de utilidad para aquellos usuarios que requieran implementarla en un proyecto de analítica de datos a gran escala.

3.3 Proceso KDD

Existen diferentes metodologías que se han creado para emplear las técnicas de minería de datos, por ejemplo, SEMMA, Crisp-DM y KDD. Esto lo expone Panizzi (2019) en un mapeo sistemático sobre minería de datos en el contexto educativo, destacando como la metodología que más se ocupa en trabajos al proceso KDD (por sus siglas en inglés Knowledge Discovery in Databases).

La metodología KDD, hace referencia al proceso general de descubrimiento de conocimiento útil y novedoso partiendo de un conjunto de datos. Este proceso fue propuesto por Fayyad, Piatetsky-Shapiro & Smyth (1996), aunque ha sido utilizado por diferentes autores en la actualidad, en cada investigación se encontrará una interpretación diferente de los pasos propuestos por estos autores, esto se debe a que esta metodología involucra numerosos pasos, sin embargo, en la Figura 4, se muestra

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

de forma gráfica las etapas o aspectos considerados durante la aplicación de minería de datos en este trabajo.

Figura 4.
Proceso KDD.



Nota: Adaptado de Fayyad, Piatetsky-Shapiro & Smyth (1996).

La figura 4, muestra en resumen los procedimientos que se llevaron a cabo a lo largo de esta investigación. Aunque el proceso KDD puede ser interpretado como un proceso de cinco o seis pasos como lo exponen Camana (2016) y Calvache-Fernández *et al.*, (2018) por mencionar algunos ejemplos. Sin embargo, esta investigación se apegó a los pasos que se proponen originalmente en esta metodología y que fueron descritos de acuerdo con su aplicación en cada una de las herramientas seleccionadas previamente.

2

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Para mayor entendimiento, en la tabla 9 se presentan las tareas realizadas con la metodología KDD clasificadas por etapas de acuerdo con la figura 4 que se expone previamente.

Tabla 9.
Clasificación por etapas de la metodología KDD.

Etapas	Tareas de la metodología KDD
Datos.	Aprender el dominio de la aplicación. Crear un conjunto de datos destino.
Selección de limpieza y transformación.	Limpieza y procesamiento de los datos. Reducción y proyección de los datos.
Patrones y creación de modelos.	Selección de funciones y algoritmos. Minería de datos.
Interpretación y evaluación del conocimiento.	Interpretación. Uso del conocimiento descubierto.

Nota: Elaboración propia.

Por lo anteriormente expuesto, la información presentada en este proyecto referente a la aplicación de la minería de datos estará organizada de acuerdo con las ocho tareas que constituyen el proceso KDD y que se presentan en lo subsecuente.

3.3.1 Aprendizaje y dominio de la aplicación

En esta primera fase del proceso se tuvo como objetivo entender y dominar las necesidades de cada una de las aplicaciones o herramientas tecnológicas propuestas para el proceso metodológico, haciendo uso de la información obtenida de diferentes fuentes bibliográficas para obtener información relevante, de modo que se pudieran conocer los diferentes entornos de trabajo en *Orange*, *WEKA* y Lenguaje R.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Previamente en el apartado de análisis y selección de las herramientas se mostraron los atributos de cada una de las herramientas previamente aludidas. Se destaca que una de las características principales en cada una de ellas es la facilidad de uso para los usuarios nuevos o con conocimientos básicos acerca del funcionamiento. Esto en principio se consideró una ventaja que facilitó el aprendizaje y dominio de las herramientas de minería de datos. Por otro lado, es importante considerar dentro del aprendizaje y dominio de la aplicación el proceso de instalación de cada uno de los *softwares*, debido a que *Orange*, *WEKA* y Lenguaje R son herramientas de instalación local en el equipo de cómputo y este aspecto puede ser un factor que contribuya para que usuario decida desistir o implementar determinada herramienta. Por lo tanto, se puede decir que el proceso de instalación en las tres herramientas no es complicado, basta con descargar los archivos ejecutables de cada *software* para instalarlos en cuestión de minutos.

2

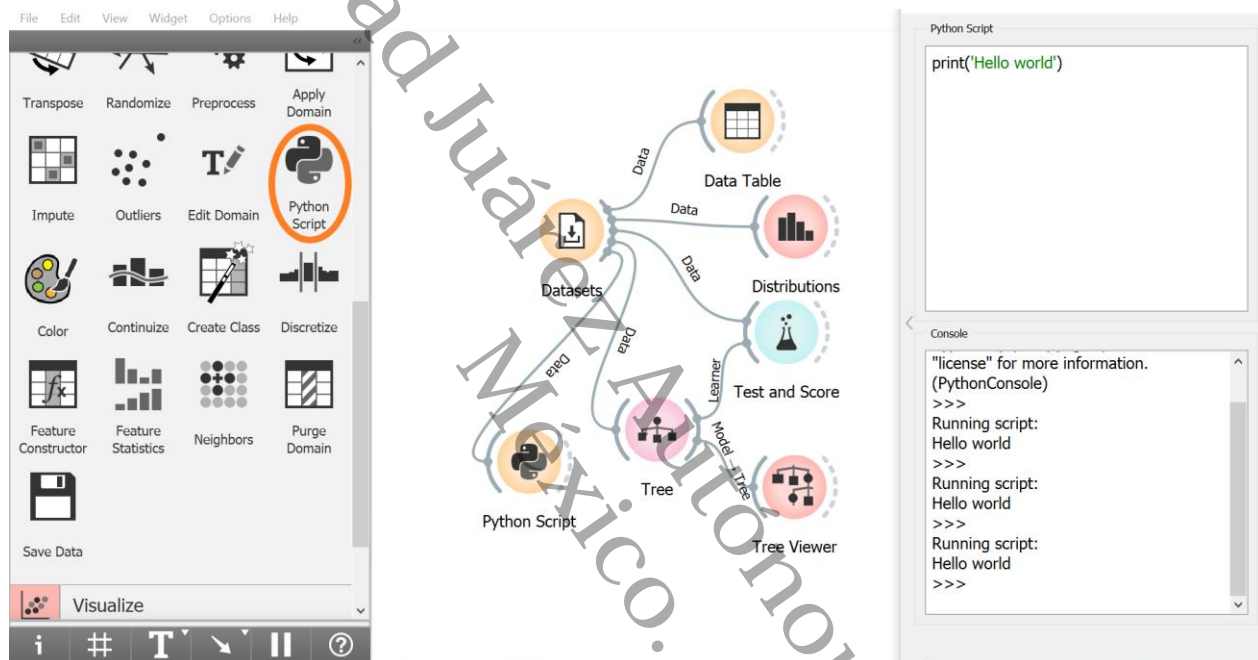
Una vez que se instalaron cada una de las aplicaciones, se procedió a interactuar con el entorno de trabajo. La primera herramienta con la que se interactuó fue *Orange* versión 3.27.1, al abrir la aplicación la primera ventana que aparece es la de bienvenida, en ella se pueden seleccionar los documentos recientes con los que se ha trabajado, revisar la documentación para aprender a usar sus instancias de trabajo o bien se puede iniciar a trabajar.

Por otro lado, en la Figura 5 se observa cómo es el acceso al entorno de trabajo en *Orange*, se presentan dos formas de acceso a la aplicación una es a través de scripts

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

de Python y la otra es mediante los *widgets* en este último caso solo se requiere arrastrar la función (*widget*) que se desea utilizar y conectarla con otras funciones para que empiece a trabajar.

Figura 5.
Interfaz Gráfica de Orange.



Nota: Elaboración propia.

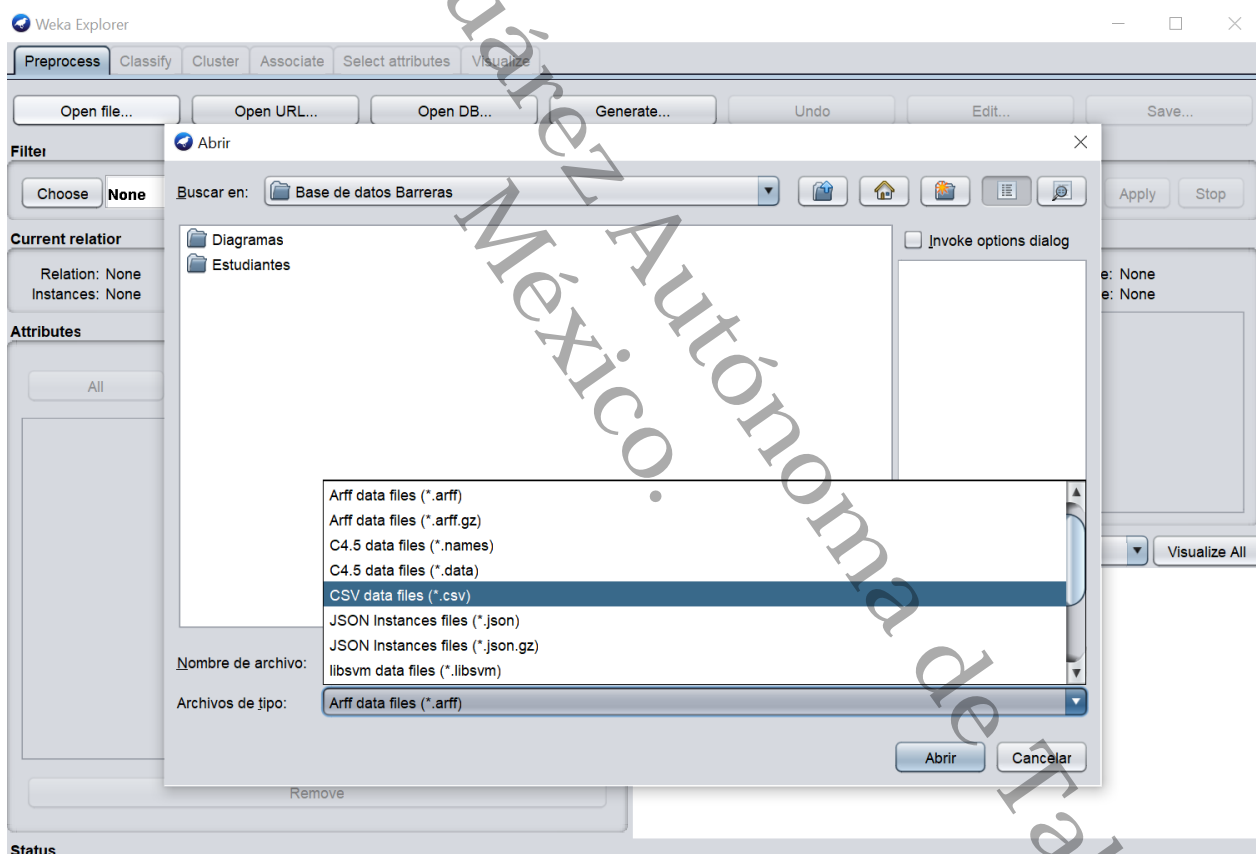
En lo que respecta a la herramienta *WEKA* en la versión 3.8.5, el entorno de trabajo presenta cinco opciones diferentes para la visualización de la interfaz gráfica: *Explorer*, *Experimenter*, *KnowledgeFlow*, *Workbench* y *SimpleCLI*. Con esta diversidad de ambientes de trabajo puede trabajar un usuario desde el nivel básico hasta el nivel más avanzado en el análisis de datos haciendo uso de conexiones para trabajar con una base de datos. Para el caso de este trabajo, se hará uso de la interfaz *Explorer*, teniendo como

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

principal tributo la facilidad de manejo para nuevos usuarios. En la Figura 6 se muestra el entorno de trabajo de WEKA desde la interfaz Explorer, se observa que el tipo de formato recomendado para importar el *dataset* es .arff sin embargo, también se tiene la opción para trabajar con archivos .csv siendo un tipo de formato de uso general en las herramientas de minería de datos.

Figura 6.
Interfaz gráfica de Explorer en WEKA.



Nota: Elaboración propia.

Por último, se describe el entorno de trabajo en Lenguaje R, para trabajar con esta herramienta se debe instalar un complemento conocido como RStudio, esto se debe a

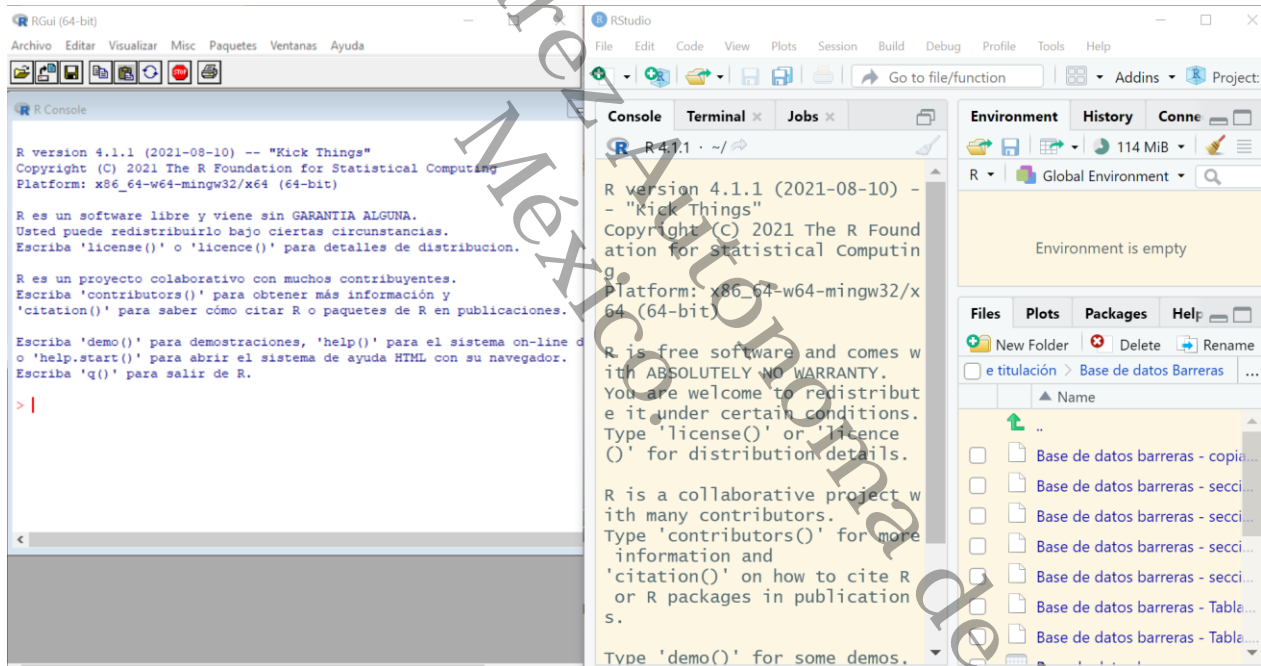
75

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

que el entorno de R es para programación y su complemento es el Entorno de Desarrollo Integrado (por sus siglas en inglés IDE). En la Figura 7 se muestra cómo es el entorno de trabajo en ambas aplicaciones, del lado izquierdo de la Figura 7 se presenta la interfaz gráfica de Lenguaje R versión 4.1.1, del lado derecho se observa la interfaz gráfica de RStudio que será el entorno de trabajo, en esta última herramienta se podrá trabajar a nivel de interfaz gráfica siendo posible la visualización del lenguaje de programación.

Figura 7.

Entorno de trabajo de Lenguaje R y RStudio.



Nota: Elaboración propia.

3.3.2 Crear un conjunto de datos destino

En este apartado lo que se buscó realizar fue la selección del conjunto de datos en el que se centró el proceso de análisis para la aplicación de variables, muestras de

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

datos y descubrimiento de datos ocultos. En esta etapa se consideró la integración del conjunto de datos a analizar y que pudo provenir de diferentes fuentes de información para crear información homogénea posteriormente.

2 Sin embargo, para el caso de este trabajo la integración de los datos no considera una selección y conjunto de información de diferentes fuentes, puesto que los datos relevantes son extraídos de la base de datos de García y Silva (2021) mencionada en apartados previos para el caso de estudio. Este *dataset* fue procesado en *Microsoft Excel* y está compuesto por un total de 1,070 registros de personas que respondieron una encuesta conformada por siete secciones y 49 cuestionamientos.

Por lo tanto, de acuerdo con el diseño de la encuesta, la selección del conjunto y subconjunto de datos, así como las variables a analizar, quedó compuesto como se muestra en la tabla 10.

Tabla 10.
Conjunto de datos.

Secciones de la encuesta	No. de variables
1. Generalidades.	6
2. Recursos disponibles.	7
3. Tiempo de uso de recursos tecnológicos.	4
4. Recursos tecnológicos utilizados por el profesor. Factores extrínsecos.	7
5. Percepciones del alumno. Factores intrínsecos.	10
6. Evaluación de actividades y profesores.	9
7. Experiencias y sugerencias.	6

Nota: Elaboración propia.

En la tabla 10 se observa el número de ítems a analizar s que conforman el *dataset*, mismos que se encuentran distribuidos en siete secciones. Por lo tanto, la

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

aplicación de las técnicas de minería de datos se llevará a cabo con 49 variables (ítems), para más información en la tabla 3 que se muestra previamente en el diseño y descripción del instrumento, se detalla cuáles son las variables de análisis.

3.3.3 Limpieza y procesamiento de datos

En esta etapa fue necesario realizar un proceso de depuración al momento de generar el conjunto de datos, con el propósito de facilitar la ejecución de los algoritmos, visualización e interpretación de la información en lo subsecuente. Por lo anterior, la limpieza de datos incluyó la eliminación de caracteres especiales y campos vacíos, por mencionar algunos ejemplos, que pudieron causar ruido al momento de hacer el análisis de datos ya que de otro modo los resultados serían imprecisos o poco claros.

En ese sentido, cabe mencionar que en el apartado 3.1.3 Visualización y exploración de los datos, el *dataset* se convirtió en una hoja de *Microsoft Excel*. Así mismo, en ese apartado se verificó la calidad de los datos y se clasificaron en cuatro criterios principales:

- Fiabilidad de los datos.
- Tipos de errores.
- Campos vacíos.
- Duplicidad de datos.

A partir los criterios mencionados previamente y que fueron retomados en esta etapa para su adecuada atención, se consideró la elaboración de una bitácora de limpieza. Para efectos de presentación de la información, esta bitácora se divide en dos

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

apartados principales, el primero consistió en la depuración de datos conflictivos y atípicos y el segundo en la revisión y resolución de problemas de ruido.

Por último, se señala que el *dataset* al estar en el formato estándar de *Microsoft Excel* (xlsx), se pudieron emplear diferentes funciones como filtros, validación de datos buscar y reemplazar, y revisión de ortografía, siendo de utilidad en este apartado.

Depuración de datos conflictivos y atípicos

En este apartado se retomaron principalmente dos aspectos, la resolución de conflictos entre datos y la eliminación de valores atípicos. Se inició con este paso porque se consideró que era necesario depurar los datos que no aportarían información valiosa al análisis de datos y sería innecesario eliminar ruido de datos que no van a ser ocupados en lo subsecuente.

Lo primero que se realizó fue la eliminación de 34 encuestas, se tomó esta decisión basada en las inconsistencias detectadas previamente (ver figura 8).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 8. Depuración de conflictos entre datos.

	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV
19	De acuerdo	Más del 50%									Regular. Fue	Regular. Com	Las actividad	Hubo algu
20	De acuerdo	Más del 50%									Regular. Fue	Regular. Com	Las actividad	Hubo algu
21	De acuerdo	El 50%									Bastante bue	Buena. La tec	Las actividad	Hubo algu
22	Ni en desac	Todos									Regular. Fue	Buena. La tec	Las actividad	Hubo algu
23	Ni en desac	Más del 50%									Excelente. Nc	Excelente. La	Las actividad	Hubo algu
24	Totalmente c	El 50%									Frustrante pc	Regular. Com	Las actividad	Hubo algu
25	De acuerdo	Ninguno									Excelente. Nc	Muy pobre p	Las actividad	La interacc
26	De acuerdo	Menos del 50%									Frustrante pc	Regular. Com	Siento que hi	La interacc
27	Totalmente c	El 50%									Excelente. Nc	Regular. Com	Las actividad	Hubo algu
28	De acuerdo	Menos del 50%									Regular. Fue	Regular. Com	Las actividad	La interacc
29	Ni en desac	Más del 50%									Bastante bue	Regular. Com	Siento que hi	La interacc
30	Totalmente c	El 50%									Frustrante pc	Regular. Com	Siento que hi	La interacc
31	En desacuer	Todos									Bastante bue	Excelente. La	Las actividad	Todo fue e
32	De acuerdo	El 50%									Frustrante pc	Muy pobre. f	Siento que hi	La interacc
33	De acuerdo	Más del 50%									Bastante bue	Regular. Com	Siento que hi	Hubo algu
34	Ni en desac	Menos del 50%									Bastante bue	Regular. Com	Siento que hi	Hubo algu
35	De acuerdo	Uno o dos	Uno o dos	Casi todos	Uno o dos	Tres o cuatro	Uno o dos	Uno o dos	Ninguno	Uno o dos	Regular. Fue	Regular. Com	Las tareas y a	La interacc
36	Totalmente c	Todos	Todos	Todos	Todos	Todos	Todos	Todos	Ninguno	Todos	Excelente. Nc	Excelente. La	Las actividad	Todo fue e
37	Ni en desac	Uno o dos	Uno o dos	Tres o cuatro	Ninguno	Uno o dos	Ninguno	Uno o dos	Tres o cuatro	Uno o dos	Regular. Fue	Muy pobre. f	Siento que hi	La interacc
38	De acuerdo	Ninguno	Ninguno	Uno o dos	Uno o dos	Ninguno	Uno o dos	Uno o dos	Uno o dos	Ninguno	Frustrante pc	Muy pobre. f	Siento que hi	La interacc
39	Totalmente c	Ninguno	Ninguno	Ninguno	Uno o dos	Ninguno	Uno o dos	Uno o dos	Casi todos	Ninguno	Frustrante pc	Muy pobre. f	Las tareas y a	La interacc
40	Totalmente c	Uno o dos	Ninguno	Ninguno	Uno o dos	Ninguno	Tres o cuatro	Uno o dos	Ninguno	Ninguno	Frustrante pc	Muy pobre p	Las tareas y a	Fue nula. J
41	De acuerdo	Uno o dos	Uno o dos	Tres o cuatro	Uno o dos	Ninguno	Uno o dos	Uno o dos	Uno o dos	Uno o dos	Regular. Fue	Muy pobre p	Las actividad	La interacc

Nota: Elaboración propia.

Como se mostró en la figura 8, las primeras 34 encuestas aplicadas a los estudiantes se tuvieron que eliminar debido a inconsistencias con modificaciones posteriores en la encuesta, las modificaciones realizadas fueron en preguntas con opciones donde el tipo de respuesta que podía seleccionar el estudiante cambió, así mismo otras preguntas fueron eliminadas de la encuesta provocando que el dataset tuviera respuestas sin pregunta o a la inversa preguntas sin respuesta lo que generó campos vacíos.

Por otro lado, se tuvo que eliminar una encuesta más, pero fue por contener valores atípicos, en la edad del alumno respondió que tenía 80 años y las respuestas dadas no ofrecían información objetiva puesto que en las preguntas con opciones se aplicó la misma respuesta para todos los cuestionamientos y en las preguntas abiertas

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

se finalizó con un comentario que decía ser una encuesta de prueba. En total las encuestas eliminadas del *dataset* durante la depuración de datos conflictivos y atípicos fueron 35.

Revisión y resolución de problemas de ruido

Este apartado consistió en detectar problemas de ruido en los datos, basado principalmente en la eliminación de caracteres especiales y la duplicidad de datos.

Como se explicó anteriormente, se hizo uso de diferentes funciones de la herramienta de *Microsoft Excel*, para la eliminar caracteres especiales. Para esta tarea lo primero que se realizó fue insertar una tabla para manipular la aplicación de filtros y empezar a corregir las inconsistencias de caracteres en preguntas con escala *Likert*. Posteriormente conforme se avanzó en la limpieza de estos caracteres se volvió más difícil su detección y se aplicaron otras funciones de *Excel* como la de reemplazar, validación de datos y ortografía, estas funciones permitieron una limpieza más eficiente en las preguntas abiertas puesto que las respuestas generadas por los alumnos eran variadas.

Por último, cabe mencionar que, durante la depuración de caracteres especiales en las preguntas abiertas, fue más fácil la detección de duplicidad en las respuestas de los encuestados puesto que había respuestas que coincidían en cada una de las preguntas por lo que se detectaron un total de 15 pares de encuestas. Se puede decir que durante la revisión y resolución de problemas de ruido también se procedió a eliminar una encuesta de cada par duplicado.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

A continuación, se presenta la figura 9 que compara el *dataset* antes y después de la ejecución de tareas de depuración y resolución de problemas de ruido.

Figura 9. Comparación del *dataset* posterior a la limpieza de datos.

49	Hombre	21	Ciencias Económico Administrativas	Licenciatura en Mercadot	49	2020/06/09 11:45	Mujer	22	Ciencias de la Salud	Licenciatura en MÃ©dic
50	Mujer	26	Ciencias BÃ¡sicas	Licenciatura en MatemÃ	50	2020/06/09 11:46	Hombre	22	IngenierÃ-a y Arquitectura	IngenierÃ-a ElÃ©ctrica y
51	Mujer	22	Ciencias Sociales y Humanidades	Licenciatura en Sociología	51	2020/06/09 11:46	Hombre	22	Ciencias BÃ¡sicas	QuÃ-mico FarmacÃ©uti
52	Hombre	20	Ciencias y Tecnologías de la Información	Ingeniería en Sistemas Co	52	2020/06/09 11:47	Mujer	20	Ciencias BiolÃ³gicas	Licenciatura en BiologÃ-
53	Mujer	26	Ciencias de la Salud	Licenciatura en Nutrición	53	2020/06/09 11:47	Hombre	22	EducaciÃ³n y Artes	Licenciatura en Idiomas
54	Hombre	24	Multidisciplinaria de Jalpa de Méndez	Licenciatura en Ingeniería	54	2020/06/09 11:47	Hombre	23	Ciencias de la Salud	Licenciatura en MÃ©dic
55	Mujer	21	Ciencias Económico Administrativas	Licenciatura en Mercadot	55	2020/06/09 11:48	Hombre	20	Ciencias Sociales y Humanid	Licenciatura en Derecho
56	Mujer	23	Ciencias de la Salud	Licenciatura en Psicología	56	2020/06/09 11:48	Hombre	25	Multidisciplinaria de los RÃ-c	Licenciatura en Administ
57	Mujer	21	Multidisciplinaria de Jalpa de Méndez	Licenciatura en Genómica	57	2020/06/09 11:48	Hombre	24	IngenierÃ-a y Arquitectura	IngenierÃ-a MecÃ¡nica E
58	Mujer	23	Educación y Artes	Licenciatura en Idiomas	58	2020/06/09 11:48	Mujer	20	Ciencias de la Salud	Licenciatura en Cirujano
59	Hombre	20	Ingeniería y Arquitectura	Ingeniería Mecánica Eléct	59	2020/06/09 11:48	Mujer	21	IngenierÃ-a y Arquitectura	IngenierÃ-a QuÃ-mica
60	Hombre	20	Ciencias de la Salud	Licenciatura en Enfermeri	60	2020/06/09 11:48	Hombre	23	EducaciÃ³n y Artes	Licenciatura en Idiomas
61	Mujer	22	Ciencias BÃ¡sicas	Químico Farmacéutico Bíc	61	2020/06/09 11:48	Hombre	20	IngenierÃ-a y Arquitectura	IngenierÃ-a ElÃ©ctrica y
62	Hombre	21	Multidisciplinaria de Jalpa de Méndez	Licenciatura en Genómica	62	2020/06/09 11:48	Mujer	21	Ciencias BÃ¡sicas	QuÃ-mico FarmacÃ©uti
63	Hombre	20	Ingeniería y Arquitectura	Ingeniería Mecánica Eléct	63	2020/06/09 11:50	Hombre	30	Ciencias Sociales y Humanid	Licenciatura en Derecho
64	Hombre	20	Ingeniería y Arquitectura	Ingeniería Química	64	2020/06/09 11:50	Hombre	21	IngenierÃ-a y Arquitectura	IngenierÃ-a Civil
65	Mujer	20	Ciencias BÃ¡sicas	Licenciatura en Actuaría	65	2020/06/09 11:50	Mujer	26	EducaciÃ³n y Artes	Licenciatura en Ciencias
66	Mujer	22	Ciencias Económico Administrativas	Licenciatura en Administr	66	2020/06/09 11:51	Hombre	20	IngenierÃ-a y Arquitectura	IngenierÃ-a QuÃ-mica
67	Hombre	20	Multidisciplinaria de Comalcalco	Licenciatura en Enfermeri	67	2020/06/09 11:51	Hombre	21	Multidisciplinaria de Jalpa de	Licenciatura en Ingenier
68	Mujer	23	Ciencias BÃ¡sicas	Ingeniería Geofísica	68	2020/06/09 11:52	Mujer	24	EducaciÃ³n y Artes	Licenciatura en Comunic
69	Mujer	22	Ciencias BÃ¡sicas	Químico Farmacéutico Bíc	69	2020/06/09 11:52	Mujer	20	Ciencias y Tecnologías de la	IngenierÃ-a en Sistemas
70	Mujer	22	Ciencias BÃ¡sicas	Químico Farmacéutico Bíc	70	2020/06/09 11:52	Mujer	20	Ciencias de la Salud	Licenciatura en Psicolog
71	Mujer	20	Ciencias BÃ¡sicas	Químico Farmacéutico Bíc	71	2020/06/09 11:52	Mujer	22	EducaciÃ³n y Artes	Licenciatura en Comunic
72	Mujer	20	Ciencias BÃ¡sicas	Licenciatura en Física	72	2020/06/09 11:54	Mujer	21	Ciencias BiolÃ³gicas	Licenciatura en BiologÃ-
73	Mujer	20	Ciencias de la Salud	Licenciatura en Psicología	73	2020/06/09 11:54	Mujer	21	Ciencias BÃ¡sicas	QuÃ-mico FarmacÃ©uti
74	Hombre	20	Ingeniería y Arquitectura	Ingeniería Química	74	2020/06/09 11:54	Hombre	20	Multidisciplinaria de los RÃ-c	IngenierÃ-a en InformÃ

Nota: Elaboración propia.

En esta etapa de limpieza y procesamiento de datos se concluyó con la eliminación 35 encuestas durante la depuración de datos conflictivos y atípicos, 15 encuestas eliminadas durante la revisión y resolución de problemas de ruido, siendo un total de 50 encuestas eliminadas por las inconsistencias detectadas. Dejando un *dataset* con 1,020 encuestas de las 1,070 que se tenían inicialmente. Cabe mencionar que a pesar de esta depuración se cuenta con un número considerable para la aplicación de minería de datos.

3.3.4 Reducción y proyección de los datos

Consiste en realizar una búsqueda de similitudes en la información, por ejemplo, respuestas obtenidas por los estudiantes que sean muy parecidas y que pueden ser representadas como datos de mosaico, haciendo uso de la reducción de dimensiones o

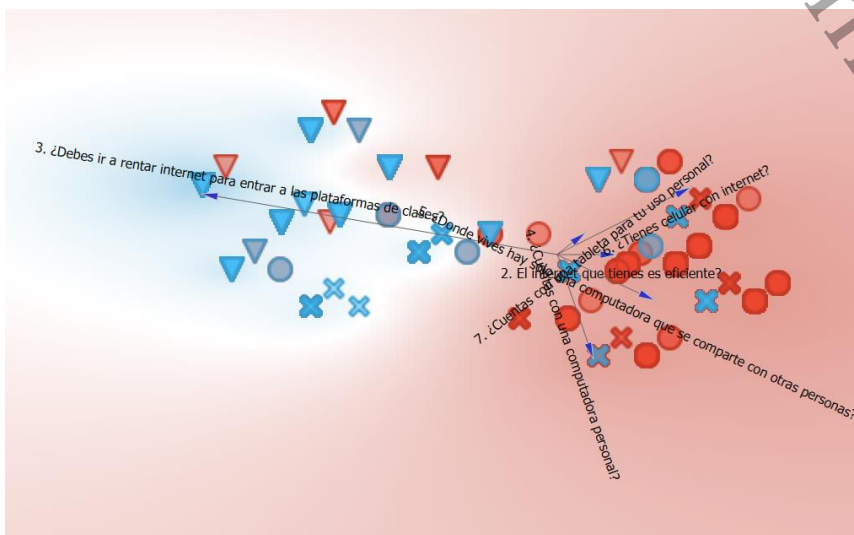
1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

bien obtener una transformación con el propósito de obtener un número de variables efectivas. En este sentido Witten *et al.*, (2017) expone que la etapa reducción y transformación de los datos va a permitir un funcionamiento eficiente de los algoritmos que se utilicen, así como una mejora en la estética del proceso de minería de datos.

Lo anterior se puede ver reflejado en la siguiente imagen donde posterior a la limpieza de los datos se realizó una proyección de los mismos con una de las herramientas que se ocuparán (ver figura 10). Se puede observar en el diagrama de prueba que se realizó, que la visualización es poco estética lo cual durante el proceso de análisis e interpretación de los datos serán más difícil su comprensión. Aunque se muestra el ejemplo de una herramienta, el propósito es ilustrar que es indispensable la parte de la transformación de los datos para lograr una proyección eficiente en cualquier herramienta de minería de datos que se implemente.

Figura 10.
Proyección de datos en Orange.



Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Nota: Elaboración propia.

Por lo anterior se tuvo que realizar una transformación del *dataset* como parte del proceso metodológico, lo primero que se realizó fue cambiar la nomenclatura de las preguntas de la encuesta (ver tabla 11).

Tabla 11.
Cambio de nomenclatura en las variables.

Variable	Cambio de nombre
Género	GEN
Edad	EDAD
División Académica	DIV_AC
Programa educativo que cursa	CARR
Año de ingreso a la universidad	ING
Asignaturas inscritas	MAT
Preguntas de la encuesta	P.(1,2,3, n...)

Nota: Elaboración propia.

Como se mostró previamente en la tabla 11, se cambió la nomenclatura desde las preguntas generales hasta las preguntas referentes a la investigación. Posteriormente se realizó un cambio en las preguntas abiertas que corresponde a los números 42 y 43 de la encuesta donde los estudiantes comparten sus opiniones y experiencias acerca de las clases virtuales. Para este caso se buscó representar estas respuestas con tres categorías:

- **Sugerencias:** para recomendaciones que realiza el alumno a la institución académica.
- **Desagrado:** Opiniones negativas referentes a sus experiencias con la modalidad de clases.
- **Conformidad:** Opiniones positivas acerca de las clases en plataforma virtual.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Por otro lado, también se modificaron las respuestas de preguntas con escala Likert, esto para reducir el número de caracteres, dejando respuestas más cortas con palabras clave, por ejemplo: Totalmente de acuerdo, de acuerdo, ni en acuerdo ni en desacuerdo en desacuerdo y Totalmente en desacuerdo.

Durante el proceso en esta etapa prevalecieron las tareas relacionadas a la transformación de datos, con respecto a la reducción de datos únicamente se eliminó una columna del *dataset* que contenía la fecha y hora en que cada estudiante realizó la encuesta porque no forma parte de la encuesta, la columna de fecha y hora es proporcionada por el formulario cuando se hace la exportación del *dataset* y no proporciona información relevante para los propósitos de este proyecto.

3.3.5 Selección de funciones y algoritmos de minería de datos

Incluye decidir cuál es el propósito del modelo que se va a diseñar y que se va a derivar de los algoritmos de minería de datos, asimismo servirán para la representación del modelo. Debido a que se busca la relación entre las barreras que impiden el uso de innovaciones tecnológicas, la función de los datos guardará relación con modelos de agrupamiento, clasificación y análisis de enlaces.

Métodos de minería de datos.

Aunque previamente se mencionan los modelos que se aplicarán durante el desarrollo de minería de datos, es necesario indicar que de acuerdo con lo expuesto por Bhatia (2019), existen dos clasificaciones o métodos para la aplicación de los algoritmos de minería de datos, mismos que se presentan a continuación:

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

- Supervisadas o predictivas: Utiliza variables o campos de una base de datos o conjunto de datos que permiten predecir valores desconocidos o futuros de manera que se logre determinar un modelo predictivo basado en conocimiento previo.
- No supervisadas o descriptivas: Encuentra patrones que permiten describir información que puede ser interpretada.

Asimismo, las técnicas de minería de datos son empleadas con el propósito de obtener información que no se puede obtener de un análisis estadístico convencional, por lo tanto, con la minería de datos se puede obtener un análisis de información oculta en grandes cantidades de datos. En este sentido de acuerdo con las aportaciones realizadas por Fayyad Piatetsky-Shapiro & Smyth (1996), Gironés *et al.*, (2017), Witten *et al.*, (2017), Zaki & Meira (2020), se describen las funciones empleadas en este proyecto de investigación:

Agrupamiento. También conocido como *clustering*, sirve para agrupar datos dentro de un número de clases aplicando criterios o algoritmos de distancia o similitud, de modo que las agrupaciones o clases que se formen de un conjunto de datos se base en atributos similares.

Clasificación. Para el caso particular de la técnica de clasificación de hará uso de árboles de decisión, que son estructuras que representan un conjunto de reglas sobre determinado conjunto de datos. Dentro de las ventajas que ofrece esta técnica se menciona que es fácil de entender, requiere de menos limpieza de datos y es útil para la

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

exploración de datos de modo que permita identificar la relevancia de las variables de entre cientos de variables.

Análisis de enlaces. En minería de datos es un modelo que determina las relaciones en la base de datos, centrando la atención en derivar correlaciones de múltiples campos que cumpla con los límites de soporte y confianza.

Las funciones previamente descritas dan origen a la elección de los algoritmos que se pueden ocupar para la aplicación de minería de datos. Aunque existen una diversidad de algoritmos que pueden aplicarse a los modelos o funciones previamente expuestas, en la tabla 10 se describen los algoritmos seleccionados para este trabajo.

Para la elección de los algoritmos influyen aspectos como la selección de métodos a aplicar, así como los modelos y parámetros que pueden ser los más indicados para el estudio, debido a que cada uno de los aspectos deben coincidir con el proceso general del método KDD.

Tabla 12.
Algoritmos de minería de datos.

Técnica	Algoritmos	Descripción
Árboles de decisión (clasificación)	CART, J48, REPTree, Tree	Es un algoritmo comúnmente utilizado para clasificar instancias dentro de un conjunto de datos. Para esto divide los datos en nodos haciendo uso del criterio de normalización conocido como ganancia de información que resulta de la elección del atributo con más información y a partir de este se derivan otras subclasificaciones (Yang, 2019).
Agrupamiento (<i>clustering</i>)	K-means	Es un método de agrupación de clústeres en un conjunto de datos, generando como resultado un nuevo conjunto de datos pero adicionando un

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Técnica	Algoritmos	Descripción
		atributo conocido como clase (Bahga & Madiseti, 2019).
Reglas de asociación	A priori Association Rules	Este tipo de algoritmo identifica afinidades dentro de un conjunto de datos. Para ello realiza una búsqueda de atributos frecuentes que pueden generar condiciones que den como resultado otras relaciones en conjuntos de datos de mayor tamaño, siempre y cuando se conserven características en común (Kantardzic, 2020).

Nota: Elaboración propia.

Otro aspecto que considerar para la selección de los algoritmos es la elección de las herramientas para minería de datos que se ocuparon en este trabajo. Cabe mencionar que, al ocupar tres herramientas distintas para el proceso metodológico, también fue necesario que los algoritmos a utilizar en las aplicaciones fuesen iguales o se aproximaran en funcionamiento.

Por lo anteriormente expuesto se destaca que la selección de funciones y algoritmos de minería de datos, así como la aplicación de los criterios del proceso KDD, tienen como propósito principal comprender y analizar los modelos generados.

3.3.6 Minería de datos

Se realiza la búsqueda e identificación de los patrones, así como la forma de representarlos, esto puede incluir reglas o árboles de decisión, agrupamiento, regresión lineal o asociación. Es necesario que todos los pasos previos se realicen de forma óptima para que la minería de datos se realice adecuadamente.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Por lo tanto, en este apartado se presenta el proceso que se llevó a cabo para la ejecución de los algoritmos seleccionados en cada una de las herramientas.

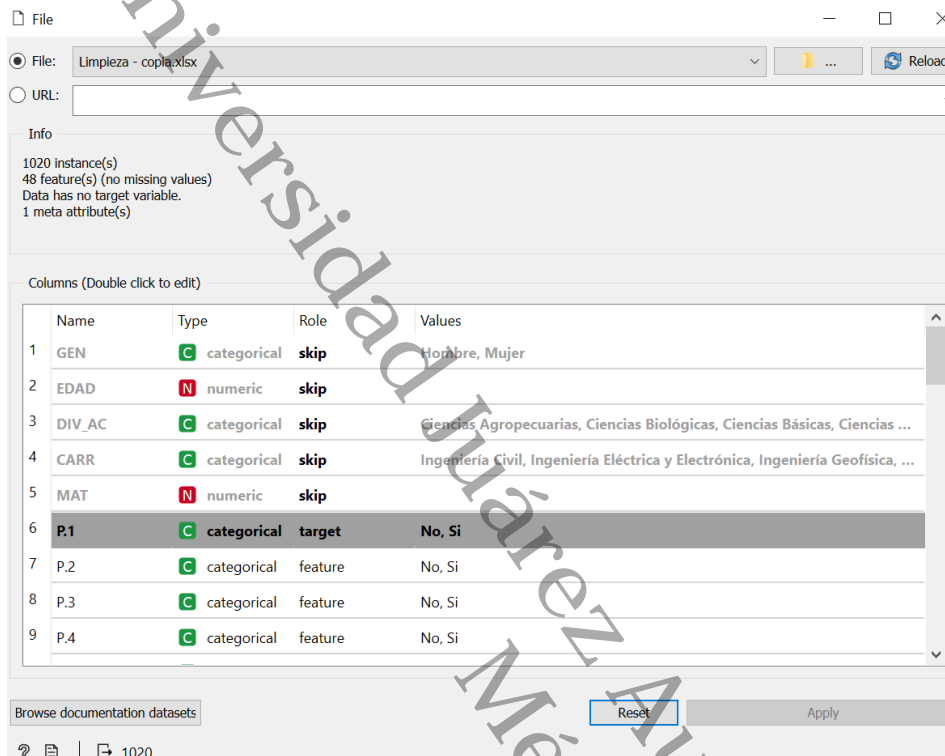
Orange.

En este apartado se describe el proceso particular de *Orange* para la aplicación de los algoritmos seleccionados y la generación de diagramas, así como la información necesaria para su representación.

Se inició el proceso de minería de datos con la selección del dataset tal y como se muestra en la figura 11, se observa que una vez que se encuentra cargado el archivo a procesar se pueden seleccionar los datos que se quieren analizar, así mismo la herramienta identifica de forma automática si las variables son de tipo numérica, categórica, *datetime* o texto. También de forma manual se puede seleccionar el tipo de variable si el usuario no está de acuerdo con el reconocimiento automático, por otro lado, se encuentra la columna de rol que identifica si la variable será considerada o no para el análisis y además si habrá una pregunta raíz, para ello se ocupará el rol *target*, para el caso de los árboles de decisión es importante determinarlo.

Figura 11.
Selección del dataset en Orange.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos



File: Limpieza - copia.xlsx

Info

1020 instance(s)
48 feature(s) (no missing values)
Data has no target variable.
1 meta attribute(s)

Columns (Double click to edit)

	Name	Type	Role	Values
1	GEN	Categorical	skip	Hombre, Mujer
2	EDAD	Numeric	skip	
3	DIV_AC	Categorical	skip	Ciencias Agropecuarias, Ciencias Biológicas, Ciencias Básicas, Ciencias ...
4	CARR	Categorical	skip	Ingeniería Civil, Ingeniería Eléctrica y Electrónica, Ingeniería Geofísica, ...
5	MAT	Numeric	skip	
6	P.1	Categorical	target	No, Si
7	P.2	Categorical	feature	No, Si
8	P.3	Categorical	feature	No, Si
9	P.4	Categorical	feature	No, Si

Browse documentation datasets

Reset Apply

1020

Nota: Elaboración propia.

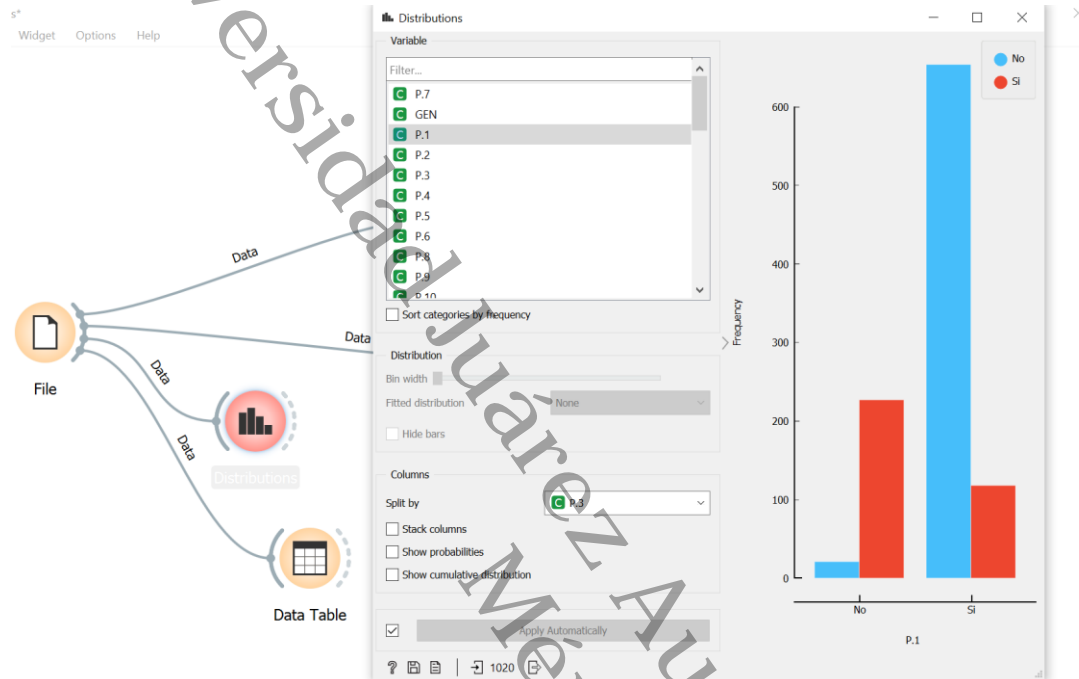
Una vez definidas las variables es importante realizar un estudio de los datos, esto es recomendable antes de aplicar cualquier algoritmo de minería de datos. Para ello se seleccionó dentro de la interfaz gráfica el elemento *distributions* y se unió a la herramienta *file* que contiene la información del *dataset*. En la figura 12 se presenta la conexión en los elementos descritos y también una representación gráfica de la función *distributions*, se puede observar una gráfica de barras que muestra la siguiente información con la selección de variable P.1 y dividir por (*split by*) P.3:

P.1 el lugar donde vives cuenta con internet=sí entonces no necesitas ir a rentar internet (sí=no).

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 12.
Representación gráfica de la función distributions.



Nota: Elaboración propia.

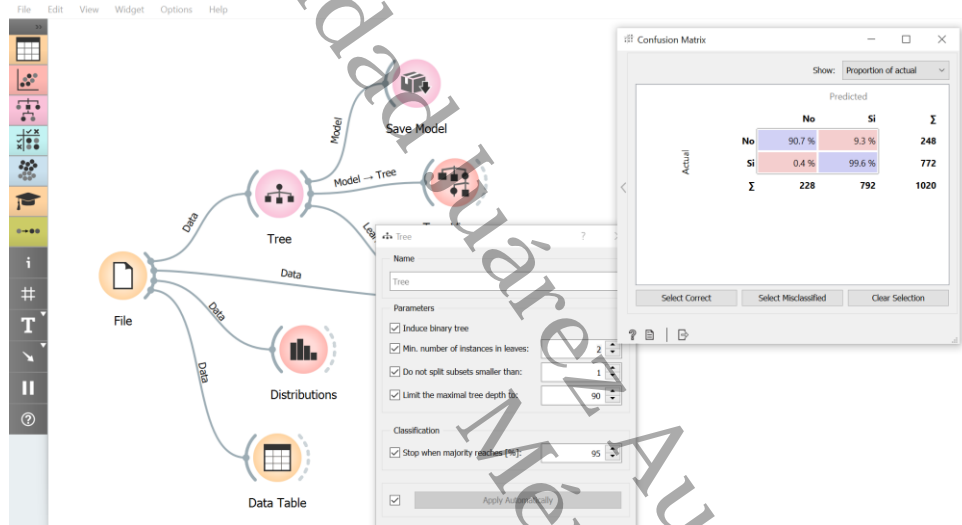
Como este ejemplo se pueden analizar todas las barras del gráfico también analizar la relación que existe entre las variables para tener una mejor comprensión de los algoritmos a integrar.

Una vez realizado el estudio de los datos se procede a la selección de los algoritmos, en este sentido se inició con los árboles de decisión que como se explicó previamente existen diferentes algoritmos como CART, J48, REPTree o ID3. La función dentro de la aplicación que lo representa se le conoce como *Tree*, para ello se realizó primero la conexión de herramientas de evaluación *Test and Score* y *Confusion Matrix* que permiten evaluar los resultados predictivos del árbol con el icono *file*, en la figura 13

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

se representa la configuración previa a visualizar el árbol de decisión, así como la función de matriz de confusión.

Figura 13.
Evaluación del árbol de decisión en Orange.



Nota: Elaboración propia.

En la figura 13 se observa que la matriz de confusión ha clasificado de forma correcta en un 90.7% el ítem "No" y 99.6% el ítem "Sí", el promedio de estos dos porcentajes da una proporción del 95.15% de reconocimiento de clasificación con la variable P.1 como pregunta raíz. Así mismo, se muestra la ventana de configuración del modelo *Tree*, donde se puede experimentar con el número de instancias, profundidad del árbol y porcentaje de aciertos, además se puede visualizar el árbol con la función de visualizador *Tree Viewer*.

Una vez realizado el proceso análisis de datos y la funcionalidad de los árboles de decisión, se procede a la aplicación del algoritmo *K-means*, se inició configurando el

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

conjunto de datos para el análisis, posteriormente se realizó una conexión con el *widget* K-means y *File*, en la figura 14 se ilustra la conexión realizada en Canvas y también la ventana de configuración para el algoritmo.

Figura 14.
Ventana de configuración de K-means en Orange.



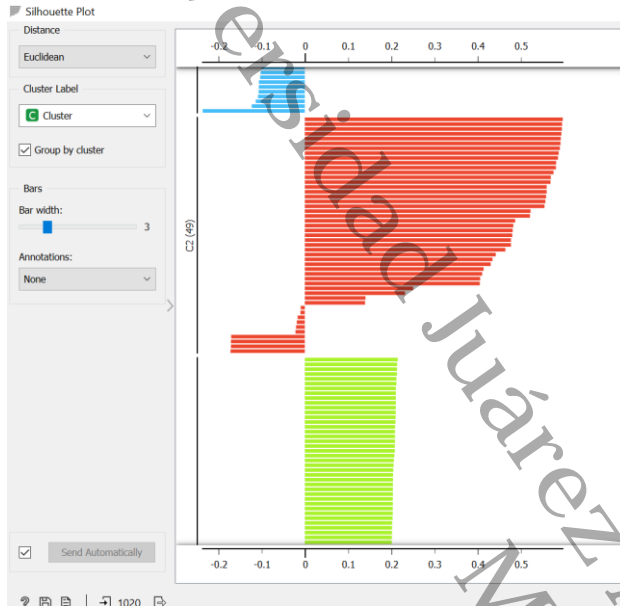
Nota: Elaboración propia.

Se observa que se puede hacer una selección automática de las agrupaciones o bien seleccionar de forma manual el intervalo de clasificaciones. Así mismo, se puede ver la opción *Silhouette Scores*, entre mayor sea la puntuación de este atributo mejor será la agrupación. También se realizó una conexión con el widget *Silhouette Plot* que permite visualizar las instancias clasificadas y permite detectar valores atípicos, los puntajes positivos representa una correcta clasificación en los *clusters*, en cuanto a los valores negativos representan instancias que no encajan correctamente (ver figura 15).

2

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 15.
Gráfica de Silueta en Orange.



Nota: Elaboración propia.

En esta primera representación se observa que se realizaron tres agrupaciones, pero, existen valores que no encajan correctamente. El algoritmo *K-means* se puede ejecutar las veces que sean necesarias para realizar la agrupación correcta de los datos.

El tercer algoritmo que se trabajó en la herramienta *Orange* fue el de reglas de asociación, para la minería de datos consiste en un conjunto de patrones frecuentes por crecimiento denominado *FP-growth*.

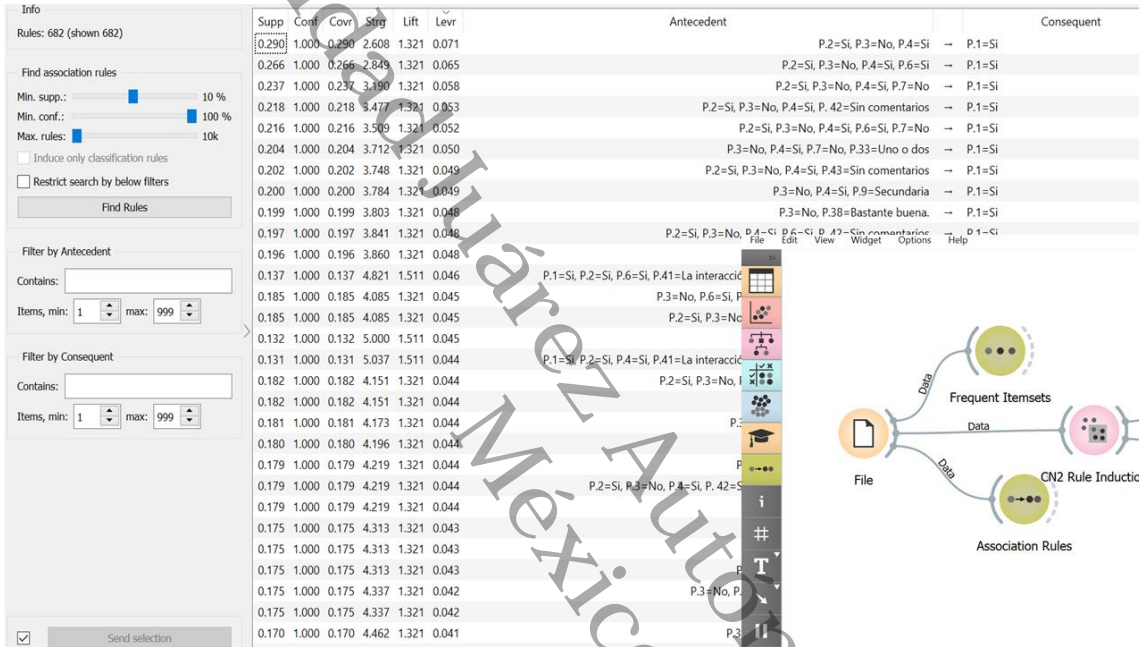
Para el uso de este algoritmo se tuvieron que descargar funciones adicionales, basta con ir al menú superior de la herramienta seleccionar *options, add-ons* y seleccionar *Orange3-Associate* versión 1.1.9, los *widgets* que se adicionan a la plataforma son *Frequent Itemsets* y *Association Rules*. Al igual que en los algoritmos anteriores, estos *widgets* deben conectarse a *File* para su aplicación y se deben seleccionar los atributos

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

que serán analizados. En la figura 16 se presenta una inducción de reglas de asociación en el dataset así como la representación de los widgets aplicados.

Figura 16. Aplicación de Reglas de asociación en Orange.



Nota: Elaboración propia.

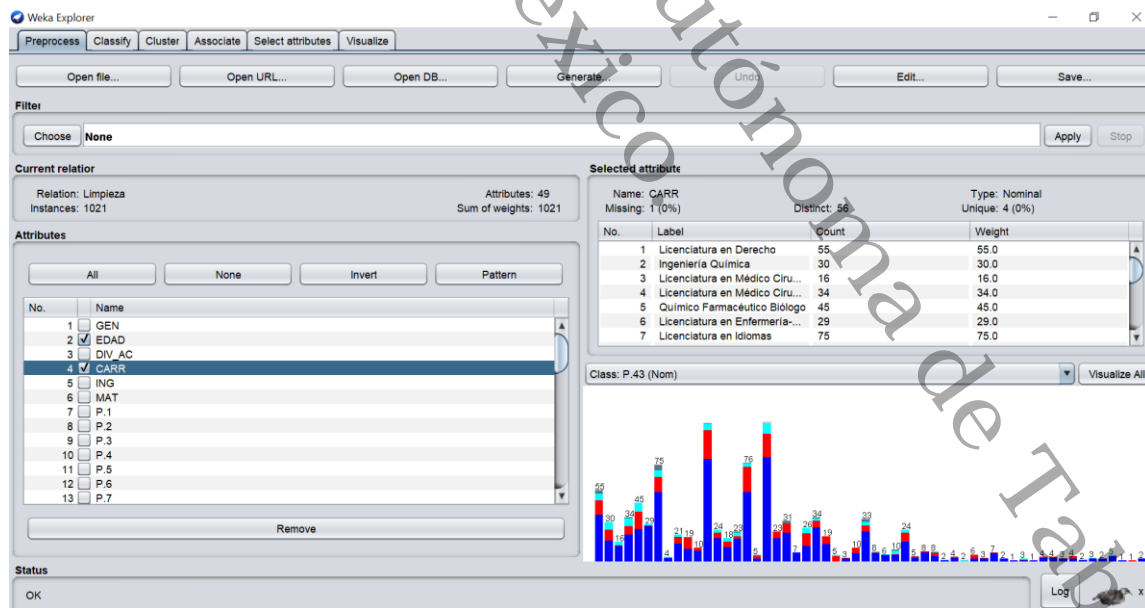
Se observa en la figura 16 que la herramienta genera todas las posibilidades de reglas dentro de los parámetros los cuales son: el soporte mínimo siendo este el porcentaje de todo el conjunto de reglas descubierto, confianza mínima y el número máximo de reglas que el sistema debe generar. Aunque la asociación de reglas ya es visible, también se puede adicionar la función de inducción de reglas con el widget CN2 Rule Induction y pueden ser visibles con CN2 Rule Viewer.

WEKA.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Con la herramienta WEKA se realizó un proceso similar a la herramienta anterior, cambian algunos de los procesos y esto se deriva del diseño de la herramienta. Se inició con el proceso de estudio de los datos para comprender cómo trabajar en el entorno *Explorer* de WEKA. En la figura 17 se muestra la pestaña del preprocesamiento de los datos, se observa que la herramienta por defecto permite visualizar gráficas de distribución por variable, incluso se puede optar por visualizar todas las variables con graficas individuales. En la sección de *Selected attribute* se puede conocer si la variable es de tipo nominal, numérica, texto o fecha, así como el nombre de la variable y las diferentes etiquetas o respuestas del *dataset*.

Figura 17.
Estudio de los datos en WEKA.



Nota: Elaboración propia.

1

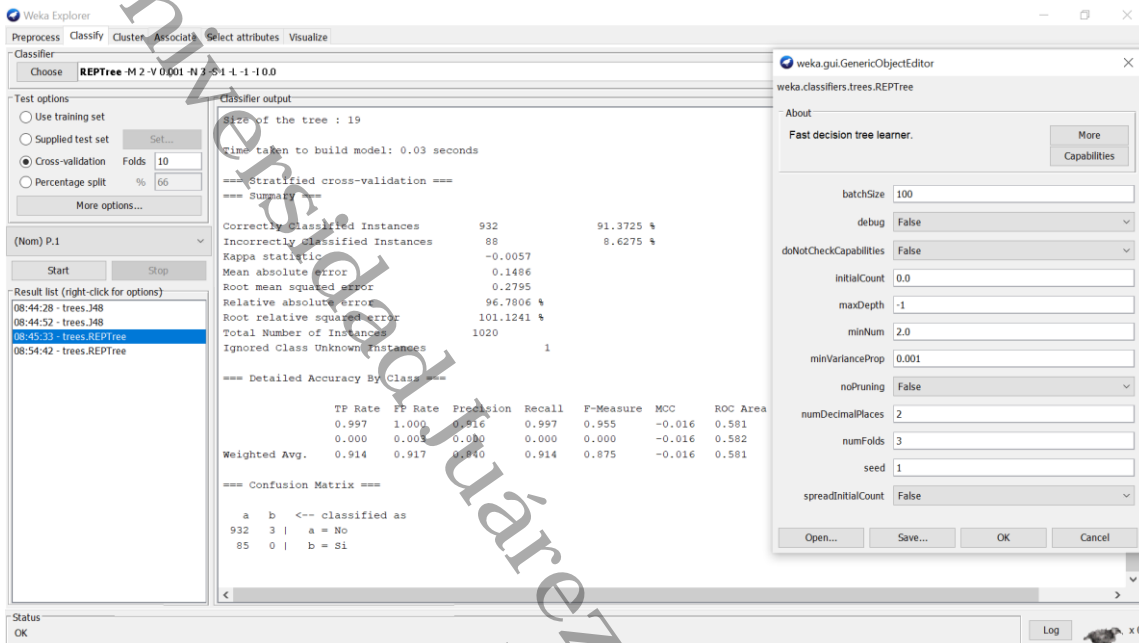
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Para hacer uso de los diferentes algoritmos se debe hacer uso de filtros que permitan seleccionar cuáles son los ítems que se aplicarán. Esto se debe a que los algoritmos requieren que el tipo de atributo a evaluar sea únicamente del tipo nominal o numérico por poner un ejemplo. En este sentido cabe mencionar que los filtros que se ocuparon son del tipo no supervisados en la sección Attribute, principalmente fueron *Remove* y *NumericToNominal*.

Una vez realizado el estudio de los datos y la aplicación de los filtros se procedió a la aplicación de los algoritmos, iniciando por el modelo de árboles de decisión. Para la aplicación se seleccionó la pestaña *Classify* en la sección Chosse, carpeta trees. En la figura 18 se observa el proceso de ejecución del árbol de decisión, la selección de la pregunta raíz o la variable con mayor ganancia de información, así como la tabla para configurar los parámetros con los que se desea trabajar.

Figura 18.
Generación del modelo árbol de decisión WEKA.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos



Nota: Elaboración propia.

Por último, se muestra que en el proceso de ejecución WEKA genera los parámetros de evaluación por defecto, aplicando una matriz de confusión y métricas de clasificación de los datos. En la parte inferior izquierda se muestran las ejecuciones realizadas que pueden ser visualizadas con click derecho sobre el modelo en la opción *Visualize Tree*.

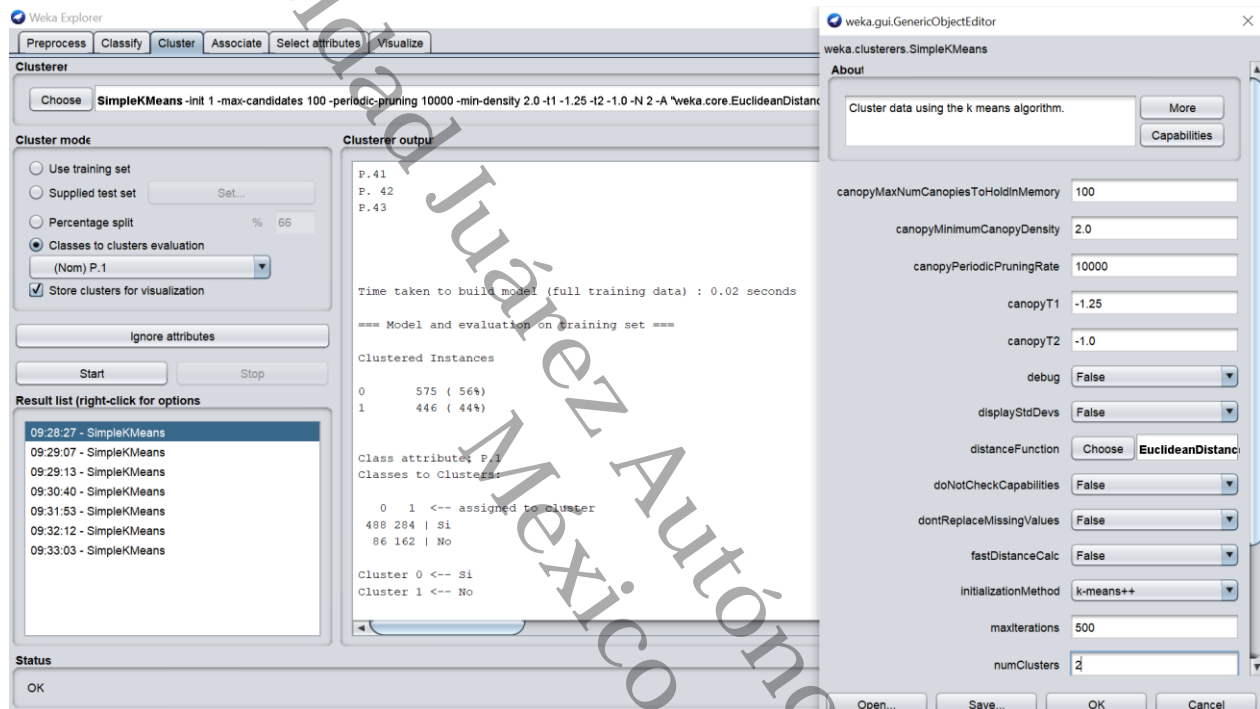
El siguiente modelo para ejecutar es el de agrupación (*cluster*), para su aplicación se selecciona la pestaña *Cluster*, se seleccionó la función *SimpleKmeans*. En la figura 19 se presenta la tabla de configuración para ejecutar el modelo, se observa que es necesario especificar un atributo base para realizar las agrupaciones, en el editor se debe especificar cuantos cluster va a generar la herramienta y el método de inicialización que puede ser Random o K-means++, en este caso se optó por utilizar el método K-means++

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

y se hicieron pruebas con dos y tres agrupaciones. Por último, se muestran los parámetros de evaluación que son representados por una matriz de confusión.

Figura 19.
Aplicación del modelo cluster en WEKA.



Nota: Elaboración propia.

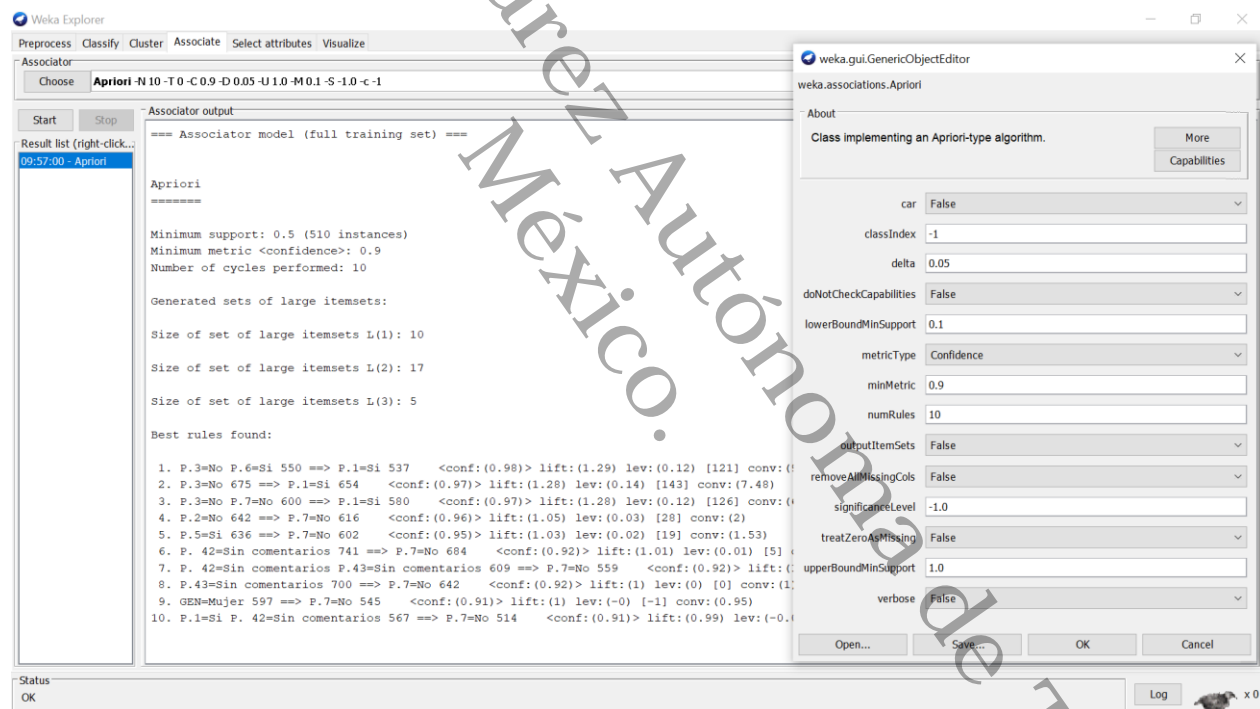
Finalmente, se presenta la aplicación del modelo de análisis de enlaces empleando reglas de asociación con la función Apriori. Este modelo se encuentra en la pestaña *Associate*, normalmente el algoritmo Apriori aparece como primera opción para ser ejecutado, al ser una recomendación que la propia herramienta realiza y por ser el algoritmo más sencillo de emplear cuando se trata de asociar variables. En la figura 20 se observa la ejecución de este algoritmo y la tabla de edición, de igual forma se muestra que este algoritmo no requiere generar una visualización adicional porque las reglas elaboradas por WEKA pueden ser visibles en el panel *Classifier output*. También se

99

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

destaca que en esta herramienta en la tabla de edición WEKA muestra las mejores diez reglas de asociación como estándar pero el usuario puede modificar ese parámetro para ejecutar las reglas que sean necesarias. En este caso, también se realizaron diferentes pruebas modificando el número de reglas, esto se debe a que WEKA me genera las mejores reglas de acuerdo con sus métricas, pero no significa que proporcionen información relevante.

Figura 20.
Aplicación del modelo de asociación WEKA.



Nota: Elaboración propia.

Lenguaje R.

La tercera herramienta que se analizó fue Lenguaje R, en la etapa de aprendizaje y dominio de la aplicación se explicó que se requiere instalar un complemento llamado R

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

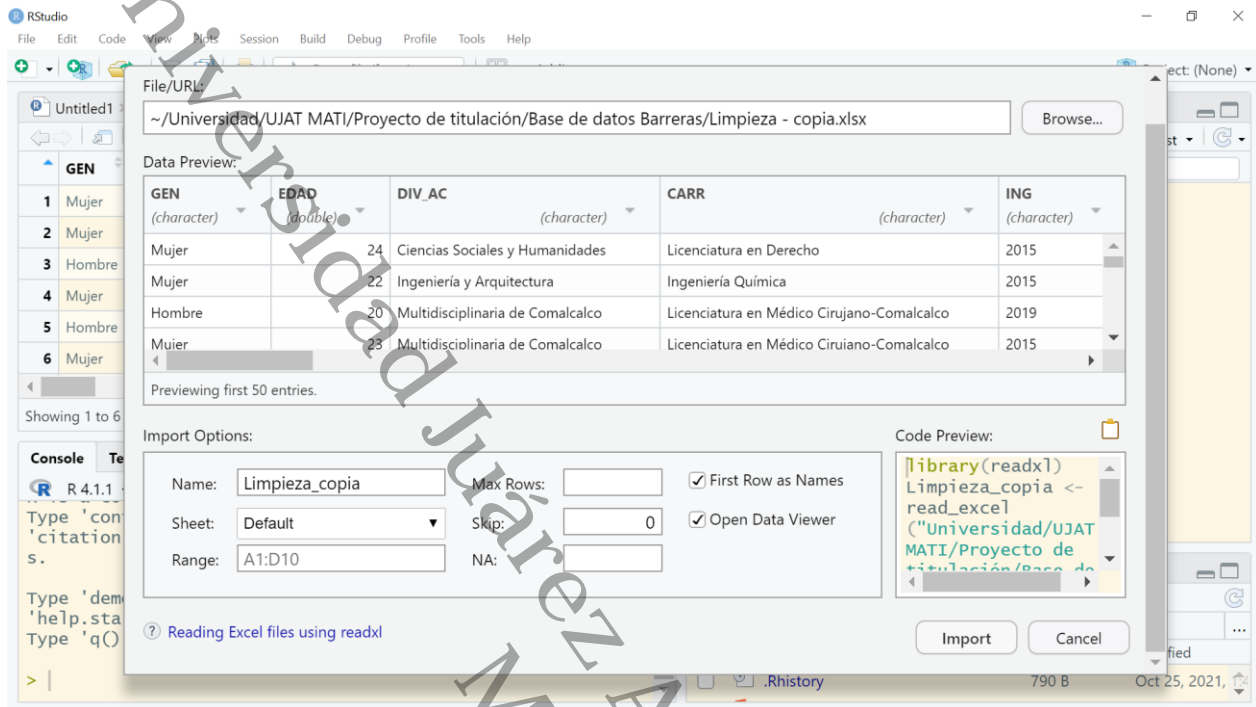
Studio que es el área de trabajo o interfaz gráfica donde se ejecutan las fórmulas y funciones de lenguaje R. A diferencia de las herramientas anteriores que ofrecen una mayor facilidad al usuario para la aplicación de funciones y algoritmos, en esta herramienta se requiere ejecutar líneas de comando para realizar los modelos de minería de datos. Sin embargo, aprender a usarlo es sencillo para un nuevo usuario.

Para iniciar la aplicación de los algoritmos se requiere la importación del *dataset* en la herramienta, hay varias opciones para realizar la importación de los datos, como líneas de comando para lectura de un archivo en Excel, CSV u otro tipo de formato como el .arff utilizado para WEKA. También se puede hacer seleccionando el ícono de importar datos y se selecciona la opción Excel.

En la figura 21 se muestra la visualización de los datos, se observa la ruta del archivo, el rango, se puede configurar el tipo de datos (texto, numérico o fecha) y ofrece la opción de saltar los atributos que no se consideren para el análisis. Una vez realizada la configuración, en la esquina inferior derecha se muestra una visualización previa del código que se generó con los cambios hechos y se puede copiar y pegar en un bloc de notas para ocuparlo posteriormente. Si se considera que se hicieron todas las configuraciones necesarias se procede a importar el archivo.

Figura 21.
Importación del dataset a Lenguaje R.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos



The screenshot shows the RStudio interface with the 'Import Data' dialog box open. The dialog is for importing an Excel file named 'Limpieza_copia.xlsx'. The 'Data Preview' section shows a table with the following data:

GEN	EDAD	DIV_AC	CARR	ING
Mujer	24	Ciencias Sociales y Humanidades	Licenciatura en Derecho	2015
Mujer	22	Ingeniería y Arquitectura	Ingeniería Química	2015
Hombre	20	Multidisciplinaria de Comalcalco	Licenciatura en Médico Cirujano-Comalcalco	2019
Mujer	23	Multidisciplinaria de Comalcalco	Licenciatura en Médico Cirujano-Comalcalco	2015

The 'Import Options' section shows the following settings:

- Name: Limpieza_copia
- Max Rows: (empty)
- First Row as Names:
- Sheet: Default
- Skip: 0
- Open Data Viewer:
- Range: A1:D10
- NA: (empty)

The 'Code Preview' section shows the following R code:

```
library(readxl)
Limpieza_copia <- read_excel("Universidad/UJAT MATI/Proyecto de titulación/Base de datos/Barreras/Limpieza_copia.xlsx")
```

The console shows the following output:

```
R 4.1.1
Type 'con'
Type 'citation'
Type 'dem'
Type 'help.sta'
Type 'q()'
> |
```

Nota: Elaboración propia.

Una vez que se tienen cargados los datos en el *software*, se pueden visualizar con las configuraciones hechas en el cuadro superior, aunque se encuentre cargado el *dataset* no se puede aplicar ninguna función sobre este a menos de que se copie el código generado durante la importación del archivo de Excel como se muestra en la figura 22 con la carga de librerías y el código de importación del *dataset*.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 22.

Descarga de librerías del árbol de decisión en Lenguaje R.

```
1 library(readxl)
2 library(rpart)
3 library(rpart.plot)
4 library(class)
5 library(tidyverse)
6 library(caret)
7 Barreras <- read_excel("~/Universidad/UJAT MATI/Proyecto de titulación/Base
8                       col_types = c("text", "skip", "skip",
9                                     "skip", "text", "text", "text", "text",
10                                    "text", "text", "text", "text", "text",
11                                    "text", "text", "text", "text",
```

Nota: Elaboración propia.

Como se muestra en la figura 22, se cargó la librería *readxl* y *tidyverse* para la importación del *dataset* en el formato estándar de *Microsoft Excel*, las librerías *rpart*, *rpart.plot* y *class* para la ejecución y visualización del árbol de decisión con el modelo de clasificación y por último la librería *caret* para generar la matriz de confusión para evaluar el modelo de árbol.

Para generar el árbol de decisión se realizaron 50 líneas de comando, integradas por diferentes tareas principales que se presentarán en lo subsecuente. Como se mostró previamente en la figura 22, se inició la creación del modelo con la descarga de librerías y la importación del *dataset*.

Posteriormente se crearon dos conjuntos de datos, el primero de entrenamiento y el segundo de testeo y para ello se asignó como pregunta raíz la variable P.1, una vez que el conjunto de datos se dividió en entrenamiento y testeo se procedió a la elaboración del árbol y su visualización (ver figura 23).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 23.
Creación del modelo de árbol de decisión en Lenguaje R.

```
22 #Dividir la data en entrenamiento y testeo#
23
24 set.seed(12345)
25 muestra = createDataPartition(Barreras$P.1, p = 0.6, list = F)
26 muestra
27 train = Barreras[muestra,]
28 test = Barreras[-muestra,]
29
30 #Crear el árbol#
31
32 arbol = rpart(P.1 ~., data = train, method = "class")
33 arbol
34 rpart.plot(arbol)
```

Nota: Elaboración propia.

Por último, se realizó la evaluación del modelo que consistió en la creación de las líneas de comando para predecir sobre el *dataset* el modelo de árbol, seguido de la elaboración de la matriz de confusión que permitió conocer la proporción de datos correctamente clasificados (ver figura 24).

Figura 24.
Modelo de evaluación del árbol de decisión en lenguaje R.

```
36 #Evaluar el modelo#
37
38 predict_arbol = predict(arbol, test, type = "class")
39 predict_arbol
40
41 #Matriz de confusion#
42
43 confusion = table(test$P.1, predict_arbol, dnn = c("real", "estimada"))
44 confusion
45 PCC = sum(diag(confusion))/sum(confusion)
46 PCC
47
48 #Matriz de confusion con libreria caret#
49
50 confusionMatrix(test$P.1, predict_arbol)
```

Nota: Elaboración propia.

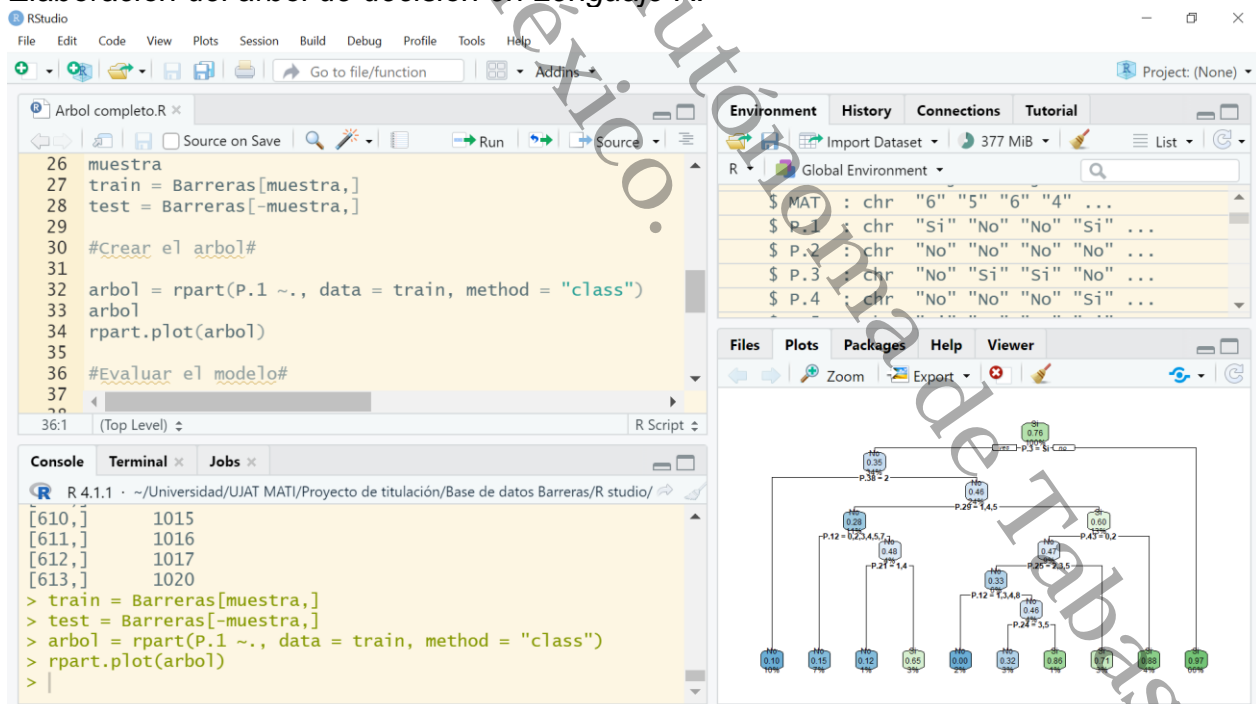
1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En la figura 24 se muestra que la matriz de confusión se puede realizar de dos maneras diferentes. La primera requiere de la creación de una tabla que contenga en conjunto de datos de testeo elaborado previamente, seguido de la elaboración del porcentaje correctamente clasificado denominado PCC. La segunda opción es más sencilla porque únicamente requiere la descarga de la librería caret y ejecutar correctamente la fórmula *confusionMatrix*.

Una vez realizadas y ejecutadas las líneas de comando se presenta la distribución de la interfaz de R al momento de generar el árbol de decisión con el fin de describir el entorno de trabajo durante el proceso (ver figura 25).

Figura 25.
Elaboración del árbol de decisión en Lenguaje R.



Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Como se presenta previamente en la figura 25, el cuadro superior izquierdo contiene las líneas de comando que genera el usuario para crear los modelos, el cuadro inferior izquierdo presenta la ejecución de los comandos desde la consola de Lenguaje R, para la ejecución de cada línea de comando se selecciona el botón *Run* y en el cuadro inferior izquierdo se muestra el diagrama generado.

El siguiente algoritmo por utilizar es K-means, para su uso se escribieron 70 líneas de comando que constan principalmente de las importación y visualización del *dataset*, aplicación del algoritmo y la búsqueda del número óptimo de clústeres y visualización de los agrupamientos.

En la figura 26 se presentan las líneas de comando uno a la cinco que corresponden a la instalación las librerías *tidyverse*, *factoextra*, *cluster* y *readxl*. Estas librerías servirán para importar, manipular, visualizar los datos, así como la aplicación del algoritmo.

Figura 26.

Descarga de librerías de K-means en Lenguaje R.

```
1 library(tidyverse)
2 library(factoextra)
3 library(cluster)
4 library(readxl)
5
6 K_means_Barreras <- read_excel("~/Universidad/UJAT MATI/Proyecto de titulación,
7                               col_types = c("numeric", "numeric", "skip",
8                               "skip", "numeric", "numeric", "num
9                               "numeric", "numeric", "numeric",
10                              "numeric", "numeric", "numeric",
11                              "numeric", "numeric", "numeric",
12                              "numeric", "numeric", "numeric",
13                              "numeric", "numeric", "numeric",
```

Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Así mismo, se tuvo que realizar la importación de los datos tal y como se hizo en los árboles de decisión, configurar las variables a ocupar y copiar el código de la importación para ocuparlo posteriormente en la ejecución de las demás líneas de comando.

Posterior a la descarga de librerías e importación de datos, en las líneas 24 a 27 se ejecutó la visualización de datos, estadística de resumen y por último el escalado de datos que permite la normalización de estos para que compartan un mismo valor medio, en la línea 28 se ejecuta el resumen estadístico posterior a la normalización de datos (ver figura 27).

Figura 27.

Escalado de datos en Lenguaje R.

```
23 View(K_means_Barreras)
24 summary(K_means_Barreras)
25
26 #Escalar la data#
27 Barrera_Kmeans = scale(K_means_Barreras, center = T, scale = T)
28 View(Barrera_Kmeans)
29
```

Nota: Elaboración propia.

Una vez realizada la normalización, el siguiente paso fue crear los clústeres. Para ello como se observa previamente en la figura 28, en la línea 32 se fijó la semilla para inicializar datos aleatorios, en la línea 33 se renombró el conjunto de datos para la aplicación de la fórmula K-means y crear el número de centroides para los agrupamientos (ver figura 28).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 28.
Creación de clústeres en Lenguaje R.

```
30 #Crear cluster#
31
32 set.seed(12345)
33 kmcluster = kmeans(Barrera_Kmeans, centers = 2, nstart = 50)
34 kmcluster
35
36 kmcluster$centers #centroides de los clusters
37 kmcluster$tot.withinss #suma de los cuadrados de las distancias
38
39 K_means_Barreras$cluster = kmcluster$cluster
40 K_means_Barreras$cluster
41 kmcluster$cluster
42
```

Nota: Elaboración propia.

De igual forma en la figura 28 se presentan las líneas 36 a la 40 que permiten la visualizar diferentes parámetros generados por el algoritmo K-means como la generación de centroides, suma de los cuadrados de las distancias y los clústeres resultantes.

Una vez creados los clústeres se pueden elaborar las gráficas para su visualización. En este sentido en la figura 29 se presentan las líneas 45 y 49 que generan gráficos de dispersión de las agrupaciones y también con variables individuales como EDAD en el eje “x” y la variable P.27 en el eje “y”.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 29.
Graficar clústeres en Lenguaje R.

```
43 #Averiguar el criterio de separacion#
44
45 ggplot(K_means_Barreras, aes(EDAD, P.27))+geom_point(aes(color = as.factor(clu
46
47 #Grafico 2 principales componentes#
48
49 fviz_cluster(kmcluster, Barrera_Kmeans)
50 kmcluster$tot.withinss
51
```

Nota: Elaboración propia.

Hasta el momento se ha presentado el proceso para ejecutar el algoritmo K-means a través de líneas de comando en la herramienta, con ello se logró realizar el agrupamiento de los datos. Con el procedimiento descrito se realizaron pruebas modificando el número de centroides. Sin embargo, también se elaboraron líneas adicionales para la búsqueda del número óptimo de clústers para apoyar la elaboración de resultados posteriores (ver figura 30).

Figura 30.
Métodos para encontrar el número óptimo de clústeres en Lenguaje R.

```
52 #Numero optimo de clusters#
53
54 TW = rep(NA, 8)
55 for( i in 1:8){
56   km = kmeans(Barrera_Kmeans, center = i, nstart = 50)
57   TW[i] = km$tot.withinss
58 }
59 TW
60 ggplot(data.frame(TW, nclust = seq(1:8)), aes(nclust, TW)) +
61   geom_line()
62
63 #Segundo metodo para encontrar el optimo de clusters#
64
65 matriz_dist = get_dist(Barrera_Kmeans, method = "euclidean")
66 fviz_nbclust(Barrera_Kmeans, FUNcluster = kmeans, k.max = 8, diss = matriz_dis
```

Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Como se presentó previamente en la figura 30, el primer método ocupado está escrito en las líneas 54 a la 61 que comprenden al método de la suma de cuadrados, también conocido como método del codo. Por otro lado, las líneas 65 y 66 se presenta la elaboración de una matriz de distancia y la representación gráfica con el método de silueta.

Por lo anterior, se procede nuevamente a generar líneas de comando para representar los clústeres partiendo del número óptimo representado en las líneas 68 a la 70 y con esto se termina el proceso de aplicación del algoritmo K-means (ver figura 31).

Figura 31.

Creación de nuevos clústeres con el número óptimo.

```
67
68 kmcluster = kmeans(Barrera_Kmeans, centers = 3, nstart = 50)
69 fviz_cluster(kmcluster, Barrera_Kmeans)
70 ggplot(K_means_Barreras, aes(EDAD, cluster)) + geom_point(aes(color = as.factor
71
72
```

Nota: Elaboración propia.

El siguiente algoritmo que queda por aplicar en esta herramienta es Apriori, para ello se ejecutaron un total de 31 líneas que conforman la construcción de las reglas de asociación. Las primeras líneas de ejecución corresponden a la descarga de las librerías, de las cuales son arules siendo este el que ejecuta el algoritmo A priori, *readxl* para la lectura de archivos tipo Excel, *plyr* para elaborar transacciones, *arulesViz*, *tidyverse* y *ggplot2* para una mejor presentación de las reglas de asociación (ver figura 32).

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 32.

Descarga de librerías del algoritmo Apriori en Lenguaje R.

```
1 library(arules)
2 library(readxl)
3 library(plyr)
4 library(arulesViz)
5 library(tidyverse)
6 library(ggplot2)
7 Archivo <- read_excel("C:\\Users\\wendy\\OneDrive\\Documentos\\")
```

Nota: Elaboración propia.

Una vez realizada la importación del *dataset*, en la línea diez de la figura 33 se muestra cómo se aplicó la función *mutate* para delimitar la pregunta *target* (objetivo) y convertirla en tipo factor, esto permitió elaborar en la línea catorce las transacciones que fueron analizadas con el algoritmo puesto que Apriori trabaja partiendo de las transacciones que se construyen con el *dataset*.

Figura 33.

Elaboración de Transacciones en Lenguaje R.

```
9 #Convertir la variable en caracter#
10 Archivo <- mutate(Archivo, P.1 = as.factor(P.1))
11 View(Archivo)
12
13 #Transformar los datos en transacciones#
14 transcciones <- ddply(Archivo, c("GEN", "EDAD", "P.1", "P.3", "P.8", "P.9",
15                               "P.10", "P.11", "P.12", "P.13", "P.14", "P.15", "P.16", "P.17", "P.18", "P.19", "P.20", "P.21", "P.22", "P.23", "P.24", "P.25", "P.26", "P.27", "P.28", "P.29", "P.30", "P.31", "P.32", "P.33", "P.34", "P.35", "P.36", "P.37", "P.38", "P.39", "P.40", "P.41", "P.42", "P.43", "P.44", "P.45", "P.46", "P.47", "P.48", "P.49", "P.50", "P.51", "P.52", "P.53", "P.54", "P.55", "P.56", "P.57", "P.58", "P.59", "P.60", "P.61", "P.62", "P.63", "P.64", "P.65", "P.66", "P.67", "P.68", "P.69", "P.70", "P.71", "P.72", "P.73", "P.74", "P.75", "P.76", "P.77", "P.78", "P.79", "P.80", "P.81", "P.82", "P.83", "P.84", "P.85", "P.86", "P.87", "P.88", "P.89", "P.90", "P.91", "P.92", "P.93", "P.94", "P.95", "P.96", "P.97", "P.98", "P.99", "P.100"),
15                               function(df1) paste(df1$P.1, collapse = ","))
16
```

Nota: Elaboración propia.

Asimismo, para la elaboración de las transacciones se deben considerar todas las variables que se desean analizar y estructurarlas en columnas para realizar la exportación de un nuevo archivo en formato CSV con las configuraciones previas, la exportación debe

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

ser en una ruta que sea fácil de localizar porque el siguiente paso es la lectura de las transacciones, pero con el nuevo archivo (ver figura 34).

Figura 34.

Exportación de transacciones generadas en Lenguaje R.

```
17 #Renombrar las columnas#
18 colnames(transcciones) <- c("GEN", "EDAD", "P.1", "P.3", "P.8", "P.9", "F
19
20 #Exportar el archivo en CSV#
21 write.csv(transcciones, file = "C:\\Users\\wendy\\OneDrive\\Documen
22
23 #Leer el archivo#
24 tr <- read.transactions(file = "C:\\Users\\wendy\\OneDrive\\Documen
25 summary(tr)
26
```

Nota: Elaboración propia.

Las últimas cuatro líneas que se generaron son el algoritmo Apriori, en la línea 28 de la figura 35 se establecen los parámetros de evaluación como el nivel de soporte y confianza de las reglas. Seguido de la línea 29 que ejecuta la función *summary* para generar todas las reglas que pueda haber con las configuraciones previas. En la línea 30 se encuentra la función *inspect* que visualiza la reglas en la consola y con la función *inspectDT* se visualizan las reglas con un formato más visual.

Figura 35.

Parametros del algoritmo Apriori Lenguaje R.

```
27 #Aplicar algoritmo Apriori#
28 reglas.asociacion <- apriori(tr,parameter = list(supp=0.01, conf=1.0
29 summary(reglas.asociacion)
30 inspect(reglas.asociacion[1:20])
31 inspectDT(reglas.asociacion[1:100])
32 |
```

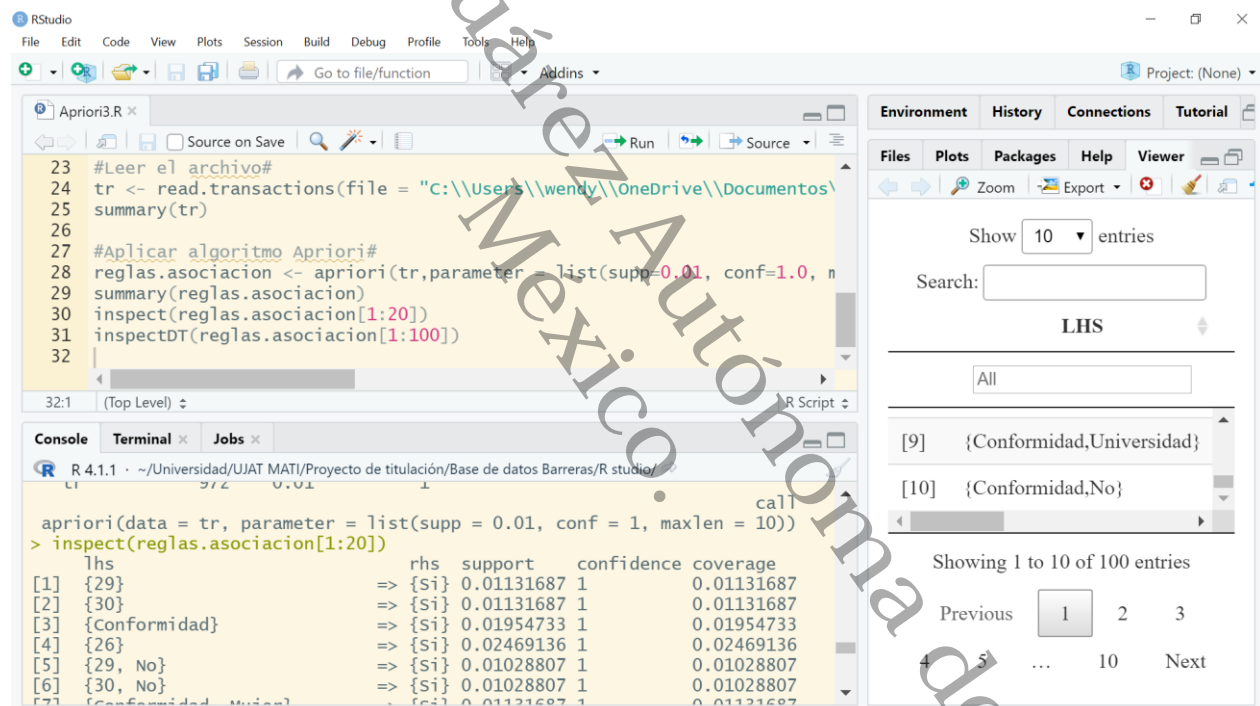
Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En la figura 36 se presenta la visualización de las reglas de asociación, se puede apreciar que en el cuadro inferior izquierdo que a nivel de consola son visibles las reglas de asociación con la función *inspect*, en cuanto el recuadro derecho con la función *inspectDT* son visibles las reglas, pero con un formato una mejor visualización para la comprensión de resultados.

Figura 36.
Visualización de reglas de asociación en Lenguaje R.



The screenshot shows the RStudio interface. The script editor contains the following code:

```
23 #Leer el archivo#
24 tr <- read.transactions(file = "C:\\Users\\wendy\\OneDrive\\Documents\\
25 summary(tr)
26
27 #Aplicar algoritmo Apriori#
28 reglas.asociacion <- apriori(tr,parameter = list(supp=0.01, conf=1.0, n
29 summary(reglas.asociacion)
30 inspect(reglas.asociacion[1:20])
31 inspectDT(reglas.asociacion[1:100])
32
```

The console output shows the results of the *inspect* function:

```
apriori(data = tr, parameter = list(supp = 0.01, conf = 1, maxlen = 10))
> inspect(reglas.asociacion[1:20])
  lhs      rhs support confidence coverage
[1] {29}    => {Si} 0.01131687 1      0.01131687
[2] {30}    => {Si} 0.01131687 1      0.01131687
[3] {Conformidad} => {Si} 0.01954733 1      0.01954733
[4] {26}    => {Si} 0.02469136 1      0.02469136
[5] {29, No} => {Si} 0.01028807 1      0.01028807
[6] {30, No} => {Si} 0.01028807 1      0.01028807
[7] {Conformidad, No} => {Si} 0.01131687 1      0.01131687
```

The Environment pane shows the output of the *inspectDT* function:

```
Files Plots Packages Help Viewer
Zoom Export
Show 10 entries
Search:
LHS
All
[9] {Conformidad,Universidad}
[10] {Conformidad,No}
Showing 1 to 10 of 100 entries
Previous 1 2 3
... 10 Next
```

Nota: Elaboración propia.

Con todo lo anteriormente presentado con la ejecución de los algoritmos en cada una de las herramientas se puede mostrar el proceso de minería de datos realizado en este paso. La aplicación de las técnicas de minería de datos es el apartado más amplio dentro del proceso KDD y de los más importantes puesto que es la etapa donde se busca

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

identificar patrones y obtener información nueva y relevante. De igual manera al ser aplicado en tres herramientas distintas permite ilustrar los diferentes procesos que se deben llevar a cabo para la visualización de los resultados.

3.3.7 Interpretación

Teniendo en cuenta todos los pasos anteriores se puede efectuar la interpretación de los patrones descubiertos, mismos que deben ser precisos, comprensibles e interesantes. Para el descubrimiento de los patrones de conducta fue necesario retomar varios de los pasos anteriores, por ejemplo, se tuvo que realizar la transformación y proyección de los datos de diferentes formas para una mejor comprensión de los gráficos, asimismo la ejecución de los algoritmos se repitió en diversas ocasiones debido a los errores o advertencias que señalaban las herramientas empleadas, en ese sentido, las diferentes ejecuciones de algoritmos condujeron a la búsqueda de soluciones para un correcto uso. Por otro lado, la interpretación de los datos también permite visualizar los patrones extraídos de modo que si se detecta información redundante se proceda a eliminar patrones repetidos.

Por lo anteriormente expuesto, se determina que este paso en particular se lleva a cabo un proceso de experimentación que resulta de cada una de las pruebas y errores generados. Todo esto con la finalidad de obtener información que sea comprensible para realizar la última etapa del proceso y también para el análisis de resultados por parte de los usuarios finales.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

La interpretación de los datos debe presentar un análisis sobre los hallazgos descubiertos con el uso de las herramientas y los resultados obtenidos. Estos resultados se dividen en tres aspectos principales:

- Pruebas realizadas por herramienta de minería de datos.
- Análisis de usabilidad de las herramientas implementadas.
- Interpretación de la información obtenida.

Sin embargo, la interpretación y análisis se describe con detalle en el capítulo de resultados donde se presentan los hallazgos obtenidos de la aplicación de minería de datos.

3.3.8 Uso del conocimiento descubierto

De acuerdo con Fayyad, Piatetsky-Shapiro & Smyth (1996), la aplicación de las técnicas de minería de datos puede ser del tipo predictiva cuando interesa principalmente el rendimiento de los algoritmos o su poder de predicción, o puede ser con fines descriptivos centrándose en la interpretación de los modelos generados y la relación existente entre las variables de estudio. Asimismo, se puede dar el caso de que una investigación combine ambos propósitos.

2

Por lo anterior, se puede decir que en esta investigación impera una tendencia con fines descriptivos, es en este sentido que cuando se llega a esta etapa del proceso es porque se obtuvo un producto que será de utilidad para la toma de decisiones de la institución.

2

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

2 En este caso se pretende reportar la información obtenida del análisis realizado y que sea de utilidad para las personas interesadas dentro de la organización, sobre la problemática planteada referente al uso de innovaciones tecnológicas en estudiantes universitarios y cómo se relacionan las barreras presentadas en esta investigación ante la situación social que se vive con las clases virtuales por la pandemia COVID-19. Con el conocimiento descubierto acerca de los patrones identificados, la institución académica puede optar por implementar medidas de mitigación y estrategias educativas ante la situación planteada en beneficio de los estudiantes que forman parte de la comunidad universitaria.

En este capítulo del desarrollo metodológico, se mostraron tres apartados, el primero consiste en la exploración del *dataset*, seguido de la selección de herramientas de minería de datos a emplear, determinando que las herramientas incorporadas fueron *Orange*, *WEKA* y Lenguaje R. Por último, el apartado del proceso KDD, que describe la aplicación de las herramientas seleccionadas y el procedimiento realizado para la generación de reportes correspondientes que serán incluidos en los resultados, haciendo uso de los modelos de minería de datos: agrupación, clasificación y análisis de enlaces.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Capítulo 4. Pruebas y Resultados

Una vez concluida la obtención de los datos y el análisis de la información, se procedió a la interpretación de resultados. Si bien en el proceso metodológico se consideraron los últimos dos pasos del proceso KDD referentes a la interpretación y uso del conocimiento descubierto, en el capítulo de resultados se exponen con detalle puesto que en ellos se describen las pruebas realizadas para obtener modelos funcionales y que durante la interpretación de los patrones detectados permitieron responder la pregunta de investigación de este proyecto, así como detectar hallazgos con información novedosa y poco obvia.

En ese sentido, este capítulo está dividido en dos apartados principales, el primero consta de las pruebas realizadas en Orange, WEKA y Lenguaje R. En cuanto al segundo apartado, se considera la interpretación de los modelos y los patrones descubiertos. Asimismo, se incorpora en este capítulo un análisis de usabilidad de las herramientas empleadas para minería de datos y que fueron seleccionadas previamente del análisis FURPS elaborado por Guzmán (2022) y retomado en este proyecto con el fin emitir una recomendación.

4.1 Pruebas realizadas por herramienta

En este apartado se presentan las pruebas realizadas por herramienta para la construcción de modelos de clasificación, agrupamiento y análisis de enlaces (reglas de asociación). En ese mismo orden se encuentra presentada la información y cabe

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

mencionar que se presentan tres pruebas por algoritmo empleado con cada uno de los modelos de minería de datos.

4.1.1 Orange

La primera herramienta presentada es *Orange*, con esta herramienta se realizaron diferentes pruebas de usabilidad para conocerla y también para interactuar con la aplicación de las técnicas de clasificación, agrupamiento y reglas de asociación con diferentes algoritmos y visualizadores gráficos.

En el caso de los árboles de decisión, se realizaron pruebas con la función *Tree* y *Random Forest* utilizando como visualizadores gráficos *Tree Viewer* y *Pythagorean Tree*.

Para el caso del modelo de clasificación, se aplicaron los algoritmos *k-Means* y *kNN*, para su representación visual se emplearon diferentes *widgets* como *Radviz*, *FreViz*, *Scatter Plot* y *t-SNE*.

Para el caso de las reglas de asociación, además de probar con el algoritmo *Association Rules*, se usó *Frequent Itemsets* y *CN2 Rule Induction* en este último caso el propósito ya no es asociar reglas sino la inducción.

La aplicación de estos algoritmos tuvo como objetivo únicamente de conocer el entorno de trabajo de la herramienta sin embargo, permitió conocer otras funciones. Aunque existe una variedad de *widgets* en la herramienta, no quiere decir que todas son apropiadas para el caso de este proyecto. Esto se debe a la cantidad de información que debe ser procesada, así como las métricas de evaluación y la interpretación.

12	Internet	journalusco.edu.co	<1%
13	Internet	revistas.upr.edu	<1%
14	Internet	ri.ujat.mx	<1%
15	Internet	www.rexe.cl	<1%
16	Publicación	Deepti Aggarwal, Sonu Mittal, Vikram Bali. "Identifying Non-Performing Students..."	<1%
17	Internet	repositorio.konradlorenz.edu.co	<1%
18	Internet	scripta.up.edu.mx	<1%
19	Internet	dpi-proceedings.com	<1%
20	Internet	scielo.sld.cu	<1%
21	Internet	repositorio.usmp.edu.pe	<1%
22	Internet	atualizacoes.emnuvens.com.br	<1%
23	Publicación	Giselle Mardones Soto. "influencia del clima escolar en el aprendizaje", Revista Re...	<1%
24	Internet	revistahorizontes.org	<1%
25	Internet	sedici.unlp.edu.ar	<1%

1

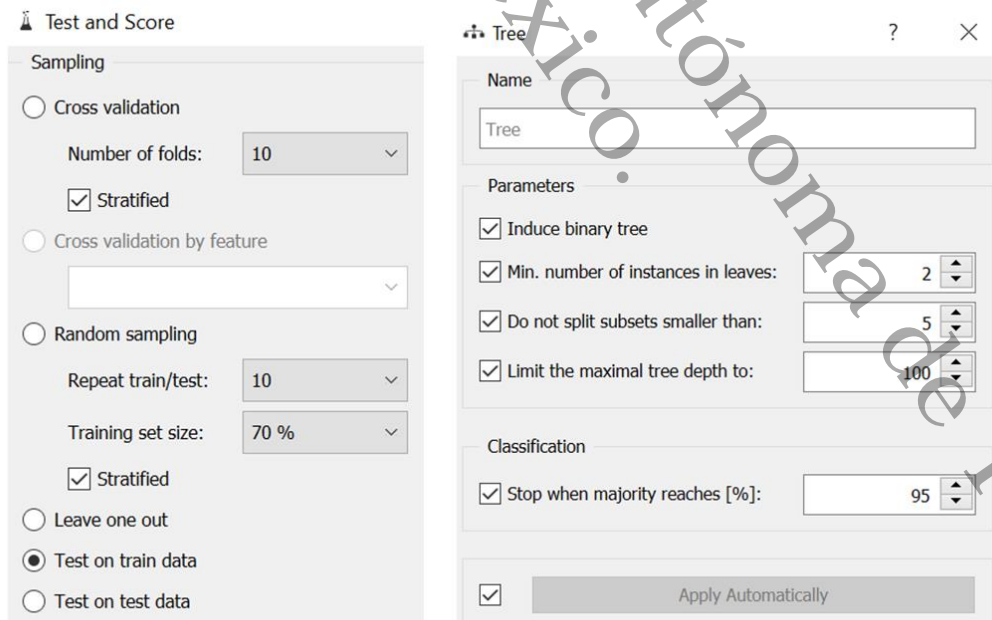
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Clasificación.

para la aplicación de pruebas con la herramienta Orange, se realizaron diferentes pruebas con las variables para seleccionar la pregunta raíz del árbol de decisión. En este sentido, se buscó emplear los ítems con mayor ganancia de información y mejores resultados obtenidos en el entrenamiento y test. Durante las pruebas se determinó que las variables P.1 ¿El lugar donde vives cuenta con internet?, P.3 ¿Debes ir a rentar internet para entrar a las plataformas de clases? y P.7 ¿Cuentas con una tableta para tu uso personal?, fueron los ítems con mejores métricas de evaluación que se presenta en lo subsecuente.

Figura 37.

Parámetros de inicio del árbol de decisión



Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Como se presenta previamente en la figura 37, además de la selección de las preguntas raíz o *target*, se realizaron configuraciones de inicio para generar el árbol de decisión, en este caso la configuración que se dejó por defecto fue el test de entrenamiento en un 70%, número mínimo de instancias en los nodos hoja y profundidad del árbol. Cabe mencionar que los parámetros establecidos desde el inicio se aplicaron a las tres pruebas realizadas.

La primera prueba realizada fue con la variable objetivo P.1, en la figura 38 se presentan las métricas del modelo del árbol de decisión como el área bajo la curva (AUC), precisión de clasificación (CA), métrica de precisión y recuperación (F1) y la precisión y proporción de verdaderos positivos. En ese sentido, la precisión de clasificación se puede interpretar en porcentajes obteniendo un 97.1% de exactitud.

Figura 38.

Resultados de métricas de evaluación en árbol de decisión con pregunta raíz P. 1

Evaluation Results					
Model	AUC	CA	F1	Precision	Recall
Tree	0.963	0.971	0.981	0.971	0.991

Nota: Elaboración propia.

En cuanto a la correcta clasificación de las variables se puede evaluar con la matriz de confusión que se presenta en la figura 39, se puede observar que los porcentajes en las casillas azules representan los datos clasificados correctamente, en este caso para el ítem “No” fue en un 90.7% y para el ítem “Sí” en un 99.1%, con estos resultados se interpreta que el modelo construido con la variable P.1 es acertado.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

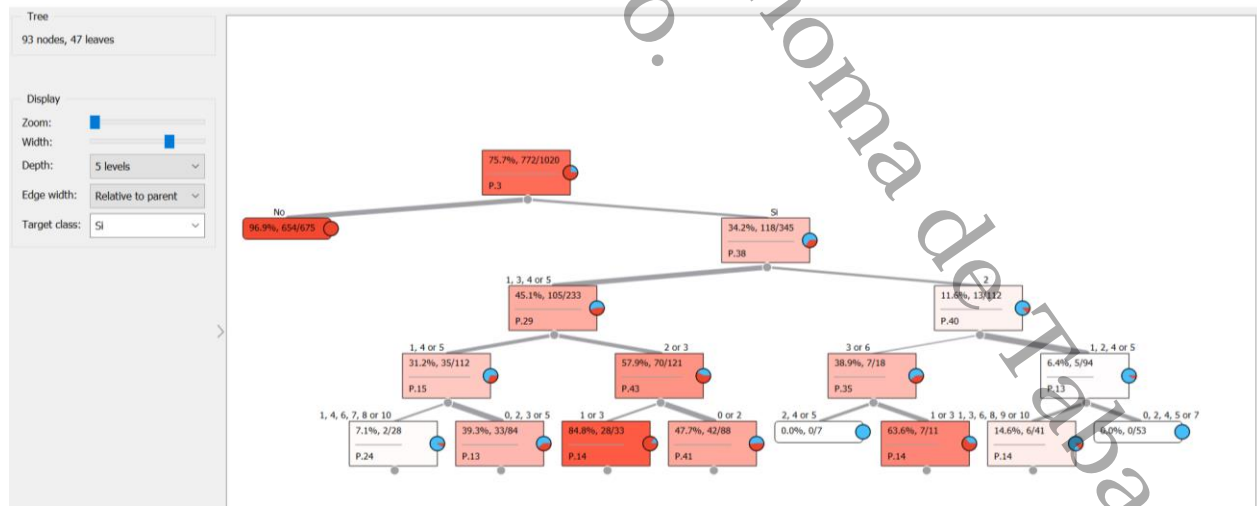
Figura 39. Matriz de árbol de decisión con pregunta raíz P.1

		Predicted		Σ
		No	Si	
Actual	No	90.7 %	9.3 %	248
	Si	0.9 %	99.1 %	772
Σ		232	788	1020

Nota: Elaboración propia.

Después de evaluar esta primera prueba de entrenamiento, la figura 40 muestra el diagrama obtenido. El árbol de decisión que generó con la variable P.1 cuenta con 93 nodos y 47 niveles, cabe mencionar que este diagrama se encuentra clasificado de acuerdo con el ítem "Si" que resulta ser el clasificador objetivo en la configuración.

Figura 40. Prueba de árbol de decisión con pregunta raíz P.1



Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En cuanto a la profundidad del árbol, Orange permite configurar los niveles para una mejor representación visual, permitiendo que el usuario determine si quiere ver todas las ramificaciones posibles o solo algunos niveles. En este caso el diagrama mostrado en la figura 40 visualiza cinco niveles de profundidad, los demás niveles que no se muestran no se eliminan, únicamente se ocultan.

La segunda prueba realizada con el modelo de clasificación es con la variable P.3, al igual que en el primer ejemplo los parámetros de inicio no fueron modificados y el modelo de entrenamiento se realizó con el 70% del conjunto de datos, los resultados obtenidos con esta variable se presentan en la figura 41, en esta se observa que el porcentaje de las métricas no muestra un cambio significativo, continúan siendo métricas mayores al 90%, en el caso de la precisión de clasificación es de 97% con la variable P.3 a diferencia de la variable P.1 que obtuvo un 97.1% de exactitud.

Figura 41.

Resultados de métricas de evaluación en árbol de decisión con pregunta raíz P.3

Model	AUC	CA	F1	Precision	Recall
Tree	0.983	0.970	0.969	0.970	0.970

Nota: Elaboración propia.

Como segunda métrica de evaluación se presenta la matriz de confusión para determinar el porcentaje de clasificación correcta, se observa en la figura 42 que el porcentaje de datos correctamente clasificados con el ítem “No” es del 98.8%, en cuanto al ítem “Sí” es del 91.9%, a diferencia de la variable P.1 en esta prueba el mayor porcentaje de clasificación lo obtuvo el ítem “No”. Sin embargo, con esta información se

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

puede interpretar que el emplear la variable P.3 como pregunta raíz también genera un modelo de clasificación correcto.

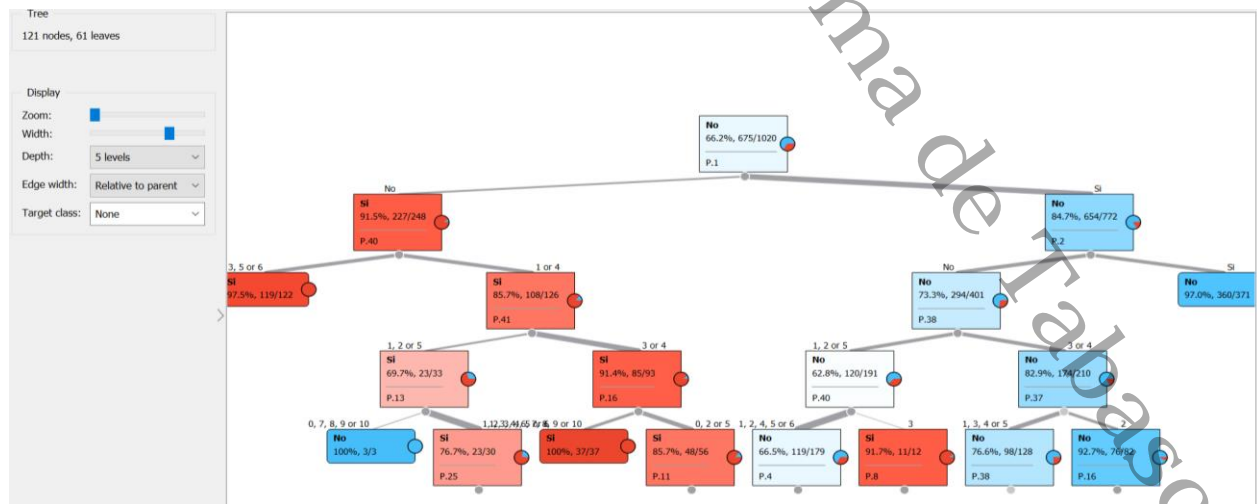
Figura 42.
Matriz de árbol de decisión con pregunta raíz P.3

		Predicted		Σ
		No	Si	
Actual	No	98.8 %	1.2 %	675
	Si	8.1 %	91.9 %	345
Σ		695	325	1020

Nota: Elaboración propia.

Después de analizar las métricas de evaluación se visualizó el diagrama obtenido, el número de nodos resultantes con este árbol fue de 121 con 61 niveles, lo que significa que el nivel de detalle fue aún mayor que en la primera prueba (ver figura 43).

Figura 43.
Prueba de árbol de decisión con pregunta raíz P.3



Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Además, con el propósito de experimentar con las métricas que ofrece Orange, para el caso del segundo diagrama, se optó por no utilizar un clasificador objetivo ya que el hacerlo también es considerado dentro de las métricas de evaluación como ocurrió en el ejemplo anterior. Asimismo, en la figura 43 previamente presentada, se aprecian las clasificaciones realizadas en el modelo con el ítem “Sí” en color rojo y el ítem “No” en color azul. De modo que es más claro el flujo de decisiones que se expone en las ramificaciones.

La tercera prueba realizada con los árboles de decisión fue con la variable P.7 como pregunta raíz, el test de entrenamiento se realizó con el 70% de los datos y los resultados obtenidos de la evaluación se presentan en la figura 44. En esta última prueba se observa que sí hubo una mayor variación en las métricas, en el caso del área bajo la curva (AUC) pasó de 96.3% y 98.3% a 86.9% lo cual fue una disminución considerable en esta última variable, de igual forma en el porcentaje de precisión con las variables P.1 y P.3 oscilaba el 97% y con la variable P.7 la precisión es de 96.4%.

Figura 44.

Resultados de métricas de evaluación en árbol de decisión con pregunta raíz P.7

Evaluation Results					
Model	AUC	CA	F1	Precision	Recall
Tree	0.869	0.965	0.961	0.964	0.965

Nota: Elaboración propia.

Si bien los porcentajes de evaluación obtenidos en las tres variables tienen porcentajes de precisión aceptables, también es cierto que el realizar estas pruebas

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

permite conocer cuál es la mejor variable por considerar para la interpretación de resultados de este modelo.

Con respecto a la validación del árbol de decisión se presenta la siguiente matriz de confusión en la figura 45, se observa que el ítem "No" se encuentra clasificado correctamente en un 99.7% sin embargo, el ítem "Si" tiene una clasificación correcta del 61.2%. A pesar de que en la variable P.7 son pocas las respuestas para el ítem "Sí" aún así no logró hacer una clasificación correcta, ocasionando que se presente confusión ya que erróneamente el 38.8% fue clasificado en el ítem "No".

Figura 45. Matriz de árbol de decisión con pregunta raíz P.7

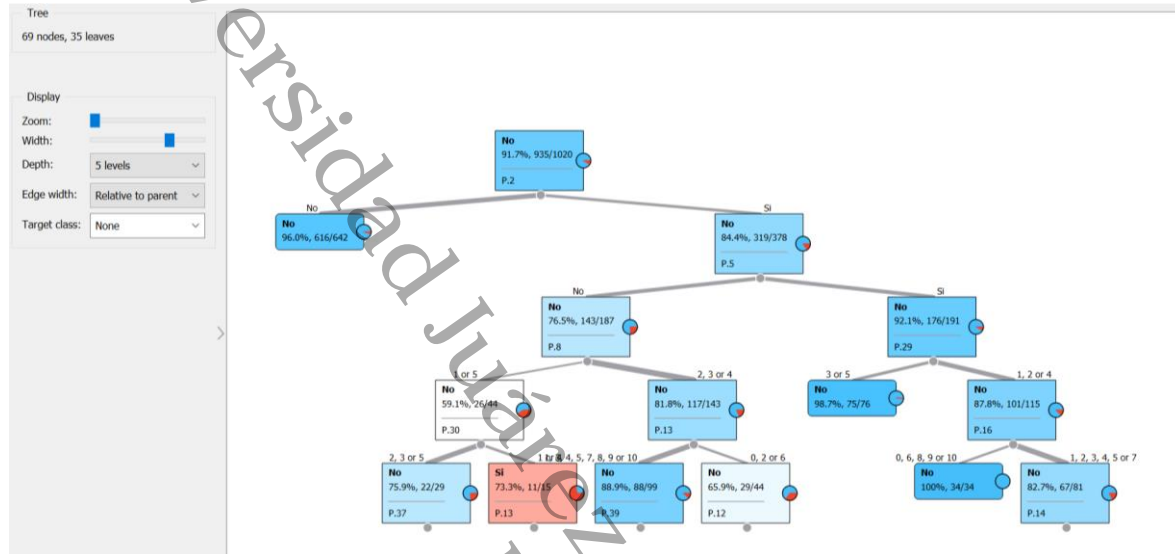
		Predicted		Σ
		No	Si	
Actual	No	99.7 %	0.3 %	935
	Si	38.8 %	61.2 %	85
Σ		965	55	1020

Nota: Elaboración propia.

Como parte de la tercera prueba realizada, en la figura 46 se presenta el diagrama de árbol generado con un total de 69 nodos y 35 niveles. De las pruebas presentadas en este apartado es el diagrama con menos flujo de información. El tener menos información en un árbol también puede ocasionar que se pierda información relevante.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 46.
Prueba de árbol de decisión con pregunta raíz P.7



Nota: Elaboración propia.

Como se observa previamente en la figura 46, aunque el diagrama no se realizó con una clasificación objetivo, en la información obtenida predomina el ítem clasificador “No”. Esto se debe a que la mayoría de los datos en la variable P.7 se encuentran clasificados con este ítem tal y como se mostró previamente en la matriz de confusión, lo cual influye en gran medida al momento de generar el flujo de decisiones.

Como se presentó previamente, aunque las tres variables contempladas para las pruebas obtuvieron resultados favorables en las métricas, los resultados obtenidos al momento de generar los diagramas son distintos. Este aspecto es importante considerarlo al momento de decidir cuál será el nodo raíz para la construcción del modelo porque de ello depende la identificación de los patrones en este trabajo.

1

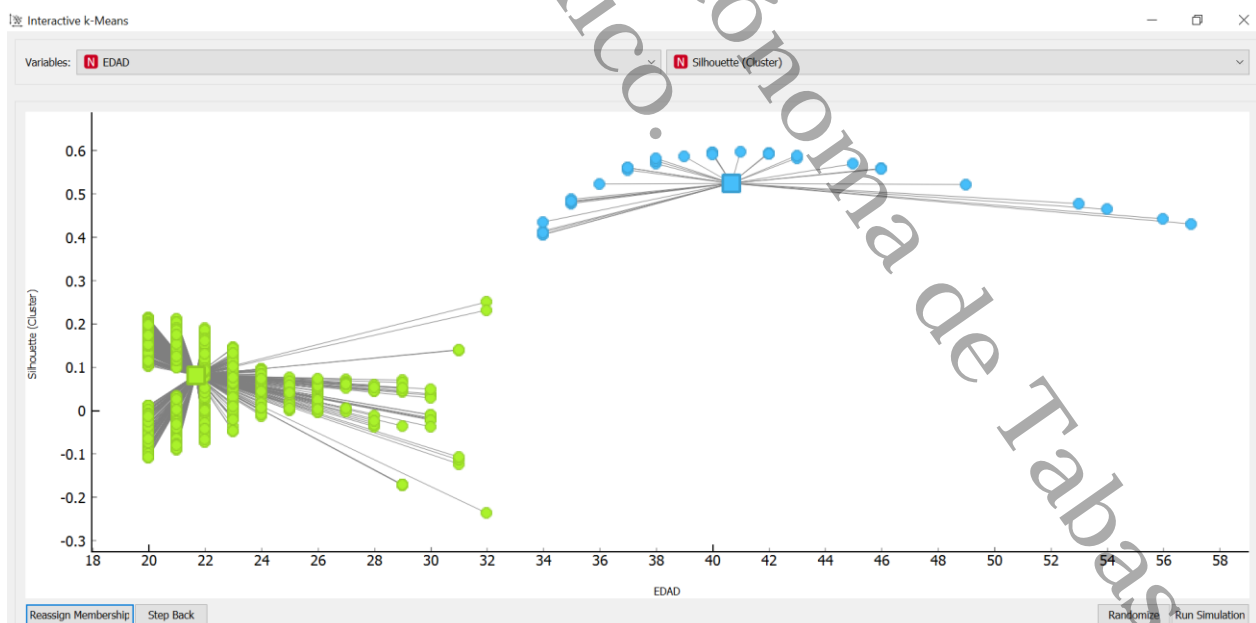
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Agrupación

El siguiente modelo de minería de datos es el de agrupación, durante las pruebas realizadas se buscó el número óptimo de clústeres con el que los datos estuviesen mejor representados, se realizaron diferentes pruebas con el algoritmo *k-Means* sin embargo, en este apartado se presentan tres pruebas con diferentes agrupaciones que fueron consideradas y fueron descartadas en el proceso.

La primera prueba realizada fue con dos agrupaciones, para su elaboración se consideró la aplicación de *Interactive k-Means* que es un *widget* que recalcula los centroides de las agrupaciones de acuerdo con el número de clústeres solicitados (ver figura 47).

Figura 47.
K-means interactivo con dos agrupaciones



Nota: Elaboración propia.

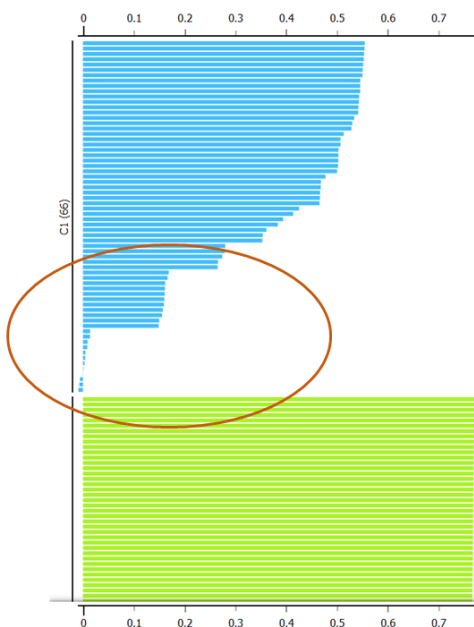
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Como se muestra previamente en la figura 47, las variables consideradas son el clúster y la EDAD, se puede observar que el clúster azul representa a los estudiantes en un rango de edad de 34 a 58 años, mientras que el clúster verde representa a estudiantes de 18 a 33 años. Sin embargo en el grupo verde se destacada una lejanía de algunos puntos al centroide, principalmente en estudiantes de 26 a 33 años, del mismo modo en el grupo azul existe un distanciamiento de los puntos en el rango de edad de 50 a 58 años.

Esto quiere decir que con dos agrupaciones los datos no están correctamente clasificados del todo y se puede ver mejor representado en el diagrama de siluetas (ver figura 48).

Figura 48.

Diagrama de silueta con dos clústeres en Orange



Nota: Elaboración propia.

1

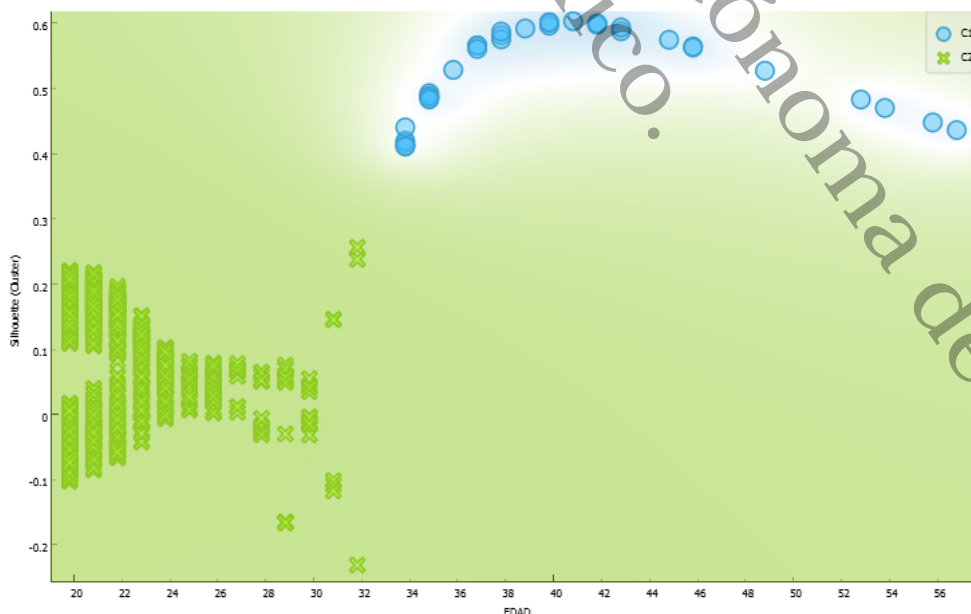
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Claramente se observa en el diagrama de siluetas de la figura 48, que el clúster uno de color azul contiene datos que no se lograron clasificar bien y se encuentran resaltados dentro de un óvalo. En esta primera prueba el puntaje de agrupación alcanzado fue de 0.70, el diagrama de siluetas lo que permite es conocer si los datos se encuentran correctamente clasificados y más cercano al puntaje de agrupación. Es decir, cuanto más cercanos se encuentren los datos al cero o al -1, más erróneo será el clúster.

En la figura 49 se presenta un gráfico de puntos que permite ver la distribución de los clústeres. Se observa que dentro del rango de 26 a 32 años existen un distanciamiento del clúster uno y se acerca más al clúster dos, esto quiere decir que aún se puede hacer un mejor agrupamiento con un mayor número.

Figura 49.

Gráfica Plot en Orange con dos agrupaciones



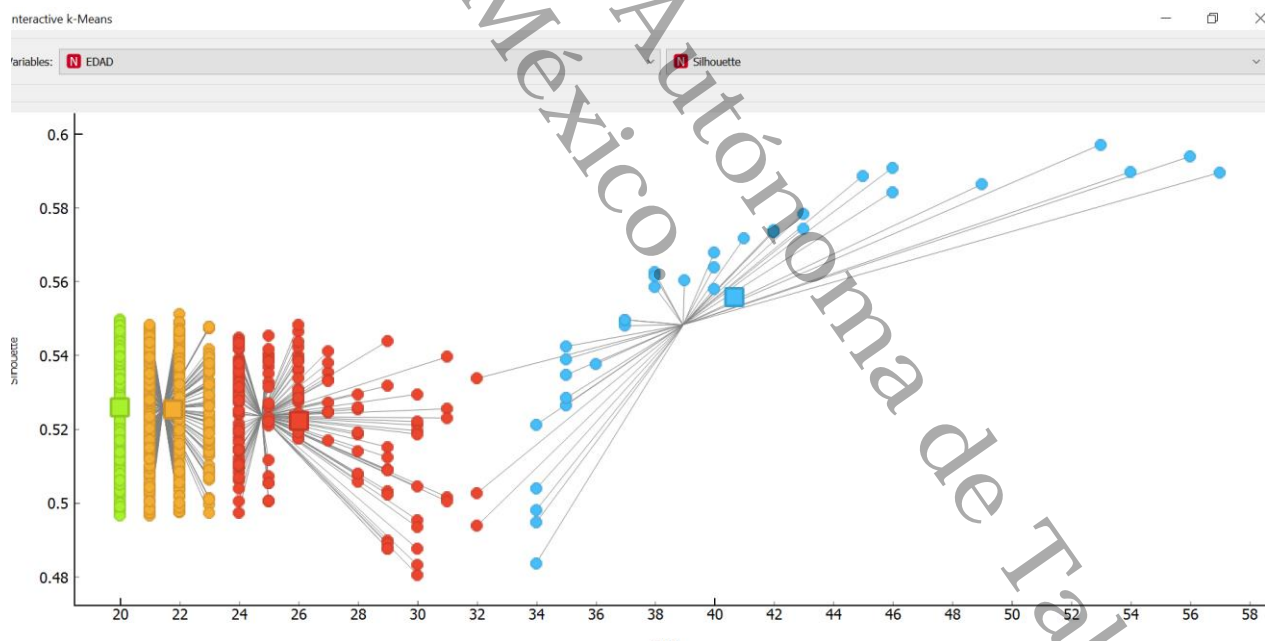
Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Por consiguiente se presentan otras dos pruebas con más agrupaciones, las siguientes pruebas también fueron realizadas con la variable EDAD puesto que se determinó que con esta variable las agrupaciones en *Orange* son más claras.

La segunda prueba que se presenta es con cuatro agrupaciones, en la figura 50 se observa el gráfico interactivo de *k-Means*, el primer clúster en color azul representa a los estudiantes en un rango de edad de 34 a 56 años, el clúster rojo son estudiantes de 24 a 32 años, el naranja pertenece a estudiantes de 21 a 23 años y el verde son estudiantes de 20 años.

Figura 50.
K-means interactivo con cuatro agrupaciones



Nota: Elaboración propia.

De acuerdo con la figura 50 se hicieron más agrupaciones en estudiantes de menores a 30 años, de igual forma esa segmentación de grupos se encuentra muy

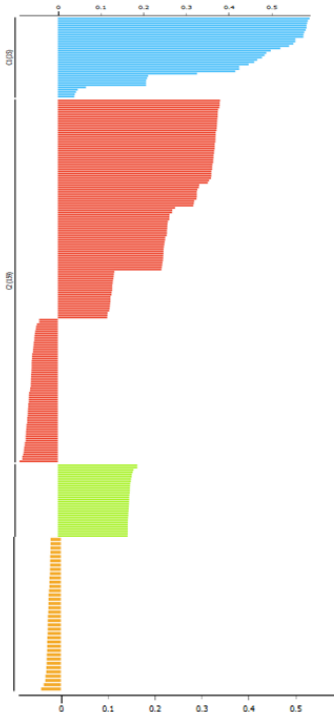
1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

cercana y podría ser parte de una misma agrupación. Para esta prueba realizada el puntaje de agrupación obtenido fue 0.60, es aún menor que el modelo de dos clústeres. En la figura 51 se muestra la gráfica de silueta de acuerdo con el puntaje alcanzado, se observa que el clúster uno en color azul quedó correctamente clasificado sin embargo, los clústeres 2, 3 y 4 carecen de una correcta agrupación, puesto que los datos tienden a aproximarse entre sí.

Figura 51.

Diagrama de silueta con cuatro agrupaciones



Nota: Elaboración propia.

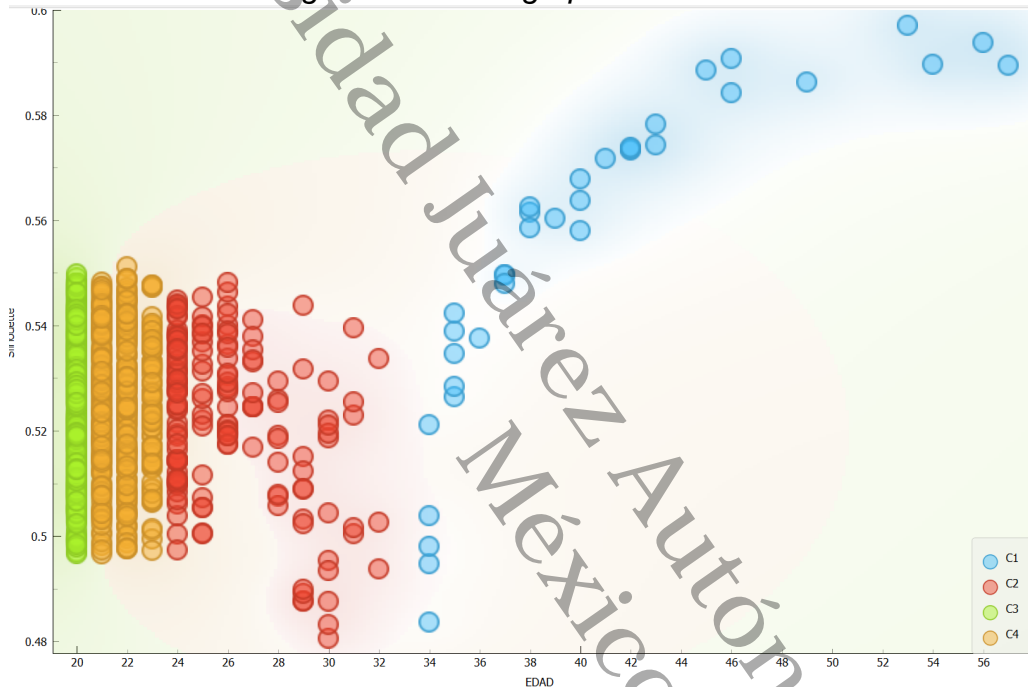
En cuanto al diagrama de dispersión que se presenta en la figura 52, presenta el comportamiento de los clústeres. Se puede observar cómo el clúster uno en color azul se encuentra segmentado sin embargo el clúster dos en color rojo tiene datos muy que están

131

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

muy cercar de los grupos tres y cuatro, pero también tiene un segmento de datos muy disperso que podría formar un grupo distinto.

Figura 52.
Gráfica Plot en Orange con cuatro agrupaciones



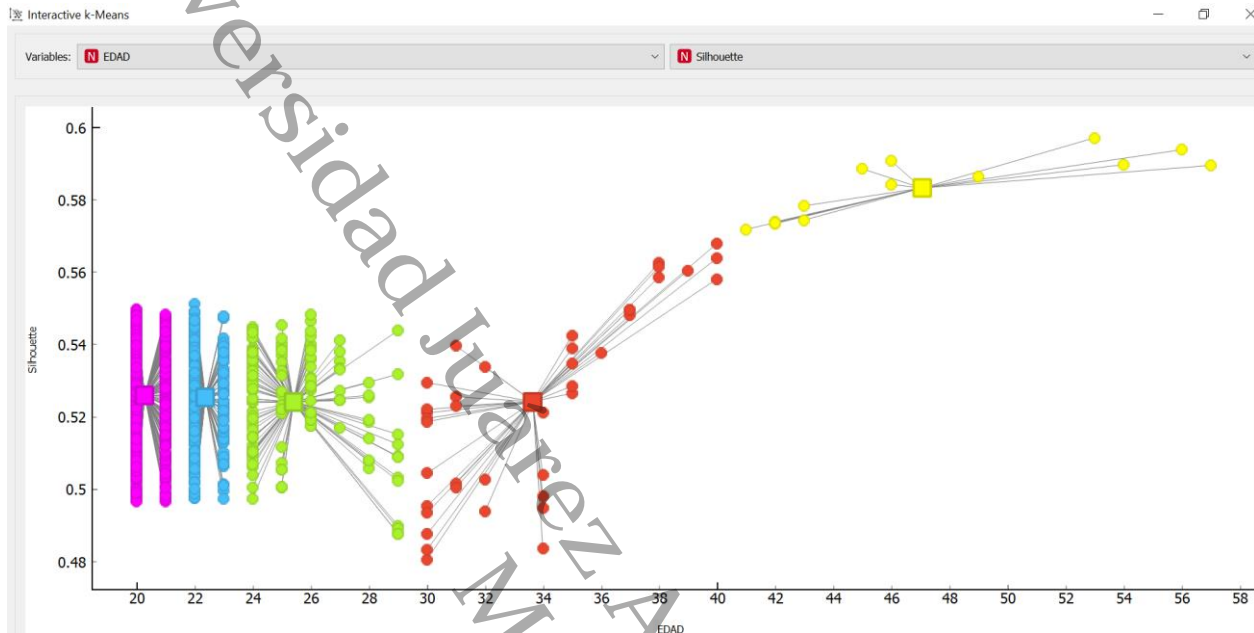
Nota: Elaboración propia.

Hasta el momento se presentó una prueba con dos clústeres que realizó un mejor que agrupamiento que en este segundo modelo de cuatro grupos. Queda por presentar una tercera prueba de agrupación con la herramienta Orange, este ejemplo se realizó con cinco grupos y se inició con la función de *k-Means* interactivo, por lo tanto, la figura 53 presenta el recalcu de centroides para cinco clústeres.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 53.
K-means interactivo con cinco agrupaciones



Nota: Elaboración propia.

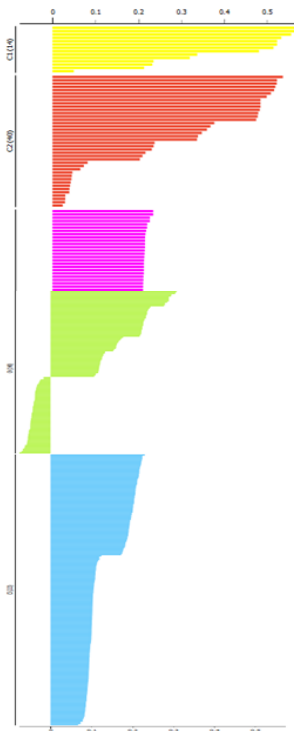
En el diagrama de la figura 53 se presentan los nuevos grupos realizados, aunque los grupos azul y rosa se encuentran muy cercanos, se puede ver una segmentación adicional que modifica las edades en el rango de 24 a 40 años, esto ocasiona que el grupo de color amarillo disminuyera. Aunque en este gráfico se observa un mejor agrupamiento, el puntaje de calidad alcanzado de acuerdo con la métrica de silueta conocida en inglés como *Silhouette* es de 0.55, siendo una puntuación aún menor que las dos primeras pruebas.

Por lo anterior, es importante presentar el diagrama de silueta porque en él se puede visualizar gráficamente si los grupos seleccionados durante tienen la calidad suficiente para ser considerado el modelo a interpretar. En la figura 54 se muestra que

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

aunque existe una agrupación aceptable en algunos grupos, también existe una clasificación incorrecta en el grupo tres de color verde ocasionando una disminución en la calidad de los clústeres.

Figura 54.
Diagrama de silueta con cuatro agrupaciones



Nota: Elaboración propia.

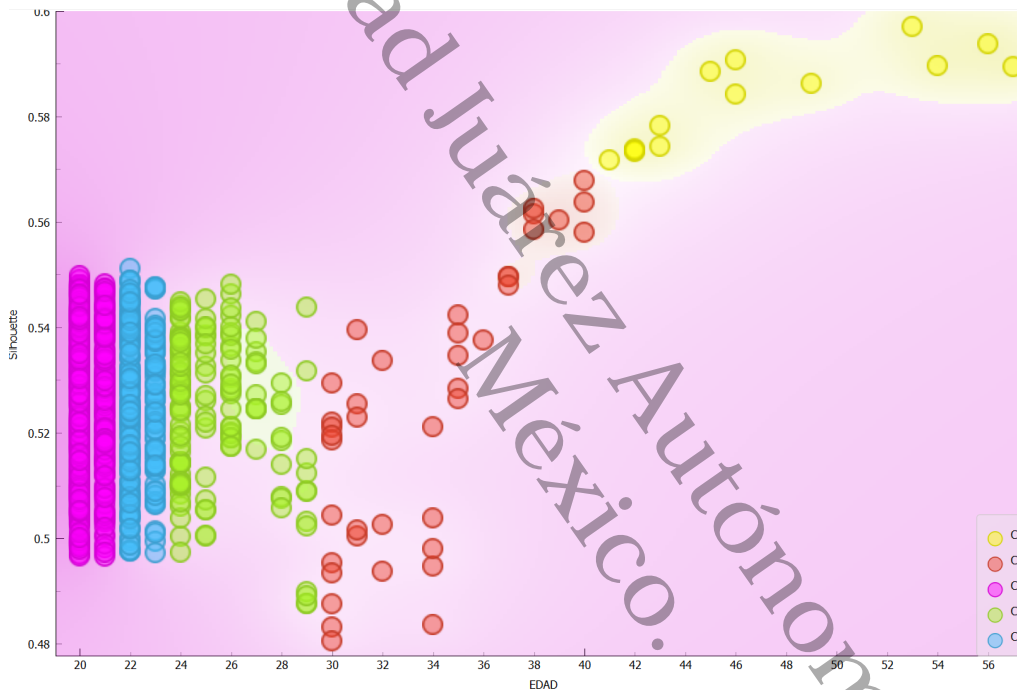
En el diagrama de dispersión representado en la figura 55, se muestra la separación de los clústeres de acuerdo con el modelo de cinco grupos obtenido en esta tercera prueba. Se puede observar que el comportamiento es muy similar al diagrama de *Interactive k-Means* es decir, los grupos tres y cuatro de color rosa y azul respectivamente, se encuentran muy juntos. También se observa que los grupos uno de

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

color amarillo y el grupo dos en rojo se logran diferenciar en cuanto a distanciamiento y por último está el grupo cuatro en color verde que sigue mostrando una relación con los grupos de edad menores a 24 en los estudiantes.

Figura 55.
Gráfica Plot en Orange con cinco agrupaciones



Nota: Elaboración propia.

Por lo anteriormente expuesto, se observa que se realizaron diversas pruebas con esta primera herramienta para agrupar los datos de forma óptima. Se detectaron similitudes en los diagramas principalmente en los grupos menores de 30 años ya que tienden a relacionarse y no permiten una buena separación de los clústeres, lo cual debe considerarse para la selección del número de agrupaciones óptima. En ese sentido

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

también se debe considerar que a mayor número de grupos la métrica de evaluación de calidad tiende a disminuir.

Reglas de asociación.

En este apartado se presentan las pruebas realizadas con la técnica de reglas de asociación. Al realizar la búsqueda Orange muestra el resultado de todas las reglas encontradas en los datos de acuerdo con los parámetros de configuración. En ese sentido, las pruebas que se realizaron fueron con el objetivo de conocer la usabilidad de los parámetros de modo que sea más eficiente la búsqueda de las principales reglas emitidas por la herramienta para la interpretación de los resultados.

La primera prueba se realizó con un soporte mínimo del 10% y un mínimo de confianza del 50% (ver figura 56)

Figura 56.

Reglas de asociación en orange con cincuenta por ciento de confianza

Info	Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	Consequent
Rules: 10000 (shown 10000)	0.393	0.519	0.757	0.531	1.292	0.089		GEN=Mujer, P.3=No
Find association rules	0.393	0.672	0.585	1.095	1.048	0.018	GEN=Mujer	P.1=Si, P.3=No
Min. supp.: 10 %	0.393	0.613	0.641	0.913	1.048	0.018	P.1=Si, P.3=No	GEN=Mujer
Min. conf.: 50 %	0.393	0.978	0.402	1.883	1.292	0.089	GEN=Mujer, P.3=No	P.1=Si
Max. rules: 10k	0.393	0.594	0.662	0.686	1.309	0.093	P.3=No	GEN=Mujer, P.1=Si
<input type="checkbox"/> Induce only classification rules	0.393	0.916	0.429	2.135	0.999	-0.000	GEN=Mujer, P. 42=Sin comentarios	P.7=No
<input type="checkbox"/> Restrict search by below filters	0.393	0.540	0.727	0.735	1.011	0.004	P. 42=Sin comentarios	GEN=Mujer, P.7=No
Find Rules	0.393	0.586	0.671	0.873	1.002	0.001	P.7=No, P. 42=Sin comentarios	GEN=Mujer
Filter by Antecedent	0.393	0.736	0.534	1.361	1.011	0.004	GEN=Mujer, P.7=No	P. 42=Sin comentarios
Contains:	0.393	0.672	0.585	1.146	1.002	0.001	GEN=Mujer	P.7=No, P. 42=Sin comentarios
Items, min: 1 max: 999	0.392	0.778	0.504	1.391	1.110	0.039	P.1=Si, P.7=No, P. 42=Sin comentarios	P.6=Si
Filter by Consequent	0.392	0.585	0.671	0.892	0.978	-0.009	P.7=No, P. 42=Sin comentarios	P.1=Si, P.6=Si
Contains:	0.392	0.895	0.438	2.092	0.976	-0.010	P.1=Si, P.6=Si, P. 42=Sin comentarios	P.7=No
Items, min: 1 max: 999	0.392	0.704	0.557	1.132	1.117	0.041	P.1=Si, P. 42=Sin comentarios	P.6=Si, P.7=No
	0.392	0.539	0.727	0.729	1.016	0.006	P. 42=Sin comentarios	P.1=Si, P.6=Si, P.7=No
	0.392	0.768	0.511	1.334	1.127	0.044	P.6=Si, P. 42=Sin comentarios	P.1=Si, P.7=No
	0.392	0.739	0.530	1.372	1.016	0.006	P.1=Si, P.6=Si, P.7=No	P. 42=Sin comentarios
	0.392	0.576	0.681	0.750	1.127	0.044	P.1=Si, P.7=No	P.6=Si, P. 42=Sin comentarios
	0.392	0.656	0.598	1.121	0.978	-0.009	P.1=Si, P.6=Si	P.7=No, P. 42=Sin comentarios
	0.392	0.518	0.757	0.611	1.120	0.042	P.1=Si	P.6=Si, P.7=No, P. 42=Sin comentarios

Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

De acuerdo con los parámetros seleccionados en la figura 56 previamente presentada se generaron un total de diez mil reglas que cuentan con un lift mayor a uno lo que indica que las reglas mostradas se repiten dentro del conjunto de datos. Sin embargo al disminuir el porcentaje de confianza el parámetro levr tiene métricas negativas que disminuyen la calidad de las reglas obtenidas y pueden dar información poco confiable. En esta primera prueba se generaron demasiadas reglas y no todas tienen información que sea relevante.

La herramienta por defecto está configurada para generar todas las reglas posibles sin filtrarlas, por lo que es necesario hacer diferentes pruebas. En la figura 57 se presenta un nuevo grupo de reglas generadas con un soporte mínimo del 15% y una confianza mínima del 100%.

Figura 57. Reglas de asociación en Orange con soporte mínimo del quince por ciento.

Association Rules

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	Consequent
0.167	1.000	0.167	4.541	1.321	0.041	P.3=No, P.4=Si, P.33=Uno o dos, P.42=Sin comentarios	P.1=Si
0.170	1.000	0.170	4.462	1.321	0.041	P.3=No, P.4=Si, P.6=Si, P.10=Universidad	P.1=Si
0.185	1.000	0.185	4.085	1.321	0.045	P.3=No, P.6=Si, P.9=Secundaria, P.42=Sin comentarios	P.1=Si
0.153	1.000	0.153	4.949	1.321	0.037	P.3=No, P.6=Si, P.9=Secundaria, P.42=Sin comentarios, P.43=Sin comentarios	P.1=Si
0.200	1.000	0.200	3.784	1.321	0.049	P.3=No, P.4=Si, P.9=Secundaria	P.1=Si
0.179	1.000	0.179	4.219	1.321	0.044	P.3=No, P.4=Si, P.6=Si, P.9=Secundaria	P.1=Si
0.165	1.000	0.165	4.595	1.321	0.040	P.3=No, P.4=Si, P.9=Secundaria, P.42=Sin comentarios	P.1=Si
0.170	1.000	0.170	4.462	1.321	0.041	GEN=Mujer, P.3=No, P.9=Secundaria	P.1=Si
0.159	1.000	0.159	4.765	1.321	0.039	P.3=No, P.4=Si, P.5=No, P.42=Sin comentarios	P.1=Si
0.150	1.000	0.150	5.046	1.321	0.036	P.3=No, P.4=Si, P.5=No, P.6=Si, P.42=Sin comentarios	P.1=Si
0.151	1.000	0.151	5.013	1.321	0.037	P.3=No, P.4=Si, P.5=No, P.43=Sin comentarios	P.1=Si
0.182	1.000	0.182	4.151	1.321	0.044	GEN=Mujer, P.2=Si, P.4=Si	P.1=Si
0.290	1.000	0.290	2.608	1.321	0.071	P.2=Si, P.3=No, P.4=Si	P.1=Si
0.180	1.000	0.180	4.196	1.321	0.044	GEN=Mujer, P.2=Si, P.3=No, P.4=Si	P.1=Si
0.160	1.000	0.160	4.736	1.321	0.039	P.2=Si, P.3=No, P.4=Si, P.5=No	P.1=Si
0.179	1.000	0.179	4.219	1.321	0.044	P.2=Si, P.3=No, P.5=Si	P.1=Si
0.170	1.000	0.170	4.462	1.321	0.041	GEN=Mujer, P.2=Si, P.4=Si, P.6=Si	P.1=Si
0.266	1.000	0.266	2.849	1.321	0.065	P.2=Si, P.3=No, P.4=Si, P.6=Si	P.1=Si
0.168	1.000	0.168	4.515	1.321	0.041	GEN=Mujer, P.2=Si, P.3=No, P.4=Si, P.6=Si	P.1=Si

Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Adicionalmente en este segundo intento se adicionó que las reglas tuviesen un mínimo de tres ítems en los datos antecedentes de modo que permita tener más información para la detección de patrones. De modo que se generaron un total de 43 reglas con esta configuración y fueron mostradas 40 que sí cumplían con todos los atributos. Por otro lado, en la figura 57 se observa de forma enfática que las reglas generadas dan como consecuencia el ítem "sí" en la variable P.1 ¿El lugar donde vives cuenta con internet?

Esto nos indica que en esta segunda prueba se pudo detectar información valiosa para la identificación de patrones en los estudiantes. Sin embargo, si se aumentan los parámetros de soporte y confianza se genera un número aún menor de reglas como se presenta en la figura 58 con la elaboración de la tercera prueba.

Figura 58. Reglas de asociación en orange con soporte mínimo del veinte por ciento

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	Consequent
0.204	1.000	0.204	3.712	1.321	0.050	P.3=No, P.4=Si, P.7=No, P.33=Uno o dos	P.1=Si
0.200	1.000	0.200	3.784	1.321	0.049	P.3=No, P.4=Si, P.9=Secundaria	P.1=Si
0.290	1.000	0.290	2.608	1.321	0.071	P.2=Si, P.3=No, P.4=Si	P.1=Si
0.266	1.000	0.266	2.849	1.321	0.065	P.2=Si, P.3=No, P.4=Si, P.6=Si	P.1=Si
0.237	1.000	0.237	3.190	1.321	0.058	P.2=Si, P.3=No, P.4=Si, P.7=No	P.1=Si
0.216	1.000	0.216	3.509	1.321	0.052	P.2=Si, P.3=No, P.4=Si, P.6=Si, P.7=No	P.1=Si
0.218	1.000	0.218	3.477	1.321	0.053	P.2=Si, P.3=No, P.4=Si, P.42=Sin comentarios	P.1=Si
0.202	1.000	0.202	3.748	1.321	0.049	P.2=Si, P.3=No, P.4=Si, P.43=Sin comentarios	P.1=Si

Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En la figura 58 se observa que el soporte mínimo es del 20% y el mínimo de confianza es del 100%, además se incluye el requisito de un mínimo de tres ítems en los datos antecedentes. Las reglas resultantes fueron un total de ocho que cuentan con un *lift* mayor a uno por lo que también puede ser de utilidad para la interpretación.

En estos ejemplos se presentó la reducción de las reglas con la finalidad de detectar un conjunto que genere información eficiente, se inició con un total de diez mil reglas y se terminó con un total de ocho. Aunque es importante obtener las reglas más importantes de la herramienta, también hay que considerar que una reducción excesiva puede generar pérdida de información relevante.

4.1.2 WEKA

Clasificación.

Se realizaron pruebas con el algoritmo REPTree con las variables P.1, P.3 y P.7, se optó por el uso de este algoritmo por ser un modelo de poda de ramificaciones del árbol de decisión. Posteriormente los resultados de clasificación correcta fueron comparados con el algoritmo J.48 que también fue utilizado para la elaboración de pruebas de árboles de clasificación, para tal caso los resultados fueron mejores con el algoritmo J.48 por lo que se optó por hacer uso de este para continuar con las pruebas y hacer la presentación de estas.

Dentro de los parámetros que se consideraron para las pruebas fueron el número máximo de ramificaciones que osciló entre los 50, 80 y 100 que permiten conocer el largo que tendrá el diagrama. Por otro lado, se consideró el porcentaje de confianza que

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

significa que a menor número mayor será la proporción de correcta clasificación, este porcentaje osciló entre el 10% al 25%, aunque se recomienda un menor porcentaje de confianza para una mejor clasificación, no significa que sea lo conveniente para los resultados, es necesario encontrar un equilibrio entre el porcentaje de clasificación correcta, la confianza y el largo del diagrama porque una poda excesiva del árbol tampoco quiere decir que se obtendrá información certera.

Por lo tanto en lo subsecuente se presentan las tres pruebas realizadas con el algoritmo J.48 y las variables P.1, P.3 y P.7 como preguntas objetivo por generar una mayor ganancia de información lo cual va a permitir una mejor clasificación de las variables dependientes para la obtención de resultados. Asimismo, el porcentaje de entrenamiento utilizado para las tres pruebas fue del 66%, siendo este el porcentaje que ofrece la herramienta por defecto.

A diferencia de la herramienta Orange que requiere de la conexión de diferentes widgets para generar la evaluación de los modelos de entrenamiento en árboles de decisión, WEKA muestra los parámetros generados en el entorno de trabajo.

Dentro de las métricas que se generan en WEKA se destaca el porcentaje de clasificación correcta, el porcentaje de precisión por clase y del modelo completo, porcentaje de recuerdo y la matriz de confusión. En la figura 59 se presentan los porcentajes obtenidos de la prueba realizada con la variable P.1, se observa que el 96.86% de instancias se clasificaron correctamente lo que da un total de 988 de las cuales 762 pertenecen a la clase "a" y 226 pertenecen a la clase "b" del *dataset*.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 59.
Métricas de evaluación en WEKA de árbol de decisión con pregunta raíz P.1

```

Correctly Classified Instances      988                96.8627 %
Incorrectly Classified Instances    32                 3.1373 %
Kappa statistic                    0.9133
Mean absolute error                 0.0564
Root mean squared error            0.1679
Relative absolute error             15.3161 %
Root relative squared error        39.1495 %
Total Number of Instances         1020
Ignored Class Unknown Instances    1

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.987   0.089   0.972     0.987   0.979     0.914   0.970    0.982    1
          0.911   0.013   0.958     0.911   0.934     0.914   0.970    0.943    2
Weighted Avg.   0.969   0.070   0.968     0.969   0.968     0.914   0.970    0.973

=== Confusion Matrix ===

  a  b  <-- classified as
762 10 |  a = 1
 22 226 |  b = 2

```

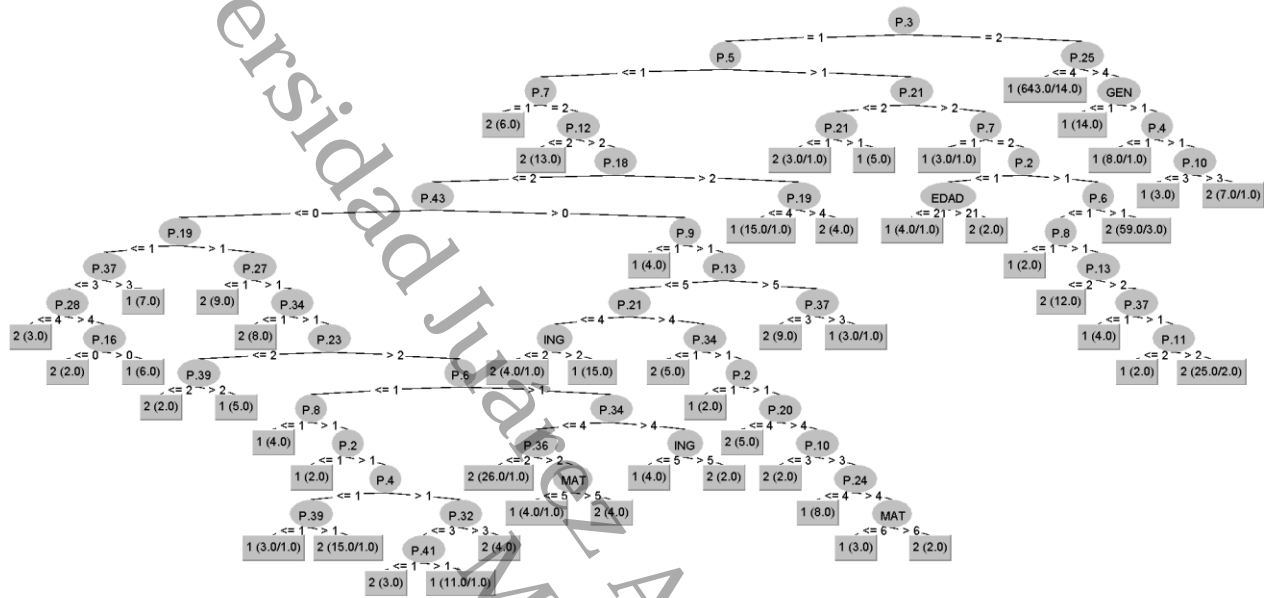
Nota: Elaboración propia.

Por otro lado el porcentaje de precisión obtenido fue del 96.8%, lo que indica que este primer modelo tiene métricas de calidad aceptables y que puede ser considerado para la interpretación de resultados.

En la figura 60 que se presenta a continuación, se observa el árbol de decisión de acuerdo con los parámetros de entrenamiento establecidos y el conjunto de datos. El diagrama generado tiene un total de 103 nodos y 52 niveles y no se pueden filtrar como el caso de *Orange*, por lo que se presenta un diagrama completo.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 60. Visualización del diagrama de árbol con pregunta raíz P.1



Nota: Elaboración propia.

De acuerdo con el diagrama se puede destacar que la clasificación “a” hace referencia al ítem “sí” que se encuentra del lado izquierdo del diagrama y siendo esta clasificación la que tiene un mayor número de nodos. En cuanto a la clasificación “b” que se refiere al ítem “no” se encuentra del lado derecho y el nivel de detalle es menor y tiene relación con el número de instancias que se asignaron a esta clase.

Con la información obtenida en esta primera prueba se puede generar una interpretación inicial de los patrones detectados, de acuerdo con las clasificaciones obtenidas con la pregunta: P.1 ¿El lugar donde vives cuenta con internet?

De los estudiantes que respondieron que no se tiene una tendencia a expresar que tienen un conocimiento tecnológico limitado para llevar clases virtuales. En cuanto a la

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

clase de alumnos que respondieron de forma positiva a contar con servicio de internet, suelen compartir computadora con otros miembros de la familia, esto último aunado a que los estudiantes no iniciaron el uso de tecnologías desde temprana edad genera un impacto negativo en su adaptación a las clases virtuales.

La segunda prueba a realizada es con la variable P.3, con la cual se obtuvo un 95% de instancias clasificadas correctamente, equivalente a 969 instancias de las cuales 314 pertenecen a la clase "a" y 655 pertenecen a la clase "b" del *dataset* (ver figura 61).

Figura 61.

Métricas de evaluación en WEKA de árbol de decisión con pregunta raíz P.3

```

Correctly Classified Instances      969          95      %
Incorrectly Classified Instances    51           5      %
Kappa statistic                    0.8874
Mean absolute error                 0.0895
Root mean squared error             0.2115
Relative absolute error             19.9845 %
Root relative squared error         44.7091 %
Total Number of Instances          1020
Ignored Class Unknown Instances     1

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.910   0.030   0.940     0.910   0.925     0.888   0.967    0.938    1
          0.970   0.090   0.955     0.970   0.963     0.888   0.967    0.973    2
Weighted Avg.   0.950   0.069   0.950     0.950   0.950     0.888   0.967    0.961

=== Confusion Matrix ===

  a  b  <-- classified as
314 31 |  a = 1
 20 655 |  b = 2

```

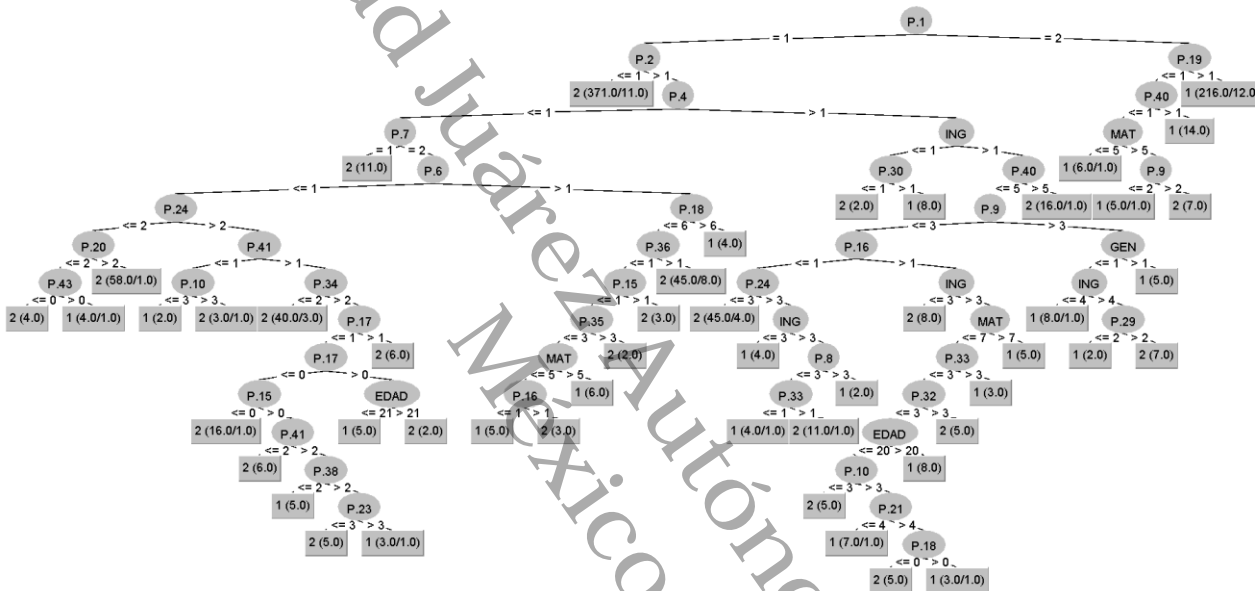
Nota: Elaboración propia.

En cuanto al porcentaje de precisión obtenido fue del 95%, lo que indica que las métricas de calidad en este modelo también son aceptables. Por lo tanto, al escoger el

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

modelo a interpretar habrá que considerar la información obtenida en los diagramas. En ese sentido se presenta el segundo diagrama generado con un total de 121 nodos y 61 niveles (ver figura 62).

Figura 62.
Visualización del diagrama de árbol con pregunta raíz P.3



Nota: Elaboración propia.

En este segundo diagrama de la figura 62, la clasificación “a” se refiere al ítem “si” que se encuentra del lado izquierdo del diagrama, en cuanto a la clasificación “b” que se refiere al ítem “no” se encuentra del lado derecho y el nivel de detalle es menor, en este ejemplo la pregunta objetivo es: P.3 ¿Debes ir a rentar internet para entrar a las plataformas de clases?

De los estudiantes que sí deben rentar internet es porque el internet que tienen en casa es ineficiente; no cuentan con una computadora personal, tableta o teléfono con

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

datos y adicionalmente consideran que fue difícil llevar clases virtuales por su escaso conocimiento en tecnologías. Con relación a la clase "b" de los estudiantes que no requieren salir de casa y rentar un servicio de internet su inconveniente radica en el tipo de asignaturas cursadas, los estudiantes con mayor carga académica entre 5 y 9 materias prefieren llevar únicamente materias teóricas en línea.

La tercera prueba se realizó con la variable P.7, con esta se obtuvo un 96.27% de instancias clasificadas correctamente, lo que es igual a 982 instancias de las cuales 52 pertenecen a la clase "a" y 930 pertenecen a la clase "b" del dataset (ver figura 63).

Figura 63. Métricas de evaluación en WEKA de árbol de decisión con pregunta raíz P.7

```

Correctly Classified Instances      982      96.2745 %
Incorrectly Classified Instances    38       3.7255 %
Kappa statistic                    0.7132
Mean absolute error                0.07
Root mean squared error            0.1871
Relative absolute error             45.6327 %
Root relative squared error         67.7017 %
Total Number of Instances          1020
Ignored Class Unknown Instances     1

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0.612   0.005   0.912     0.612   0.732     0.730  0.851   0.665    1
          0.995   0.388   0.966     0.995   0.980     0.730  0.853   0.975    2
Weighted Avg.   0.963   0.356   0.961     0.963   0.959     0.730  0.853   0.949

=== Confusion Matrix ===

  a  b  <-- classified as
52 33 |  a = 1
 5 930 |  b = 2

```

Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En esta misma clasificación existe otro patrón que se relaciona con tener internet en casa sin embargo, este patrón concuerda con la descripción expuesta en la primera prueba por lo que no ofrece información nueva.

En cuanto a las personas que no cuentan con una tableta, se encuentra relacionado con la edad de personas mayores a 28 años.

Agrupación.

El segundo modelo de minería de datos por presentar con la herramienta WEKA es el de agrupación, durante las pruebas realizadas se buscó número agrupaciones para la mejor representación de los datos realizando diferentes pruebas con el algoritmo *SimpleKMeans*. En este apartado se presentan tres pruebas con diferente número de clústeres que fueron descartadas durante el proceso de acuerdo con la métrica de calidad que ofrece el *software* con el algoritmo aplicado. Las métricas de evaluación y calidad obtenidas en WEKA se realizan a través de una matriz de confusión que ofrece un porcentaje de instancias agrupadas incorrectamente, siendo este el parámetro considerado durante los ejemplos.

Por otro lado, esta herramienta ofrece diferentes gráficas para visualizar diferentes combinaciones de las variables del *dataset* de acuerdo con el número de agrupaciones solicitadas por lo que durante los ejemplos también se consideran distintos diagramas.

La primera prueba realizada fue con dos agrupaciones, en la figura 65 se presenta que el modelo clasificó en 513 instancias en el grupo cero y 508 instancias en el grupo

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

uno, sin embargo el porcentaje de instancias clasificadas incorrectamente fueron 322 siendo equivalente a un 31.53%.

Figura 65. Modelo de entrenamiento de dos clústeres en WEKA

```

=== Model and evaluation on training set ===
                                0  1  <-- assigned to cluster
                                481 291 | 1
Clustered Instances              31 217 | 2

0          513 ( 50%)
1          508 ( 50%)

Cluster 0 <-- 1
Cluster 1 <-- 2

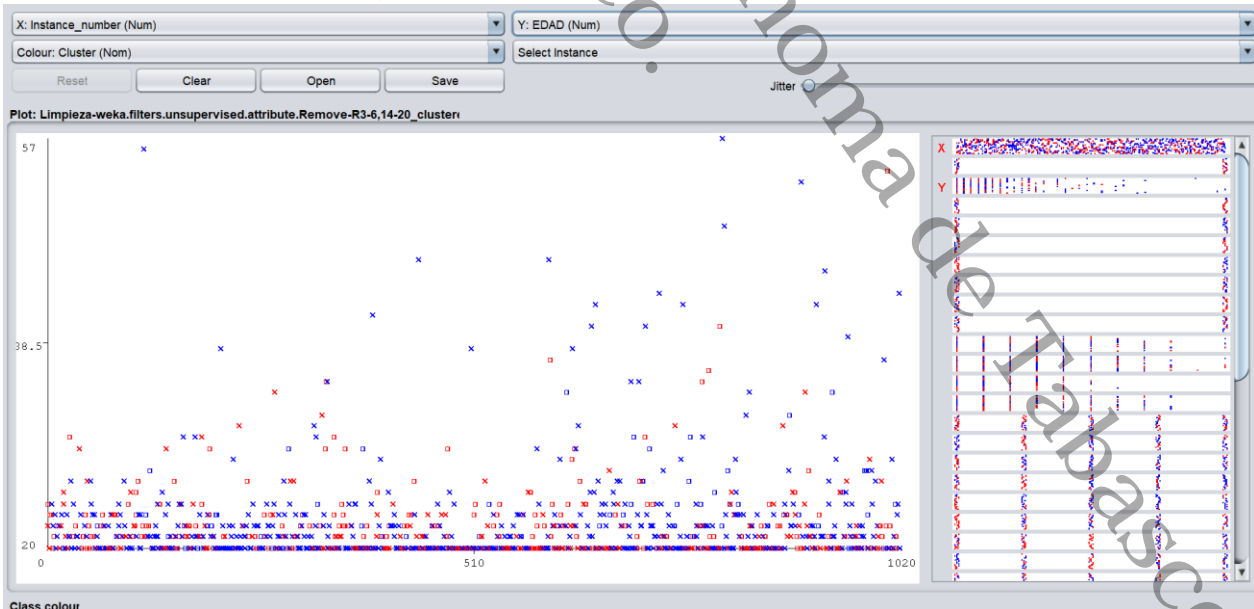
Incorrectly clustered instances :      322.0      31.5377 %

```

Nota: Elaboración propia.

En la figura 66 se presenta una gráfica realizada con la variable *Instance_number* y la variable EDAD. Se observa una mayor densidad de los datos en la parte inferior de la gráfica debido a que los estudiantes en su mayoría son menores de 25 años.

Figura 66. K-means prueba de dos agrupaciones por edad



Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Como se observa previamente del lado derecho de la figura 66, se muestran diferentes opciones de gráficos con otras variables del dataset con dos agrupaciones y que serán presentadas en las pruebas siguientes.

La segunda prueba realizada fue con cuatro agrupaciones, de acuerdo con la figura 66 la clasificación de instancias se presenta de la siguiente manera: 217 en el grupo cero, 307 en el grupo uno, 234 del grupo dos y 263 pertenecen al grupo tres. Asimismo se menciona que 533 instancias fueron clasificadas incorrectamente siendo igual al 56.12%. De acuerdo con esto, la calidad de las agrupaciones disminuyó en comparación con la prueba realizada con dos agrupaciones (ver figura 67).

Figura 67.

Modelo de entrenamiento de cuatro clústeres en WEKA

```

=== Model and evaluation on training set ===
Classes to Clusters:
Clustered Instances
0      217 ( 21%)
1      307 ( 30%)
2      234 ( 23%)
3      263 ( 26%)

0 1 2 3 <-- assigned to cluster
192 104 232 244 | 1
25 203 2 18 | 2
Cluster 0 <-- No class
Cluster 1 <-- 2
Cluster 2 <-- No class
Cluster 3 <-- 1

Incorrectly clustered instances :      573.0      56.1214 %

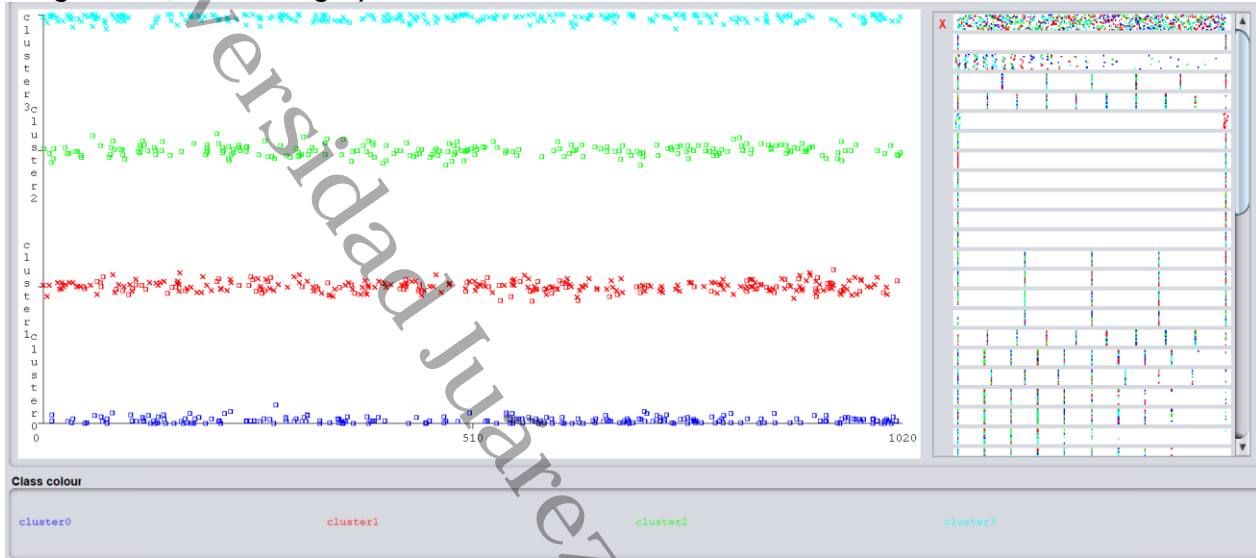
```

Nota: Elaboración propia.

En este segundo gráfico se presenta otro ejemplo de visualización de los clústeres, las variables utilizadas fueron *Instance_number* y clústeres, este tipo de diagrama lo que permite observar es la separación de los cuatro grupos (ver figura 68).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 68. Diagrama de cuatro agrupaciones en WEKA



Nota: Elaboración propia.

La tercera prueba fue con cinco agrupaciones, de acuerdo con la figura 69 el agrupamiento fue realizado de la siguiente manera: 186 instancias clasificadas en el grupo cero, 208 en el grupo uno, 214 del grupo dos, 186 en el grupo tres y 227 pertenecientes al grupo cuatro.

Figura 69. Entrenamiento de cinco clústeres en WEKA

```

=== Model and evaluation on training set ===
Classes to Clusters:
Clustered Instances
0      186 ( 18%)
1      208 ( 20%)
2      214 ( 21%)
3      186 ( 18%)
4      227 ( 22%)

0  1  2  3  4  <-- assigned to cluster
180 69 212 106 205 | 1
6 139 2 80 21 | 2

Cluster 0 <-- No class
Cluster 1 <-- 2
Cluster 2 <-- 1
Cluster 3 <-- No class
Cluster 4 <-- No class

Incorrectly clustered instances :      669.0      65.524 %

```

Nota: Elaboración propia.

1

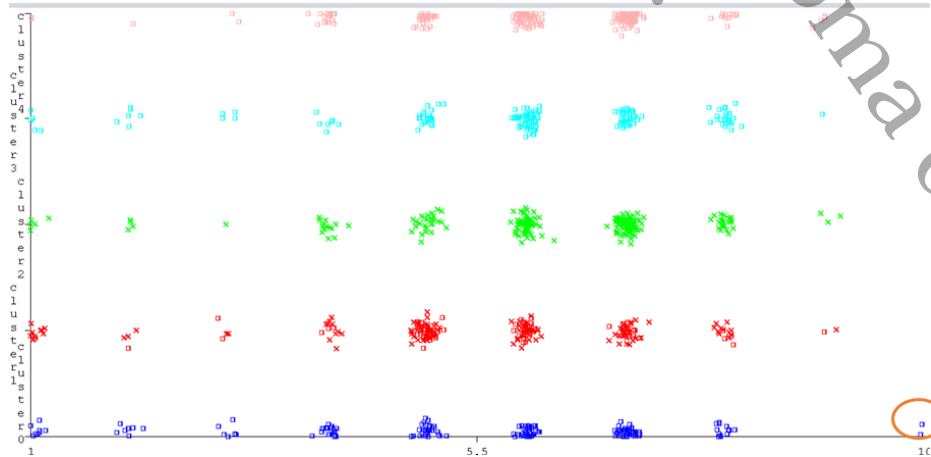
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Como se presentó previamente, en esta última agrupación los clústeres cero y tres son similares en cuanto a distribución. Por otro lado, el porcentaje de clasificación incorrecta es del 65.22% lo que indica que solo 351 instancias lograron agruparse correctamente. De acuerdo con las métricas de calidad, esta última prueba fue la que obtuvo los resultados más bajos.

Sin embargo, al igual que en las agrupaciones anteriores también se presentan ejemplos de diagramas que fueron considerados para su presentación. Por lo tanto se muestra el siguiente diagrama con las variables MAT y Clúster, este gráfico lo que presenta es el número de materias cursadas por el estudiante de acuerdo con su agrupación. En este sentido se observa que los estudiantes en su mayoría cursaron de cuatro a ocho materias durante clases virtuales y solo en el clúster 0 se presenta el caso de estudiantes con 10 materias (ver figura 70).

Figura 70.

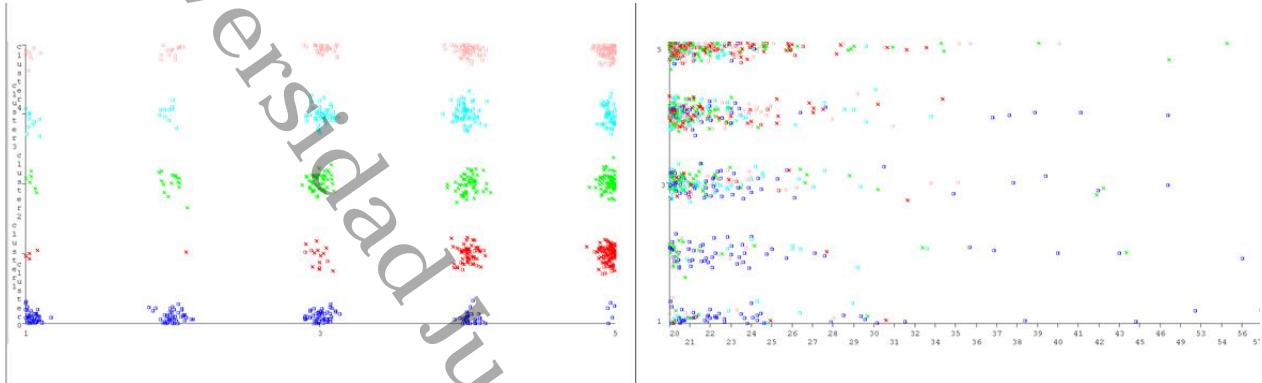
K-means prueba de cinco agrupaciones por materia en WEKA



Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 71.
Diagramas de cinco agrupaciones con la variable P.22 en WEKA



Nota: Elaboración propia.

En la figura 71 se presentan gráficos con información adicional, en el diagrama de la izquierda las variables ocupadas fueron P.22 que se refiere al estrés y ansiedad de trabajar con tecnología y el Clúster asignado. El gráfico derecho ocupa nuevamente la variable P.22 con la EDAD, conservando el color de las agrupaciones.

Como se muestra en esta tercera prueba, aunque se ocupa una misma variable en común se presenta de diferente manera lo que permite complementar la información generada con la figura 69 para ser interpretado.

Durante las pruebas de agrupamiento en WEKA se presentaron pruebas solo de la calidad de los modelos para la selección del número óptimo, también se mostró la funcionalidad de las gráficas que ofrece la herramienta con el algoritmo *SimpleKMeans*.

Reglas de asociación.

Al igual que en las muestras presentadas previamente, en este apartado se exponen tres ejemplos de pruebas realizadas con el modelo de reglas de asociación con

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

la herramienta *WEKA*. Los modelos realizados con el algoritmo Apriori incluyeron variaciones en las métricas de calidad, principalmente con el parámetro confianza y soporte.

El primer ejemplo se obtuvieron 613 instancias con un soporte mínimo del 60%, una confianza del 50% y un total de ocho ciclos requeridos por el algoritmo para su ejecución. Aunque la herramienta encontró determinado número de instancias con los parámetros establecidos, únicamente muestra las principales reglas de acuerdo con la solicitud del usuario, para tal caso se requirieron las diez reglas principales de las cuales se detectaron 11 patrones de tamaño uno y siete patrones de tamaño dos (ver figura 72).

Figura 72.

Métricas del algoritmo Apriori en WEKA con cincuenta por ciento de confianza

```
Apriori
=====

Minimum support: 0.6 (613 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 8

Generated sets of large itemsets:

Size of set of large itemsets L(1): 11

Size of set of large itemsets L(2): 7
```

Nota: Elaboración propia.

Como se observa en la figura 73, el algoritmo generó las diez reglas solicitadas previamente, la primera regla obtenida nos dice que los estudiantes que no necesitan rentar una computadora en un café internet son porque cuentan con el servicio en casa.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En este primer ejemplo el antecedente a la regla se cumple en 675 instancias del *dataset*, en cuanto a la regla completa se cumple en 654 instancias del conjunto de datos.

Figura 73.

Reglas de asociación en WEKA con cincuenta por ciento de confianza

Best rules found:

```

1. P.3=No 675 ==> P.1=Si 654 <conf:(0.97)> lift:(1.28) lev:(0.14) [143] conv:(7.48)
2. P.2=No 642 ==> P.7=No 616 <conf:(0.96)> lift:(1.05) lev:(0.03) [28] conv:(2)
3. P.17=0 733 ==> P.7=No 682 <conf:(0.93)> lift:(1.02) lev:(0.01) [10] conv:(1.19)
4. P. 42=Sin comentarios 742 ==> P.7=No 684 <conf:(0.92)> lift:(1.01) lev:(0) [4] conv:(1.06)
5. P.43=Sin comentarios 700 ==> P.7=No 642 <conf:(0.92)> lift:(1) lev:(0) [0] conv:(1)
6. P.1=Si 772 ==> P.7=No 695 <conf:(0.9)> lift:(0.98) lev:(-0.01) [-11] conv:(0.83)
7. P.6=Si 715 ==> P.7=No 643 <conf:(0.9)> lift:(0.98) lev:(-0.01) [-11] conv:(0.83)
8. P.1=Si 772 ==> P.3=No 654 <conf:(0.85)> lift:(1.28) lev:(0.14) [143] conv:(2.2)
9. P.7=No 935 ==> P.1=Si 695 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-11] conv:(0.95)
10. P.7=No 935 ==> P. 42=Sin comentarios 684 <conf:(0.73)> lift:(1.01) lev:(0) [4] conv:(1.01)

```

Nota: Elaboración propia.

Sin embargo, en este primer ejemplo también se observa que aunque el *lift* es alto las métricas de *leverage* (soporte precedente y consecuente) son negativas lo cual les resta calidad a las reglas obtenidas.

Por lo anterior, la segunda prueba realizada fue con una confianza mínima del 70%, esto generó un total de 562 instancias un soporte mínimo del 55% y un total de 9 ciclos requeridos para la ejecución del algoritmo. Con estos parámetros se solicitaron las principales 25 reglas, al aumentar el número de reglas requeridas dio como resultado un mayor número de patrones (ver figura 74).

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 74. Métricas del algoritmo Apriori en WEKA con setenta por ciento de confianza

```
Minimum support: 0.55 (562 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 9
```

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 15

Size of set of large itemsets L(3): 1

Nota: Elaboración propia.

En la figura 74 se muestran los parámetros previamente descritos, de las 25 reglas solicitadas se detectaron un total de 12 patrones o reglas con tamaño uno que significa que fue construida con una variable, 15 patrones con dos variables y un patrón con tres variables.

Figura 75. Reglas de asociación en WEKA con setenta por ciento de confianza

```
1. P.3=No 675 ==> P.1=Si 654 <conf:(0.97)> lift:(1.28) lev:(0.14) [143] conv:(7.48)
2. P.3=No P.7=No 600 ==> P.1=Si 580 <conf:(0.97)> lift:(1.28) lev:(0.12) [126] conv:(6.97)
3. P.2=No 642 ==> P.7=No 616 <conf:(0.96)> lift:(1.05) lev:(0.03) [28] conv:(2)
4. P.5=Si 636 ==> P.7=No 602 <conf:(0.95)> lift:(1.03) lev:(0.02) [19] conv:(1.53)
5. P.17=0 733 ==> P.7=No 682 <conf:(0.93)> lift:(1.02) lev:(0.01) [10] conv:(1.19)
6. P.42=Sin comentarios 742 ==> P.7=No 684 <conf:(0.92)> lift:(1.01) lev:(0) [4] conv:(1.06)
7. P.43=Sin comentarios 700 ==> P.7=No 642 <conf:(0.92)> lift:(1) lev:(0) [0] conv:(1)
8. P.18=0 649 ==> P.7=No 595 <conf:(0.92)> lift:(1) lev:(0) [0] conv:(0.99)
9. P.18=0 649 ==> P.17=0 592 <conf:(0.91)> lift:(1.27) lev:(0.12) [126] conv:(3.16)
10. P.1=Si 772 ==> P.7=No 695 <conf:(0.9)> lift:(0.98) lev:(-0.01) [-11] conv:(0.83)
11. P.6=Si 715 ==> P.7=No 643 <conf:(0.9)> lift:(0.98) lev:(-0.01) [-11] conv:(0.83)
12. P.4=Si 647 ==> P.7=No 577 <conf:(0.89)> lift:(0.97) lev:(-0.02) [-15] conv:(0.77)
13. P.3=No 675 ==> P.7=No 600 <conf:(0.89)> lift:(0.97) lev:(-0.02) [-18] conv:(0.75)
14. P.1=Si P.3=No 654 ==> P.7=No 580 <conf:(0.89)> lift:(0.97) lev:(-0.02) [-18] conv:(0.73)
15. P.43=Sin comentarios 700 ==> P.42=Sin comentarios 610 <conf:(0.87)> lift:(1.2) lev:(0.1) [101] conv:(2.1)
16. P.3=No 675 ==> P.1=Si P.7=No 580 <conf:(0.86)> lift:(1.26) lev:(0.12) [120] conv:(2.25)
17. P.6=Si 715 ==> P.1=Si 610 <conf:(0.85)> lift:(1.13) lev:(0.07) [69] conv:(1.65)
18. P.1=Si 772 ==> P.3=No 654 <conf:(0.85)> lift:(1.28) lev:(0.14) [143] conv:(2.2)
19. P.1=Si P.7=No 695 ==> P.3=No 580 <conf:(0.83)> lift:(1.26) lev:(0.12) [120] conv:(2.03)
20. P.42=Sin comentarios 742 ==> P.43=Sin comentarios 610 <conf:(0.82)> lift:(1.2) lev:(0.1) [101] conv:(1.75)
21. P.17=0 733 ==> P.18=0 592 <conf:(0.81)> lift:(1.27) lev:(0.12) [126] conv:(1.88)
22. P.1=Si 772 ==> P.6=Si 610 <conf:(0.79)> lift:(1.13) lev:(0.07) [69] conv:(1.42)
23. P.42=Sin comentarios 742 ==> P.1=Si 568 <conf:(0.77)> lift:(1.01) lev:(0.01) [6] conv:(1.03)
24. P.1=Si 772 ==> P.3=No P.7=No 580 <conf:(0.75)> lift:(1.28) lev:(0.12) [126] conv:(1.65)
25. P.7=No 935 ==> P.1=Si 695 <conf:(0.74)> lift:(0.98) lev:(-0.01) [-11] conv:(0.95)
```

Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Con los resultados obtenidos en esta prueba, en la figura 75 se observa que la regla uno es la misma que se generó en la primer prueba pero en un orden invertido, adicionalmente se muestran otras reglas generadas de una sola variable que no aportan información que sea útil o novedosa. Esto se debe a que nuevamente la métrica *leverage* genera resultados negativos, siendo esta métrica la que permite conocer si el número de transacciones generadas emite información relevante sobre el contexto.

Por lo tanto, aunque se aumentó en porcentaje de confianza y se halló un mayor número de patrones, las reglas principales no emiten información oportuna por lo que es necesario aumentar el valor de los parámetros de calidad como se muestra en la tercera prueba.

Figura 76.

Métricas del algoritmo Apriori en WEKA con noventa por ciento de confianza

```
Apriori
=====

Minimum support: 0.4 (408 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 12

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20
Size of set of large itemsets L(2): 41
Size of set of large itemsets L(3): 31
Size of set of large itemsets L(4): 6
```

Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

De acuerdo con la figura 76 presentada previamente, el tercer ejemplo se realizó considerando un 90% de confianza como métrica de la calidad, lo que dio como resultado un total de 408 instancias del dataset con un soporte mínimo del 40% y un total de 12 ciclos requeridos por el algoritmo para su ejecución. Con los parámetros descritos se obtuvieron un total de 20 patrones generados con una variable, 41 con dos variables, 31 con tres variables y seis reglas de tamaño cuatro.

De acuerdo con los descrito en las pruebas anteriores se observa que a mayor porcentaje de confianza, tiende a ser mayor el número de patrones que se detectan con el algoritmo. Sin embargo, en esta tercera prueba se optó por mostrar las 20 mejores reglas de acuerdo con la herramienta (ver figura 77).

Figura 77.

Reglas de asociación en WEKA con noventa por ciento de confianza

Best rules found:

1. P.3=No P.4=Si P.7=No 428 ==> P.1=Si 423 <conf:(0.99)> lift:(1.31) lev:(0.1) [99] conv:(17.4)
2. P.3=No P.4=Si P.6=Si 421 ==> P.1=Si 416 <conf:(0.99)> lift:(1.31) lev:(0.1) [97] conv:(17.11)
3. P.3=No P.4=Si 494 ==> P.1=Si 488 <conf:(0.99)> lift:(1.31) lev:(0.11) [114] conv:(17.21)
4. P.3=No P.6=Si 550 ==> P.1=Si 537 <conf:(0.98)> lift:(1.29) lev:(0.12) [121] conv:(9.58)
5. P.3=No P.6=Si P.7=No 482 ==> P.1=Si 470 <conf:(0.98)> lift:(1.29) lev:(0.1) [105] conv:(9.04)
6. P.3=No P.42=Sin comentarios 502 ==> P.1=Si 489 <conf:(0.97)> lift:(1.29) lev:(0.11) [109] conv:(8.74)
7. P.3=No P.7=No P.42=Sin comentarios 452 ==> P.1=Si 439 <conf:(0.97)> lift:(1.28) lev:(0.1) [97] conv:(7.87)
8. P.3=No P.43=Sin comentarios 475 ==> P.1=Si 461 <conf:(0.97)> lift:(1.28) lev:(0.1) [101] conv:(7.72)
9. P.3=No 675 ==> P.1=Si 654 <conf:(0.97)> lift:(1.28) lev:(0.14) [143] conv:(7.48)
10. P.3=No P.7=No P.43=Sin comentarios 424 ==> P.1=Si 410 <conf:(0.97)> lift:(1.28) lev:(0.09) [89] conv:(6.89)
11. P.3=No P.7=No 600 ==> P.1=Si 580 <conf:(0.97)> lift:(1.28) lev:(0.12) [126] conv:(6.97)
12. P.2=No P.42=Sin comentarios 455 ==> P.7=No 437 <conf:(0.96)> lift:(1.05) lev:(0.02) [20] conv:(2.02)
13. P.2=No 642 ==> P.7=No 616 <conf:(0.96)> lift:(1.05) lev:(0.03) [28] conv:(2)
14. P.2=No P.5=Si 445 ==> P.7=No 426 <conf:(0.96)> lift:(1.05) lev:(0.02) [18] conv:(1.87)
15. P.2=No P.43=Sin comentarios 433 ==> P.7=No 413 <conf:(0.95)> lift:(1.04) lev:(0.02) [16] conv:(1.74)
16. P.5=Si P.42=Sin comentarios 465 ==> P.7=No 442 <conf:(0.95)> lift:(1.04) lev:(0.02) [16] conv:(1.63)
17. P.5=Si 636 ==> P.7=No 602 <conf:(0.95)> lift:(1.03) lev:(0.02) [19] conv:(1.53)
18. P.1=Si P.5=Si 495 ==> P.7=No 467 <conf:(0.94)> lift:(1.03) lev:(0.01) [13] conv:(1.44)
19. P.33=Uno o dos 479 ==> P.7=No 450 <conf:(0.94)> lift:(1.03) lev:(0.01) [11] conv:(1.34)
20. P.1=Si P.4=Si P.6=Si 445 ==> P.3=No 416 <conf:(0.93)> lift:(1.41) lev:(0.12) [121] conv:(5.03)

Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

A diferencia de las pruebas presentadas previamente, en la figura 77 se observa que las reglas generadas obtuvieron métricas de calidad positivas, en el caso de los parámetros *lift* y *conviction* los puntajes alcanzados son superiores a uno lo que permite considerar que la información mostrada de los patrones encontrados contiene información relevante. Asimismo, se observa que la primera regla ya no es la misma que se generó en pruebas anteriores. Por otro lado, en esta ocasión la regla número dos expone que los estudiantes que cuentan con internet en casa no necesitan rentar una computadora en un café internet, cuentan con una computadora personal y tienen celular con internet. A partir de esta nueva información, se puede realizar una interpretación de los patrones detectados.

4.1.3 Lenguaje R

Clasificación.

En este apartado se presentan las pruebas realizadas con la herramienta Lenguaje R, aplicando el modelo de árboles de decisión para la clasificación de datos. Al igual que en los modelos presentados en Orange y WEKA, se requirió realizar sets de entrenamiento con 60% del total de los datos empleando las variables P1, P.3 y P.7 con el propósito de comprobar la eficacia de los modelos con diferentes preguntas objetivo a fin de obtener predicciones correctas y puedan ser comparados.

La primera prueba fue realizada con la variable P.1, en el siguiente esquema se presenta el árbol de clasificación generado. Cada inciso o línea son los nodos que

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

conforman el diagrama, para este ejemplo son un total de 179 nodos, también se muestran las reglas de clasificación construidas con los datos (ver figura 78).

Figura 78.

Árbol de decisión Lenguaje R con pregunta raíz P.1

n= 613

```
node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 613 149 si (0.24306688 0.75693312)
 2) P.3=si 209 73 No (0.65071770 0.34928230)
 4) P.38=2 62 6 No (0.90322581 0.09677419) *
 5) P.38=1,3,4,5 147 67 No (0.54421769 0.45578231)
10) P.29=1,4,5 65 18 No (0.72307692 0.27692308)
 20) P.12=0,2,3,4,5,7 40 6 No (0.85000000 0.15000000)
 21) P.12=6,8 25 12 No (0.52000000 0.48000000)
 42) P.21=1,4 8 1 No (0.87500000 0.12500000) *
 43) P.21=2,3,5 17 6 Si (0.35294118 0.64705882) *
11) P.29=2,3 82 33 si (0.40243902 0.59756098)
 22) P.43=0,2 57 27 No (0.52631579 0.47368421)
 44) P.25=2,3,5 36 12 No (0.66666667 0.33333333)
 88) P.12=1,3,4,8 10 0 No (1.00000000 0.00000000)
 89) P.12=5,6,7,9 26 12 No (0.53846154 0.46153846)
 178) P.24=3,5 19 6 No (0.68421053 0.31578947) *
 179) P.24=2,4 7 1 si (0.14285714 0.85714286) *
 45) P.25=1,4 21 6 si (0.28571429 0.71428571) *
 23) P.43=1,3 25 3 si (0.12000000 0.88000000) *
```

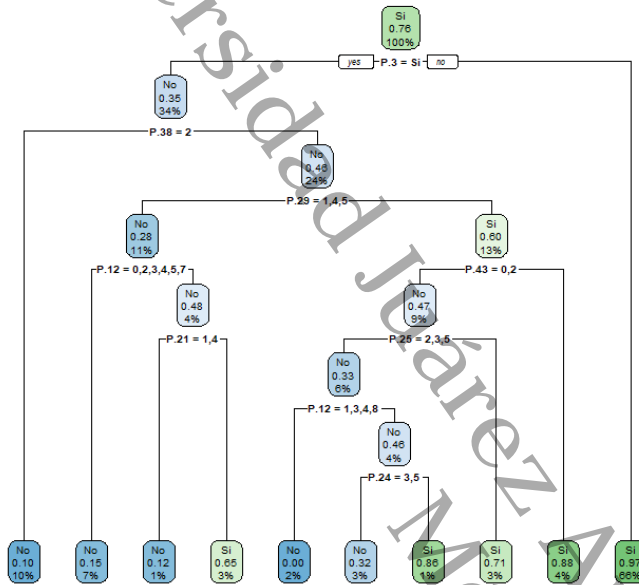
Nota: Elaboración propia.

Aunque en estructura de la figura 78 se puede interpretar la información, Lenguaje R permite generar el gráfico de árbol que ofrece un resultado visualmente más claro y atractivo. En este tipo de diagrama cada nodo se encuentra identificado con el color al que pertenece de acuerdo con la categoría de los datos que clasifica.

Para esta primera prueba, las clasificaciones se encuentran identificadas en color verde para el ítem “Si” y en color azul para el ítem “No”, otorgando una mayor categorización de color azul haciendo referencia a los estudiantes que no cuentan con internet en casa (ver figura 79).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 79. Visualización del diagrama de árbol con pregunta raíz P.1



Nota: Elaboración propia.

Asimismo, se detectan reglas en las hojas finales uno, dos y cuatro de derecha a izquierda con 97%, 88% y 86% respectivamente de clasificaciones correctas. Lo que estas reglas en particular nos dicen es que los estudiantes sí cuentan con internet en casa pero su conocimiento en tecnologías y plataformas virtuales es limitado.

Aunque el diagrama generado nos permite conocer la calidad de las reglas resultantes, fue necesario evaluar el modelo para conocer la calidad de las predicciones. En este caso se aplicó una matriz de confusión cruzando la predicción con los datos reales.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 80.
Métricas de evaluación en Lenguaje R de árbol de decisión con pregunta raíz P.1

```
Reference
Prediction No Si
No 70 29
Si 28 280

Accuracy : 0.86
95% CI : (0.8224, 0.8922)
No Information Rate : 0.7592
P-Value [Acc > NIR] : 3.424e-07

Kappa : 0.6183

McNemar's Test P-Value : 1

Sensitivity : 0.7143
Specificity : 0.9061
Pos Pred Value : 0.7071
Neg Pred Value : 0.9091
Prevalence : 0.2408
Detection Rate : 0.1720
Detection Prevalence : 0.2432
Balanced Accuracy : 0.8102

'Positive' class : No
```

Nota: Elaboración propia.

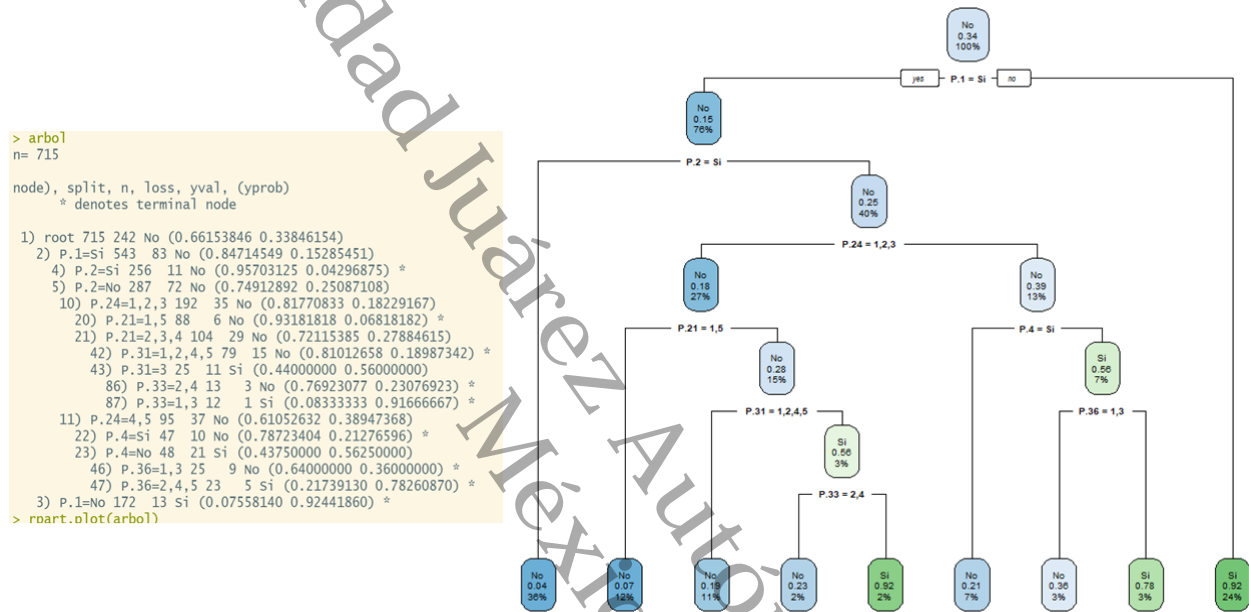
En la figura 80 se presenta las métricas generadas para la evaluación del primer modelo, se observa que la precisión obtenida (*Accuracy*) es del 86%, que nos indica que la clasificación realizada es aceptable y clasifica correctamente la mayor parte de los datos. Aunque esta primera prueba puede ser útil para su interpretación, se consideraron otras dos pruebas más modificando la variable, ya que los grupos resultantes tendrán una separación distinta de modo que generen nuevas reglas de clasificación.

Por lo tanto, se generó un segundo árbol utilizando la variable P.3 como pregunta raíz, se obtuvieron un total de 87 nodos, las clasificaciones se encuentran identificadas en color verde para el ítem “Si” y en color azul para el ítem “No”, otorgando una mayor

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

categorización de color azul haciendo referencia a los estudiantes que no requieren ir a rentar internet para sus clases virtuales (ver figura 81).

Figura 81. Árbol de decisión Lenguaje R con pregunta raíz P.3



Nota: Elaboración propia.

Por otro lado, se detectan reglas en las hojas finales uno y cinco de derecha a izquierda con 92% respectivamente de clasificaciones correctas. Lo que estas reglas en particular nos dicen es que los estudiantes que no requieren rentar internet sí cuentan con internet en casa pero es ineficiente aunado a que casi todos los profesores solicitan exposiciones hace que las clases virtuales sean más pesadas que las presenciales.

En cuanto a las métricas de evaluación, la figura 82 presenta la matriz de confusión generada que da como resultado una precisión del 84.92%, aunque en esta segunda

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

prueba el porcentaje de clasificación bajo la clasificación realizada sigue siendo aceptable.

Figura 82.

Métricas de evaluación en Lenguaje R de árbol de decisión con pregunta raíz P.3

```
Confusion Matrix and Statistics

      Reference
Prediction No  Si
      No 189  13
      Si  33  70

      Accuracy : 0.8492
      95% CI : (0.804, 0.8874)
      No Information Rate : 0.7279
      P-Value [Acc > NIR] : 3.357e-07

      Kappa : 0.646

      Mcnemar's Test P-Value : 0.005088

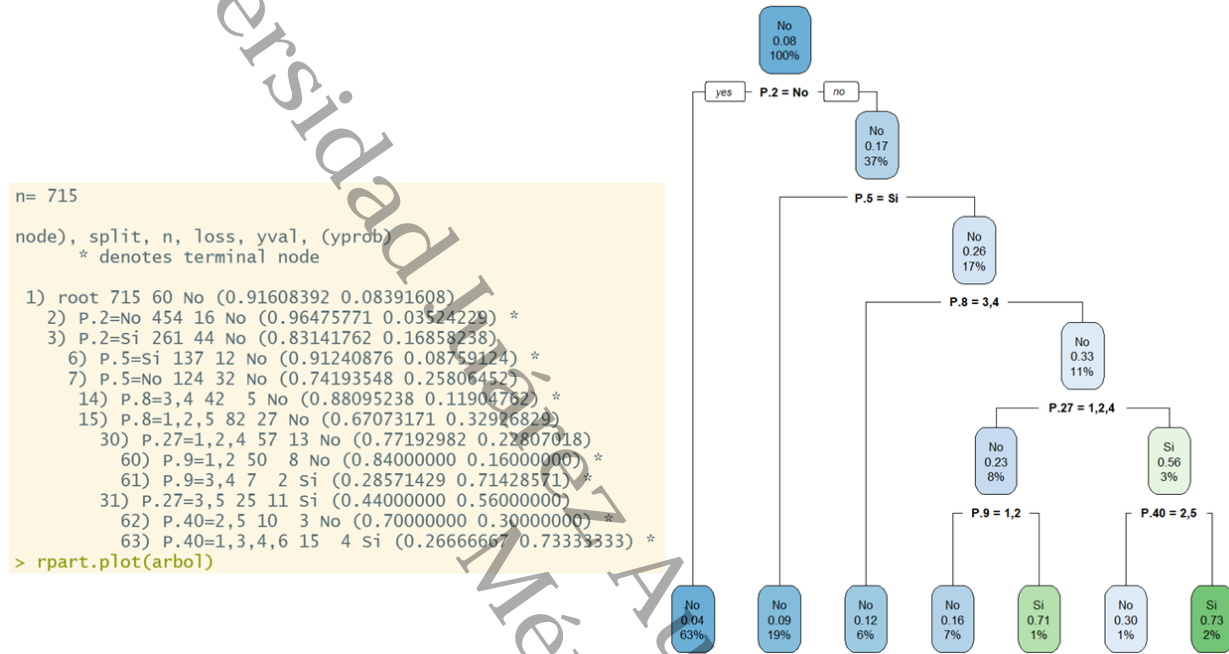
      Sensitivity : 0.8514
      Specificity : 0.8434
      Pos Pred Value : 0.9356
      Neg Pred Value : 0.6796
      Prevalence : 0.7279
      Detection Rate : 0.6197
      Detection Prevalence : 0.6623
      Balanced Accuracy : 0.8474
```

Nota: Elaboración propia.

La tercera prueba se generó con la variable P.7 como pregunta objetivo, como resultado el tamaño del árbol fue de 63 nodos, siendo este el árbol con el menor tamaño de los tres modelos presentados. Las clasificaciones realizadas se identifican en color verde para el ítem “Si” y en color azul para el ítem “No”, otorgando una mayor categorización de color azul que se refiere a los estudiantes que no con una tableta para uso personal (ver figura 83).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 83. Árbol de decisión Lenguaje R con pregunta raíz P.7



Nota: Elaboración propia.

Con respecto a las reglas generadas en el diagrama, detectan reglas en las hojas finales uno y tres de derecha a izquierda con 73% respectivamente de clasificaciones correctas, lo que significa que la calidad en las reglas generadas disminuyó considerablemente.

Por otro lado, las dos mejores reglas construidas en el último modelo dicen que los estudiantes que sí cuentan con tableta para uso personal comenzaron a utilizar recursos de internet a partir de la primaria y secundaria. Sin embargo también consideran que las actividades encargadas por el profesor fueron excesivas y aportaron poco conocimiento.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Por último se presentan las métricas para evaluar el tercer modelo, la figura 83 se muestra como resultado una precisión del 88.85%, en esta tercera prueba el porcentaje de clasificación fue mayor que en los modelos anteriores. Sin embargo, en la matriz de confusión los datos del ítem "Si" fueron clasificados incorrectamente, esto quiere decir que solo una clase tuvo una precisión alta generando un porcentaje de evaluación mayor que en otros modelos.

Figura 84.

Métricas de evaluación en Lenguaje R de árbol de decisión con pregunta raíz P.7

Confusion Matrix and Statistics		
	Reference	
Prediction	No	Si
No	269	11
Si	23	2
Accuracy : 0.8885		
95% CI : (0.8477, 0.9216)		
No Information Rate : 0.9574		
P-Value [Acc > NIR] : 1.00000		
Kappa : 0.0521		
McNemar's Test P-Value : 0.05923		
Sensitivity : 0.9212		
Specificity : 0.1538		
Pos Pred Value : 0.9607		
Neg Pred Value : 0.0800		
Prevalence : 0.9574		
Detection Rate : 0.8820		
Detection Prevalence : 0.9180		
Balanced Accuracy : 0.5375		

Nota: Elaboración propia.

En esta sección se presentaron las pruebas de árboles de decisión con Lenguaje R, aunque los porcentajes de evaluación fueron favorables en los tres ejemplos, se debe

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

hacer una selección del modelo que genere no solo la métrica de calidad más alta sino la mejor representación de la información.

Agrupación.

En este apartado se presentan los modelos de agrupación realizados con el algoritmo *Kmeans*, los ejemplos realizados son con agrupaciones de dos, cuatro y cinco clústeres realizados con la herramienta para encontrar el agrupamiento de mejor calidad. Este es un algoritmo del tipo no supervisado que tiene como objetivo otorgarles un clúster a los datos, por lo que se recomienda su uso cuando no se tiene una clasificación para el conjunto de datos que se desea analizar.

En ese sentido para valorar la calidad de los agrupamientos, se aplicarán métricas de calidad internas o generales que se obtienen con la ejecución del número de particiones solicitadas. Por otro lado, se presentan gráficos que generaron durante las pruebas para conocer las agrupaciones y su comportamiento con la variable EDAD.

La métrica de calidad aplicada fue la suma de las distancias al cuadrado de cada observación con relación al promedio general de la muestra, esta métrica nos indica el porcentaje de variación o separación en las particiones. Es decir, que tan distintos son los datos agrupados dentro de un mismo clúster. Por otro lado, se ofrecen otros parámetros como el tamaño del agrupamiento, clúster, centroides y número de interacciones.

La primera prueba realizada fue con dos agrupaciones con una asignación de 546 instancias en el clúster uno y 474 instancias en el clúster dos. En la figura 85 se observan

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

los vectores generados con el primer modelo, asimismo se pueden apreciar que el porcentaje de disimilitud entre las instancias es del 9.3%, en ese sentido el porcentaje de similitud se puede interpretar como del 90.7%, lo que indica que hubo una correcta agrupación de los datos en esta primera prueba.

Figura 85.

Métricas generales de calidad con dos agrupamientos en Lenguaje R

```
[925] 1 1 1 1 1 2 2 1 1 1 2 2 1 2 1 2 1 2 2 2 1 2 1 1 1 1 1 2 1
[953] 2 1 2 2 2 2 1 2 1 2 2 1 2 2 1 1 1 2 2 2 1 2 1 2 1 1 1 1
[981] 1 2 2 1 2 2 1 2 1 1 2 2 2 1 1 2 1 1 1 2
[ reached getOption("max.print") -> omitted 20 entries ]

Within cluster sum of squares by cluster:
[1] 22230.63 21209.51
(between_SS / total_SS = 9.3 %)
```

Available components:

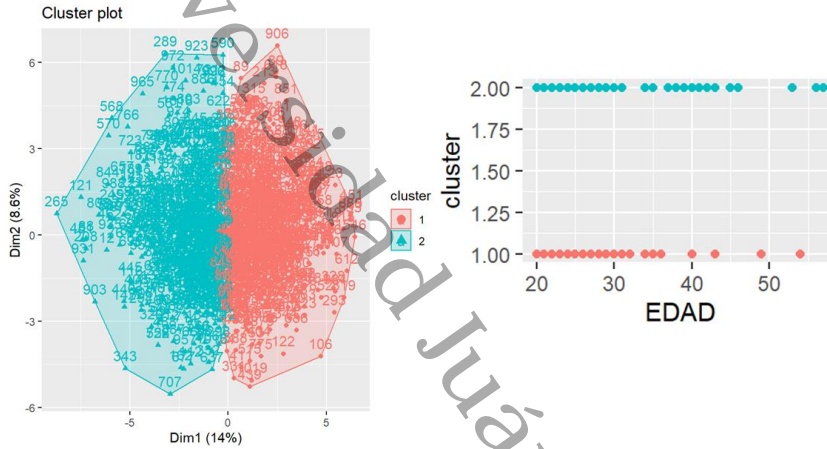
```
[1] "cluster"      "centers"      "totss"        "withinss"
[5] "tot.withinss" "betweenss"    "size"         "iter"
[9] "ifault"
```

Nota: Elaboración propia.

Para representar los clústeres de acuerdo con la métrica obtenida se realizaron dos gráficos Lenguaje R, el primer diagrama presenta las agrupaciones realizadas, en este se detecta que la mayoría de los datos dentro de cada grupo se encuentran relacionados y son pocos los datos dispersos dentro de cada clúster, asimismo se aprecia que los grupos se encuentran separados entre sí haciendo más clara su clasificación. Por otro lado, el diagrama de la derecha está representado por los clústeres y la variable EDAD, este diagrama permite identificar que el grupo dos de color azul contiene una gran cantidad de estudiantes mayores de 30 años, este dato puede influir en los patrones de comportamiento (ver figura 86).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 86. Diagramas de dos agrupaciones en Lenguaje R



Nota: Elaboración propia.

Aunque en este primer ejemplo la calidad de los grupos es aceptable, se debe considerar aumentar el número de clústeres para conocer su comportamiento.

La segunda prueba realizada fue con cuatro agrupaciones con una asignación de 320 instancias en el clúster uno, 208 en el dos, 280 en el clúster tres y en el cuarto 212 instancias. En la figura 87 aprecian los vectores generados en este modelo, en cuanto el porcentaje de disimilitud entre las instancias es del 16.5%, por lo tanto el porcentaje de similitud es del 83.5%.

Figura 87. Vectores con cuatro agrupamientos en Lenguaje R

```
[953] 2 2 4 4 4 4 3 2 2 1 4 1 2 1 2 3 3 2 1 2 3 2 2 4 1 3 3 3
[981] 3 4 2 3 2 4 1 4 2 2 2 2 2 3 1 2 1 3 2 2
[ reached getOption("max.print") -- omitted 20 entries ]

within cluster sum of squares by cluster:
[1] 11088.198 8831.087 10859.017 9208.367
(between_SS / total_SS = 16.5 %)
```

Nota: Elaboración propia.

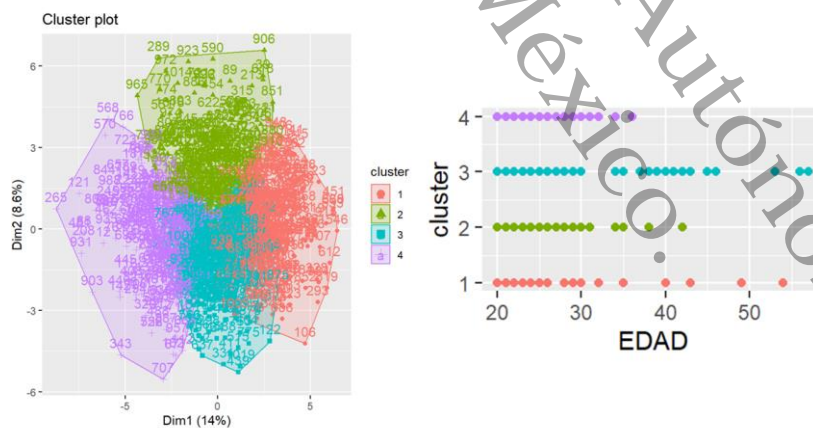
1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Aunque la métrica generada indica que hubo una correcta agrupación de los datos, la calidad del agrupamiento disminuyó y esto se puede apreciar en la figura 88 con el diagrama de agrupaciones realizadas *Cluster plot*, se puede apreciar que los grupos uno de color rojo y tres de color azul se encuentran muy cercanos de modo que los datos pueden confundir su agrupación. Respecto al diagrama de la derecha que representa a los clústeres y la variable EDAD, se identifica que los grupos uno y tres concentran una mayor población de alumnos mayores de 30 años, en cuanto al clúster cuatro se encuentra definido por alumnos en su mayoría menores de 30 años.

Figura 88.

Diagramas de cuatro agrupaciones en Lenguaje R



Nota: Elaboración propia.

En este segundo modelo se observa que cuatro agrupaciones no sería lo óptimo para la interpretación de los resultados, puesto que podrían generarse interpretaciones poco precisas.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

La tercera prueba se realizó con cinco agrupaciones, el tamaño de los grupos se encuentra distribuido de la siguiente manera: 99 instancias en el clúster uno, 229 en el dos, 207 en el clúster tres, en el cuarto 290 instancias y 195 instancias en el último grupo.

En la figura 89 se muestran las últimas líneas de los vectores generados en este modelo, en cuanto el porcentaje de disimilitud entre las instancias es del 18.2%, por lo tanto el porcentaje de similitud es del 81.8%.

Figura 89.

Vectores con cinco agrupamientos en Lenguaje R

```
[953] 5 5 3 3 1 3 2 5 5 4 3 4 5 4 5 2 2 5 4 5 2 5 5 3 4 1 1 1
[981] 2 3 5 2 5 3 4 3 5 5 5 5 5 2 4 5 4 2 5 5
[ reached getOption("max.print") -- omitted 20 entries ]

within cluster sum of squares by cluster:
[1] 4063.263 8411.508 8790.890 9661.544 8229.885
(between_SS / total_SS = 18.2 %)
```

Nota: Elaboración propia.

La métrica generada indica que se obtuvo agrupación aceptable de los datos, pero que la calidad del agrupamiento es aún menor que en los dos modelos previos.

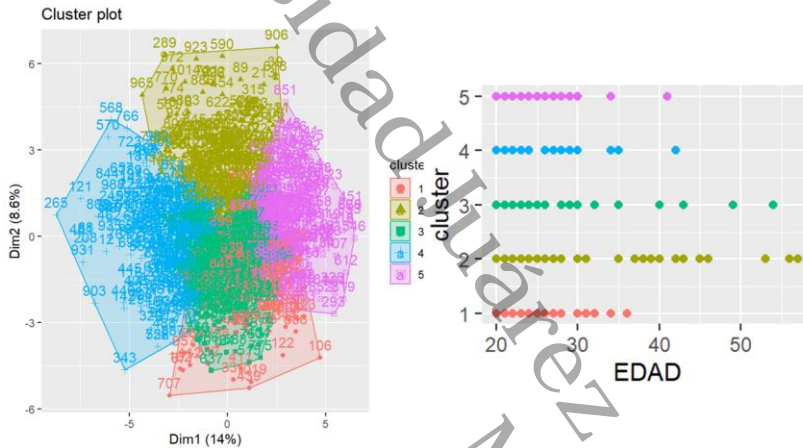
Para representar los clústeres de este último modelo, se presentan los gráficos empleados previamente. El diagrama *Cluster plot* muestra las agrupaciones realizadas, en este se aprecia que los grupos dos y cuatro se encuentran separados entre sí haciendo más clara su clasificación, a diferencia de los grupos uno y tres que se interponen porque no lograron identificar su grupo correctamente. El diagrama de la derecha representado por los clústeres y la variable EDAD, permite identificar que el

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

grupo dos de color mostaza y el grupo tres de color verde contienen una gran cantidad de estudiantes mayores de 30 años, con relación a los otros tres grupos (ver figura 90).

Figura 90.
Diagramas de cinco agrupaciones en Lenguaje R



Nota: Elaboración propia.

Como se presentó previamente, los modelos fueron perdiendo calidad conforme se aumentaba el número de agrupaciones. Aunque estas métricas permitieron conocer la calidad del agrupamiento con relación a la separación entre clústeres se debe complementar con métodos que generen métricas adicionales para obtener el agrupamiento óptimo.

Reglas de asociación.

En este apartado se presentan las pruebas realizadas con el modelo de reglas de asociación aplicando el algoritmo Apriori. Se presentan tres ejemplos con variación en el parámetro de confianza (*confidence*). Aunque las reglas generadas son visibles en la

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

consola de Lenguaje R, se optó por mostrar los resultados en formato *dataframe* para una mejor presentación.

Además de las métricas de soporte y confianza se consideró el parámetro *lift*, que permite medir la calidad de las reglas. El parámetro *lift* es preferible que sea mayor a uno, ya que esto significa que se realizan no por una cuestión aleatoria o de azar, sino que la regla representa un patrón real.

En la primera prueba se obtuvieron un total de 143,060 reglas con una confianza mínima del 50% y un soporte mínimo de 0.1, la mayoría de las reglas generadas (56,192) fueron construidas por cinco *ítems*, en ese sentido las reglas formadas por ítems individuales son los de menor tamaño (ver figura 91).

Figura 91.

Métricas del algoritmo Apriori en Lenguaje R con cincuenta por ciento de confianza
set of 143060 rules

```
rule length distribution (lhs + rhs): sizes
  1      2      3      4      5      6      7      8      9
  4    254   5024 29932 56192 39229 11328  1088    9

mining info:
data ntransactions support confidence
tr          971      0.01      0.5
```

Nota: Elaboración propia.

En esta primera prueba, de todas las reglas generadas se solicitaron únicamente las 10 mejores reglas que de acuerdo con la herramienta son las que ofrecen las mejores métricas y que aparecen con mayor frecuencia sin hacer distinciones en el tamaño de los *itemsets*.

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

De acuerdo con la figura 92, se observa que las primeras cuatro reglas que se generaron están construidas con un solo ítem, en estas reglas no se obtiene información relevante puesto que son datos que se puede obtener en una consulta simple del *dataset*.

En lo que respecta de las reglas cinco a la 10 fueron construidas por un ítem antecedente (*LHS*) y un ítem consecuente (*RHS*).

Figura 92.
Reglas de asociación en Lenguaje R con cincuenta por ciento de confianza

	LHS	RHS	support	confidence	coverage	lift	count
	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>
[1]	{}	{Mujer}	0.582	0.582	1.000	1.000	565.000
[2]	{}	{P.3 No}	0.659	0.659	1.000	1.000	640.000
[3]	{}	{P.43 Sin comentarios}	0.680	0.680	1.000	1.000	660.000
[4]	{}	{P.1 Si}	0.754	0.754	1.000	1.000	732.000
[5]	{P.43 Conformidad}	{Mujer}	0.010	0.556	0.019	0.955	10.000
[6]	{P.43 Conformidad}	{P.3 No}	0.013	0.722	0.019	1.096	13.000
[7]	{P.43 Conformidad}	{P.1 Si}	0.015	0.833	0.019	1.105	15.000
[8]	{P.10 Primaria}	{P.8 Primaria}	0.011	0.579	0.020	4.291	11.000
[9]	{P.10 Primaria}	{P.9 Primaria}	0.010	0.526	0.020	3.872	10.000
[10]	{P.10 Primaria}	{Hombre}	0.010	0.526	0.020	1.262	10.000

Nota: Elaboración propia.

Con lo expuesto previamente, en esta primera prueba con los parámetros establecidos aunque la métrica de *lift* es favorable, la información obtenida no fue relevante y se encuentra limitada.

En la segunda prueba realizada un total de 74,603 reglas aumentado la confianza mínima al 70%, en cuanto al soporte mínimo se mantuvo de 0.1, la mayoría de las reglas

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

construidas (28,365) prevalecen con un tamaño de cinco ítems. Con respecto a las reglas formadas por ítems individuales solo se encuentra una regla (ver figura 93).

Figura 93.

Métricas del algoritmo Apriori en Lenguaje R con setenta por ciento de confianza

```
set of 74603 rules

rule length distribution (lhs + rhs):sizes
  1      2      3      4      5      6      7      8      9
  1      84    1857  12724  28365  23335  7454   777    6
mining info:
data ntransactions support confidence
tr          971      0.01      0.7
```

Nota: Elaboración propia.

En esta segunda prueba, los resultados no muestran las mejores 10 reglas, se realizó una solicitud de los mejores 1800 *itemsets*, esta solicitud considera las reglas con una combinación de hasta tres *ítems*. Con las reglas obtenidas se realizó un filtro en el ítem consecuente, mostrando únicamente combinaciones con la variable P.1, se consideró un segundo filtro que mostrara reglas con un *lift* mayor a 1.13 y un tercer filtro en el parámetro *count* para presentar asociaciones con una frecuencia mínima de 56 veces en el conjunto de datos.

De acuerdo con lo previamente expuesto, se presentan las ocho reglas que se obtuvieron con los parámetros establecidos y realizando filtros en *dataframe*. Se destaca que los resultados obtenidos mejoraron en comparación con la primera prueba y se puede realizar una interpretación de la información obtenida. Por ejemplo, en la regla 588 de los estudiantes que sí cuentan con internet en casa no necesitaron ir a un café internet para

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

tomar sus clases virtuales y consideran que su experiencia con sus profesores y compañeros fue excelente, esta regla se repite 56 veces en el *dataset* (ver figura 94).

Figura 94. Reglas de asociación en Lenguaje R con setenta por ciento de confianza

	LHS	RHS	support	confidence	coverage	lift	count
	All	{P.1 €	All	All	All	1.13€	56 ... 732
[18]	{P.20 En desacuerdo}	{P.1 Si}	0.061	0.868	0.070	1.151	56
[30]	{P.8 Primaria}	{P.1 Si}	0.118	0.878	0.135	1.164	732
[32]	{P.9 Primaria}	{P.1 Si}	0.123	0.902	0.136	1.196	119.000
[83]	{P.3 No}	{P.1 Si}	0.639	0.969	0.659	1.285	620.000
[533]	{P.3 No,P.41 Todo fue excelente. Se interactuó con los profesores y los compañeros a través de diferentes estrategias.}	{P.1 Si}	0.058	0.982	0.059	1.303	56.000
[660]	{P.3 No,P.30 Todos}	{P.1 Si}	0.060	0.951	0.063	1.261	58.000
[692]	{P.3 No,P.9 Universidad}	{P.1 Si}	0.062	0.923	0.067	1.224	60.000
[741]	{P.19 Totalmente en desacuerdo,P.3 No}	{P.1 Si}	0.077	0.904	0.085	1.199	75.000

Showing 1 to 8 of 8 entries (filtered from 800 total entries) Previous 1 Next

Nota: Elaboración propia.

Sin embargo, se consideró una tercera prueba para realizar una depuración mayor de reglas como se muestra en la figura 95 con un total de 23,822 reglas y una confianza mínima del 90%.

Figura 95. Métricas del algoritmo Apriori en Lenguaje R con noventa por ciento de confianza set of 23822 rules

```
rule length distribution (lhs + rhs):sizes
  2   3   4   5   6   7   8   9
  4  231 2514 8025 9104 3538 402  4
mining info:
data ntransactions support confidence
tr          971      0.01      0.9
```

Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

De acuerdo con lo anterior, la mayoría de las reglas construidas (9,104) tiene un tamaño de seis ítems, se depuraron las reglas de un ítem y se conservan solo cuatro reglas de dos ítems.

En esta última prueba se solicitó generar las mejores 2000 reglas para obtener *itemsets* con tamaño cuatro. En la figura 96 se muestran diez reglas obtenidas que no requirieron el uso de filtros para su búsqueda. Con esta última prueba se obtuvo más información que es susceptible de ser interpretada. Por ejemplo, se observa que en la columna del ítem consecuente (*RHS*) dominan resultados con la variable P.1 “Si”.

Figura 96.
Reglas de asociación en Lenguaje R con noventa por ciento de confianza

	LHS	RHS	support	confidence	coverage	lift	count
	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>	<input type="text" value="All"/>
[1991]	{P.11 Preparatoria,P.20 De acuerdo,P.30 Tres o cuatro}	{P.1 Si}	0.022	0.955	0.023	1.266	21.000
[1992]	{P.11 Preparatoria,P.30 Tres o cuatro,P.8 Secundaria}	{P.1 Si}	0.026	0.926	0.028	1.228	25.000
[1993]	{P.11 Preparatoria,P.30 Tres o cuatro,P.9 Secundaria}	{P.1 Si}	0.030	0.935	0.032	1.241	29.000
[1994]	{P.10 Universidad,P.11 Preparatoria,P.30 Tres o cuatro}	{P.3 No}	0.018	0.944	0.019	1.433	17.000
[1995]	{P.10 Universidad,P.11 Preparatoria,P.30 Tres o cuatro}	{P.1 Si}	0.018	0.944	0.019	1.253	17.000
[1996]	{Hombre,P.11 Preparatoria,P.30 Tres o cuatro}	{P.1 Si}	0.023	0.917	0.025	1.216	22.000
[1997]	{P.11 Preparatoria,P.30 Tres o cuatro,P.41 La interacción se centró más en el envío de tareas.}	{P.3 No}	0.026	0.926	0.028	1.405	25.000
[1998]	{P.11 Preparatoria,P.30 Tres o cuatro,P.41 La interacción se centró más en el envío de tareas.}	{P.1 Si}	0.027	0.963	0.028	1.277	26.000
[1999]	{P.11 Preparatoria,P.20 Totalmente de acuerdo,P.30 Tres o cuatro}	{P.1 Si}	0.019	0.900	0.021	1.194	18.000
[2000]	{P.11 Preparatoria,P.3 No,P.30 Tres o cuatro}	{P.1 Si}	0.053	0.962	0.055	1.276	51.000

Showing 1,991 to 2,000 of 2,000 entries

Previous 1 ... 196 197 198 199 **200** Next

Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

4.1.4 Análisis de usabilidad de las herramientas

Con base en la experiencia obtenida en las pruebas realizadas para la elaboración de modelos en las herramientas Orange, WEKA y Lenguaje R se realizó un análisis de usabilidad. Para ello se consideraron los cinco principios básicos de usabilidad en *software*, expuestos por Chanchí *et al.*, (2020) y que serán presentados en los subsecuente.

Facilidad de aprendizaje

En este apartado se describe qué tan rápido es aprender la funcionalidad del sistema, de modo que el usuario sea capaz de realizar correctamente las tareas que requiera. En este sentido se puede decir que la herramienta con una mayor aportación a las técnicas de minería de datos es lenguaje R, debido a la diversidad de opciones para la ejecución de algoritmos y su visualización gráfica. Esta herramienta está especializada para el análisis de datos, sin embargo, requiere de tiempo en capacitación para su uso, si bien es cierto que una persona con poco o nulo conocimiento en programación puede hacer uso de la herramienta, también es cierto que se requiere de una inversión de tiempo para su documentación así como para la ejecución de los algoritmos.

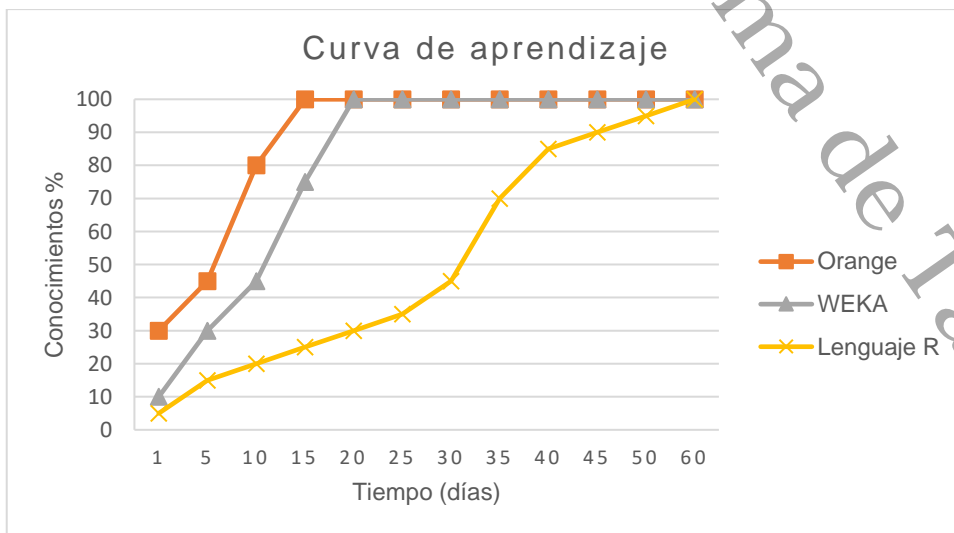
Por otro lado WEKA es la herramienta con mayor uso para el análisis de datos, por ser de uso gratuito, a pesar de ser necesario invertir tiempo en capacitación para ser empleada, su ejecución es más interactiva para el usuario por ser un *software* que no requiere de conocimientos en programación.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En última instancia se tiene a la herramienta Orange, de las tres herramientas es la que más destaca por su facilidad de uso, la documentación para su aplicación es de comprensión sencilla y rápida puesto que se pueden encontrar videos tutoriales para la aplicación de los algoritmos. Además se le atribuye ser un programa muy visual convirtiéndola es una herramienta recomendable para un usuario principiante respecto al tema de análisis y minería de datos.

Por lo anteriormente expuesto, en la figura 97 se presenta el siguiente gráfico con relación al tiempo requerido para el aprendizaje y dominio de la aplicación. Se observa que la aplicación que más tiempo requiere para su aprendizaje es lenguaje R, cabe destacar que en la curva de aprendizaje de R hay un descenso y eso se debe al estancamiento originado por errores en el sistema lo que requirió de tiempo para solucionarlo y avanzar con su uso.

Figura 97.
Curva de aprendizaje de herramientas para minería de datos



1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Nota: Elaboración propia.

Por otro lado, en el caso de *Orange* y *WEKA* una vez que se aprende a usar la aplicación el dominio se vuelve constatable. Sin embargo de las tres herramientas empleadas, este primer parámetro favorece a *Orange*.

Eficiencia.

El número de transacciones que el usuario puede realizar por unidad de tiempo, esto hace referencia a la velocidad de realización de tareas solicitadas por el usuario. En el caso *Orange* y *WEKA* al ser herramientas de uso directo en Canvas la ejecución de los algoritmos la hace en cuestión de segundos y se puede visualizar los gráficos generados para su interpretación. Sin embargo, en el caso de *Orange* los parámetros de evaluación como se mostró en las pruebas realizadas los genera en ventanas separadas y hay que juntas diferentes *widgets* para su funcionamiento a diferencia de *WEKA* que los parámetros de evaluación los ofrece en la consola en un mismo informe.

Para el caso de R, al ser un lenguaje de programación se es más el tiempo que se invierte para la elaboración de un modelo porque hay que escribir las líneas de comando para importar el dataset a la herramienta, configurar los atributos, ejecutar el algoritmo, visualizar los gráficos y elaborar las métricas de evaluación. Requiere invertir varias horas que pueden estar dispersas en días para que las líneas se ejecuten correctamente, pero una vez que se obtiene un modelo funcional se ejecución la realiza en segundos.

Recuerdo del tiempo.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En este atributo se analiza la capacidad de uso del sistema sin tener que hacer un recordatorio exhaustivo partiendo de cero. El recuerdo del tiempo está relacionado con la facilidad de aprendizaje de la herramienta, puesto que entre menos documentación y capacitación se requiera para su uso da como resultado un uso más orgánico del software y una mejor comprensión aunque se deje de usar por días o incluso semanas.

Por lo anterior se puede decir que en el caso de *Orange* y *WEKA* al ser herramientas interactivas una vez que se aprenden se pueden dejar de usar por algún tiempo y al abrir las plataformas reconocer cuáles son las funciones y filtros por utilizar, de igual forma en estas herramientas es fácil recordar cómo cargar un dataset y configurarlo.

En el caso de lenguaje R sí es indispensable llevar una bitácora o apuntes que permitan hacer un recordatorio del por qué se escribió determinada línea de comando y cuál es su función. Sin embargo, el tiempo de preparación previa o recordatorio no es demasiado, solo se requiere unos minutos para realizarlo.

Tasa de errores.

Este parámetro hace referencia al número de errores producidos por el usuario durante la ejecución de tareas para el funcionamiento de los algoritmos de minería de datos en cada herramienta y el modo de solucionarlos.

En el caso de Orange surgieron errores en el proceso de instalación puesto que no permitía instalar el complemento anaconda para el uso del lenguaje de programación

1

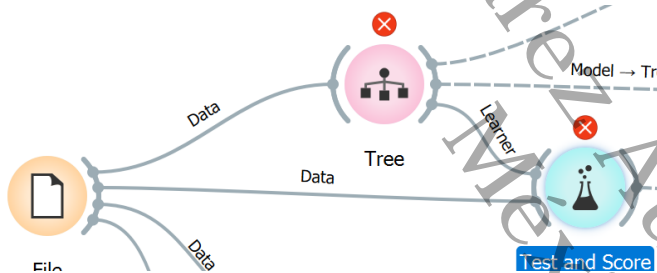
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Python. La solución a este problema fue desinstalar la herramienta, reiniciar el equipo y verificar que las rutas de instalación estuviesen en carpetas no estuviesen ocultas.

Asimismo, cuando se comienza a trabajar con una herramienta surgen detalles en el proceso de ejecución de los algoritmos, como las advertencias en el uso de variables o por ejemplo, cuando no se puede ejecutar el árbol de decisión porque no se seleccionó una pregunta raíz (ver figura 98).

Figura 98.

Error en Orange: no se asignó el nodo raíz



De igual forma existen *widgets* que no se pueden conectar entre sí porque no son compatibles. Este tipo de inconvenientes se resuelven conforme se va aprendiendo el dominio de la aplicación puesto que no son errores del sistema sino la falta de conocimiento del usuario.

En el caso de WEKA, no hubo errores de instalación o inconvenientes con el uso de la plataforma, únicamente aparecían advertencias cuando se utilizaban atributos numéricos y se requería que fuesen nominales. Este tipo de advertencias se solucionan con la aplicación de filtros de tipo no supervisado o usando ítems que contenga el tipo de atributo solicitado por la herramienta.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Sin embargo, con lenguaje R sí se detectaron errores, principalmente en la instalación de librerías que no reconoce el sistema (ver figura 99).

Figura 99.

Error de lenguaje R: no se encontró la función repart.plot

```
Error in repart.plot(arbol) : could not find function "repart.plot"
> plot.repart(arbol)
```

Nota: Elaboración propia.

Para ello se ha requerido instalar manualmente las librerías en la carpeta de instalación del programa y deshabilitar la casilla de modo lectura en la carpeta que resguarda las librerías de R. Con este proceso se han podido ejecutar las librerías dentro de la consola.

Otro error detectado es cuando una línea de comando no fue bien ejecutada por una mala sintaxis al momento de escribir la función (ver figura 99).

Figura 100.

Error de lenguaje R: Error en el argumento escrito

```
> fviz_nbclust(Barrera_Kmeans, FUNCcluster = kmeans, k.max = 8,
diss = matriz_dist, nstart = 50, method = "silhouette")
Error in match.arg(method) :
'arg' should be one of "silhouette", "wss", "gap_stat"
```

Nota: Elaboración propia.

Esto suele ocurrir porque se escribió mal alguna palabra dentro del argumento, se añadieron espacios innecesarios o se ingresaron caracteres especiales como las comillas que no fueron solicitadas dentro de la fórmula. Este tipo de errores suelen aparecer con la sugerencia de sintaxis correcta.

Satisfacción del usuario.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Este atributo puede llegar a ser el más subjetivo en el análisis puesto que depende de la percepción del investigador y la respuesta estriba en la experiencia adquirida. Si bien en las tres herramientas se logró la ejecución de los algoritmos para clasificar, agrupar y asociar reglas con métricas de evaluación superiores al 85%, hubo atributos individuales de las herramientas que se destacan y también son consideradas para optar por una herramienta más que por otra.

En el caso de Orange es una herramienta muy visual, desde la aplicación de los *widgets* hasta la construcción de los gráficos y presentación de reportes. Los parámetros de evaluación durante las pruebas fueron superiores al 95% en los algoritmos ejecutados, aunque, la gama de algoritmos para aplicar reglas de asociación y árboles de decisión es limitada tiene diferentes alternativas para su presentación.

Por otro lado WEKA al tener variedad de algoritmos para realizar árboles de decisión permite tener más alternativas al momento de realizar las pruebas y evaluar la mejor opción para su presentación. Sin embargo, carece de atractivo visual en su interfaz y también en los gráficos que genera.

Por último lenguaje R, ofrece gran variedad de gráficos muy vistosos para su presentación, en el caso de las reglas de asociación ofrece un informe que visualmente permite una mejor comprensión para la exposición de resultados, en el caso de las métricas de evaluación de los modelos también tiene diferentes opciones para su ejecución y presentación que permite al usuario seleccionar de acuerdo con sus

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

preferencias. El inconveniente es el tiempo invertido en pruebas para lograr la construcción de los modelos.

De acuerdo con las pruebas realizadas para la apropiada selección y configuración de las herramientas, en el siguiente apartado se presentará la interpretación de los modelos que fueron seleccionados para la identificación de patrones en los estudiantes. Asimismo, con los atributos descritos en el análisis de usabilidad derivado de las pruebas técnicas realizadas, se presenta una matriz de evaluación de las herramientas y los resultados de esta.

4.2 Interpretación de resultados

En el apartado previo se presentaron las pruebas realizadas, siendo de utilidad para la selección de los modelos que proporcionan evaluaciones de calidad favorables, además de obtener información relevante para discusión de las conclusiones.

La interpretación de resultados se presenta estructurada de acuerdo con las técnicas de minería de datos empleadas: clasificación, agrupación y reglas de asociación. Con el propósito de comparar los resultados obtenidos en cada herramienta de acuerdo con los algoritmos utilizados.

4.2.1 Clasificación

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 P.3 ¿Debes ir a rentar internet para entrar a las plataformas de clases?, por lo tanto, en este apartado se presentan los

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

modelos del árbol de decisión obtenidos en *Orange*, *WEKA* y lenguaje R con la variable P.3 como pregunta raíz.

Orange

De acuerdo con el modelo de árbol de decisión generado en *Orange*, se obtuvieron métricas de evaluación mayores al 90%, destacando que la precisión de clasificación (CA) fue del 98.3% de exactitud (ver figura 101).

Figura 101.

Resultados de métricas de evaluación del modelo de árbol de decisión en Orange

Model	AUC	CA	F1	Precision	Recall
Tree	0.985	0.983	0.983	0.984	0.983

Nota: Elaboración propia.

En ese sentido se presenta en la figura 102 la matriz de confusión para evaluar la correcta clasificación de las variables, las casillas azules representan los datos clasificados correctamente, en este caso para el ítem “No” fue en un 91.5% y para el ítem “Sí” en un 99.7%, los resultados obtenidos demuestran que el modelo tiene una calidad aceptable en su clasificación.

Figura 102.

Matriz de confusión del árbol de decisión en Orange

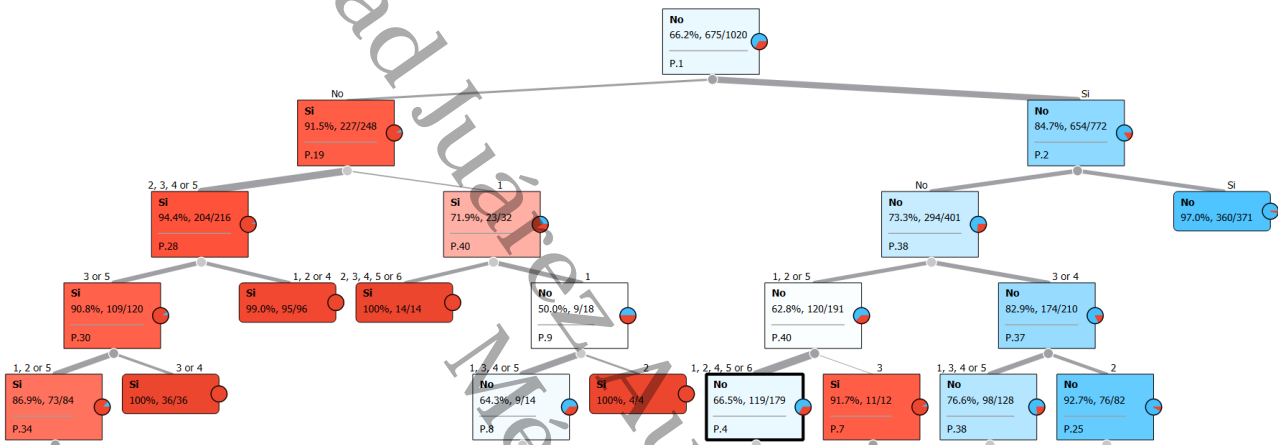
		Predicted		Σ
		No	Si	
Actual	No	99.7 %	0.3 %	675
	Si	4.3 %	95.7 %	345
Σ		688	332	1020

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Nota: Elaboración propia.

De los resultados obtenidos en la evaluación, se presenta el diagrama obtenido con el modelo de árbol de decisión, cuenta con 145 nodos y 73 niveles (ver figura 103).

Figura 103.
Representación gráfica de árbol decisión en Orange



Nota: Elaboración propia.

De acuerdo con el diagrama presentado, las clasificaciones se encuentran identificadas en color rojo para el ítem “Si” y en color azul para el ítem “No”, otorgando una mayor categorización de color azul haciendo referencia a los estudiantes que tienen internet en casa pero necesitan ir a rentar internet para tomar sus clases en línea.

Con los resultados obtenidos en el diagrama, se crea la tabla 13 para una mayor optimización de los resultados y su interpretación:

Tabla 13.
Decisiones del árbol generado en Orange

No.	Clase	Variables	Decisión
1	Si	{P.19= 2, 3, 4 o 5}, {P.28= 3 o 5}, {P.30= 1, 2 o 5}	P.1 No

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

No.	Clase	Variables	Decisión
2	Si	{P.19= 2, 3, 4 o 5}, {P.28= 3 o 5}, {P.30= 3 o 4}	P.1 No
3	Si	{P.19= 2, 3, 4 o 5}, {P.28= 1, 2 o 4}	P.1 No
4	Si	{P.19= 1}, {P.40= 1}	P.1 No
5	No	{P.2=Si}	P.1 Si
6	No	{P.2=No}, {P.38= 3}, {P.37= 1 o 3}	P.1 Si
7	No	{P.2=No}, {P.38= 4}, {P.37= 4 o 5}	P.1 Si
8	No	{P.2=No}, {P.38= 3 o 4}, {P.37= 2}, {P.25= 1 o 2}	P.1 Si
9	No	{P.2=No}, {P.38= 1, 2 o 5}, {P.40= 1, 2, 4, 5 o 6}	P.1 Si

Nota: Elaboración propia.

Las decisiones detectadas en los resultados conforme a la clasificación otorgada permiten generar la siguiente categorización de los patrones identificados:

Clase Si. Estudiantes que requieren acudir a rentar internet para ingresar a clases virtuales.

En esta clasificación se detectaron un total de 345 estudiantes que se identificaron con los patrones enlistados con los números del uno al cuatro de acuerdo con la tabla 13 previamente expuesta, de se describe que acuden a rentar internet porque no cuentan con servicio de internet en casa.

En ese sentido, cabe destacar que las decisiones identificadas como 1, 2 y 3 tienen en común que los estudiantes perciben a las materias teóricas aptas para las clases en línea y consideran el trabajo en equipo complicado en la modalidad virtual. Aunado a esta situación en la decisión uno los estudiantes respondieron que uno o dos profesores e incluso ninguno retroalimentaron sus clases a distancia.

Asimismo, de acuerdo con la decisión dos, hubo estudiantes en la misma situación que respondieron que entre tres o cuatro profesores e incluso casi todos retroalimentaron sus actividades a distancia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En lo que respecta a la decisión 4, las variables relacionadas exponen que se encuentran en desacuerdo en cuestión de que solo las asignaturas teóricas son aptas para clases en línea, también sintieron que su experiencia con tareas y actividades en línea fue negativa porque hubo demasiada carga académica que no comprendían y requería mayor explicación.

Clase No. Estudiantes que no rentaron internet para ingresar a clases virtuales.

En esta clasificación se detectaron un total de 675 estudiantes, de los cuales la decisión describe que no acudieron a un ciber café a tomar sus clases virtuales porque contaba con el servicio de internet en casa. Sin embargo, de esta decisión se desprendieron dos ramificaciones más, la primera con 360 estudiantes se encuentra representada con la decisión 5, significa que cuentan con internet eficiente y en la segunda que desprende más patrones identificados en las decisiones 6, 7, 8 y 9 derivado de que los estudiantes cuentan con internet pero es ineficiente, en esta rama se encontraron.

De acuerdo con lo anterior, en la decisión 6 se describe que los estudiantes consideraron su experiencia con el internet durante clases virtuales regular y el total de profesores con los que sí aprendieron fue de 3 o 4.

A diferencia de la decisión 7 en la que se expone que aunque el internet fue ineficiente la experiencia con el internet en clases virtuales fue buena y consideran que aprendieron con casi todos los profesores. Sin embargo, este patrón solo se encuentra en 29 estudiantes.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En cuanto a la decisión 8 a pesar de haber tenido una experiencia entre regular y buena con el internet, respondieron solo con uno o dos profesores aprendieron durante sus clases en línea. Asimismo, perciben que no tienen conocimientos limitados para llevar sus clases virtuales.

Por último, la decisión 9 describe que la experiencia de los estudiantes con el internet fue frustrante por la inestabilidad de la red y sienten que hubo demasiada carga académica que no comprendían y requería mayor explicación. Esto provocó que su uso en clases virtuales con relación a tareas y actividades fuese negativo.

WEKA

Con las métricas obtenidas en *WEKA*, en la figura 104 se presentan los porcentajes obtenidos del modelo de árbol de decisión, se observa que el 94.80% de instancias se clasificaron correctamente lo que da un total de 967 de las cuales 317 pertenecen a la clase "a" que hace referencia a los estudiantes que requieren ir a rentar internet para tomar sus clases en línea y 650 pertenecen a la clase "b" solo los estudiantes que cuentan con internet en casa y por lo tanto no necesitan rentar internet de acuerdo con la variable P.3 como el nodo raíz empleado.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 104. Métricas de evaluación del modelo de árbol de decisión en WEKA

```

Correctly Classified Instances      967          94.8039 %
Incorrectly Classified Instances    53           5.1961 %
Kappa statistic                    0.8837
Mean absolute error                0.0922
Root mean squared error            0.2147
Relative absolute error            20.5818 %
Root relative squared error        45.3724 %
Total Number of Instances         1020
Ignored Class Unknown Instances    1

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.919   0.037   0.927     0.919   0.923     0.884   0.968    0.933    1
          0.963   0.081   0.959     0.963   0.961     0.884   0.968    0.975    2
Weighted Avg.   0.948   0.066   0.948     0.948   0.948     0.884   0.968    0.961

=== Confusion Matrix ===

  a  b  <-- classified as
317 28 |  a = 1
 25 650 |  b = 2

```

Nota: Elaboración propia.

En la figura 105 que se presenta a continuación, se observa el árbol de decisión de acuerdo con los parámetros de entrenamiento establecidos y el conjunto de datos. El diagrama generado tiene un total de 105 nodos y 62 niveles, por lo tanto se presenta el diagrama completo.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Las decisiones detectadas en los resultados conforme a la clasificación otorgada permiten otorgar la siguiente clasificación de los patrones identificados:

Clase "a". Estudiantes que requieren acudir a rentar internet para ingresar a clases virtuales.

En esta clasificación se encuentran un total de 317 estudiantes, de los cuales la decisión describe que acuden a rentar internet porque no cuentan con servicio de internet en casa. Los patrones identificados están enlistados con los números uno y dos de acuerdo con la tabla 14 previamente expuesta.

El patrón enlistado como la decisión uno, se refiere a los estudiantes que consideran que solo las materias teóricas son aptas para las clases a distancia.

Por otro lado, el segundo patrón que se identificó en esta clasificación fue menciona que además de que los estudiantes están de acuerdo con el hecho de que solo las asignaturas teóricas son aptas para las clases a distancia, también se menciona que en cuanto a recursos tecnológicos tienen un celular con internet, sin embargo la interacción con profesores mediante redes sociales con los profesores inició en la etapa universitario, con respecto a las evaluaciones expresaron que ninguno de sus profesores aclaró cómo evaluarían.

Clase "b". Estudiantes que no rentaron internet para ingresar a clases virtuales.

En esta clasificación se detectaron un total de 650 estudiantes, de los cuales la decisión describe que no acudieron a un ciber café a tomar sus clases virtuales porque contaba con el servicio de internet en casa. En ese sentido se hallaron dos ramificaciones,

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

la primera relacionada con los estudiantes que cuentan con internet eficiente en casa asociado al patrón identificado como número 3 y del cual forman parte 371 estudiantes.

Por otro lado la segunda ramificación relacionada con los estudiantes que tienen internet ineficiente en casa y del cuál se generaron otros patrones identificados por los números 4, 5, 6 y 7.

En el patrón identificado como 4, se detectaron estudiantes que tuvieron una experiencia frustrante con el internet derivado de la inestabilidad en la red, aunque cuentan con equipo de cómputo personal su experiencia con las materias no fue favorable puesto que sintieron que hubo demasiadas tareas que no comprendieron y requerían mayor explicación. En consecuencia los estudiantes respondieron que las clases en línea son más pesadas que las presenciales.

En lo concerniente al patrón 5, describe a los estudiantes que mencionaron haber tenido una experiencia frustrante con el internet derivado de la inestabilidad en la red, cuentan con equipo de cómputo personal e iniciaron a interactuar con ambientes virtuales de aprendizaje en la preparatoria.

De acuerdo con el patrón 6, los estudiantes tuvieron una experiencia regular con el internet, consideran que aprendieron con uno o dos profesores y tuvieron que compartir equipo de cómputo con otros miembros de la familia.

Por último, en el patrón 7 se identificaron estudiantes con una experiencia buena del internet, en lo que respecta a la evaluación de actividades y profesores respondieron que uno o dos evaluaron sólo con un trabajo final.

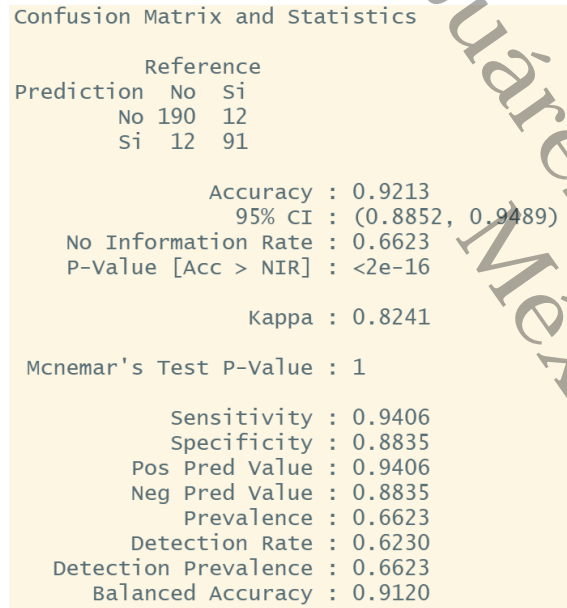
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Lenguaje R.

El último modelo de clasificación a presentar es con lenguaje R, en la figura 106 se presentan las métricas obtenidas de la evaluación, se observa que la precisión obtenida (*Accuracy*) es del 92.13%, que nos indica que clasifica correctamente la mayoría de los datos con una tasa alta de precisión.

Figura 106.

Métricas de evaluación del modelo de árbol de decisión en lenguaje R



Nota: Elaboración propia.

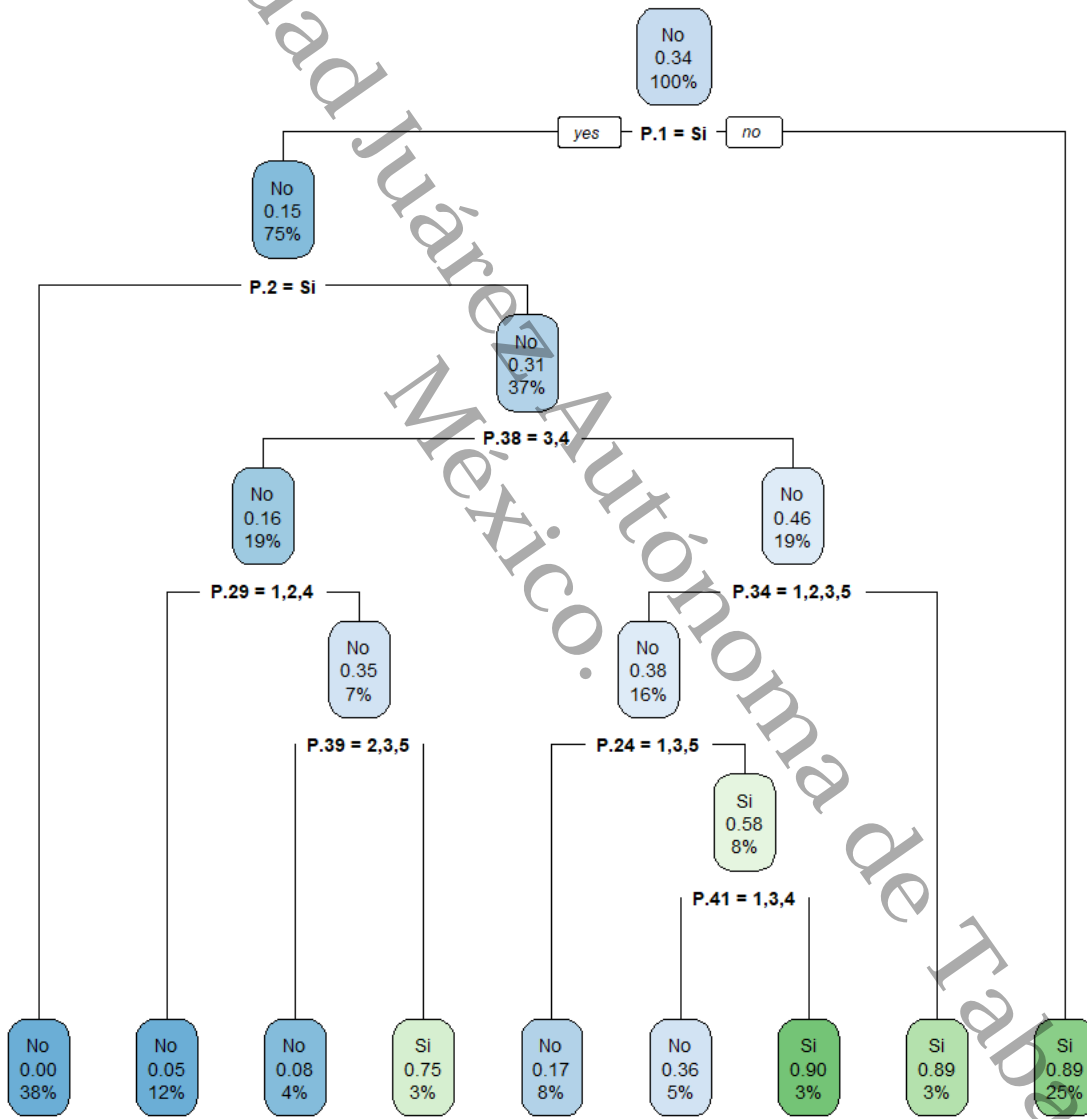
En la figura 107 se presenta el árbol de decisión de acuerdo con los parámetros de entrenamiento establecidos y el conjunto de datos. El diagrama generado tiene un total de 181 nodos, en este diagrama las clasificaciones se encuentran identificadas en color verde para el ítem “Si” y en color azul para el ítem “No”, otorgando una mayor

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

categorización de color azul con relación a los estudiantes que cuentan con internet en casa pero necesitan acudir a un café internet para tomar sus clases en línea.

Figura 107.
Diagrama de árbol Lenguaje R



Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Con los resultados obtenidos en el diagrama, se consideraron las siguientes decisiones presentadas en la tabla 15 para la interpretación de los resultados:

Tabla 15.
Decisiones del árbol generado en lenguaje R

No.	Clase	Variable	Decisión
1	Si	{}	P.1 Si
2	No	{P.2=Si}	P.1 Si
3	No	{P.2=No}, {P.38 = 3 o 4}, {P.29= 1, 2 o 4}	P.1 Si
4	No	{P.2=No}, {P.38 = 3 o 4}, {P.29= 3 o 5}, {P.39= 2, 3 o 5}	P.1 Si
5	No	{P.2=No}, {P.38 = 1, 2 o 5}, {P.34= 1, 2, 3 o 5}, {P.24= 1, 3 o 5}	P.1 Si
6	No	{P.2=No}, {P.38 = 1, 2 o 5}, {P.34= 1, 2, 3 o 5}, {P.24= 2 o 4}, {P.41= 1, 3 o 4}	P.1 Si

Nota: Elaboración propia.

Las decisiones detectadas en los resultados conforme a la clasificación otorgada permiten otorgar la siguiente clasificación de los patrones identificados:

Clase Si. Estudiantes que requieren acudir a rentar internet para ingresar a clases virtuales.

Para el caso de esta clasificación únicamente se considera el patrón de la decisión 1 en el cual se describen a los estudiantes que acuden a rentar internet son los que no cuentan con el servicio en casa.

Clase No. Estudiantes que no rentaron internet para ingresar a clases virtuales.

En esta clasificación se encuentran los estudiantes que no necesitaron acudir a un sitio a rentar internet porque contaban con el servicio en casa. Sin embargo, se detectaron estudiantes con internet eficiente de acuerdo con la descripción del patrón descrito en la decisión 2. En cuanto a los estudiantes que no cuentan con internet eficiente se tuvieron cuatro patrones descritos en las decisiones 3, 4, 5 y 6.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En cuanto a la decisión 3, describe que los estudiantes tuvieron una experiencia que se encuentra entre regular o buena de acuerdo con las respuestas otorgadas y con respecto al envío de planeación de actividades por parte de los profesores mencionaron que solo uno o dos lo realizaban.

A diferencia de la decisión 4 que aunque coincide con una experiencia de entre regular o buena con el internet, difiere con el envío de planeación de actividades por parte de los profesores puesto que las respuestas dadas corresponden a las opciones “todos” o “3 o 4” profesores que cumplieron con ese aspecto. Aunado a esto consideraron una experiencia regular con las asignaturas.

Por otro lado, las decisiones 5 y 6 muestran patrones que coinciden con el nodo de la experiencia con el internet durante las clases en línea, en la cual sus respuestas exponen que fue frustrante por la inestabilidad de la red. Asimismo coinciden en que solo uno o dos profesores realizaron al menos una evaluación parcial en las asignaturas.

Sin embargo, la decisión 5 difiere con la última variable puesto que los estudiantes que caen en este patrón no estuvieron de acuerdo ni en desacuerdo en cuanto a percibir un escaso conocimiento en tecnologías de su parte que dificultara el tomar clases mediante ambientes virtuales.

Con respecto a la decisión 6 están de acuerdo en que su escaso conocimiento en tecnologías dificultó tomar clases a distancia, además en este último patrón los estudiantes experimentaron una interacción con compañeros y profesores limitada al envío de tareas.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

4.2.2 Agrupamiento

Con el fin de detectar patrones el comportamiento en los estudiantes de la IES caso de estudio, y que se encuentran asociados a la apropiación de tecnologías durante las clases virtuales por la pandemia COVID-19, se presentan las agrupaciones obtenidas con el algoritmo K-means, con ello se pretende abordar la interpretación de los resultados obtenidos.

Como se mostró previamente, se examinaron diferentes alternativas de agrupamiento que van desde el número de clústeres y variables a analizar, esto con el objetivo de identificar cuál era la mejor opción para el diseño del modelo de agrupamiento para este proyecto de investigación.

Sin embargo, aunque es de utilidad llevar a cabo diversas pruebas con cada una de las variables, también fue necesario considerar métricas de validación que recomienden el número óptimo de clústeres con el algoritmo K-means, para tomar una decisión con mayor fundamento.

A diferencia de lenguaje R y Orange que ofrecen métricas de calidad internas para evaluar el clúster óptimo, WEKA solo me permite conocer si el modelo funciona correctamente por medio de métricas de evaluación externa como la matriz de confusión y eso solo es posible después de realizar diferentes pruebas de entrenamiento y ver los porcentajes de datos agrupados. Por lo que en este apartado se presentan los puntajes alcanzados con modelos de tres agrupaciones de acuerdo con las métricas de evaluación que ofrece cada modelo.

1

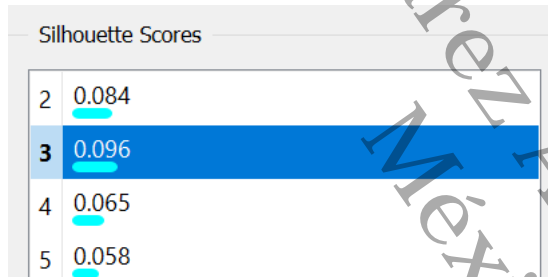
Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Orange

Con la herramienta Orange se trabajó con métricas de calidad interna basados en los puntajes obtenidos con el método de silueta. En la figura 108 se muestran los puntajes alcanzados utilizando como parámetro de dos a cinco agrupamientos. De acuerdo con la información obtenida, el agrupamiento óptimo se obtuvo con tres grupos, incluso se muestra que la misma herramienta resalta el número de clústeres óptimos.

Figura 108.

Métricas de calidad del algoritmo *k*-Means en Orange

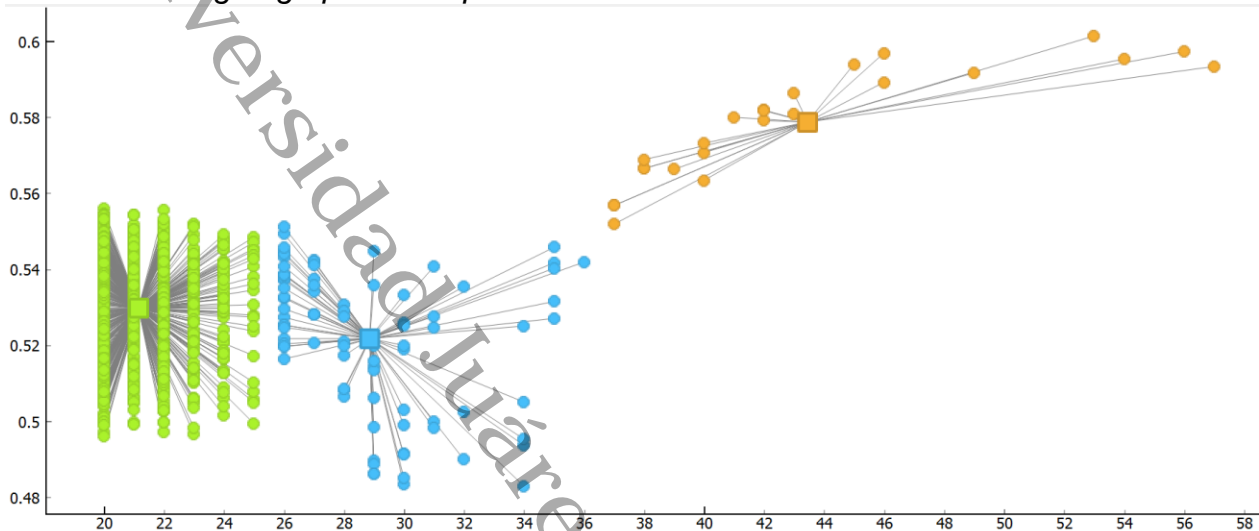


Nota: Elaboración propia.

De acuerdo con las métricas presentadas previamente, se muestra el diagrama de *Interactive k-Means* considerando las variables clúster y EDAD para su representación. Se puede observar que los agrupamientos son claramente identificables y se encuentran distribuidos de la siguiente manera: el clúster verde representa a los estudiantes en un rango de edad de 20 a 25 años, el clúster azul representa a estudiantes de 26 a 36 años y el clúster naranja son estudiantes de 37 años en adelante (ver figura 109).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 109.
K-means Orange agrupamiento por edad



Nota: Elaboración propia.

Con base en la información anterior, las agrupaciones fueron organizadas por la edad de los estudiantes para una mejor comprensión de los patrones de comportamiento y cuál es su relación con las barreras para el uso de innovaciones tecnológicas:

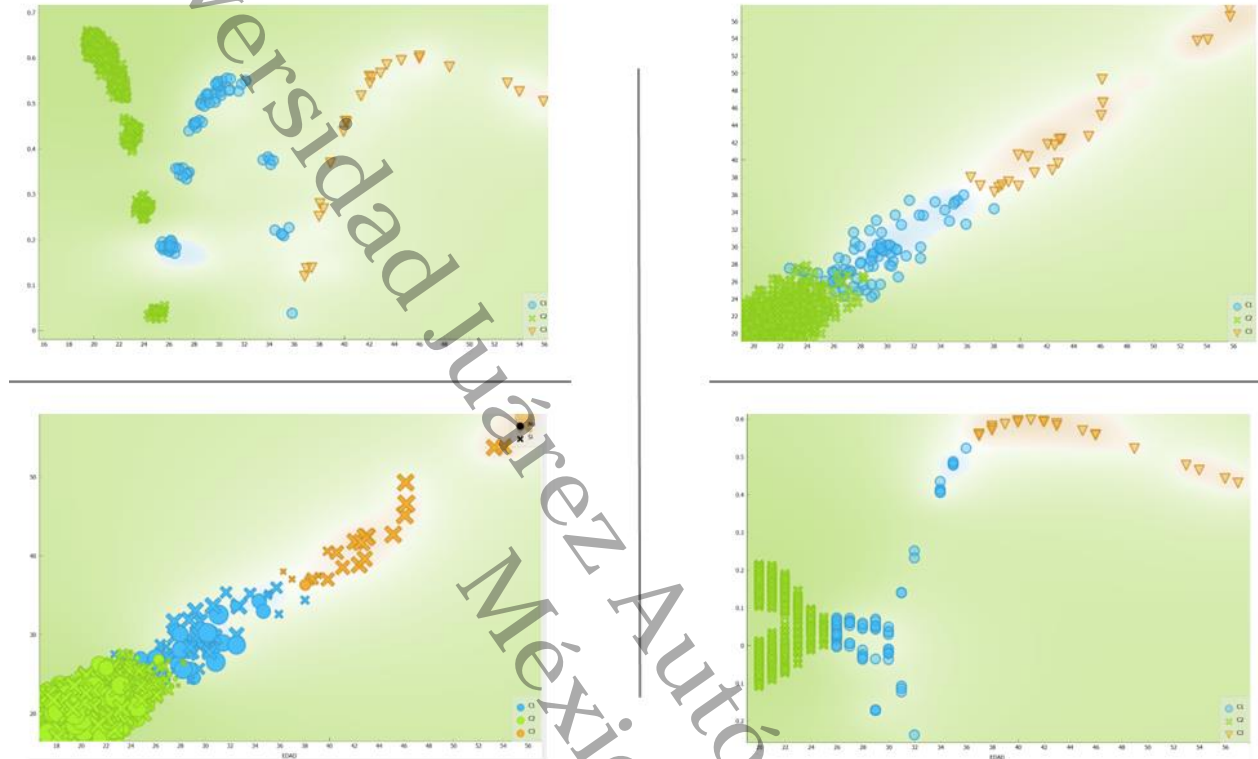
- Uso de TI óptimo.
- Uso de TI promedio.
- Uso de TI bajo.

En la figura 110, se muestran diferentes diagramas de dispersión obtenidos donde se observa la separación y densidad de los clústeres de acuerdo con el modelo de tres grupos presentados con el uso de la variable EDAD de los estudiantes y la variable Clúster siendo esta última una variable producto de las pruebas realizadas y la cual detecta las agrupaciones obtenidas de acuerdo con las métricas establecidas.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 110.
K-means Orange gráficos de dispersión de acuerdo con el clúster



Nota: Elaboración propia.

De acuerdo con los diagramas presentados en la figura 110, existe una mayor densidad en el grupo estudiantes jóvenes de 20 a 25 años, sin embargo, logran ser identificados y tienen características particulares. En cuando al grupo de adulto joven de 26 a 36 años, este grupo guarda características que forman parte del grupo uno y grupo tres, aunque esto no quiere decir que los datos de este grupo permitan confusión en su agrupamiento. En cuanto al grupo de adultos maduros de 37 años en adelante, su lejanía con los primeros grupos también se interpreta como una agrupación que comparte características particulares y distintivas.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Con base en los diagramas obtenidos y el informe de clústeres generado en Orange, se realizó la interpretación de los agrupamientos que permiten identificar los patrones obtenidos:

C3 Uso de TI óptimo. Son estudiantes de 37 a 57 años, que cuentan con internet y una computadora de uso personal para acceder a las clases virtuales. Con referencia al estrés y la ansiedad es un factor neutral en esta clasificación puesto que las respuestas dadas en este cuestionamiento son de acuerdo, desacuerdo y neutral. Por otro lado, aunque se consideran capaces de llevar clases en plataformas virtuales, también opinan que las clases virtuales son más pesadas que las clases presenciales. Asimismo, recomiendan que la institución académica debe capacitar a los estudiantes para el uso de plataformas académicas.

C1 Uso de TI promedio. Esta clase los estudiantes de 26 a 36 años, cuentan con internet y computadora de uso personal en la mayoría de los casos. Por otro lado, consideran que la interacción con sus compañeros y profesores fue escasa y se limitó a la solicitud de tareas. A diferencia de la clasificación anterior, los estudiantes en la mayoría de los casos consideran que las clases virtuales son motivo de estrés y ansiedad por lo que prefieren las clases presenciales que virtuales. Por último también sugieren que la institución académica debe capacitar a los estudiantes para el uso de tecnologías.

C2 Uso de TI bajo. En esta última clase, los estudiantes tienen una edad de 20 a 25 y como se mostró previamente hay una gran parte de la población estudiantil que se identifica en este grupo y guarda relación con el hecho de que es el rango de edad

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

promedio para estudiar una licenciatura. Este grupo se destaca principalmente porque no cuentan con internet o tienen un servicio ineficiente, cuentan con computadora pero es de uso compartido con otros miembros de la familia por lo que tienen que recurrir a otros

2

dispositivos como celular o tableta (en pocos casos). Casi todos los encuestados tienen estrés y ansiedad por la carga de responsabilidades académicas, tuvieron poca interacción con sus profesores y compañeros además de considerar las clases virtuales más pesadas que las presenciales.

Con estas agrupaciones se encontraron comportamientos similares entre los clústeres, principalmente la falta de una buena conexión a internet, los niveles de estrés y ansiedad por la carga académica o la poca interacción con los profesores y compañeros. Pero también se encontraron comportamientos distintos, que guardan relación con los cambios generacionales y que a su vez deriva en la actitud que se toma hacia las circunstancias que se viven por las clases virtuales y la pandemia COVID-19. Haciendo que el uso de innovaciones tecnológicas se perciba de forma negativa como en el caso de un estudiante joven, o de forma resiliente y positiva como en el caso de un estudiante con edad madura.

2

Como se mostró en la interpretación realizada con Orange, los patrones identificados se encuentran relacionados principalmente con las barreras económica, acceso y ambiente académico. No es lo mismo el comportamiento de un estudiante joven en la cual su generación tiene mayor estrés y ansiedad ocasionado por querer las cosas de forma inmediata a diferencia de un adulto mayor que tiende a ser más paciente y

2

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

adaptarse a su contexto. En el caso de un adulto joven que también por su edad tiene comportamientos de ambos grupos. Y estas mismas conductas son las que permiten o no poder llevar clases de forma virtual.

WEKA.

En el caso de la herramienta WEKA, se empleó la matriz de confusión como métricas de evaluación y calidad del agrupamiento. Los resultados obtenidos con tres agrupaciones, se presenta de la siguiente manera: 316 en el clúster cero, 321 en el clúster uno y 384 del clúster dos. Asimismo se menciona que 12 instancias fueron clasificadas incorrectamente siendo igual al 1.17%. De acuerdo con esto, la calidad de las agrupaciones es confiable (ver figura 111).

Figura 111.

Métricas de calidad del algoritmo SimpleKMeans en WEKA

```

Clustered Instances      Classes to Clusters:
0      316 ( 31%)         0  1  2  <-- assigned to cluster
1      321 ( 31%)         310  0  1  | cluster0
2      384 ( 38%)         0 320  4  | cluster1
                               6  1 379  | cluster2

Class attribute: Cluster Cluster 0 <-- cluster0
                        Cluster 1 <-- cluster1
                        Cluster 2 <-- cluster2

Incorrectly clustered instances :      12.0      1.1753 %
    
```

Nota: Elaboración propia.

En la figura 112 se presentan diferentes gráficos realizados con la variable clúster en el eje “X”, en cuanto al eje “Y” se emplearon las variables P.25, P.22, P.1, P.38, P.39 y EDAD. En este caso el clúster 0 se encuentra identificado de color azul, el clúster 1 de rojo y el clúster 3 de color verde. De acuerdo con esta información se expone que en el

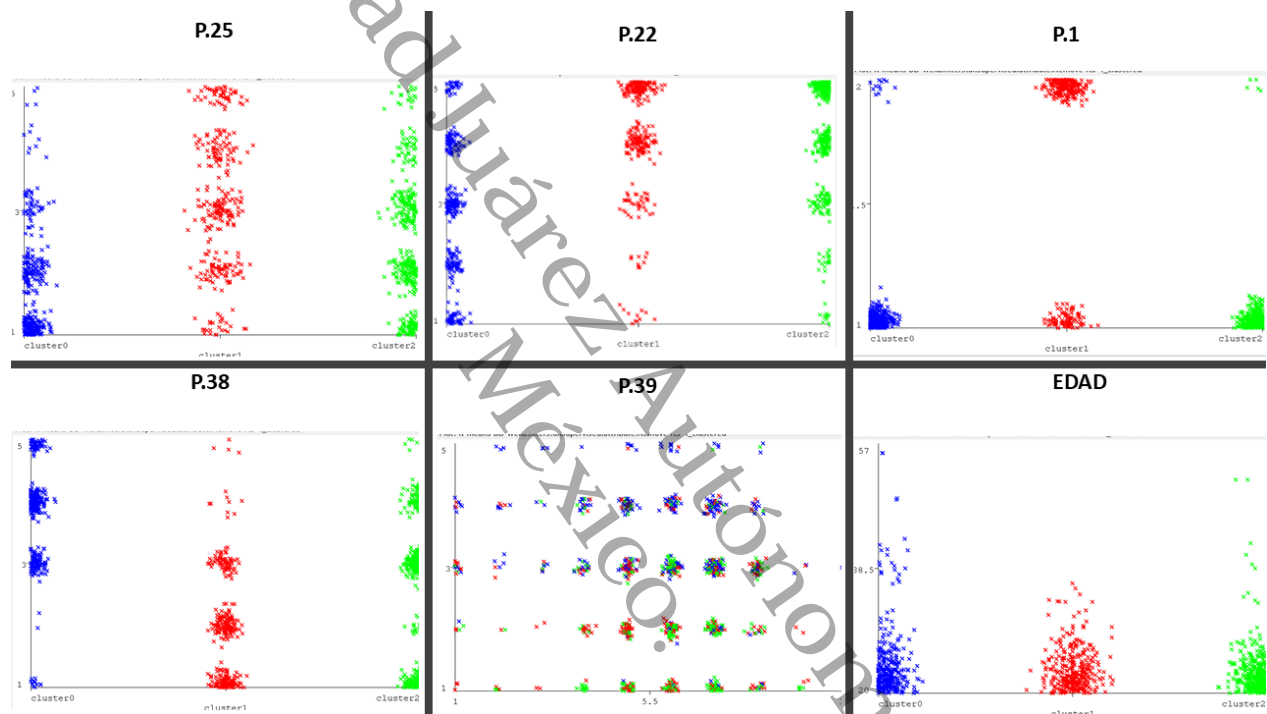
1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

clúster rojo hay una mayor población de estudiantes que consideran tuvieron problemas de accesibilidad para tomar sus clases y eso provocó una experiencia que oscila entre lo regular y frustrante por falta de internet.

Figura 112.

SimpleKMeans WEKA gráficos de dispersión de acuerdo con el clúster



Nota: Elaboración propia.

Por otro lado, en el clúster azul se encuentra la población de estudiantes que consideró haber tenido una experiencia de regular a excelente con el internet y esto también generó una experiencia positiva con sus asignaturas.

A partir de las observaciones realizadas en los diagramas obtenidos y de los clústeres generados en esta herramienta, se realizó la interpretación de los patrones identificados:

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

C0 Uso de TI óptimo. En esta categoría los estudiantes cuentan con servicio de internet eficiente en el lugar donde viven. Por lo tanto, no requieren de ir a un ciber para entrar a sus clases, además cuentan con computadora ya sea personal o de uso compartido para ingresar a las clases virtuales. Sin embargo, perciben que las asignaturas prácticas deben enseñarse presencialmente y que la institución debe la capacitarlos para tomar clases virtuales.

Por otro lado, consideran que alrededor de tres o cuatro profesores son quienes retroalimentan las actividades, mayormente los profesores solicitan a los estudiantes exponer en línea y consideran que sí han aprendido con tres o cuatro de sus profesores.

En cuanto al ambiente de enseñanza virtual, a pesar de que su experiencia con las asignaturas fue regular tienen sentimientos neutrales con respecto al estrés y la ansiedad, consideran que hubo interacción en el aula virtual puesto que los profesores los ponían a trabajar en equipo y aplicaban diversas estrategias de enseñanza y la calidad del internet fue bueno por lo cual se sienten capaces de aprender en línea. (este es el clúster más pequeño y en el que se concentran la mayoría de los adultos maduros y jóvenes adultos).

En este primer grupo, se detectaron diferentes patrones de conducta que se logran relacionar con las barreras pedagógica y del ambiente académico. Esto se debe a que la barrera del acceso y la económica no fueron obstáculos para estos estudiantes puesto que cuentan con elementos como internet eficiente y computadora para ingresar a sus

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

clases en línea. Esto genera en consecuencia que la barrera pedagógica y del ambiente académico se encuentren presentes de forma atenuada.

C1 Uso de TI bajo. En este grupo existen alumnos que no cuentan con internet en casa y deben ir a un ciber para tomar sus clases. Así como los alumnos que sí tienen servicio de internet pero es ineficiente. Indistintamente del servicio de internet, solo cuentan con un equipo de cómputo de uso compartido.

Consideran que las materias prácticas deben ser impartidas de forma presencial y perciben la enseñanza virtual más pesada que la presencial. Por lo tanto, creen que es necesaria la capacitación para clases virtuales.

Por otro lado, son pocos los profesores que retroalimentan las actividades y que solicitan exposiciones en línea. Solo uno o dos profesores han realizado una evaluación parcial y solo con uno o dos profesores han aprendido durante las clases virtuales.

En lo que respecta a sus experiencias en clases virtuales, la calidad del del internet fue frustrante por ser inestable, el aprendizaje en las asignaturas fue pobre debido a su abordaje, con respecto a las tareas y actividades no fue buena debido a la carga de tareas que no fueron comprendidas y requerían explicación, la interacción con los compañeros y profesores fue escasa y se limitó al envío de tareas, sienten estrés y ansiedad aunque tienen una tendencia neutral con respecto a sus capacidades de ser autodidactas.

A diferencia del grupo C1, en este segundo clúster, se logra identificar claramente que las barreras del acceso y económica son obstáculos relevantes que detonan falta de inclusión de los estudiantes a sus clases virtuales, provocando un aumento significativo

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

en las barreras pedagógica y del ambiente académico por la falta de asistencia a las clases y de interacción con los profesores y compañeros.

C2 Uso de TI promedio. En este grupo se encuentran los estudiantes que cuentan con internet ineficiente, optan por no ir a un ciber para tomar sus clases y comparten equipo de cómputo.

Al igual que en los grupos anteriores, perciben que las asignaturas prácticas deben darse presencialmente, asimismo la enseñanza virtual es más pesada que la presencial ocasionando que aprendan menos. Por lo tanto, coinciden en que es necesaria la capacitación para tomar clases virtuales.

En lo que respecta a la evaluación de actividades y profesores, perciben que aprendieron con pocos profesores porque la mayoría de ellos no aclaran cómo será el proceso de evaluación. Solo uno o dos docentes retroalimentaron las actividades, solicitaron exponer en línea y evaluaron trabajos finales.

Por último, en cuanto a sus experiencias en clases virtuales, el servicio de internet y el aprendizaje en las asignaturas fue regular, sienten que hubo demasiadas tareas que no comprendieron y requerían más explicación, y que la interacción en el aula se centró en el envío de tareas. Sin embargo, a pesar de sus experiencias, se consideran capaces de llevar clases virtuales.

Este último grupo se le denominó promedio por ser el caso de la mayoría de los estudiantes encuestados. En este clúster se identifica principalmente que la barrera del acceso, la económica y la del ambiente académico sí se encuentran presentes pero de

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

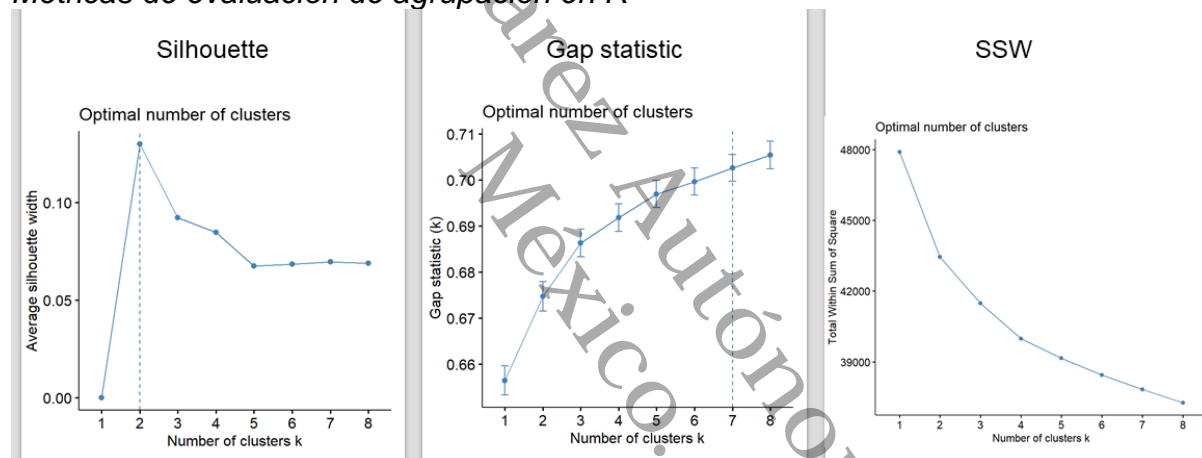
forma atenuada, sin embargo, la barrera pedagógica sigue siendo un obstáculo relevante para el aprendizaje de los alumnos.

Lenguaje R.

Con la herramienta Lenguaje R se revisaron las métricas de calidad interna *SSW* (*Total Within Sum of Square*), *Silhouette* y el método estadístico Gap, para la determinación del clúster óptimo (ver figura 113).

Figura 113.

Métricas de evaluación de agrupación en R



Nota: Elaboración propia.

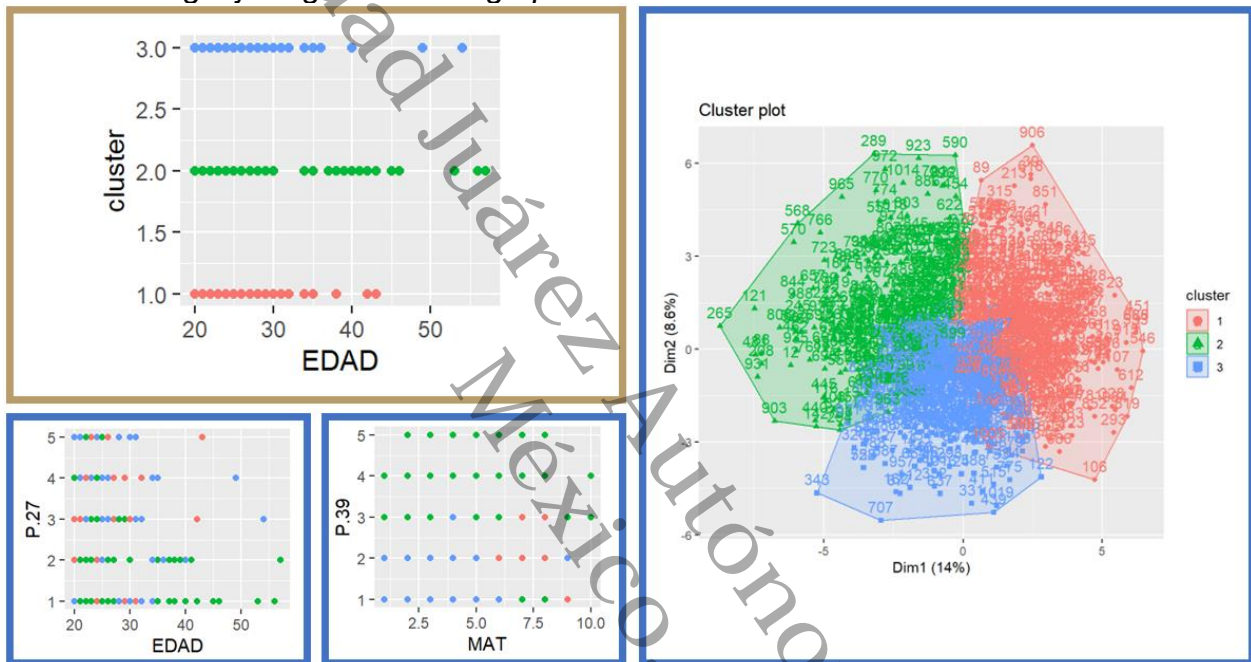
De acuerdo con la información se consideró que el agrupamiento óptimo se obtuvo con tres grupos aplicando la métrica de calidad basada en el método *Silhouette*, obteniendo 0.82 como puntaje de tres clústeres.

Para complementar la evaluación, en la figura 88 se muestra que el porcentaje de disimilitud entre las instancias es del 13.4%, por lo tanto el porcentaje de similitud es del 86.6%, lo que significa que hubo una correcta agrupación de los datos.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En la figura 114 se presentan diferentes gráficos generados con este modelo con el propósito de visualizar el agrupamiento realizado por la herramienta y analizar el comportamiento de diferentes variables.

Figura 114.
KMeans Lenguaje R gráficos de agrupamiento



Nota: Elaboración propia.

Por lo anterior, se puede identificar el clúster 1 de color rojo, clúster 2 de color verde y el clúster 3 de color azul. Asimismo, en el diagrama clúster plot se identifica la separación de los grupos haciendo más clara su clasificación. Por otro lado, se muestra que en el clúster 2 hay una mayor población de estudiantes mayores de 30 años y en este mismo grupo los estudiantes se consideran capaces de aprender solos.

27

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

A partir de las observaciones realizadas en los diagramas obtenidos y de los clústeres generados en esta herramienta, se realizó la interpretación de los patrones identificados:

Clúster 1 Uso de TI promedio. En este grupo los estudiantes cuentan con internet eficiente y no requieren ir a rentar internet porque tienen computadora de uso compartido y celular con internet. En la secundaria comenzaron a utilizar computadora y recursos de internet para realizar tareas. Sin embargo, el contacto con los ambientes virtuales fue durante la universidad.

De los profesores asignados durante el periodo escolar en la mitad de los casos usaron *Microsoft Teams* para mantener comunicación y entre uno o dos usaron otra plataforma virtual, correo electrónico o alguna red social como medio de comunicación principal. Consideran que trabajar con tecnología les produce ansiedad y estrés, no consideran tener escaso conocimiento en tecnologías aunque tienen una postura neutral respecto a ser dependientes o no de la ayuda del profesor.

En cuanto a sus tareas y evaluaciones alrededor de uno o dos profesores enviaron planeación de la materia, retroalimentaron las actividades o revisaron las tareas. Además los alumnos percibieron que fue muy complicado trabajar en equipo virtualmente, Esto generó en consecuencia que los estudiantes consideren que solo con uno o dos aprendieron a distancia.

Por otro lado sus experiencias en el ambiente virtual fueron regulares con el internet así como la interacción con compañeros y profesores. En cuanto a sus tareas y

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

actividades, sintieron que hubo demasiadas tareas que no entendías y requerían mayor explicación.

Clúster 2 Uso de TI óptimo. Los estudiantes sí cuentan con internet eficiente computadora personal y celular con internet. Iniciaron el uso de computadora en la preparatoria, uso de internet en la secundaria y plataformas virtuales a partir de la universidad.

Todos los profesores asignados durante su periodo escolar usaron *Microsoft Teams*, dejando nulo el uso otra plataforma, correo electrónico o redes sociales como medio de interacción principal.

Perciben que no tienen conocimiento tecnológico limitado para las clases virtuales por lo que trabajar con tecnología no les produce ansiedad y estrés.

En cuanto a sus actividades y evaluaciones todos sus profesores enviaron planeación, hubo retroalimentación por parte de los profesores en las actividades, se revisaron las tareas enviadas pocos profesores solicitaron exponer en línea y consideran haber aprendido con casi todos sus profesores.

En cuanto a su experiencia con el internet fue buena, con las asignaturas fue buena porque las tareas y actividades eran moderadas y comprensibles. Esto favoreció su interacción con compañeros y profesores porque generó una buena experiencia.

Clúster 3 Uso de TI bajo. Este grupo de estudiantes sí cuentan con internet pero no es eficiente, por lo tanto, existen estudiantes que requieren ir a un ciber para tomar sus clases en línea. Comparten equipo de cómputo con otros miembros de la familia, la

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

mitad de este grupo no cuenta con teléfono celular con internet. Aunque empezaron a usar computadora e internet a partir de la secundaria fue a partir de la pandemia que interactuaron con ambientes virtuales.

De los profesores asignados en ese periodo solo la mitad de sus profesores utilizaban Teams y solo un profesor utilizaba otras alternativas como correo, redes sociales u otro ambiente virtual como medio principal de comunicación. Perciben que su escaso conocimiento en plataformas virtuales dificultó las clases a distancia por lo que consideran que trabajar con tecnología les produce ansiedad y estrés.

Con respecto a sus actividades de evaluaciones solo uno o dos profesores enviaron planeación de sus actividades, retroalimentaron actividades o aclararon la forma de evaluación. Por lo tanto consideran haber aprendido solo con uno o dos profesores.

En cuanto a su experiencia con el internet fue frustrante por inestabilidad de la red el aprendizaje en las asignaturas tuvo un abordaje pobre en el contenido. Por último sintieron que hubo demasiadas tareas y actividades que no comprendían y requerían mayor explicación.

4.2.3 Reglas de asociación

En este apartado se presentan las reglas de asociación generadas en las herramientas con las métricas que obtuvieron mejores resultados, de acuerdo con las evaluaciones realizadas en cada herramienta se seleccionaron 22 reglas resultantes con la ejecución de los algoritmos de asociación.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

A partir de los resultados obtenidos, se elaboró un diagrama de Venn para detectar las reglas que coincidieron en *Orange*, *WEKA* y lenguaje R. A partir de esta información, se presentan los patrones identificados con esta técnica.

Orange

En este apartado se presentan los *itemsets* generados en *Orange*, para ello se empleó un soporte mínimo del 15% y una confianza mínima del 100%, 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.

De modo que se generaron un total de 22 reglas con la configuración descrita y fueron con un mínimo de dos ítems en los datos antecesores. En ese sentido cabe mencionar que las reglas generadas dan como consecuencia el ítem “sí” en la variable P.1 ¿El lugar donde vives cuenta con internet?, esto significa que los patrones identificados se encuentran relacionados con la barrera del acceso, siendo esto un factor importante para que los alumnos logren interactuar en sus clases virtuales (ver figura 115).

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 115. Reglas de asociación Orange

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent	Consequent
0.170	1.000	0.170	4.462	1.321	0.041	P.3=No, P.4=Si, P.6=Si, P.10=Universidad	→ P.1=Si
0.200	1.000	0.200	3.784	1.321	0.049	P.3=No, P.4=Si, P.9=Secundaria	→ P.1=Si
0.179	1.000	0.179	4.219	1.321	0.044	P.3=No, P.4=Si, P.6=Si, P.9=Secundaria	→ P.1=Si
0.290	1.000	0.290	2.608	1.321	0.071	P.2=Si, P.3=No, P.4=Si	→ P.1=Si
0.160	1.000	0.160	4.736	1.321	0.039	P.2=Si, P.3=No, P.4=Si, P.5=No	→ P.1=Si
0.179	1.000	0.179	4.219	1.321	0.044	P.2=Si, P.3=No, P.5=Si	→ P.1=Si
0.266	1.000	0.266	2.849	1.321	0.065	P.2=Si, P.3=No, P.4=Si, P.6=Si	→ P.1=Si
0.154	1.000	0.154	4.917	1.321	0.037	P.2=Si, P.3=No, P.4=Si, P.5=No, P.6=Si	→ P.1=Si
0.152	1.000	0.152	4.981	1.321	0.037	P.2=Si, P.3=No, P.5=Si, P.6=Si	→ P.1=Si
0.160	1.000	0.160	4.736	1.321	0.039	P.2=Si, P.3=No, P.9=Secundaria	→ P.1=Si
0.162	1.000	0.162	4.679	1.321	0.039	P.2=Si, P.33=Uno o dos	→ P.1=Si
0.158	1.000	0.158	4.795	1.321	0.038	P.2=Si, P.3=No, P.33=Uno o dos	→ P.1=Si
0.159	1.000	0.159	4.765	1.321	0.039	P.2=Si, P.3=No, P.41=La interacción se centró más en el envío de tareas.	→ P.1=Si
0.196	1.000	0.196	3.860	1.321	0.048	P.3=No, P.6=Si, P.11=Universidad	→ P.1=Si
0.162	1.000	0.162	4.679	1.321	0.039	P.3=No, P.4=Si, P.6=Si, P.11=Universidad	→ P.1=Si
0.175	1.000	0.175	4.313	1.321	0.043	P.3=No, P.4=Si, P.25=En desacuerdo	→ P.1=Si
0.150	1.000	0.150	5.046	1.321	0.036	P.3=No, P.4=Si, P.6=Si, P.25=En desacuerdo	→ P.1=Si
0.199	1.000	0.199	3.803	1.321	0.048	P.3=No, P.38=Bastante buena.	→ P.1=Si
0.151	1.000	0.151	5.013	1.321	0.037	P.2=Si, P.3=No, P.38=Bastante buena.	→ P.1=Si
0.165	1.000	0.165	4.595	1.321	0.040	P.3=No, P.4=Si, P.38=Bastante buena.	→ P.1=Si
0.175	1.000	0.175	4.313	1.321	0.043	P.3=No, P.6=Si, P.38=Bastante buena.	→ P.1=Si
0.154	1.000	0.154	4.917	1.321	0.037	P.3=No, P.21=Ni en desacuerdo ni de acuerdo	→ P.1=Si

Nota: Elaboración propia.

Como se presenta previamente el efecto de que los estudiantes cuenten con internet en casa puede influir en las demás variables relacionadas en cada una de las reglas generadas en Orange.

WEKA

En esta herramienta se obtuvieron 153 instancias con un soporte mínimo del 15%, una confianza del 100% y un total de 17 ciclos requeridos por el algoritmo para su ejecución. Las reglas generadas por el algoritmo fueron construidas con un ítem como tamaño mínimo y por seis ítems como tamaño máximo, cabe señalar que la mayoría de los ítemsets se encuentran clasificados entre las reglas con tamaño dos y tres (ver figura 116).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 116.
Resultados de métricas de calidad del algoritmo Apriori en WEKA

```
Apriori
=====

Minimum support: 0.15 (153 instances)
Minimum metric <confidence>: 1
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 91
Size of set of large itemsets L(2): 689
Size of set of large itemsets L(3): 1031
Size of set of large itemsets L(4): 582
Size of set of large itemsets L(5): 131
Size of set of large itemsets L(6): 10
```

Nota: Elaboración propia.

Aunque la herramienta encontró determinado número de instancias con los parámetros establecidos, únicamente se muestran las principales reglas de acuerdo con la solicitud del usuario, para tal caso se requirieron las 22 mejores reglas con un total de dos a seis ítems en los datos antecesores y con la variable P.1 como ítem consecuente en todas las reglas. Por otro lado, en el parámetro *lift* se alcanzó puntaje superior a uno (1.32) en todas las reglas, lo que permite considerar que la información mostrada en los patrones identificados contiene información relevante. (ver figura 117).

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 117.
Resultados de reglas de asociación en WEKA

1. P.2=Si P.3=No P.4=Si 296 ==> P.1=Si 296 <conf:(1)> lift:(1.32) lev:(0.07) [72] conv:(72.19)
2. P.2=Si P.3=No P.4=Si P.6=Si 271 ==> P.1=Si 271 <conf:(1)> lift:(1.32) lev:(0.06) [66] conv:(66.09)
3. P.2=Si P.3=No P.4=Si P.7=No 242 ==> P.1=Si 242 <conf:(1)> lift:(1.32) lev:(0.06) [59] conv:(59.02)
4. P.2=Si P.3=No P.4=Si P.6=Si P.7=No 220 ==> P.1=Si 220 <conf:(1)> lift:(1.32) lev:(0.05) [53] conv:(53.65)
5. P.3=No P.4=Si P.7=No P.33=Uno o dos 208 ==> P.1=Si 208 <conf:(1)> lift:(1.32) lev:(0.05) [50] conv:(50.73)
6. P.3=No P.4=Si P.9=Secundaria 204 ==> P.1=Si 204 <conf:(1)> lift:(1.32) lev:(0.05) [49] conv:(49.75)
7. P.3=No P.38=Bastante buena. 203 ==> P.1=Si 203 <conf:(1)> lift:(1.32) lev:(0.05) [49] conv:(49.51)
8. P.3=No P.6=Si P.11=Universidad 200 ==> P.1=Si 200 <conf:(1)> lift:(1.32) lev:(0.05) [48] conv:(48.78)
9. P.3=No P.4=Si P.7=No P.9=Secundaria 185 ==> P.1=Si 185 <conf:(1)> lift:(1.32) lev:(0.04) [45] conv:(45.12)
10. P.2=Si P.3=No P.5=Si 183 ==> P.1=Si 183 <conf:(1)> lift:(1.32) lev:(0.04) [44] conv:(44.63)
11. P.3=No P.4=Si P.6=Si P.9=Secundaria 183 ==> P.1=Si 183 <conf:(1)> lift:(1.32) lev:(0.04) [44] conv:(44.63)
12. P.3=No P.4=Si P.25=En desacuerdo 179 ==> P.1=Si 179 <conf:(1)> lift:(1.32) lev:(0.04) [43] conv:(43.65)
13. P.3=No P.6=Si P.38=Bastante buena. 179 ==> P.1=Si 179 <conf:(1)> lift:(1.32) lev:(0.04) [43] conv:(43.65)
14. P.3=No P.7=No P.38=Bastante buena. 179 ==> P.1=Si 179 <conf:(1)> lift:(1.32) lev:(0.04) [43] conv:(43.65)
15. P.3=No P.4=Si P.5=No P.7=No 178 ==> P.1=Si 178 <conf:(1)> lift:(1.32) lev:(0.04) [43] conv:(43.41)
16. P.3=No P.4=Si P.6=Si P.7=No P.33=Uno o dos 178 ==> P.1=Si 178 <conf:(1)> lift:(1.32) lev:(0.04) [43] conv:(43.41)
17. P.3=No P.4=Si P.6=Si P.10=Universidad 173 ==> P.1=Si 173 <conf:(1)> lift:(1.32) lev:(0.04) [42] conv:(42.19)
18. P.3=No P.6=Si P.7=No P.11=Universidad 172 ==> P.1=Si 172 <conf:(1)> lift:(1.32) lev:(0.04) [41] conv:(41.95)
19. P.3=No P.4=Si P.38=Bastante buena. 168 ==> P.1=Si 168 <conf:(1)> lift:(1.32) lev:(0.04) [40] conv:(40.97)
20. P.2=Si P.3=No P.5=Si P.7=No 168 ==> P.1=Si 168 <conf:(1)> lift:(1.32) lev:(0.04) [40] conv:(40.97)
21. P.3=No P.4=Si P.5=No P.6=Si P.7=No 166 ==> P.1=Si 166 <conf:(1)> lift:(1.32) lev:(0.04) [40] conv:(40.48)
22. P.3=No P.4=Si P.6=Si P.7=No P.9=Secundaria 166 ==> P.1=Si 166 <conf:(1)> lift:(1.32) lev:(0.04) [40] conv:(40.48)

Nota: Elaboración propia.

En esta segunda presentación de métricas de calidad con la herramienta *WEKA* se observa que los resultados coinciden con los parámetros obtenidos en *Orange*, asimismo se detectó con los patrones en las reglas resultantes también coinciden con la variable P.1 relacionado con el acceso a internet en los estudiantes.

Lenguaje R

Para la generación de resultados en la herramienta Lenguaje R se obtuvieron un total de 33 reglas con una confianza mínima del 100% y un soporte mínimo de 0.15, la mayoría de las reglas generadas fueron construidas por cinco ítems y las reglas formadas por tres ítems son las de menor tamaño (ver figura 118).

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 118. Resultados de métricas de calidad del algoritmo Apriori en Lenguaje R set of 33 rules

```
rule length distribution (lhs + rhs):sizes
 3  4  5  6
 3 11 14  5
mining info:
 data ntransactions support confidence
  tr          971      0.15          1
```

Nota: Elaboración propia.

En ese sentido en la figura 119 se presentan diez de las 22 reglas generadas en dataframe y que de acuerdo con la herramienta son las que ofrecen las mejores métricas y que aparecen con mayor frecuencia sin hacer distinciones en el tamaño de los itemsets.

Figura 119. Resultados de reglas de asociación en Lenguaje R

Show 10 entries Search:

	LHS	RHS	support	confidence	coverage	lift	count
[5]	{P.3 No,P.38 Bastante buena.,P.6 Si}	{P.1 Si}	0.170	1.000	0.170	1.327	165.000
[18]	{P.11 Universidad,P.3 No,P.6 Si,P.7 No}	{P.1 Si}	0.170	1.000	0.170	1.327	165.000
[6]	{P.3 No,P.38 Bastante buena.,P.7 No}	{P.1 Si}	0.169	1.000	0.169	1.327	164.000
[17]	{P.11 Universidad,P.3 No,P.4 Si,P.6 Si}	{P.1 Si}	0.167	1.000	0.167	1.327	162.000
[21]	{P.2 Si,P.3 No,P.5 Si,P.7 No}	{P.1 Si}	0.165	1.000	0.165	1.327	160.000
[3]	{P.2 Si,P.33 Uno o dos}	{P.1 Si}	0.162	1.000	0.162	1.327	157.000
[4]	{P.3 No,P.38 Bastante buena.,P.4 Si}	{P.1 Si}	0.158	1.000	0.158	1.327	153.000
[9]	{P.2 Si,P.3 No,P.9 Secundaria}	{P.1 Si}	0.158	1.000	0.158	1.327	153.000
[11]	{P.2 Si,P.3 No,P.33 Uno o dos}	{P.1 Si}	0.158	1.000	0.158	1.327	153.000
[19]	{P.2 Si,P.3 No,P.4 Si,P.5 No}	{P.1 Si}	0.158	1.000	0.158	1.327	153.000

Nota: Elaboración propia.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

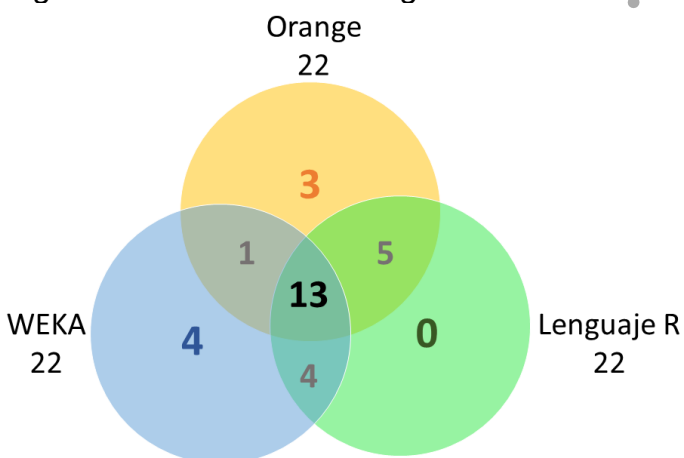
Con los resultados expuestos en esta última herramienta se obtuvo más información que es susceptible de ser interpretada. Asimismo se detectaron coincidencias con los resultados generados en *Orange* y *WEKA* por ejemplo, se observa que en la columna del ítem consecuente dominan resultados con la variable P.1 “Si” y el *lift* obtenido en los *ítemsets* es igual a 1.327 como en resultados presentados previamente con esta técnica.

Interpretación de las reglas obtenidas.

Con base en las 22 reglas generadas en cada una de las herramientas, se buscaron diferencias y similitudes que permitieran una selección e interpretación eficiente de la información. En ese sentido la relación entre las reglas presentadas en Orange, WEKA y Lenguaje R se presenta en un diagrama de Venn que muestra el número de reglas que coincidieron en dos o tres herramientas (ver figura 120).

Figura 120.

Diagrama de Venn con las reglas de asociación



Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Como se muestra previamente en la figura 120, se obtuvieron un total de 13 reglas que coinciden en Orange, WEKA y lenguaje R así como se detectan un total de 10 reglas que coinciden en dos de las tres herramientas. Cabe destacar que todas las reglas generadas en lenguaje R coincidieron con al menos una herramienta.

Con base en lo anterior se seleccionaron 16 reglas que fueron consideradas por la información que aportan presentan en la tabla para su interpretación. Para tal caso se omitirá interpretar la variable consecuente de las reglas porque en todos los casos que se presentan es la variable P.1 "Sí" donde se asume que el alumno cuenta con internet en casa:

Reglas generadas en Orange, WEKA y lenguaje R.

1. $\{P.3=\text{No}, P.4=\text{Si}, P.9=\text{Secundaria}\} = \{P.1=\text{Si}\}$

Son 204 ítems que cumplen con esta regla, son estudiantes que no tienen que ir a rentar internet, cuentan con computadora personal y comenzaron a utilizar recursos de internet en la secundaria para realizar sus actividades académicas.

2. $\{P.3=\text{No}, P.4=\text{Si}, P.6=\text{Si}, P.9=\text{Secundaria}\} = \{P.1=\text{Si}\}$

Esta regla cuenta con las mismas variables del primer ítem sin embargo, se adiciona que además cuentan con la variable P.6 que significa que también tienen un celular con internet. Por lo tanto disminuye el número de estudiantes en estas circunstancias puesto que 183 ítems cumplen con la regla.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

$$3. \{P.2=Si, P.3=No, P.4=Si\} = \{P.1=Si\}$$

Este es el patrón que más se repite con una frecuencia de 296 veces, son estudiantes que cuentan con internet eficiente y computadora para ingresar a clases virtuales por lo tanto, no requieren ir a un café internet a rentar una computadora.

$$4. \{P.2=Si, P.3=No, P.5=Si\} = \{P.1=Si\}$$

Son 183 ítems que cumplen con la regla, son estudiantes que cuentan con internet eficiente en casa, no requiere acudir a un café internet para rentar una computadora aunque comparten una misma computadora con otros miembros de la familia.

$$5. \{P.2=Si, P.3=No, P.4=Si, P.6=Si\} = \{P.1=Si\}$$

Son 271 ítems que cumplen con la regla, son estudiantes que cuentan con internet eficiente en casa como recurso disponible por lo que no requieren ir a un café internet. En lo que respecta a recursos tecnológicos afirmaron tener computadora y celular con internet.

$$6. \{P.2=Si, P.3=No, P.9=Secundaria\} = \{P.1=Si\}$$

Son estudiantes que tienen internet eficiente en casa y no requieren ir a un café internet. Por otro lado comenzaron a utilizar recursos de internet en la secundaria para realizar sus actividades académicas. Esta regla se presentó en 153 encuestas realizadas por los alumnos.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

7. {P.2=Si, P.33=Uno o dos} = {P.1=Si}

En esta regla cumplen 157 ítems. Los alumnos coincidieron en tener como recurso disponible internet eficiente en casa aunque también argumentaron que entre uno o dos profesores solicitaron exponer en línea como parte de la evaluación de actividades.

8. {P.3=No, P.6=Si, P.11=Universidad} = {P.1=Si}

En este caso 200 ítems cumplen con la regla, los estudiantes se encuentran en la situación de no necesitar acudir a un café internet para tomar clases en línea, disponen de celular con internet y a partir de la universidad comenzaron a interactuar por redes sociales con sus profesores para actividades escolares.

9. {P.3=No, P.4=Si, P.6=Si, P.11=Universidad} = {P.1=Si}

Esta regla cuenta con las mismas variables del *ítemset* anterior sin embargo, se adiciona la variable P.4 que significa que también tienen un equipo de cómputo. Al adicionar esta última variable disminuyó el número de estudiantes en estas circunstancias puesto que esta regla se presentó en 162 encuestas.

10. {P.3=No, P.4=Si, P.25=En desacuerdo} = {P.1=Si}

Son estudiantes que disponen de internet en casa y equipo de cómputo por lo tanto, no requieren acudir a un café internet. Además se perciben que tienen conocimiento tecnológico para llevar clases en plataformas virtuales. En este caso la regla se encuentra 179 veces dentro del conjunto de datos.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

11. {P.3=No, P.38=Bastante buena} = {P.1=Si}

Son 203 ítems que cumplen con la regla, son estudiantes que cuentan con internet en casa y no necesitan acudir a un ciber o café internet para rentar equipo de cómputo. Por lo tanto, consideran que su experiencia con el internet fue bastante buena.

12. {P.3=No, P.4=Si, P.38=Bastante buena} = {P.1=Si}

En este caso 168 ítems cumplen con la regla, son estudiantes que no necesitan acudir a un ciber o café internet para tomar sus clases en línea porque cuentan con internet en casa y equipo de cómputo. Por lo tanto, consideran que su experiencia con el internet fue bastante buena.

13. {P.3=No, P.6=Si, P.38=Bastante buena} = {P.1=Si}

Son 179 ítems que cumplen con la regla y expone que los estudiantes no necesitan acudir a un ciber o café internet para tomar sus clases en línea porque disponen de recursos como internet en casa y celular con datos. Al igual que en la regla doce, los alumnos consideran que su experiencia con el internet fue bastante buena. Cabe señalar que la regla trece no significa que los estudiantes que cuentan con celular no cuentan con equipo de cómputo. La regla se limita a exponer que se identificaron estudiantes que caen en la situación presentada.

A continuación se presentan tres reglas en las que únicamente coinciden dos de las tres herramientas empleadas:

Orange y lenguaje R.

14. {P.2=Si, P.3 No, P.41=La interacción se centró en el envío de tareas} = {P.1=Si}

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Este patrón tiene 149 ítems que cumplen con la regla y presenta a estudiantes que cuentan con internet eficiente en casa y no necesitan acudir a un café internet para tomar sus clases en línea. Sin embargo esto no significa que estén satisfechos con su experiencia tomando clases virtuales puesto que consideraron que la interacción en el aula virtual se centró en el envío de tareas.

Orange y WEKA

15. {P.3=No, P.4=Si, P.6=Si, P.10=Universidad} = {P.1=Si}

Son estudiantes que disponen de recursos como internet, equipo de cómputo y celular con datos para acceder a sus clases virtuales por lo que no acuden a un café internet. En lo que respecta al tiempo de uso de aulas virtuales, los estudiantes iniciaron su interacción a partir de la universidad. Esta regla se repite 173 veces dentro del conjunto de datos.

WEKA y lenguaje R

16. {P.2=Si, P.3=No, P.5=Si, P.7=No} = {P.1=Si}

Este caso tiene 166 ítems que cumplen con la regla, son estudiantes que disponen con internet eficiente en casa y no requieren acudir a un café internet, el equipo de cómputo es de uso compartido con otros miembros de la familia y no cuentan con una tableta como dispositivo adicional para atenuar esta situación.

De acuerdo con lo expuesto previamente, los patrones identificados se encuentran presentes dentro del conjunto de datos analizados en un intervalo de 162 a 296 repeticiones sobre determinado suceso.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Para el caso de las reglas de asociación se logra identificar que la barrera del acceso se encuentra atenuada dentro de los patrones hallados, esto se debe a que domina el hecho de que los estudiantes sí cuentan con internet en sus casas como se presentó en todas las reglas. De forma particular se identificó que los estudiantes han tenido acceso a internet desde temprana edad (secundaria) para llevar a cabo sus actividades escolares, este antecedente permitió que a pesar de que la interacción con ambientes virtuales inició durante su etapa universitaria, no fuese un obstáculo para el uso de plataformas académicas. Esta circunstancia se puede encontrar en la mayoría de las reglas (1-3, 5-6, 8-9 y 15) lo que se puede interpretar de forma favorable para el estudiante.

Sin embargo, también se identificó que la ausencia o atenuación de una de las barreras como el caso del acceso, no significa que las barreras restantes como la económica, pedagógica y del ambiente académico no afecten el uso de innovaciones tecnológicas en los estudiantes.

Por ejemplo, se encontró que la regla 7 se relaciona con la barrera pedagógica de forma negativa con el factor extrínseco porque a pesar de que el alumno no tiene problemas de accesibilidad, consideró que solo uno o dos profesores solicitaron actividades en línea como el caso de las exposiciones. A diferencia de la regla 10 que también guarda relación con la barrera pedagógica de forma favorable con el factor intrínseco es decir, desde la percepción del alumno, aquí se destaca que el disponer con

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

recursos tecnológicos favoreció para que los alumnos consideren tener los conocimientos necesarios para llevar clases virtuales.

Asimismo, se menciona que la barrera del acceso tuvo un efecto positivo en la barrera del ambiente académico, esta afirmación se demuestra con las reglas 11,12 y 13 donde los estudiantes al disponer de internet y algún dispositivo tecnológico como celular o computadora lograron tener una experiencia buena con respecto al internet y sus clases virtuales. Por otro lado en la regla 14 se detectaron casos de estudiantes que a pesar de contar con acceso a internet su experiencia dentro del ambiente virtual se limitó al envío de tareas, esto refleja que la barrera del ambiente académico afectó la interacción entre los alumnos y sus profesores.

Por último, se encontró que las reglas 4 y 16 están asociadas con la barrera económica, en estos casos aunque los estudiantes cuentan con acceso eficiente de internet, son alumnos que comparten equipo de cómputo e incluso no cuentan con una tableta como dispositivo alternativo para tomar sus clases virtuales. Esto quiere decir que la barrera económica no se atenúa con el hecho de disponer de recursos tecnológicos como el acceso a internet y una computadora, ya que intervienen aspectos como los ingresos familiares que inducen al estudiante a compartir dispositivos, esto da lugar a interpretar que el aspecto económico influyó de tal modo que se consideró el servicio de internet por encima de adquirir un equipo de cómputo adicional.

De acuerdo con la identificación de los patrones con la técnica de reglas de asociación y su relación con las barreras: acceso, económica, pedagógica y del ambiente

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

académico, queda expuesto que existen reglas donde se demuestra la afectación de estas barreras en los estudiantes. Sin embargo, estas reglas que reflejan una influencia negativa son solo cuatro, por lo que las doce reglas restantes muestran una atenuación de las barreras y esto se debe a la importancia del acceso a recursos tecnológicos, principalmente el servicio de internet.

39

4.2.4 Resultados del análisis de usabilidad

En este apartado se presentan los resultados del análisis de usabilidad realizado, como se muestra en la tabla 16 se requirió definir una escala de evaluación para calificar a las herramientas de acuerdo con los cinco atributos de usabilidad: facilidad de aprendizaje, eficiencia, recuerdo en el tiempo, tasa de errores y satisfacción. Conforme a la descripción realizada en el apartado de pruebas realizadas por herramienta.

Tabla 16.

Escala de ponderación de las herramientas

Escala	Descripción	Puntaje asignado
Alto	El funcionamiento de la herramienta en el parámetro evaluado es correcto y facilita al usuario su aplicación.	3
Medio	El funcionamiento de la herramienta en el parámetro evaluado contiene áreas de mejora que requieren ser atendidas pero no impiden el uso del software.	2
Bajo	El funcionamiento de la herramienta en el parámetro evaluado contiene deficiencias que requieren ser atendidas, aunque no impiden el uso del software, puede convertirse en un problema de comprensión y usabilidad.	1

Nota: Elaboración propia.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

En la tabla 17, se presenta la matriz de calificación de las herramientas, para tal caso los cinco atributos considerados para evaluación contienen características que sirvieron de referencia. En ese sentido, cabe señalar que el total de características fueron trece, considerando que el puntaje más alto de acuerdo con la escala es tres, entonces la puntuación máxima que pueden obtener las herramientas es 39.

Tabla 17.
Matriz de evaluación para las herramientas

Parámetros	Características	Herramientas		
		Orange	WEKA	Lenguaje R
<i>Facilidad de aprendizaje</i>	Entendimiento de los datos a ingresar y los resultados obtenidos.	3	2	2
	Cierre de la aplicación.	3	2	3
	Se identifican fácilmente las figuras, las tablas, las zonas activas y el tipo de acción que se debe ejecutar.	3	3	2
<i>Eficiencia</i>	La herramienta no requiere volver a escribir la información solicitada anteriormente.	3	2	3
	Si una tarea tiene opciones por defecto, están a disposición del usuario.	3	3	1
	En caso de que un algoritmo requiera varios pasos es posible volver a los anteriores para modificar los datos.	3	3	3
<i>Recuerdo en el tiempo</i>	No se necesita recurrir a la ayuda en caso de no recordar el uso de la herramienta.	3	3	2
	Es fácil de recordar como ejecutar una tarea en la herramienta.	3	3	2
<i>Tasa de errores</i>	El mensaje de error propone una solución.	2	3	2
	La herramienta está diseñada para necesitar el mínimo de ayuda en la resolución de errores.	3	3	1

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Parámetros	Características	Herramientas		
		Orange	WEKA	Lenguaje R
	El acceso a la ayuda está en una zona visible y reconocible.	3	1	3
	Se requiere poco tiempo de documentación para solucionar el error.	3	3	1
<i>Satisfacción</i>	La apariencia de la herramienta es agradable y sencilla.	3	2	2

Nota: Elaboración propia con base en datos de Chanchí *et al.*, (2020).

De acuerdo con las características calificadas en la matriz presentada previamente, en la tabla 18 se muestran los resultados clasificados por parámetros, de acuerdo con la información obtenida de las tres herramientas analizadas se destaca a *Orange* como la herramienta con mejor experiencia en usabilidad.

Tabla 18.
Resultados obtenidos de la evaluación por herramienta

Parámetros	Herramientas		
	Orange	WEKA	Lenguaje R
Facilidad de aprendizaje	9	7	7
Eficiencia	9	8	7
Recuerdo en el tiempo	6	6	4
Tasa de errores	11	10	7
Satisfacción	3	2	2
Total	38	33	27

Nota: Elaboración propia.

En este capítulo se presentaron las pruebas de usabilidad con las herramientas *Orange*, *WEKA* y lenguaje *R*. Permitiendo la elaboración de un análisis para emitir una recomendación en el siguiente capítulo. Asimismo, las pruebas realizadas también

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

contribuyeron a la selección de los modelos que sirvieron para la interpretación de los resultados.

Durante la exposición de resultados se presentó información sobre las métricas de calidad así como los patrones identificados en cada herramienta aplicando técnicas de clasificación, agrupación y reglas de asociación. Durante el desarrollo de la interpretación de los resultados, se detectó información relevante en cada herramienta destacando coincidencias en las agrupaciones, reglas o decisiones generadas así como diferencias que permitieron complementar el conocimiento descubierto. Con los hallazgos identificados se presentarán las conclusiones de este trabajo de investigación así como las propuestas de trabajos futuros.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Capítulo 5. Conclusiones, recomendaciones y trabajos futuros

5.1 Conclusiones

Las percepciones expuestas por los estudiantes de licenciatura de la universidad caso de estudio y que se encuentran relacionadas con barreras para el uso de innovaciones tecnológicas, confirman una idea presentada en este documento, toda investigación de carácter científico y académico debe ser delimitado. Para tal caso, en este trabajo se expusieron las barreras: acceso, económica, pedagógica y ambiente académico. Sin embargo, también es importante mencionar que la elaboración de una investigación conlleva un proceso para la generación de nuevo conocimiento ya sea partiendo de estudios previos o de una idea diferente.

A través de la formulación de las diferentes etapas de este estudio, como lo son las generalidades, el marco de la investigación, el desarrollo metodológico y los resultados, expuestos respectivamente en cada capítulo, se presentan las conclusiones abordando tres elementos principales: el cumplimiento del objetivo general y los objetivos específicos, así como la exposición de los hallazgos expuestos en los resultados que permiten dar respuesta a la pregunta de investigación.

1

El objetivo del estudio fue identificar patrones de conducta, relacionados con las barreras para el uso de innovaciones tecnológicas en estudiantes de licenciatura de una Institución de Educación Superior (IES), empleando técnicas de minería de datos. Asimismo, tres de los objetivos específicos se encuentran relacionados con la aplicación de las técnicas de minería de datos: árboles de decisión, agrupamiento y reglas de

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

asociación. Con respecto al último objetivo específico, establece el analizar los patrones de compartimiento identificados en los estudiantes.

De acuerdo con lo anterior, se dio cumplimiento a los objetivos específicos, puesto que se llevó a cabo el uso de la técnica de clasificación aplicando árboles de decisión, también la aplicación de la técnica de agrupamiento fue empleada con el algoritmo K-Means logrando hacer tres agrupamientos con cada herramienta, por último se aplicó la técnica de análisis de enlaces o reglas de asociación, hallando coincidencias en las reglas resultantes. Cabe señalar que, cada técnica empleada fue ejecutada en *Orange*, *WEKA* y lenguaje R. Con los resultados obtenidos se llevó a cabo el análisis e interpretación de los patrones identificados. Por lo anterior, también se expone que el objetivo general ha sido completado.

Conforme a los resultados presentados se destaca que en *Orange*, *WEKA* y Lenguaje R, al emplear el algoritmo de árboles de decisión los resultados coincidieron en el tipo de clasificación otorgado, partiendo de que el alumno cuenta o no con internet en casa. Como se observó en los resultados, existe un 35% de estudiantes que no contaron con el servicio de internet para llevar sus clases virtuales, es decir, casi 350 alumnos que tuvieron problemas de accesibilidad, esto sin contar los estudiantes que tenían el servicio pero no era eficiente. Estos resultados cuantitativos fueron coincidentes en las tres herramientas y aunque son resultados que se pueden sacar de una encuesta convencional o de un cálculo estadístico, a partir de las técnicas de minería de datos y la aplicación de los árboles de decisión, estos datos se convirtieron en información relevante

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

en las herramientas *Orange*, *WEKA* y lenguaje R, porque derivado de estos datos y de la aplicación de esta técnica se descubrió que este porcentaje de alumnos al no contar con internet tuvieron que asistir a un ciber café, tuvieron dificultades referentes al factor pedagógico porque no comprendían las explicaciones del docente, percibieron que las clases virtuales únicamente funcionan con materias teóricas, también obstaculizó la comunicación con los profesores porque los medios electrónicos como las redes sociales requieren de internet. En ese sentido, respecto al factor del ambiente académico relacionado con las experiencias del alumno, el no contar con internet, los trabajos en equipo así como la relación entre compañeros y la relación profesor-alumno, se vieron afectadas y fueron percibidas por el estudiante como experiencias negativas.

Por otro lado, dentro de los estudiantes que cuentan con internet en casa pero es ineficiente se tiene una media de 300 estudiantes de acuerdo con los datos generados por las herramientas, estos estudiantes a pesar de contar con el servicio también percibieron una experiencia regular o frustrante en su uso para las clases virtuales.

Aunque en este aspecto *Orange*, *WEKA* y Lenguaje R coincidieron, cada herramienta generó información distinta que se logra complementar, consiguiendo generar una mejor interpretación. Por ejemplo, en el caso de *Orange* los estudiantes que cuentan con internet ineficiente a pesar de sus experiencias negativas no se consideran con habilidades limitadas para el uso de tecnologías en los ambientes virtuales, en el caso de *WEKA* aunque los estudiantes mencionaron que cuentan con recursos tecnológicos como un equipo de cómputo, también se expone que es de uso compartido

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

con otros familiares lo que hace más complejo que el estudiante atienda sus actividades escolares si no tiene recursos tecnológicos para su uso individual y si a esto se le suman aspectos pedagógicos como el caso de que solo uno o dos profesores realicen evaluaciones o retroalimenten las tareas entonces se vuelve más compleja la labor de estudiante. Por último, en el caso de lenguaje R, complementa lo previamente mencionado puesto que se menciona interacción limitada a tareas y actividades con profesores y compañeros.

De acuerdo con lo anterior, más de la mitad de los estudiantes encuestados tienen problemas con el internet ya sea porque el servicio es ineficiente o no cuentan con este.

Aunque en este primer algoritmo se encuentran presentes las barreras económicas, pedagógica y del ambiente académico, son en consecuencia a la barrera del acceso la cual muestra una mayor afectación entre los alumnos.

Por otro lado, con el algoritmo K-means se identificaron patrones clasificados por uso de tecnologías en los estudiantes, mismos que fueron obtenidos a través de la técnica de agrupación de minería de datos, si bien los agrupamientos realizados en las herramientas contienen la misma etiqueta de grupo: uso de TI promedio, uso de TI óptimo y uso de TI bajo, se logró identificar información particular de cada aplicación así como información coincidente.

Por ejemplo en el caso de la herramienta *Orange*, de los diferentes cuestionamientos que respondieron los estudiantes, algunos se encuentran asociados con el acceso a internet y dispositivos electrónicos, mientras que en otros casos se

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

relacionaron a aspectos pedagógicos y de ambiente académico por ejemplo, el trato con los profesores y compañeros de clases o las actividades y tareas escolares.

2 De acuerdo con las agrupaciones realizadas con *Orange*, se pudo observar que influye mucho la edad que tengan los estudiantes, en el caso de los recursos tecnológicos como una computadora, requieren contar con poder adquisitivo que les permita adquirir un equipo para uso personal, eso es más factible en estudiantes que trabajan como un adulto de 26 años o más.

2 En ese sentido, también se muestra que la diferencia de edad influye en las actitudes del estudiante para afrontar los niveles de estrés por la carga académica, así como de sus relaciones interpersonales entre compañeros y profesores, esto se debe a que cuanto mayor sea el alumno tendrá mayor madurez para afrontar la carga académica permitiendo tener un mayor dominio del estrés y la ansiedad. Sin embargo, en lo que respecta al factor pedagógico en su faceta extrínseca como las evaluaciones, tareas y la cátedra del profesor puede ser percibido por los estudiantes de forma negativa indistintamente a la edad, este es el factor que más afecta la educación en modalidad virtual derivado de la falta de una secuencia didáctica en los docentes o aptitudes deficientes para transmitir su materia mediante plataformas tecnológicas.

En el caso de WEKA, se identificó que los estudiantes no tenían un rango de edad tan definido como en *Orange*, sin embargo dentro de los clústeres sí se detectaron edades que dominaban dentro de los agrupaciones, por ejemplo, en el caso de los estudiantes considerados con un uso óptimo de TI se contaba con población de

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

estudiantes mayores de treinta años, lo que influyó en los patrones identificados porque son estudiantes que logran tener un mayor control del estrés y la ansiedad, así mismo al disponer de recursos tecnológicos este grupo logró tener una mayor interacción y participación en el aula virtual. A diferencia de los estudiantes con uso de TI promedio o bajo que sí se enfrentaron a obstáculos económicos y de accesibilidad. Asimismo, se destaca que la barrera pedagógica y del ambiente académico se encuentran presentes en las tres agrupaciones, principalmente en estudiantes con uso bajo de TI, puesto que no solo tuvieron problemas de conectividad sino también en la falta de comunicación y retroalimentación con los profesores.

En lo que respecta a lenguaje R, se identificó que los estudiantes con un uso promedio de TI o un uso óptimo a pesar de iniciar interacción con tecnologías en nivel secundaria fue hasta la universidad que comenzaron a usar plataformas virtuales. Sin embargo, esto no fue impedimento para ingresar a sus clases virtuales puesto que se consideraron capaces para el uso de tecnologías, por otro lado, no experimentaron obstáculos aludidos a la barrera pedagógica en su forma extrínseca puesto que todos sus profesores retroalimentaban tareas o enviaban planeaciones, aunque se encontró presente la barrera del ambiente académico por una interacción regular con el internet no fue impedimento para lograr interactuar con profesores y compañeros.

Así como se mostraron las diferencias y hallazgos particulares en cada herramienta con la técnica de agrupamiento, también se exponen las principales

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

coincidencias que se encontraron presentes dentro de los clústeres en *Orange*, *WEKA* y lenguaje R:

- Los alumnos consideran necesario que la institución académica brinde capacitación para el uso de plataformas de educación virtual.
- Por otro lado, consideran que las asignaturas prácticas no son aptas para ser enseñadas a distancia.

En lo que respecta al clúster de uso de TI bajo, en las tres herramientas se detecta que es el grupo con mayor afectación puesto que se encuentran presentes las cuatro barreras presentadas en este trabajo. Este clúster se encuentra representado por un 31% de la población encuestada, en el caso del clúster con uso de TI promedio aunque también influyen las barreras del acceso y económica, las que afectan principalmente son la barrera pedagógica y la barrera del ambiente académico, cabe señalar que este último clúster se encuentra representado por un 38% de los estudiantes encuestados.

Así mismo, se manifiesta que los estudiantes con un uso óptimo de tecnologías a pesar de ser el grupo con mejores condiciones también muestran afectaciones de la barrera pedagógica y del ambiente académico pero de forma atenuada.

Como se presentó previamente con los árboles de decisión y la técnica de agrupación, se hace un mayor énfasis en las afectaciones que tuvieron los estudiantes durante clases virtuales, derivado principalmente por el acceso a internet y la limitada disponibilidad de recursos tecnológicos, que hacen referencia a la barrera del acceso y la barrera económica y cómo estos obstáculos influyen en los aspectos pedagógicos y

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

del ambiente académico, causando afectación por las cuatro barreras establecidas en esta investigación. Sin embargo, a través de las reglas de asociación también se descubrió información relevante enfocada en los estudiantes que sí disponen de recursos tecnológicos y servicio de internet, de modo que a pesar de que el aspecto económico y de accesibilidad no fueron inconvenientes, sí afrontaron obstáculos.

En el caso de las reglas de asociación, aunque se empleó el algoritmo FP Growth en *Orange* y el algoritmo A priori en *WEKA* y lenguaje R, la mayoría de las reglas resultantes fueron coincidentes en los *softwares* empleados. Destacando dos aspectos, primero el caso de los estudiantes que tuvieron limitaciones del tipo pedagógicas y en el ambiente académico porque solo con uno o dos profesores se realizaron tareas o actividades relacionadas con la asignatura, además experimentaron poca interacción y comunicación en el aula virtual porque las clases se limitaban al envío de tareas. De modo que los estudiantes quedaban con interrogantes sobre temas específicos en las asignaturas.

En segundo lugar, se encuentra un escenario óptimo en el cual los estudiantes comenzaron a usar tecnologías desde el nivel secundaria, facilitando su adaptación a los ambientes virtuales y contribuyendo a una percepción positiva en cuanto su aprendizaje en las asignaturas. Este segundo escenario se encuentra relacionado a doce reglas de las 16 presentadas en este trabajo, es decir en la mayoría de los *ítemsets*. Aunque este segundo contexto es favorecedor no quiere decir que no representa información relevante para el análisis, por el contrario, estas circunstancias reafirman la información presentada

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

a través del algoritmo K-means en los clústeres con la etiqueta de uso óptimo de TI, siendo este grupo el que presenta una menor cantidad de estudiantes en condiciones favorables. Por lo tanto, cabe señalar que si existen alumnos que tienen circunstancias óptimas de estudio por adquirir habilidades tempranas y disponer de recursos, no significa que no existan barreras para el uso de tecnologías. Por el contrario, confirma que las barreras propuestas en este trabajo son una realidad que dificultan el aprendizaje de los estudiantes en los ambientes virtuales en el 68% de los datos analizados.

Esto quiere decir que aun contando con el servicio de internet, persisten obstáculos en los estudiantes con respecto al uso de tecnologías y que se encuentran relacionados con la barrera pedagógica y del ambiente académico, sin importar la atenuación de la barrera del acceso y económica como en el caso de los grupos de estudiantes con uso óptimo o promedio de TI. Por otro lado cabe mencionar que las barreras expuestas se encuentran mayormente asociadas a los estudiantes menores de 25 años.

De acuerdo con lo anterior, resulta que la barrera del acceso impide a los estudiantes un aprovechamiento adecuado de las TI para fines académicos, sin acceso a internet no se puede hacer uso de plataformas digitales y por ende no se puede tener interacción con profesores y compañeros. Y aun contando con servicio de internet no se garantiza que la calidad del servicio sea eficiente. Esta situación se encuentra reflejada en diferentes patrones, principalmente los relacionados con los clústeres de estudiantes con un uso bajo de tecnologías de la información, así como en las clasificaciones de los

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

árboles de decisión en donde un 33.8% de los estudiantes no cuentan con internet en casa y deben acudir a un ciber café, y el 34.38% son estudiantes que tienen internet en casa pero no es eficiente. En cuanto a las reglas de asociación también se detectó que la mayoría de las reglas resultantes se obtuvieron relaciones con las diferentes barreras pero en todos los casos se contaba con internet en casa. Sin embargo, si bien existen estudiantes con recursos a su disposición aun así tuvieron obstáculos que afrontar para el uso de tecnologías.

Por lo tanto, se puede decir que los patrones de comportamiento en estudiantes que se encuentran relacionados con las barreras: acceso, económica, pedagógica y del ambiente académico, presentados en este trabajo demuestran que influyen en el uso de innovaciones tecnológicas en los estudiantes. De los principales patrones relacionados a las barreras, mismos que fueron identificados con base en las técnicas de minería de datos de clasificación, agrupación y reglas de asociación empleados son los siguientes (ver tabla 19):

Tabla 19.
Principales patrones identificados

Patrones	Barreras
Falta de internet o servicio ineficiente.	Acceso.
Recursos tecnológicos compartidos o ausencia de estos.	Económica.
Necesidad de capacitación para el uso de tecnologías.	Pedagógica. Factor intrínseco.
Estrés y ansiedad por el uso de tecnologías y la carga académica.	Pedagógica. Factor intrínseco.
Falta de seguimiento académico por parte de los docentes. El seguimiento académico se encuentra relacionado a las tareas, actividades, evaluaciones, retroalimentación y planeación de la asignatura.	Pedagógica. Factor extrínseco.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Patrones	Barreras
Experiencias negativas relacionadas a la interacción con profesores y compañeros, carga académica y trabajos en equipo.	Ambiente académico.

Nota: Elaboración propia.

En algunos casos se presentaron patrones que contenían una atenuación de una o dos barreras sin embargo, se encontraban presentes las demás. Así como se mostraron patrones donde las cuatro barreras se encontraban presentes.

De acuerdo con las barreras expuestas en este trabajo, cabe señalar que continuarán siendo una amenaza para los estudiantes universitarios de la Institución de Educación Superior (IES) que fue considerada como caso de estudio. Esto se debe a que siempre existirá población en rezago por diversas circunstancias y bajo diferentes contextos, porque no es un problema que se pueda eliminar por completo. Las barreras para que los estudiantes usen tecnologías existirán, así como existe la pobreza; sin embargo, lo que sí se puede hacer es atenuar estos obstáculos, buscando beneficiar cada vez a una mayor parte de la población y hacer que disminuyan los porcentajes de estudiantes que se ven perjudicados por las barreras económica, acceso, pedagógica y ambiente académico.

2

Por lo anterior, se concluye que las técnicas de minería de datos permitieron obtener información relevante a través de los patrones de conducta identificados en los estudiantes. Se puede decir que los resultados obtenidos serán de beneficio para la IES caso de estudio.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

5.2 Recomendaciones

Derivado de la experiencia generada durante el desarrollo de la investigación, en este apartado se presenta un conjunto de buenas prácticas para que los usuarios de los productos generados en el estudio u otros investigadores que repliquen el uso de métodos, tecnologías, datos, entre otros, las observen convenientemente. Para tal caso, se presentan recomendaciones de tres tipos: el primero consta de recomendaciones a nivel técnico sobre las herramientas implementadas en el estudio, en el segundo, se exponen recomendaciones a partir de la información descubierta con los datos analizados, en tercer orden se encuentran recomendaciones relacionadas con el *dataset* empleado.

Conforme a los puntajes obtenidos en el análisis de usabilidad realizado con las herramientas, *Orange* resultó ser el *software* mejor evaluado. Por lo tanto, se recomienda esta herramienta principalmente porque puede ser empleada por aquellos usuarios que no son especialistas en análisis y minería de datos, requiere poca documentación para su uso y la curva de aprendizaje no es amplia.

De acuerdo con lo anterior, se presentan los siguientes escenarios para la aplicación de *Orange*:

- En el ámbito educativo puede ser implementada en un curso de capacitación o asignatura a nivel licenciatura ya que permite que el usuario logre adaptarse e interactuar con la herramienta en poco tiempo, lo que resulta ser benéfico para

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

conocer las diferentes técnicas y métricas de evaluación para el análisis de información y la construcción de modelos.

- En el área comercial profesionales de diferentes áreas pueden hacer uso de esta herramienta por contar con presentaciones visualmente atractivas y que pueden ser de utilidad para la interpretación de información publicitaria, financiera o de procesos operativos y humanos dentro de la organización.

En lo que respecta a las herramientas *WEKA* y lenguaje *R*, son herramientas con amplia variedad de algoritmos, sin embargo su recomendación va dirigida a un usuario con conocimiento previo en minería de datos. Para tal caso, el incursionar con una herramienta previa como es *Orange* permitirá al usuario un mayor entendimiento y dominio de estas y otras herramientas.

Como última consideración a nivel tecnológico, se recomienda que la técnica de agrupamiento sea empleada con más de una herramienta. Esto se debe a que cada *software* puede realizar agrupaciones distintas de la información logrando detectar coincidencias entre las herramientas pero también información diferente y relevante en cada una, lo que permite complementar la generación de nuevo conocimiento.

Con respecto al análisis realizado de los patrones de comportamiento en estudiantes, se recomiendan los siguientes aspectos a considerar por la universidad caso de estudio:

- Capacitación de docentes y estudiantes para la implementación de herramientas tecnológicas como ambientes virtuales, herramientas ofimáticas,

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

aplicaciones para diseño y presentación de trabajos escolares que sean de uso básico para el desarrollo elemental de las funciones académicas.

- En caso de que la institución académica decida implementar ambientes virtuales en los planes de estudios a distancia, en esquemas híbridos o bien de forma emergente, sería pertinente invitar a los alumnos a realizar una revisión de recursos mínimos como equipo de cómputo, dispositivo móvil, servicio de internet de modo que sean estudiantes aptos para llevar a cabo modalidades de estudio de este tipo.
- Asimismo, en caso de los alumnos que deseen hacer uso de una educación a distancia, sería de utilidad que la institución solicite como parte de los requisitos de inscripción o del curso propedéutico, una constancia o capacitación para el uso de la plataforma que vayan a emplear en el semestre. De modo que sea un referente de que los alumnos son conscientes de las implicaciones que conlleva tomar clases en esta modalidad de estudios.
- Por otro lado, queda claro en este proyecto que el uso de internet se ha vuelto indispensable para el uso de tecnología, por esta razón se recomienda a la institución contar con la infraestructura necesaria para abastecer de un servicio de internet funcional dentro de las instalaciones de la universidad.

Con estas recomendaciones a nivel institucional no se busca la eliminación del problema, pero pueden formar parte de la solución y una forma de mitigación.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Por último, aunque se obtuvo información relevante de esta investigación a partir de los datos proporcionados, se recomienda incluir preguntas relacionadas a aspectos socioeconómicos del estudiante, por ejemplo: los ingresos mensuales familiares, si el estudiante trabaja y estudia, zona en la que se encuentra su domicilio etc., esto queda a consideración ya sea para la modificación del *dataset* proporcionado y un nuevo análisis.

5.3 Trabajos futuros

En este apartado se presentan nuevos desafíos derivados de la investigación realizada. Pueden ser nuevos estudios a partir de lo que no se pudo abordar en la investigación o para aplicar la misma perspectiva a otras problemáticas o disciplinas.

Derivado de esta investigación se pueden realizar las siguientes investigaciones como posibles trabajos a futuro y que pueden tener una especialización del tipo educativo o tecnológico:

- El estudio de las barreras asociadas al uso de innovaciones tecnológicas, no únicamente se limita a los factores presentados en esta investigación, la clasificación de estos factores dependerá de cada autor, asimismo, el analizar el comportamiento del individuo como el caso de los estudiantes es una tarea que requiere de atención, por lo tanto se propone retomar el estudio de factores relacionados al uso de tecnologías desde un enfoque cualitativo.
- Por otro lado, este trabajo fue realizado a partir de datos recopilados de una encuesta aplicada durante la pandemia, en ese sentido es importante que se

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

realice de nueva cuenta un análisis para identificar patrones de comportamiento pero con estudiantes en el nuevo esquema híbrido o postpandemia.

- Aunque este estudio trata de las barreras que influyen en el uso de tecnologías en estudiantes, este trabajo se limita a la percepción de los estudiantes. Por lo tanto, puede ser retomado partiendo de la percepción de los profesores para conocer una nueva clasificación de las barreras desde un punto de vista diferente.
- A partir de los resultados y discusión, presentados en la investigación, se pueden elaborar nuevas estrategias educativas. Este es un trabajo que debe ser realizado por especialistas en materia de educación de la institución, con el fin de implementar nuevas políticas educativas que permitan la atenuación del problema planteado en este estudio.

Desde el punto de vista tecnológico surgen las siguientes propuestas de trabajos a futuro:

- En este trabajo se emplearon las técnicas de agrupación, clasificación y análisis de enlaces como técnicas de minería de datos. Con estos modelos se aplicaron algoritmos como árboles de decisión, K-means, FP-Growth y A priori. Sin embargo, al emplear minería de datos para la identificación de patrones de conducta en estudiantes, existen diferentes algoritmos que pueden ser empleados con las mismas técnicas, por ejemplo Naive Bayes, Random Forest, KNN y reglas de inducción.

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

- Con el paso del tiempo la tecnología avanza de forma acelerada, lo que permite la generación de nuevas herramientas que pueden ser usadas en la minería de datos, por lo tanto, se puede realizar un nuevo estudio aplicando herramientas distintas a las presentadas en este trabajo.
- Este trabajo deja abierta un área de oportunidad para la realización de una base de datos, en ella se pueden considerar registros de estudiantes para estudiar su comportamiento nuevamente partiendo de un mayor registro de estudiantes de modo que la minería de datos no solo se realice con datos estáticos sino también dinámicos. Inclusive a partir de la creación de una base de datos, considerar nuevas áreas de interés en el ámbito académico que resuelva otro tipo de problemáticas dentro de la institución.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Referencias

29 Aguilar, R. (2004). La guía didáctica, un material educativo para promover el aprendizaje autónomo, Evaluación y mejoramiento de su calidad en la modalidad abierta y a distancia de la UTPL. *RIED*, 7(1-2), 179-192.

<http://revistas.uned.es/index.php/ried/article/view/1082/998>

Alom, M. & Courtney, M. (2018). Educational Data Mining: A Case Study Perspectives from Primary to University Education in Australia. *Modern Education and Computing Science*, 10(2), 1-9. DOI: 10.5815/ijitcs.2018.02.01

Álvarez, A., Alarcón, A. y Callejas, M. (2016). Comparación de modelos y estándares de evaluación de calidad para una plataforma de aprendizaje virtual. En Serna, E. (Ed), *Desarrollo e Innovación en Ingeniería* (pp. 401-413).

<https://doi.org/10.5281/2613946>

5 Álvarez-Silva, L.A., Gallegos-Luna, R. M. y Herrera-López, P. S. (2018). Estrés académico en estudiantes de Tecnología Superior. *UNIVERSITAS*, (28), 193-209.

<http://dx.doi.org/10.17163/uni.n28.2018.10>

Amazon AWS (s.f.) *Amazon Web Services (AWS)* <https://aws.amazon.com/es/>

Andrés, G. (2014). Una aproximación conceptual a la "apropiación social" de TIC. *Question*, 1(43), 17-31.

<http://perio.unlp.edu.ar/ojs/index.php/question/article/view/2227>

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Araujo, O. (2020). Modelo FURPS aplicado al análisis de calidad de un software desarrollado con Sencha Ext JS. *REDDI*, 5(1), 1-10.
<https://reddi.unlam.edu.ar/index.php/ReDDi/article/view/121/223>

2 Arellano-Becerril, E. y González-Zermeño, M. (2021). Análisis de la brecha digital e inclusión digital vs. Educación virtual en tiempos de pandemia COVID-19. En Escudero-Nahón, A. y Palacios-Díaz, R. (Coords), *Tecnología y contingencias*, (72-80). Editorial Transdigital. <https://www.editorial-transdigital.org/libros/tecnologia-y-contingencias/>

Bahga, A. & Madiseti, V. (Eds.). (2019). *Big Data Science & Analytics*.
<http://202.62.95.70:8080/jspui/handle/123456789/14472>

Ballesteros, A., Sánchez-Guzmán, D. y Salcedo, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. *Latin-American Journal of Physics Education*, 7(4), 662-668.
http://www.lajpe.org/dec13/22-LAJPE_814_bis_Alejandro_Ballesteros.pdf

17 Balmori, E., Garza, M. y Guzmán, E. (2013). Diseño y validación de un instrumento para determinar las variables de deserción en los institutos tecnológicos. *Pistas Educativas*, (101), 291-302. <http://pistaseducativas.itc.mx>

13 Bayrón, C. (2012). Teoría social cognitiva y teoría de retención de Vincent Tinto: Marco teórico para el estudio y medición de la auto-eficacia académica en estudiantes universitarios. *Revista Griot*, 5(1), 28-49.
<https://revistas.upr.edu/index.php/griot/article/view/1774/1567>

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Bean, J. & Metzner, B. (1985). A conceptual model of nontraditional undergraduate student attrition. *Review of Educational Research*, 55(4), 485-540.

<https://doi.org/10.2307/1170245>

Benalcázar, J. (2017). *Análisis comparativo de metodologías de minería de datos y su aplicabilidad en la industria de servicios* [tesis de maestría, Universidad de Las

Américas Ecuador] <http://dspace.udla.edu.ec/handle/33000/7547>

Bernal, J., Rodríguez, J. & Portella, J. (2021). DBMS and ORACLE Data Mining. *Preprints*.

doi:10.20944/preprints202103.0640.v1

Bhatia, P. (2019). Beginning with Weka and Language R. *Data Mining and Data Warehousing Principles and Techniques* (pp. 28-54). Cambridge University Press.

<https://doi.org/10.1017/9781108635592.004>

Busu, N. (2017). *Unstructured Big Data Processing in Cloud Computing Environment by Using Amazon Elastic MapReduce* [Master's thesis, University Putra Malaysia].

<http://psasir.upm.edu.my/id/eprint/67852/1/FSKT%202017%2024%20IR.pdf>

Cabero, J. (2005). Las TIC y las universidades: retos, posibilidades y preocupaciones.

Revista de la educación Superior. 34(135), 77-100.

<http://resu.anuies.mx/ojs/index.php/resu/issue/archive/2>

Cabrera, A., Nora, A. & Castañeda, M. (1992). The role of finances in the process: A structural model. *Research in Higher Education*, 33(5), 303-336.

<https://doi.org/10.1007/BF00973759>

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Callejas-Cuervo, M., Alarcón-Aldana, A. y Álvarez-Carreño, A. (2017). Modelos de calidad del software, un estado del arte. *Ingeniería y Tecnología*, 13(1), 236-250.
<http://dx.doi.org/10.18041/entramado.2017v13n1.25125>

2 Calvache-Fernández, L., Álvarez-Vallejo, V. y Triviño-Arbeláez, J. (2018). Proceso KDD como apoyo a las estrategias del proceso SARA (Sistema de Acompañamiento para el Rendimiento Académico). *Revista Educación e Ingeniería*, 13(6), 82-89.
<http://dx.doi.org/10.26507/rei.v13n26.916>

Camana, R. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnológica ESPOL*, 29(1), 170-183.
<http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/464>

Cámara de Diputados del Honorable Congreso de la Unión. (2010, 5 de julio). *Ley Federal de Protección de Datos Personales en Posesión de los Particulares*. Diario Oficial de la Federación. <http://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>

3 Capurro, R. (2002). Perspectivas de la cultura digital en Latinoamérica. *DataGramZero Revista de Ciência da Informação*, 3(2), A01.
<http://www.brapci.inf.br/index.php/article/download/7515>

2 Castro-Villagrán, A., Cosgaya-Barrera, B. R., Sosa-González, W. E. y Ceballos, L. (2016). Los estudiantes universitarios y el uso de las TIC como herramienta de aprendizaje. Una revisión de estudios realizados en México. *IC INVESTIG@CCION*, (9), 36-48.
<http://revistaic.instcamp.edu.mx/volumenes/volumen9#revista09-5>

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Chanchi, G., Gómez, M. y Campo, W. (2020). Propuesta de una herramienta de inspección según los atributos de usabilidad de Nielsen. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (26), 448-460.

<http://www.risti.xyz/issues/ristie26.pdf>

26 Comisión Económica para América Latina y el Caribe. (2020). *La educación en tiempos de la pandemia de COVID-19*.

https://repositorio.cepal.org/bitstream/handle/11362/45904/1/S2000510_es.pdf

Constanzo, M. (2014). Comparación de modelos de calidad, factores y métricas en el ámbito de la ingeniería de software. *ICT-UNPA*, 6(1), 1-36.

<http://dx.doi.org/10.22305/ict-unpa.v6i1.89>

Crespo, M. y Palaguachi, M. (2020). Educación con tecnología en una pandemia: Breve análisis. *Revista Scientific*, 5(17), 292-310.

34 <https://doi.org/10.29394/Scientific.issn.2542-2987.2020.5.17.16.292-310>

8 Covi, D. (2007). Dimensión social del acceso, uso y apropiación de las TIC. *Contratexto*, (16), 65-79. <https://doi.org/10.26439/contratexto2008.n016.784>

Covi, D. y López, R. (2011). Tejiendo voces: jóvenes universitarios opinan sobre la apropiación del internet en la vida académica. *Revista Mexicana de Ciencias Políticas y Sociales*, 56(212), 69-80.

<http://www.revistas.unam.mx/index.php/rmcpys/article/view/30421/28244>

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Cruz, M. (2016). Factores que influyen en el rendimiento académico de los estudiantes.

Escenarios: empresa y territorio, 5(5), 93-118.

<http://revistas.esumer.edu.co/index.php/escenarios/article/view/66/61>

Diao, J., Wang, Y., Liu, Y. & Yang, H. (2018, 19-20 august). *Research on the Application*

of Educational Data Mining Technology Based on Association Rule in Students'

Grades Analysis (Taking the Computer Science and Technology Major of HZNU

as an Example) [conference]. International Conference on Humanities and

Advanced Education Technology (ICHAET 2018), Guangzhou, China. DOI:

10.12783 / dtssehs / ichae2018 / 25634

Díaz, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena.

Estudios Pedagógicos, 34(2), 65-86.

<http://revistas.uach.cl/pdf/estped/v34n2/art04.pdf>

Do Carmo, G., Manhes, E.& Terra, M. (2018). Pistas sobre a permanencia estudiantil

como virada epistemológica: iniciativas de um núcleo de pesquisa por uma

sociologia da permanencia na educacao a partir de Vincent Tinto. *Synthesis*, 11(1),

54-63. [https://doi.org/10.12957/\(syn\)thesis.2018.54571](https://doi.org/10.12957/(syn)thesis.2018.54571)

Espinoza, E. y Ricaldi, M. (2018). El tutor en entornos virtuales de aprendizaje.

Universidad y Sociedad, 10(1), 201-210. <http://rus.ucf.edu.cu/index.php/rus>

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). The KDD Process for Extracting

Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39(11), 27-

34. <https://doi.org/10.1145/240455.240464>

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Fernández, M. y Bonilla, R. (2020). Bibliominería, datos y el proceso de toma de decisiones. *Revista Iberoamericana de Bibliotecnología*, 43(2), e18/1-e18/12.

<https://doi.org/10.17533/udea.rib.v43n2e18>

Flores, G., Cadena, J., Quinatoa, E. y Villa, M. (2019). Minería de datos como herramienta estratégica. *RECIMUNDO*, 3(1), 955-970. DOI:

10.26820/recimundo/3.(1).enero.2019.955-970

Fonseca, G. (2016). Articulación teórico metodológica para el estudio de la retención estudiantil universitaria. *Revista pertinencia académica*, (1), 25-36. <http://revista-academica.utb.edu.ec/index.php/pertacade/article/view/3>

<http://revista-academica.utb.edu.ec/index.php/pertacade/article/view/3>

Fonseca, G. y García, F. (2016). Permanencia y abandono de estudios en estudiantes universitario: un análisis desde la teoría organizacional. *Revista de la Educación Superior*, 45(179), 25-39. <http://doi.org/10.1016/j.resu.2016.06.004>

<http://doi.org/10.1016/j.resu.2016.06.004>

Frontline Systems. (s.f.). *Analytic Solver para Excel*. <https://www.solver.com/analytic-solver->

[platform?utm_source=Google&utm_medium=PPC&utm_campaign=Tier1ConversionalOptimization&utm_term=&qclid=Cj0KCQjwwY-](https://www.solver.com/analytic-solver-platform?utm_source=Google&utm_medium=PPC&utm_campaign=Tier1ConversionalOptimization&utm_term=&qclid=Cj0KCQjwwY-LBhD6ARIsACvT72PjcOPXVlrcT_GkvFTkZqns_XVyrq-XrCQ22zbhzl3xLcPI-yijwGMAoSyEALw_wcB)

[LBhD6ARIsACvT72PjcOPXVlrcT_GkvFTkZqns_XVyrq-XrCQ22zbhzl3xLcPI-](https://www.solver.com/analytic-solver-platform?utm_source=Google&utm_medium=PPC&utm_campaign=Tier1ConversionalOptimization&utm_term=&qclid=Cj0KCQjwwY-LBhD6ARIsACvT72PjcOPXVlrcT_GkvFTkZqns_XVyrq-XrCQ22zbhzl3xLcPI-yijwGMAoSyEALw_wcB)

[yijwGMAoSyEALw_wcB](https://www.solver.com/analytic-solver-platform?utm_source=Google&utm_medium=PPC&utm_campaign=Tier1ConversionalOptimization&utm_term=&qclid=Cj0KCQjwwY-LBhD6ARIsACvT72PjcOPXVlrcT_GkvFTkZqns_XVyrq-XrCQ22zbhzl3xLcPI-yijwGMAoSyEALw_wcB)

Fundación R (s.f.) ¿Qué es R? <https://www.r-project.org/about.html>

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

García, L. (2019). El problema del abandono en estudios a distancia. Respuestas desde el diálogo didáctico mediado. *RIED*, 22(1), 245-270.
<https://doi.org/10.5944/ried.22.1>

9 García, L., Ortega, J., Peña, M., Ruano, I., y Ortiz, A. (2010). La calidad de la docencia virtual: la importancia de la guía de estudio. *Pixel-Bit*, 37, 77-92.
<https://recyt.fecyt.es/index.php/pixel/article/view/61401/37414>

García, V. y Silva, M. (2021). *Estudiantes-Barreras para el uso de innovaciones tecnológicas de estudiantes universitarios, ante la contingencia COVID 19* [Base de datos]. <https://docs.google.com/forms/d/17IIHOALaluKqPaKU5Ulv4Ka48XwmM2WHlvqJ8eMfSo/edit?ts=5fd7a893&qxids=7628>

23 Garrido, M., Jiménez, N., Landa, A., Páez, E. y Ruíz, M. (2013). Factores que influyen en el rendimiento académico: La motivación como papel mediador en las estrategias de aprendizaje y clima escolar. *Revista electrónica de investigación Docencia Creativa*, 2(), 17-25. doi.org/10.30827/Digibug.27620

9 Gaytan, J. (2015). Comparing faculty and student perceptions regarding factors the affect student retention in online education, *American Journal of Distance Education*, 29(1), 56-66. doi:10.1080/08923647.2015.994365.

4 Giraldo, J. y Vargas, F. (2017). Aplicación de la técnica regresión logística de la minería de datos en el proceso de descubrimiento de conocimiento (KDD) en bases de datos operativas o transaccionales. *PERSPECTIV@S*, 14(13), 51-55.
<http://revistas.uigv.edu.pe/index.php/perspectiva/article/view/568>

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Gironés, J., Casa, J., Minguillón, J. y Caihuelas, R. (2017). Introducción a la minería de datos. *Minería de datos Modelos y algoritmos* (pp. 25-35). Editorial UOC.

<http://www.editorialuoc.com>

GNU (2019, 15 de septiembre). *¿Qué es el software libre?*

<https://www.gnu.org/philosophy/free-sw.es.html>

Gómez, A., Alvarado, R., Martínez, M. y Díaz, C. (2018). La brecha digital: una revisión conceptual y aportaciones metodológicas para su estudio en México.

Entreciencias: Diálogos de la sociedad del conocimiento. 6(16), 49-64.

<http://revistas.unam.mx/index.php/entreciencias/article/view/62611/57630>

Google Cloud (s.f.) *Google Cloud: Cloud Computing Services* <https://cloud.google.com/>

Guzmán, Z. (2022). *Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos* [tesis de maestría, Universidad Juárez Autónoma de Tabasco de Tabasco México]

Hernández, C. y Carpio, N. (2019). Introducción a los tipos de muestreo. *ALERTA*, 2(1), 75-79. <https://doi.org/10.5377/alerta.v2i1.7535>

Hussain, S., Dahan, N., Ba-Alwib, F. & Ribata, N. (2017). Educational data mining and analysis of students academic performance using WEKA. *Indonesian Journal of*

Electrical Engineering and Computer Science, 9(2), 447-459.

DOI:10.11591/ijeecs.v9.i2.pp447-459

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

IBM México. (s.f.) *IBM SPSS Software* <https://www.ibm.com/mx-es/analytics/spss-statistics-software>

Infraestructura de Datos Espaciales para el Distrito Capital. (2019). *Metodología para la analítica de datos. Guía de buenas prácticas para analítica.* https://www.ideca.gov.co/sites/default/files/MetodologiaAnaliticaDatos_0.pdf

2 Instituto Nacional de Estadística y Geografía. (2021, 23 de marzo). *Resultados de la encuesta para la medición del impacto COVID-19 en la educación (ECOVID-ED) 2020.* [Comunicado de prensa]. https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2021/OtrTemEcon/E-COVID-ED_2021_03.pdf

Islas, C. y Franco, S. Detección de patrones en competencias digitales manifestadas por estudiantes universitarios. *Revista Electrónica de Tecnología Educativa*, (64), 51-67. <https://doi.org/10.21556/edutec.2018.64.1079>

5 Jaramillo, A. y Paz-Arias, H. (2015). Aplicación de Técnicas de Minería de Datos para Determinar las Interacciones de los Estudiantes en un Entorno Virtual de Aprendizaje. *Revista tecnológica ESPOL*, 28(1), 64-90. <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/351/229>

Kantardzic, M. (2020). Association Rules. En Hossain, E. (Ed), *Data Mining Concepts, Models, Methods, and Algorithms* (pp. 335-355). IEEE Press Editorial Board. <https://onlinelibrary.wiley.com/doi/book/10.1002/9781119516057>

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Marín, F. V., Inicarte, A., Hernández, H. G. y Pitre, R. C. (2017). Estrategias de las Instituciones de Educación Superior para la Integración de las Tecnologías de la Información y la Comunicación y de la Innovación en los procesos de Enseñanza. Un Estudio en el Distrito de Barranquilla, Colombia. *Formación Universitaria*, 10(6), 29-38. <http://dx.doi.org/10.4067/S0718-50062017000600004>

2 Martínez, M. (2021). La desigualdad digital en México: un análisis de las razones para el no acceso y el no uso de internet. *Revista de Tecnología y Sociedad*, (19), 1-19. <http://dx.doi.org/10.32870/Pk.a10n19.519>

Martínez, M., Meléndez, M., e Hidalgo, S. (2019). Implementación de minería de datos en la identificación de patrones de comportamiento estudiantil con el uso de dispositivos móviles impactando el rendimiento académico. *Revista Educarnos*. (36), 173-184. <https://revistaeducarnos.com/wp-content/uploads/2020/01/articulo-martha.pdf>

15 Martínez-Abad, F. y Hernández-Ramos, J. (2018). Técnicas de minería de datos con software libre para la detección de factores asociados al rendimiento. *Revista de Estudios y Experiencia en Educación*, (3), 135-145. <http://www.rexe.cl/ojournal/index.php/rexe/article/view/555/448>

MathWorks. (s.f.). MATLAB & Simulink. <https://la.mathworks.com/products/matlab.html?requestedDomain=>

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Medina, F. y Gómez, C. (2014). Funcionalidades de la minería de datos. *Revista Ingeniería y Región*. 12(), 31-40.

<https://journalusco.edu.co/index.php/iregion/article/view/728/1394>

Medina, G. (2021). La tecnología educativa presente en la formación científica: uso de la plataforma Microsoft Teams. En Medina, G. y Aquino, S. (Coords), *La tecnología educativa en tiempos de pandemia* (pp. 14-30). Gradus.

<https://es.graduseditora.com/la-tecnologia-educativa>

Menacho, C. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26-33.

https://revistas.lamolina.edu.pe/index.php/acu/article/view/811/pdf_43

Microsoft. (2019, 2 de mayo). *Analytic Solver*. <https://appsource.microsoft.com/en-us/product/office/wa200000018?tab=overview>

Microsoft. (s.f.). Microsoft 365 Personal. <https://www.microsoft.com/es-mx/microsoft-365/p/microsoft-365-personal/cfq7tc0k5bf?rtc=1&activetab=pivot:overviewtab>

Microsoft Azure (s.f.) *Microsoft Azure: Servicios de informática en la nube*.

<https://azure.microsoft.com/es-mx/>

Microsoft Support. (s.f.). *Tareas básicas en Excel*. [https://support.microsoft.com/es-es/office/tareas-b%C3%A1sicas-en-excel-dc775dd1-fa52-430f-9c3c-](https://support.microsoft.com/es-es/office/tareas-b%C3%A1sicas-en-excel-dc775dd1-fa52-430f-9c3c-d998d1735fca)

[d998d1735fca](https://support.microsoft.com/es-es/office/tareas-b%C3%A1sicas-en-excel-dc775dd1-fa52-430f-9c3c-d998d1735fca)

Mojica, Y. (2015). Algoritmo basado en estimación de distribución (EDA) para determinar causas de la deserción escolar. [tesis de maestría, Universidad Distrital José de

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Caldas de Bogotá Colombia]

<https://repository.udistrital.edu.co/handle/11349/4391>

Molina, J. y García, J. (2012). Técnicas de minería de datos basadas en aprendizaje automático. En *Técnicas de análisis de datos. Aplicaciones prácticas utilizando Microsoft*

Excel y Weka (pp. 96-266). Universidad Carlos III de Madrid.

<http://ocw.uc3m.es/ingenieria-informatica/analisis-de-datos/libroDataMiningv5.pdf>

Moncada, L. (2014). La integración académica de los estudiantes universitarios como factor determinante del abandono a corto plazo. Un análisis del sistema de educación superior a distancia del Ecuador. *RIED*, 17(2), 173-196.

<https://doi.org/10.5944/ried.17.2.12683>

Morales, A., Zacatenco, J. D., Luna, M., García, R. Z. e Hidalgo, C. (2020). Acceso y actitud del uso de Internet entre jóvenes de educación universitaria. *Revista Digital de Investigación en Docencia Universitaria*, 14(1), 1-11.

<https://doi.org/10.19083/ridu.2020.1174>

Mosquera, M. (2008). De la etnografía antropológica a la etnografía virtual. *FERMENTUM*, 18(53), 532-549.

<http://www.saber.ula.ve/handle/123456789/28656>

Na, W. (2020). A Data Mining Method for Students Behavior Uderstanding. *Paper*, 15(6),

18-32. <https://doi.org/10.3991/ijet.v15i06.13175>

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Naik, A. & Samant, L. (2016). Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Orange and Knime. *Procedia Computer Science*, 85(), 662-668. <https://doi.org/10.1016/j.procs.2016.05.251>

Nordman, J. (2020). *Google Cloud and solution for industrial* [Master's thesis, University of Turku] <http://urn.fi/URN:NBN:fi-fe2020041719044>

Oracle Latinoamérica (s.f.) *Aplicaciones en la nube y servicios de plataforma integrados* <https://www.oracle.com/mx/database/what-is-database/>

Ozgun, C., Colliau, T., Rogers, G., Hughes, Z. & Myer-Tyson, E. (2017). MatLab vs. Python vs. R. *Journal of Data Science*, 15(3), 355-372. [https://doi.org/10.6339/JDS.201707.15\(3\).001](https://doi.org/10.6339/JDS.201707.15(3).001)

Padilla, M. (2008). Opiniones y experiencias en el desempeño de la dirección escolar de las mujeres en Andalucía. *RELIEVE*, 14(1), 1-27. <https://ojs.uv.es/index.php/RELIEVE/article/view/4194/3798>

Panizzi, M. (2019). Establecimiento del estado del arte sobre la Minería de Datos Educativa en el Nivel Superior: Un Estudio de Mapeo Sistemático. *Revista de Investigaciones Científicas de la Universidad de Morón*, (4), 51-60. <https://repositorio.unimoron.edu.ar/handle/10.34073/164>

3

Pérez-López, E., Vázquez, A. y Cambero, S. (2021). Educación a distancia en tiempos de COVID-19: Análisis desde la perspectiva de los estudiantes universitarios. *RIED. Revista Iberoamericana de Educación a Distancia*. 24(1), 331-350. <http://revistas.uned.es/index.php/ried/article/view/27855>

2

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Pierrakeas, C., Xenon, M., Panagiotakopoulos, C. & Park, C. (2004). A comparative study of dropout rates and causes two different distance education courses. *The international review of research open and distributed learning*, 5(2), 1-15.

<https://doi.org/10.19173/irrodl.v5i2.183>

Prieto Rivero, N. (2018). Visión Formativa de la Metodología Cuantitativa en la Educación. *Revista Cientific*, 2(6), 352-365. <https://doi.org/10.29394/scientific.issn.2542-2987.2017.2.6.19.352-365>

Ramírez, C. A. (2019). Factores incidentes en la construcción de confianza de los estudiantes hacia las autoridades escolares en la educación media superior en México. *Diálogos sobre Educación. Temas actuales en investigación educativa*. 10(19), 1-21

Ramos Secaira, F. (2019). *Evaluación de calidad del software en las empresas de servicios tecnológicos situadas en la ciudad de Santo Domingo* [tesis de maestría, Universidad Tecnológica Empresarial de Guayaquil de Ecuador]

<http://181.39.139.68:8080/handle/123456789/1084>

RapidMiner (s.f.) *Best Data Science & Machine Learning Platform*. <https://rapidminer.com/>

Riquelme, J., Ruíz, R. y Gilbert K. (2006). Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial*, 10(29), 11-18.

[http://journaldocs.iberamia.org/articles/479/article%20\(1\).pdf](http://journaldocs.iberamia.org/articles/479/article%20(1).pdf)

1 Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Rivadeneira Rodríguez, E. (2017). Lineamientos teóricos y metodológicos de la investigación cuantitativa en ciencias sociales. *In Crescendo*, 8(1), 115-121.
<https://doi.org/10.21895/incres.2017.v8n1.11>

7 Rodríguez, C. y Sandoval, D. (2017). Estratificación digital, acceso y usos de las TIC en la población escolar chilena. *REDIE*, 19(1), 21-34.
<https://redie.uabc.mx/redie/article/view/902/1504>

24 Rodríguez, R. y Espinoza, L. (2017). Trabajo colaborativo y estrategias de aprendizaje en entornos virtuales de jóvenes universitarios. *RIDE*, 7(14), 86-109.
<http://dx.doi.org/10.23913/ride.v7i14.274>

Rodríguez, Y. y Díaz, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 3(3-4), 73-80.
<https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=78&path%5B%5D=70>

2 Rosado, A. y Verjel, A. (2017). Aplicación de la minería de datos en la educación en línea. *Revista Colombiana de Tecnologías Avanzadas*, 1(29), 92-98.
http://www.unipamplona.edu.co/unipamplona/portallG/home_40/recursos/revistas/21082017/revista_29.jsp

11 Saiz, M., Marticorena, R., Arnaiz, A., Escolar, M. y Queiruga, M. (2018). Detección del alumno en riesgo en titulaciones de Ciencias de la Salud: aplicación de técnicas de Learning Analytics. *European Journal of Investigation in Health*, 8(3), 129-142.
<https://formacionasunivep.com/ejihpe/index.php/journal/article/view/273>

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

25 Salas-Rueda, R. A., Salas-Rueda, R.D. y Salas-Silis, J. A. (2019). Análisis de la percepción de los estudiantes sobre el uso de las tecnologías de la información y comunicación en el proceso educativo por medio de la ciencia de datos. *Revista electrónica calidad en la educación superior*, 10(1), 96-124.

<http://dx.doi.org/10.22458/caes.v10i1.2463>

Salinas, J. (2002). ¿Qué aportan las tecnologías de la información y la comunicación a las universidades convencionales? Algunas consideraciones y reflexiones. *Revista Educación y Pedagogía*, 14(33).

<https://revistas.udea.edu.co/index.php/revistaeyp/article/view/5573/4996>

Sánchez, J., Sánchez, P. y Ramos, F. (2012). Usos pedagógicos de Moodle en la docencia universitaria desde la perspectiva de los estudiantes. *Revista Iberoamericana de Educación*, (60), 15-38.

<https://rieoei.org/historico/documentos/rie60a01.pdf>

2 Sánchez, S. & Luján, S. (2020). Design, implementation and evaluation of MOOCs to improve inclusion of diverse learners. In Management Association (Ed.), *Accesibility and Diversity in Education: Breakthroughs in Research and Practice*

(pp. 52-79). IGI Global. <http://doi:10.4018/978-1-7998-1213-5.ch004>

SAS. (s.f.). *Data mining software*. https://www.sas.com/es_ar/software/enterprise-miner.html

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Simpson, O. (2013). Student retention in distance education: are failing our students? *Open learning: The journal of open, distance and e-learning*, 28(2), 105-119. doi: 10.1080/02680513.2013.847363

2 Sotelo, M., Vales, J. y García, R. (2016). Actitud de estudiantes universitarios hacia los cursos en la modalidad mixta. *Educación y ciencia*, 5(45), 66-75. http://www.educacionyciencia.org/index.php/educacionyciencia/article/view/375/pdf_30

Stančin, I. & Jović, A. (2019, 20-24 may). *An overview and comparison of free Python libraries for data mining and big data analysis*. [conference] 42nd International Convention MIPRO 2019, Opatija, Croatia. doi: 10.23919 /MIPRO.2019.8757088.

Swan, K. (2001). Virtual interaction: desing factors affecting student statisfaction and perceived learning in asynchronous online courses. *Distance Education*, 22(2), 306-331. doi: 10.1080/0158791010220208

Tinto, V. (1975). Dropout from higher education: A theoretical Synthesis of recent research. *Journal of higher education*, (45), 89-125. <https://doi.org/10.3102/00346543045001089>

Tinto, V. (1989). Definir la deserción: una cuestión de perspectiva. *Revista de Educación Superior*, 71(18), 1-9. <http://publicaciones.anuies.mx/revista/71/1/3/es/definir-la-desercion-una-cuestion-de-perspectiva>

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Torres, C. y Moreno, G. (2013). Inclusión de las TIC en los escenarios de aprendizaje universitario. *Apertura*, 5(1), 48-65.

<http://www.udgvirtual.udg.mx/apertura/index.php/apertura/article/view/369/308>

2 Troche de Trevisan, B. G. (2019). *Estudio del Rendimiento Académico del Estudiante en Línea como Variable Predictiva del Abandono en Educación Superior: El caso de la Universitat Oberta de Catalunya* [tesis de doctorado, Universitat Oberta de Catalunya Barcelona España]

[http://openaccess.uoc.edu/webapps/o2/bitstream/10609/100346/1/Troche de Trevisan Cap 2 2019.pdf](http://openaccess.uoc.edu/webapps/o2/bitstream/10609/100346/1/Troche_de_Trevisan_Cap_2_2019.pdf)

Vaca Sierra, T. (2017). *Modelos de calidad de software aplicado al módulo de talento humano del sistema informático integrado universitario UTN* [tesis de maestría, Universidad Técnica del Norte Ecuador]

<http://repositorio.utn.edu.ec/handle/123456789/7457>

6 Vicentini, I. (2020). *La educación superior en tiempos de COVID-19: Aportes de la segunda Reunión del Diálogo Virtual con Rectores de Universidades Líderes de América Latina*. Banco Interamericano de Desarrollo.

<https://publications.iadb.org/publications/spanish/document/La-educacion-superior-en-tiempos-de-COVID-19-Aportes-de-la-Segunda-Reunion-del-Di%C3%A1logo-Virtual-con-Rectores-de-Universidades-Lideres-de-America-Latina.pdf>

1

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Villegas, W. y Luján-Mora, S. (2016). Análisis de las herramientas de minería de datos para la mejora del E-Learning en plataformas LMS. En Durán, F. y Durán, I. (Coords), *TIC Actualizadas para una nueva docencia universitaria* (pp. 761-774). <http://rd.unir.net/sisi/research/resultados/1524045566Libro%206.%20Segundas.%209788448612658.pdf>

Winocur, R. (2006). Internet en la vida cotidiana de los jóvenes. *Revista Mexicana de Sociología*, 68(3), 551-558. <http://dx.doi.org/10.22201/iis.01882503p.2006.003>

Witten, I., Frank, E., Hall, M. & Pal, C. (2017). Data Transformations. *Data Mining Practical Machine Learning Tool and Techniques* (4ta. Ed.). Morgan Kauffman. <https://doi.org/10.1016/B978-0-12-804291-5.00008-8>

Yang, X. (2019). Data mining techniques. En Lutz, M. (Ed), *Introduction to Algorithms for Data Mining and Machine Learning* (pp. 109-128). Academic Press. <https://doi.org/10.1016/B978-0-12-817216-2.00013-2>

Yugan, A., Morales, C., Delgado, J. y Espinoza, S. (2019). Modelo FURPS para el análisis de rendimiento de frameworks JSF. *3C TIC. Cuadernos de desarrollo aplicado a las TIC*, 8(4), 65-83. <http://doi.org/10.17993/3ctic.2019.84.65-83>

Zaki, M. & Meira, W. (2020). *Data mining and machine learning: fundamental concepts and algorithms* (2da. Ed.). Cambridge University Press. <https://doi.org/10.1017/9781108564175>

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Zimmerman, T. (2012). Exploring learner to content interaction as succes factor in online courses. *Review of research in open and distributed learning*, 13(4), 152-165.
[http:// dx.doi.org/10.19173/irrodl.v13i4.13](http://dx.doi.org/10.19173/irrodl.v13i4.13)

Universidad Juárez Autónoma de Tabasco
México.

Identificación de patrones de conducta, asociadas con las barreras para el uso de innovaciones tecnológicas en estudiantes de educación superior, aplicando técnicas de minería de datos

Glosario

C

CAD: Diseño Asistido por Computadora

CAM: Fabricación Asistida por Computadora

CRM: Administración Basada en la Relación con los Clientes

D

DOF: Diario Oficial de la Federación

DENUE: Directorio Estadístico Nacional de Unidades Económicas

Wendy Guadalupe Azuara García.pdf

 Universidad Juárez Autónoma de Tabasco

Detalles del documento

Identificador de la entrega

trn:oid:::3117:582710555

Fecha de entrega

24 abr 2026, 2:00 p.m. GMT-6

Fecha de descarga

24 abr 2026, 2:38 p.m. GMT-6

Nombre del archivo

Wendy Guadalupe Azuara García.pdf

Tamaño del archivo

10.8 MB

292 páginas

54.543 palabras

357.680 caracteres




9% 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 15 palabras)
- ▶ Abstract

Fuentes principales

- 9%  Fuentes de Internet
- 1%  Publicaciones
- 0%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad




N.º de alerta de integridad para revisión

-  **Caracteres reemplazados**
164 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
- 1%  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	
archivos.ujat.mx		4%
2	Internet	
dilemascontemporaneoseduccionpoliticayvalores.com		3%
3	Internet	
www.researchgate.net		<1%
4	Internet	
rte.espol.edu.ec		<1%
5	Internet	
hdl.handle.net		<1%
6	Internet	
revistasimbiosis.org		<1%
7	Internet	
redie.uabc.mx		<1%
8	Internet	
revistacomunicar.com		<1%
9	Internet	
www.coursehero.com		<1%
10	Internet	
www.dgzero.org		<1%
11	Internet	
www.ubu.es		<1%

26	Publicación	Santiago Andrés Rodríguez, Mónica López Ramírez, María Isabel Belausteguigoiti...	<1%
27	Internet	rclimatol.eu	<1%
28	Internet	www.learning-analytics.info	<1%
29	Internet	doi.org	<1%
30	Internet	revistas.unap.edu.pe	<1%
31	Internet	dialnet.unirioja.es	<1%
32	Internet	sontusdatos.org	<1%
33	Internet	libros-utp.com	<1%
34	Internet	portalinvestigacion.uniovi.es	<1%
35	Internet	repositorio.uptc.edu.co	<1%
36	Internet	www.ptolomeo.unam.mx:8080	<1%
37	Publicación	Yadira Robles Aranda, Anthony R. Sotolongo. "Integración de los algoritmos de m...	<1%
38	Internet	aws.amazon.com	<1%
39	Internet	ddd.uab.cat	<1%

40

Internet

www.lineaoutlet.com.mx

<1%