



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN



**IDENTIFICACIÓN DE PATRONES DE CONDUCTA, ASOCIADAS
CON PROCRASTINACIÓN Y AUTORREGULACIÓN ACADÉMICA
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:

Zully Kristel Guzmán Caraveo

Directores de Trabajo Recepcional:

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

Jurado Revisor:

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

Cuerpo Académico o Grupo de Investigación de los directores:
Gestión de Tecnologías de la Información

Línea de Generación y Aplicación del Conocimiento de la Maestría que
alimenta la investigación:

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

Cunduacán, Tabasco

Febrero de 2022



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS
Y TECNOLOGÍAS DE LA INFORMACIÓN



Cesión de Derechos

Cunduacán, Tabasco., a 04 de febrero de 2022

A quien corresponda:

La abajo firmante, en mi calidad de alumno de la DACyTI, en esta ciudad de Cunduacán, Tabasco, declaro que el trabajo recepcional desarrollado en la Universidad Juárez Autónoma de Tabasco titulado, **"Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica 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 **"Dra. Martha Patricia Silva Payró** y al **Dr. Guillermo de los Santos Torres"**, quienes participaron como asesores en el desarrollo del mismo y a los cuales relevo de cualquier sanción y asumo responder a cualquier reclamo de derechos de autor ante las autoridades competentes.

Atentamente

Autor:

Zully Kristel Guzmán Caraveo	R/A el Rio, Jalpa de Méndez, Tabasco.	
Nombre	Domicilio	Firma autógrafa



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS
Y TECNOLOGÍAS DE LA INFORMACIÓN



Carta 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 grado denominada, **"Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos"**, de la cual soy autor y titular de los Derechos de Autor.

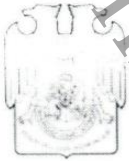
La finalidad del uso por parte de la Universidad Juárez Autónoma de Tabasco de la tesis antes, mencionada, será única y exclusivamente para difusión, educación y sin fines de lucro; autorización que se hace de manera enunciativa más no limitativa para subirla a la Red Abierta de Bibliotecas Digitales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

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

Se firma la presente autorización en la ciudad de Cunduacán, Tabasco a los 04 días del mes de febrero del año 2022.

Autorizo

Zully Kristel Guzmán Caraveo



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



Cunduacán, Tabasco a 31 de agosto de 2020

Asunto: Dirección 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 Directora de la tesis titulada "IDENTIFICACIÓN DE PATRONES DE CONDUCTA, ASOCIADAS CON PROCRASTINACIÓN Y AUTORREGULACIÓN ACADÉMICA EN ESTUDIANTES DE EDUCACIÓN SUPERIOR, APLICANDO TÉCNICAS DE MINERÍA DE DATOS", a realizar por la **C. Zully Kristel Guzmán Caraveo**, para obtener el grado de Maestro en Administración de Tecnologías de la Información.

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

Atentamente


MTE Oscar Alberto González González
Director



C.c.p. Coordinación de Posgrado
Alumno, C. Zully Kristel Guzmán Caraveo
Archivo.
Consecutivo.

[Firma manuscrita]

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



UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIA EN LA DUDA. ACCIÓN EN LA FE"



Cunduacán, Tabasco a 31 de agosto de 2020

Asunto: Dirección 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 PROCRASTINACIÓN Y AUTORREGULACIÓN ACADÉMICA EN ESTUDIANTES DE EDUCACIÓN SUPERIOR, APLICANDO TÉCNICAS DE MINERÍA DE DATOS"**, a realizar por la **C. Zully Kristel Guzmán Caraveo**, 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

UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN

C. c. p. Coordinación de Posgrado
Alumno, C. Zully Kristel Guzmán Caraveo
Archivo.
Consecutivo.



UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO

DIVISIÓN ACADÉMICA DE CIENCIAS
Y TECNOLOGÍAS DE LA INFORMACIÓN



F7: Respuesta de Jurado

Cunduacán, Tabasco., a 17 de enero de 2022

MTE. OSCAR 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 procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos**", realizada por el C. Zully Kristel Guzmán Caraveo, 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 correspondientes a la obtención del grado.

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

Atentamente integrantes del Jurado

Dr. Pablo Payró Campos

Dr. Gerardo Arceo Moheno

Dr. Pablo Pancardo García

Dra. Verónica García Martínez

Mtro. Rafael Mena de la Rosa

C.C.P. Dr. Eddy Arquímedes García Alcocer – Encargado del despacho de la Coordinación de Posgrado, DACYTI
Estudiante





UNIVERSIDAD JUÁREZ
AUTÓNOMA DE TABASCO

"ESTUDIO EN LA DUDA. ACCIÓN EN LA FE"



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS
DE LA INFORMACIÓN



PRECURSOR DE LA REVOLUCIÓN MEXICANA

"2022, Año de Ricardo Flores Magón"

Cunduacán, Tabasco a 31 de enero de 2022
Oficio No. 099/DACYTI/CP/2022

Asunto: Autorización de impresión de Tesis

C. Zully Kristel Guzmán Caraveo
Matricula: 192H19006

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 PROCRASTINACIÓN Y AUTORREGULACIÓN ACADÉMICA 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. Oscar Alberto González González
Director

UNIVERSIDAD JUÁREZ AUTÓNOMA DE TABASCO



DIVISIÓN ACADÉMICA DE
CIENCIAS Y TECNOLOGÍAS DE LA INFORMACIÓN

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

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

Agradecimientos

En primer lugar, expreso mi agradecimiento a mis directores de tesis, a la Dra. Martha Patricia Silva Payró y al Dr. Guillermo de los Santos Torres, por la oportunidad de trabajar y desarrollar esta investigación bajo su dirección, que no hubiera sido posible sin sus conocimientos, sugerencias e ideas. De igual manera, agradecer el apoyo que me han brindado, su paciencia y consejos, los cuales fueron siempre útiles cuando no salían de mi pensamiento las ideas para escribir lo que hoy he logrado, por sus palabras de aliento para seguir adelante y la confianza depositada en mí. Muchas gracias por sus orientaciones.

Gracias a mi universidad, la Universidad Juárez Autónoma de Tabasco (UJAT), por volverme a aceptar y abrirme las puertas nuevamente para seguirme forjando como profesionista en mi alma mater, pero en esta ocasión con la oportunidad de estudiar y obtener el grado de Maestro de Administración en Tecnología de la Información (MATI).

Quiero agradecer de igual forma, a todos los docentes por compartir sus conocimientos de manera profesional e invaluable, y fueron parte de este proceso integral de formación, que el día de hoy se ve relegado con la culminación de mis estudios. Gracias por su dedicación, perseverancia, paciencia y tolerancia.

Igualmente agradecer a los miembros del jurado, al Dr. Pablo Payró Campos, el Dr. Gerardo Arceo Moheno, al Dr. Pablo Pancardo García, la Dra. Verónica García Martínez y al Mtro. Rafael Mena de la Rosa, por sus valiosas contribuciones realizadas a la tesis,

con sus observaciones y comentarios, y agradezco su tiempo dedicado para la revisión de la misma.

Agradezco a mis compañeros y amigos con los cuales conviví dentro y fuera de las aulas, compartiendo experiencias y conocimientos obtenidos a lo largo de estos dos años de estudios, gracias por su apoyo y amistad.

Por último, agradezco hoy y siempre a mis padres y hermanos, los cuales siempre han estado conmigo acompañándome y animándome para no rendirme, pero sobre todo agradezco y doy gracias a mis padres por ser los principales motivadores en mi vida, por haberme forjado como la persona que soy actualmente, por sus consejos, valores y principios que me han inculcado, gracias por estar siempre a mi lado cuidándome y guiándome en cada momento.

A todos, mis agradecimientos y respetos.

Dedicatorias

A Dios:

Por darme salud, fortaleza, sabiduría y entendimiento para superar las dificultades que se me presentan en la vida, así como disfrutar los momentos buenos que he tenido a lo largo de estos años. Doy gracias por haberme regalado la dicha de concretar esta ilusión que parecía inalcanzable, por ser parte de mi en todo momento y guiar mis pasos en el sendero de la vida.

A mis padres:

Los cuales son una parte importante en mi vida, a mi papá José del Carmen Guzmán Cornelio y mi mamá Leticia Caraveo Chablé, quienes me han heredado el tesoro más valioso que puede dársele a una hija: Amor. Ustedes que con cariño y apoyo me han enseñado a luchar y cumplir las metas y objetivos que uno se proponga y no dejarse vencer por ninguna adversidad por muy difícil que sea, sino, al contrario, hay que seguir y mira hacia el futuro.

Dedico este logro, a ustedes papá y mamá, que sin en escatimar esfuerzo alguno, han sacrificado gran parte de su vida para formarme y educarme, para convertirme en una persona de provecho, y nunca podre pagar todos sus desvelos ni aun con las riquezas más grandes del mundo, por eso y más gracias. Los quiero mucho.

A mis hermanos:

Mayra y Armando los quiero mucho, gracias por su cariño, consejos y siempre brindarme su ayuda y apoyo incondicional en cada momento, importante en mi vida para cumplir

mis sueños y metas, pero sobre todo durante todo este proceso, donde estoy cumpliendo uno de mis grandes objetivos. Gracias por ser además de mis hermanos, mis amigos y cómplices de vida.

Muchas gracias por todo, los quiero.

Zully Kristel Guzmán Caraveo

México.

Universidad Juárez Autónoma de Tabasco.

Resumen

La procrastinación y autorregulación académica son patrones de comportamiento que se dan entre los jóvenes, principalmente los estudiantes. Por un lado, la procrastinación, es el aplazamiento de la realización de actividades y la autorregulación académica, es la forma de aprendizaje que el estudiante ejerce para aumentar su rendimiento académico. El objetivo de esta investigación es identificar los patrones de conducta, relacionados con procrastinación y autorregulación académica de estudiantes de licenciatura de la Universidad Juárez Autónoma de Tabasco (UJAT), empleando técnicas de minería de datos; las cuales son: arboles de decisión, agrupamiento y reglas de asociación. el enfoque de investigación es de tipo cuantitativo. La metodología utilizada fue una metodología de tres fases, la primera fase fue analizar la información que contiene el *dataset*, la segunda fase fue la selección de las herramientas para aplicar minería de datos y la tercera fase, la utilización de la metodología *KDD* (Descubrimiento de Conocimiento en Base de Datos), la cual en uno de sus pasos está la aplicación de la minería de datos. Mediante la realización y aplicación de las técnicas de minería de datos en las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*, se identificaron patrones, similitudes y asociaciones entre los datos y las variables que conforma el *dataset*; de igual manera, con la técnica de árbol de decisión se detectaron una serie de variables que no intervienen en la clasificación de la clase del estudiante, con la técnica de agrupamiento se observaron instancias que aunque la clase sea autorregulador o procrastinador se agruparon en el mismo *clúster* y con la técnica de

reglas de asociación, se encontraron 13,215 reglas asociadas con las 29 variables y 746 instancias que lo conforman, además, se realizaron tres pruebas por cada técnica de minería de datos y realizadas en las tres herramientas mencionadas anteriormente.

Universidad Juárez Autónoma de Tabasco.
México.

Introducción

El presente trabajo consistió en la identificación de patrones de conducta, asociados con procrastinación y autorregulación académica en estudiantes de licenciatura de la Universidad Juárez Autónoma de Tabasco (UJAT), aplicando técnicas de minería de datos, cuyo proceso se describe en cinco capítulos, tres apéndices y seis anexos.

En el capítulo uno, se tratan las generalidades de la investigación, los cuales son los antecedentes, planteamiento del problema, donde se incluye la definición del problema, delimitación de la investigación, objetivo general y específicos; de igual forma en este capítulo, se incluye la justificación y por último la metodología utilizada, donde se explicó la metodología utilizada para el desarrollo de la investigación y el enfoque de investigación utilizada.

En el capítulo dos se encuentra el marco teórico de la investigación, el cual se conforma por la descripción del marco referencial que contiene información de artículos científicos y tesis de licenciatura, de maestría y doctorado, acerca de investigaciones sobre procrastinación, autorregulación académica y minería de datos. El marco conceptual que presenta conceptos e investigaciones relacionadas sobre procrastinación, técnicas de minería de datos, proceso *KDD*, entre otros conceptos.

De igual forma, en el capítulo dos, se encuentra el marco tecnológico, donde se describen los programas o herramientas para realizar minería de datos, de las cuales están seccionadas por: herramientas *framework*, lenguajes de programación, herramientas de Microsoft y servicio en la nube y, por último, el marco legal.

El capítulo tres contiene el desarrollo de la metodología implementada para realizar la investigación, donde la primera fase consistió en analizar la información que conforma el *dataset* utilizado, la segunda fase, que consistió en la selección de la herramienta para aplicar minería de datos, donde se seleccionaron las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*, y la última fase que consistió en emplear la metodología del proceso KDD para realizar la limpieza de los datos, ejecución de las técnicas de minería de datos en las herramientas y la interpretación de la información descubierta.

El capítulo cuatro describe las pruebas realizadas al *dataset* con las técnicas de minería de datos, utilizando diferentes valores para los parámetros de los mismos y los resultados de la implementación de las técnicas de árbol de decisión, agrupamiento y reglas de asociación en las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*.

Finalmente, en el capítulo cinco se presentan las conclusiones, recomendaciones y trabajos futuros, que se pueden realizar más adelante derivados de esta investigación y sobre las variables de procrastinación y autorregulación académica.

De igual manera, los tres anexos que conforman este trabajo de investigación se identifican con las letras "A", "B" y "C". En el anexo A se presenta las descripciones de las variables y datos que conforman el *dataset* utilizado en la investigación, el apéndice B y C son las reglas de asociación obtenidas en las pruebas uno y dos respectivamente, con la herramienta de *Microsoft Azure Machine Learning*.

Por último, en los apéndices se identifican con las letras “A” a la “E”. El apéndice A, trata acerca de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares: capítulo uno, disposiciones generales; el apéndice B, es el cuestionario que fue aplicado a los estudiantes de la UJAT, donde se obtuvo el *dataset* en cuestión; y finalmente los apéndices de la C a la E, contiene el código fuente utilizado para la implementación de los algoritmos de las técnicas de árbol de decisión, agrupamiento y reglas de asociación en la herramienta de Lenguaje R.

Universidad Juárez Autónoma de Tabasco.
México.

Índice general

Índice de tablas	xxiv
Índice de figuras	xxvi
Capítulo 1. Generalidades	1
1.1 Antecedentes	1
1.1.1 Procrastinación	1
1.1.2 Autorregulación académica	3
1.1.3 Minería de datos	4
1.2 Planteamiento del problema	6
1.2.1 Definición del problema	6
1.2.2 Delimitación de la investigación	8
1.3 Objetivos	9
1.3.1 Objetivo general	9
1.3.2 Objetivos específicos	9
1.4 Justificación	10
1.5 Metodología utilizada	13
1.5.1 Enfoque de investigación	14
1.5.2 Fuentes de investigación	14
1.5.3 Técnicas de recolección de datos	15
1.5.4 Modelo de desarrollo	15
1.5.5 Tipo de programación	17

Capítulo 2. Marco teórico.....	18
2.1 Marco referencial.....	18
2.1.1 Análisis del impacto de la procrastinación en el rendimiento académico de estudiantes universitarios usando técnicas de minería de datos	19
2.1.2 El papel de la procrastinación académica como factor de la deserción universitaria.....	20
2.1.3 Gestión del tiempo y procrastinación en la educación superior	21
2.1.4 Procrastinación, adicción al internet y rendimiento académico de estudiantes universitarios ecuatorianos	21
2.1.5 Procrastinación académica como predictor en el rendimiento académico en jóvenes de educación superior.....	22
2.1.6 Evidencias de validez y confiabilidad de la escala de procrastinación académica en una población estudiantil.....	23
2.1.7 Procrastinación académica entre estudiantes de cursos en línea. Validación de un cuestionario.....	24
2.1.8 La autorregulación académica como varianza explicativa de los procesos de aprendizaje universitarios.....	25
2.1.9 Teorías de autorregulación educativa: una comparación y reflexión teórica.	25
2.1.10 Autorregulación y rendimiento académico en la transición secundaria- universidad.....	26

2.1.11 Evaluación de autorregulación académica en estudiantes de psicología en modalidad en línea	27
2.1.12 Comparación de estrategias de estudio y autorregulación en universitarios	28
2.1.13 The <i>KDD</i> Process for extracting useful knowledge from volumes of data	29
2.1.14 Análisis comparativo usando minería de datos en la predicción del rendimiento académico de adolescentes basados en emociones.....	30
2.1.15 Análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos.....	31
2.1.16 Análisis de deserción-permanencia de estudiantes universitarios utilizando técnicas de clasificación en minería de datos	32
2.1.17 Adicción a las nuevas tecnologías y a las redes sociales en jóvenes: un nuevo reto	33
2.1.18 Gestión del tiempo y uso de las TIC en estudiantes universitarios	33
2.1.19 Adicción a internet: desarrollo y validación de un instrumento en escolares adolescentes de Lima, Perú.....	34
2.2 Marco conceptual.....	35
2.2.1 Procrastinación.....	35
2.2.2 Autorregulación académica	35
2.2.3 Proceso <i>KDD</i>	36
2.2.4 Minería de datos.....	38

2.2.5	Técnicas de minería de datos	38
2.2.6	Métodos de aprendizaje en las técnicas de minería de datos	40
2.2.7	Uso de internet	42
2.2.8	Modelo FURPS	42
2.3	Marco tecnológico	44
2.3.1	Herramientas <i>framework</i> para minería de datos	44
2.3.2	Lenguaje de programación para minería de datos	46
2.3.3	Herramientas de Microsoft para minería de datos.....	47
2.3.4	Servicios en la nube para minería de datos	48
2.4	Marco legal	51
2.4.1	Licencia de software libre GPL.....	51
2.4.2	Licencia de software libre AGPL	52
2.4.3	Licencia de software propietario.....	53
2.4.4	Ley Federal de Protección de Datos Personales en Posesión de los Particulares	54
Capítulo 3.	Aplicación de la metodología y desarrollo	55
3.1	Análisis de la información del dataset.....	55
3.1.1	Cuantificación de la población de estudio	56
3.1.2	Depuración de la población	57
3.1.3	Diseño de instrumento	57
3.1.4	Aplicación del instrumento	59

3.1.5	Visualización y exploración de los datos del <i>dataset</i>	59
3.2	Selección de las herramientas para aplicar minería de datos.....	62
3.2.1	Aplicación de la metodología FURPS	64
3.2.2	Características de las herramientas seleccionadas	69
3.3	Metodología del proceso <i>KDD</i>	76
3.3.1	Aprendizaje del dominio de la aplicación	78
3.3.2	Creación de un conjunto de datos de destino	84
3.3.3	Limpieza de datos y procesamiento previo	86
3.3.4	Reducción y proyección de datos.....	91
3.3.5	Elección de función de los datos de extracción.....	92
3.3.6	Elección de algoritmo(s) de minería de datos	94
3.3.7	Minería de datos.....	97
3.3.8	Interpretación	162
3.3.9	Conocimiento descubierto.....	164
Capítulo 4.	Pruebas y Resultados.....	165
4.1	Pruebas de las técnicas de minería de datos	165
4.1.1	Pruebas de Árboles de decisión.....	166
4.1.2	Pruebas de Agrupamiento.....	191
4.1.3	Pruebas de Reglas de asociación	216
4.2	Resultados.....	232
4.2.1	Resultados de Árboles de decisión	232

4.2.2 Resultados de Agrupamiento	251
4.2.3 Resultados de Reglas de asociación	261
Capítulo 5. Conclusiones, recomendaciones y trabajos futuros	280
5.1 Conclusiones	280
5.2 Recomendaciones	283
5.3 Trabajos futuros	284
Referencias	286
Glosario	304
Apéndice A. Catálogo y descripción de las variables y datos del <i>dataset</i>	308
Apéndice B. Reglas de asociación obtenidas en la prueba uno en la herramienta <i>Microsoft Azure Machine Learning</i>	318
Apéndice C. Reglas de asociación obtenidas en la prueba dos en la herramienta <i>Microsoft Azure Machine Learning</i>	320
Anexo A. Ley Federal de Protección de Datos Personales en Posesión de los Particulares: Capítulo I Disposición Generales	322
Anexo B. Cuestionario aplicado a los estudiantes para la recolección de datos	326
Anexo C. Código para la creación del modelo de árbol de decisión en Lenguaje R ...	330
Anexo D. Código para la creación del modelo de agrupamiento en Lenguaje R	333
Anexo E. Código para la creación del modelo de reglas de asociación en Lenguaje R	336

Índice de tablas

Tabla 1 Dificultades de autorregulación con la procrastinación.....	20
Tabla 2 Comparación de teorías de autorregulación.....	26
Tabla 3 Distribución de participantes por División Académica.	57
Tabla 4 Ítems del instrumento para medir la procrastinación en estudiantes universitarios.....	58
Tabla 5 Datos nulos o vacíos identificados por cada ítem.....	61
Tabla 6 Factores y criterios de la metodología FURPS.....	63
Tabla 7 Factores, criterios y ponderación de la metodología FURPS para la selección de las herramientas para minería de datos.....	64
Tabla 8 Clasificación de las herramientas analizadas con el modelo FURPS.....	65
Tabla 9 Análisis de la selección de las herramientas para minería de datos aplicando la metodología FURPS.	67
Tabla 10 Ponderación de las herramientas de acuerdo con la metodología FURPS.	69
Tabla 11 Características de las categorías e instancias de trabajo de Microsoft Azure Machine Learning.....	73
Tabla 12 Instancias y costo de las instancias de trabajo de Microsoft Azure Machine Learning.	74
Tabla 13 Instancia de trabajo seleccionada en Microsoft Azure Machine Learning	75
Tabla 14 Conjunto de datos del dataset.	85
Tabla 15 Modificación del nombre de las variables del dataset.....	89

Tabla 16 Extracto de la homogeneización y numeración de los datos del campo Ciclo Escolar.	90
Tabla 17 Reglas de asociación obtenidas en Microsoft Azure Machine Learning.	267
Tabla 18 Reglas de asociación obtenidas en Weka.	274
Tabla 19 Catálogo de las variables del dataset.	308

Índice de figuras

Figura 1 Pasos que constituyen el proceso KDD.	29
Figura 2 Árbol de decisión simple con las pruebas en los atributos X e Y.	39
Figura 3 Ejemplo de aprendizaje supervisado.....	41
Figura 4 Ejemplo de aprendizaje no supervisado.....	42
Figura 5 Dataset utilizado en la aplicación de minería de datos.....	56
Figura 6 Filtro para identificar datos nulos o vacíos en el dataset.	60
Figura 7 Identificación de datos nulos o vacíos en el dataset.....	61
Figura 8 Técnicas y algoritmos para minería de datos que ofrece Microsoft Machine Learning.	72
Figura 9 Metodología del proceso KDD.....	77
Figura 10 Interfaz principal de la herramienta Lenguaje R.	79
Figura 11 Interfaz principal de la herramienta Weka.....	80
Figura 12 Interfaz principal de Explorer de la herramienta Weka.....	81
Figura 13 Interfaz principal de Microsoft Azure.	82
Figura 14 Interfaz para la creación del área de trabajo en Microsoft Azure Machine Learning.....	83
Figura 15 Área de trabajo creada en Microsoft Azure Machine Learning.....	84
Figura 16 Tabla generada en Weka en relación con el campo Ciclo escolar.....	87
Figura 17 Preprocesamiento del dataset en las herramientas.....	88
Figura 18 Renombre y modificaciones de los datos del dataset.....	91

Figura 19 Clasificación de las técnicas de minería de datos.	93
Figura 20 Tipos de algoritmos y Algoritmo C4.5 o J48 de árbol de decisión.	95
Figura 21 Tipos de algoritmos y Algoritmo K-means de agrupamiento.	96
Figura 22 Tipos de algoritmos y Algoritmo A priori de reglas de asociación.	97
Figura 23 Lectura del dataset en RStudio para árbol de decisión.	99
Figura 24 Librería cargada C50 para el árbol de decisión en RStudio.	100
Figura 25 Cambio del tipo de dato de la variable categórica ClaseEstudiante para árbol de decisión en RStudio.	100
Figura 26 Creación de sets de entrenamiento y prueba del modelo de árbol de decisión en RStudio.	101
Figura 27 Código para la creación del árbol de decisión en RStudio.	102
Figura 28 Modelo e información del Árbol de decisión en RStudio.	103
Figura 29 Árbol de decisión en RStudio.	104
Figura 30 Verificación de la capacidad de predicción del árbol de decisión en RStudio.	105
Figura 31 Matriz de confusión y porcentaje de aciertos de árbol de decisión en RStudio.	105
Figura 32 Librerías utilizadas para la creación del modelo de agrupamiento en RStudio.	106
Figura 33 Normalización de los datos para el modelo de agrupamiento en RStudio. .	107

Figura 34 Cálculo de la matriz de distancia utilizando el método euclídea en RStudio.....	108
Figura 35 Matriz de distancias de los datos en RStudio.....	108
Figura 36 Estimación de número de cluster para el modelo de agrupamiento en RStudio.....	109
Figura 37 Estimación del número óptimo de clústeres a implementar en los datos en RStudio.....	110
Figura 38 Creación del modelo de agrupamiento en RStudio.....	111
Figura 39 Información obtenida por el modelo de agrupamiento en RStudio.....	112
Figura 40 Segunda forma de visualizar los resultados del componente del modelo de agrupamiento en RStudio.....	113
Figura 41 Representación gráfica del modelo de agrupamiento en RStudio.....	114
Figura 42 Revisión de los datos en función y medias en función de los cluster del modelo de agrupamiento en RStudio.....	115
Figura 43 Ejemplo de revisión de datos en función al modelo de agrupamiento en RStudio.....	115
Figura 44 Visualización de la asignación de los cluster al dataset del modelo de agrupamiento en RStudio.....	117
Figura 45 Resumen de las variables del dataset con tipo de dato carácter en RStudio.....	118
Figura 46 Creación del modelo A priori de reglas de asociación en RStudio.....	119

Figura 59 Entrenamiento del modelo y selección de la variable a entrenar en el árbol de decisión en Azure Machine Learning Studio.....	130
Figura 60 Modelo de árbol de decisión en Azure Machine Learning Studio.....	131
Figura 61 Ejecución del modelo de árbol de decisión en Azure Machine Learning Studio.	132
Figura 62 Ejecución completa de árbol de decisión en Azure Machine Learning Studio.	133
Figura 63 Árbol de decisión en Azure Machine Learning Studio.	134
Figura 64 Normalizar datos en el modelo de agrupamiento en Azure Machine Learning Studio.....	135
Figura 65 Selección de las variables a normalizar para el modelo de agrupamiento en Azure Machine Learning Studio.....	136
Figura 66 Configuración de parámetros del algoritmo K-means en Azure Machine Learning Studio.....	137
Figura 67 Modelo de agrupamiento en Azure Machine Learning Studio.	138
Figura 68 Ejecución completa de agrupamiento en Azure Machine Learning Studio..	139
Figura 69 Asignación de cluster con la técnica de agrupamiento en Azure Machine Learning Studio.....	140
Figura 70 Métricas obtenidas con la técnica de agrupamiento en Azure Machine Learning Studio.....	141

Figura 71 Creación del plan de servicio web de Machine Learning Studio (clásico) para la técnica de reglas de asociación.....	142
Figura 72 Creación del área de trabajo de Machine Learning Studio (clásico) para la técnica de reglas de asociación.....	143
Figura 73 Creación de cuenta de almacenamiento para el área de trabajo de Machine Learning Studio (clásico) para la técnica de reglas de asociación.	144
Figura 74 Selección del plan de servicio web para el área de trabajo de Machine Learning Studio (clásico) para la técnica de reglas de asociación.	144
Figura 75 Revisión y creación del área de trabajo de Machine Learning Studio (clásico) para la técnica de reglas de asociación.....	145
Figura 76 Subir el conjunto de datos a Machine Learning Studio (clásico) para la técnica de reglas de asociación.....	146
Figura 77 Importar el modelo de Discover association rules en Machine Learning Studio (clásico) para la técnica de reglas de asociación.	147
Figura 78 Selección del dataset para ejecutar la regla de asociación en Machine Learning Studio (clásico).	148
Figura 79 Código para identificar reglas de asociación en Machine Learning Studio (clásico).	149
Figura 80 Modelo de reglas de asociación en Machine Learning Studio (clásico).	150
Figura 81 Reglas de asociación obtenidas con el algoritmo A priori en Machine Learning Studio (clásico).	151

Figura 82 Cargar y abrir dataset en Weka.....	152
Figura 83 Visualización del dataset en Weka.....	153
Figura 84 Selección del algoritmo J48 para árbol de decisión en Weka.	154
Figura 85 Minería de datos con árbol de decisión en Weka.....	155
Figura 86 Visualizar el árbol de decisión en Weka.	156
Figura 87 Modelo clasificador del árbol de decisión realizado en Weka.	156
Figura 88 Selección del algoritmo SimpleK-means en Weka.	157
Figura 89 Minería de datos con agrupamiento en Weka.	158
Figura 90 Visualización de agrupamiento en Weka.....	159
Figura 91 Modelo de asignación de agrupamiento de los datos realizados en Weka.	159
Figura 92 Carga del dataset con datos nominales para implementar el algoritmo A piori en Weka.....	160
Figura 93 Cuadro de diálogo para modificar parámetros del algoritmo A pirori en Weka.	161
Figura 94 Reglas de asociación encontradas con el algoritmo A priori en Weka.	162
Figura 95 Árbol de decisión obtenido en la prueba uno con Lenguaje R.	168
Figura 96 Diagrama del árbol de decisión obtenido en la prueba uno con Lenguaje R.....	169
Figura 97 Matriz de confusión y porcentaje de acierto de la prueba uno con Lenguaje R.....	170
Figura 98 Árbol de decisión obtenido en la prueba dos con Lenguaje R.....	171

Figura 99 Diagrama del árbol de decisión obtenido en la prueba dos con Lenguaje R.....	172
Figura 100 Matriz de confusión y porcentaje de acierto de la prueba dos con Lenguaje R.....	173
Figura 101 Árbol de decisión obtenido en la prueba tres con Lenguaje R.	174
Figura 102 Diagrama del árbol de decisión obtenido en la prueba tres con Lenguaje R.....	175
Figura 103 Matriz de confusión y porcentaje de acierto de la prueba tres con Lenguaje R.	175
Figura 104 Predicción de la clase de acuerdo con el árbol de decisión obtenido en la prueba uno con Microsoft Azure Machine Learning.....	176
Figura 105 Conjunto de métricas obtenidas de la prueba uno del árbol de decisión con Microsoft Azure Machine Learning.....	177
Figura 106 Matriz de confusión obtenida de la prueba uno del árbol de decisión con Microsoft Azure Machine Learning.	178
Figura 107 Predicción de la clase de acuerdo con el árbol de decisión obtenido en la prueba dos con Microsoft Azure Machine Learning.....	179
Figura 108 Conjunto de métricas obtenidas de la prueba dos del árbol de decisión con Microsoft Azure Machine Learning.....	180
Figura 109 Matriz de confusión obtenida de la prueba dos del árbol de decisión con Microsoft Azure Machine Learning.	181

Figura 110 Predicción de la clase de acuerdo con el árbol de decisión obtenido en la prueba tres con Microsoft Azure Machine Learning.....	182
Figura 111 Conjunto de métricas obtenidas de la prueba tres del árbol de decisión con Microsoft Azure Machine Learning.....	183
Figura 112 Matriz de confusión obtenida de la prueba tres del árbol de decisión con Microsoft Azure Machine Learning.	183
Figura 113 Árbol de decisión obtenido en la prueba uno con Weka.....	184
Figura 114 Diagrama del árbol de decisión obtenido en la prueba uno con Weka.....	185
Figura 115 Matriz de confusión de la prueba uno con Weka.....	186
Figura 116 Árbol de decisión obtenido en la prueba dos con Weka.....	187
Figura 117 Diagrama del árbol de decisión obtenido en la prueba dos con Weka.	188
Figura 118 Matriz de confusión de la prueba dos con Weka.....	188
Figura 119 Árbol de decisión obtenido en la prueba tres con Weka.....	189
Figura 120 Diagrama del árbol de decisión obtenido en la prueba uno con Weka.....	190
Figura 121 Matriz de confusión de la prueba tres con Weka.....	190
Figura 122 Modelo de agrupamiento con tres clústeres de la prueba uno en Lenguaje R.	192
Figura 123 Gráfico de los tres agrupamientos de la prueba uno en Lenguaje R.....	193
Figura 124 Visualización de la asignación de las instancias en los clústeres de acuerdo con la variable ClaseEstudiante con la prueba uno en Lenguaje R.....	194

Figura 125 Visualización de los grupos de agrupamiento de la prueba uno en Lenguaje R.	195
Figura 126 Modelo de agrupamiento con cuatro clústeres de la prueba dos en Lenguaje R.	196
Figura 127 Gráfico de los cuatro agrupamientos de la prueba dos en Lenguaje R.	197
Figura 128 Visualización de la asignación de las instancias en los clústeres de acuerdo con la variable ClaseEstudiante con la prueba tres en Lenguaje R.	198
Figura 129 Visualización de los grupos de agrupamiento de la prueba dos en Lenguaje R.	198
Figura 130 Modelo de agrupamiento con dos clústeres de la prueba tres en Lenguaje R.	199
Figura 131 Gráfico de los dos agrupamientos de la prueba tres en Lenguaje R.	200
Figura 132 Visualización de la asignación de las instancias en los clústeres de acuerdo con la variable Especie con la prueba tres en Lenguaje R.	201
Figura 133 Visualización de los grupos de agrupamiento de la prueba tres en Lenguaje R.	201
Figura 134 Agrupamiento de los tres clústeres de la prueba uno en Microsoft Azure Machine Learning.	202
Figura 135 Evaluación de la asignación de clúster con la prueba uno con Microsoft Azure Machine Learning.	204

Figura 136 Agrupamiento de los cuatro clústeres de la prueba dos en Microsoft Azure Machine Learning.....	204
Figura 137 Evaluación de la asignación de clúster con la prueba dos con Microsoft Azure Machine Learning.....	206
Figura 138 Agrupamiento de los dos clústeres de la prueba tres en Microsoft Azure Machine Learning.....	207
Figura 139 Evaluación de la asignación de clúster con la prueba tres con Microsoft Azure Machine Learning.....	208
Figura 140 Modelo de agrupamiento con tres clústeres de la prueba uno en Weka... ..	209
Figura 141 Visualización de la asignación de las instancias en los clústeres y asignación, de acuerdo con la variable ClaseEstudiante con la prueba uno en Weka.	210
Figura 142 Gráfico de los tres agrupamientos de la prueba uno en Weka.....	210
Figura 143 Modelo de agrupamiento con cuatro clústeres de la prueba dos en Weka.	211
Figura 144 Visualización de la asignación de las instancias en los clústeres y asignación, de acuerdo con la variable ClaseEstudiante con la prueba dos en Weka.	212
Figura 145 Gráfico de los cuatro agrupamientos de la prueba dos en Weka.....	213
Figura 146 Modelo de agrupamiento con dos clústeres de la prueba tres en Weka... ..	214

Figura 147 Visualización de la asignación de las instancias en los clústeres y asignación, de acuerdo con la variable clase con la prueba tres en Weka.	215
Figura 148 Gráfico de los cuatro agrupamientos de la prueba dos en Weka.	215
Figura 149 Modelo de reglas de asociación obtenido en la prueba uno en Lenguaje R.	218
Figura 150 Reglas con mayor confianza de las reglas de asociación obtenidas en la prueba uno con Lenguaje R.	218
Figura 151 Modelo de reglas de asociación obtenido en la prueba dos en Lenguaje R.	220
Figura 152 Reglas con mayor confianza de las reglas de asociación obtenidas en la prueba dos con Lenguaje R.	220
Figura 153 Modelo de reglas de asociación obtenido en la prueba tres en Lenguaje R.	222
Figura 154 Reglas con mayor confianza de las reglas de asociación obtenidas en la prueba tres con Lenguaje R.	222
Figura 155 Modelo de reglas de asociación obtenido en la prueba uno en Machine Azure Machine Learning.	224
Figura 156 Modelo de reglas de asociación obtenido en la prueba dos en Machine Azure Machine Learning.	225
Figura 157 Modelo de reglas de asociación obtenido en la prueba tres en Machine Azure Machine Learning.	226

Figura 158 Modelo de reglas de asociación obtenido en la prueba uno en Weka.....	227
Figura 159 Reglas con mayor confianza de las reglas de asociación obtenido en la prueba uno con Weka.....	228
Figura 160 Modelo de reglas de asociación obtenido en la prueba dos en Weka.....	229
Figura 161 Reglas con mayor confianza de las reglas de asociación obtenido en la prueba dos con Weka.....	230
Figura 162 Modelo de reglas de asociación obtenido en la prueba tres en Weka.....	231
Figura 163 Resultado del árbol de decisión obtenido en Lenguaje R.....	235
Figura 164 Matriz de confusión de los resultados del árbol de decisión obtenido en Lenguaje R.	239
Figura 165 Diagrama del árbol de decisión de los resultados obtenido con Lenguaje R.....	240
Figura 166 Resultado del árbol de decisión obtenido en Microsoft Azure Machine Learning.....	241
Figura 167 Conjunto de métricas del árbol de decisión obtenido en Microsoft Azure Machine Learning.	242
Figura 168 Matriz de confusión de los resultados del árbol de decisión obtenido en Microsoft Azure Machine Learning.	243
Figura 169 Resultado del árbol de decisión obtenido en Weka.....	245
Figura 170 Matriz de confusión de los resultados del árbol de decisión obtenido en Weka.	249

Figura 171 Diagrama del árbol de decisión de los resultados obtenido con Weka.....	250
Figura 172 Resultados de agrupamiento obtenido en Lenguaje R.....	253
Figura 173 Gráfico del resultado obtenido con la técnica de agrupamiento en Lenguaje R.....	254
Figura 174 Visualización de la asignación de las instancias en los clústeres de acuerdo con la variable ClaseEstudiante en Lenguaje R.	255
Figura 175 Resultados de agrupamiento obtenido en Microsoft Azure Machine Learning.....	256
Figura 176 Evaluación de la asignación de clúster con Microsoft Azure Machine Learning.....	257
Figura 177 Resultados de agrupamiento obtenido en Weka.....	258
Figura 178 Visualización de la asignación de las instancias en los clústeres y la asignación de acuerdo con la variable ClaseEstudiante en Weka.....	259
Figura 179 Gráfico del resultado obtenido con la técnica de agrupamiento en Weka.	260
Figura 180 Resultados del modelo de reglas de asociación obtenido en Lenguaje R.	262
Figura 181 Resultados de las reglas de asociación con Lenguaje R.	263
Figura 182 Resultados del modelo de reglas de asociación obtenido en Microsoft Azure Machine Learning.....	267
Figura 183 Modelo de reglas de asociación obtenido en Weka.	273
Figura 184 Resultados de las reglas de asociación obtenidas con Weka.	274

Capítulo 1. Generalidades

1.1 Antecedentes

1.1.1 Procrastinación

La procrastinación -también conocida como dilación- es entendida como un patrón de comportamiento que se caracteriza por aplazar voluntariamente la realización de actividades que deben ser entregadas en un momento establecido (Ferrari & Diaz-Morales, 2007; Ferrari & Tice, 2000; Steel, 2007). Según Quant y Sánchez (2012) la identificación de conductas relacionadas con la procrastinación se asocia a comportamientos que se presentan de forma crónica, como una baja autoestima, déficit en autoconfianza, déficit de autocontrol, depresión, comportamientos de desorganización y en algunos casos perfeccionismo, impulsividad disfuncional y ansiedad (Ferrari & Diaz-Morales, 2007; Senecal, Koestner & Vallerand, 1995; Solomon & Rothblum, 1984; Spada, Hiou & Nikcevic, 2006).

Es claro que un estudiante universitario, enfrenta una serie de retos desde el momento que inicia sus estudios superiores hasta que los concluye, los cuales no son únicamente de índole académica, sino familiar, social, económica, entre otras; así como un sinnúmero de retos personales que engloban un conjunto de aspectos emocionales que en la mayoría de los casos no son resueltos y que por ende inciden sobre todo en el rendimiento académico, base de la formación del futuro profesional (Chan, 2011).

Según Ramos-Galarza, Jadán-Guerrero, Paredes-Núñez, Bolaños-Pasquel y Gómez-García (2017) el ideal de estudiante universitario es aquel que cumple de manera

satisfactoria con todas las demandas del medio educativo superior, sin embargo, ésta no es una realidad absoluta, ya que existe un elevado porcentaje de estudiantes que no son capaces de tener éxito académico, porque presentan, entre otros fenómenos, un comportamiento procrastinador (Wilson & Bguyen, 2012).

Steel (2007) señala que el primer análisis histórico sobre la procrastinación fue escrito por Milgram en 1992, quien argumentó que las personas usualmente adquieren varios compromisos que deben cumplir en periodos de tiempo cortos, situación que conlleva a que se presente la procrastinación. Ferrari, Johnson & McCown (1995) señalan que la procrastinación se ha presentado con frecuencia a lo largo de la historia, considerándose como un aspecto usual en el comportamiento humano que se asocia a connotaciones negativas (Quant y Sánchez, 2012).

Takás (como se citó en Angarita, 2012) clasifica a las personas que procrastinan en siete tipos: (1) perfeccionista, quien concibe las responsabilidades o tareas como aplazables debido a que no abarcan sus estándares personales, (2) soñador, se caracteriza por ser divagador y poco realista, (3) preocupado, aquel sujeto que se percibe afectado en los momentos que siente que las cosas están fuera de su control, (4) generador de crisis, el disfruta de interacciones asociadas con el aplazamiento, (5) desafiante, utiliza la agresión como excusa de posible control por parte de terceros, (6) ocupado, realiza varias actividades al mismo tiempo, sin embargo, no culmina ninguna; y (7) relajado, el tipo de persona que evita situaciones y actividades que le generan algún tipo de estrés y compromiso.

La procrastinación es un comportamiento común que se presenta en gran medida en la vida universitaria del estudiante, puede afectar a la persona en sus relaciones sociales debido a los pensamientos de desamparo e ineficacia, pues no le permiten la consecución y el desarrollo de la actividad académica. Posibles factores a este comportamiento son el sentimiento de baja responsabilidad o poca concentración, ansiedad y miedo a no tener éxito por las percepciones negativas y factores de la personalidad que influyen en la dilación académica (Balkis & Duru, 2009).

1.1.2 Autorregulación académica

La construcción del aprendizaje a una actividad intencional y consciente orientada a conseguir una construcción significativa para el sujeto que aprende, por tanto, es una conducta estratégica y en consecuencia de un aprendizaje estratégico; el cual debe ser reflexivo y con el fin de autonomía de la persona y la autodirección (Pozo y Monereo, 2000, como se citó en García, 2012).

La autorregulación académica se entiende como la acción reguladora que una persona ejerce en los distintos momentos de su proceso de aprendizaje. Para que esta acción reguladora sea posible es necesario primero el conocimiento de eso que hacemos y conocemos, conocimiento al que accedemos por medio de la metacognición (Monereo, 2001, como se citó en García, 2012).

Un alumno difícilmente será autorregulado académicamente si no posee conocimiento sobre sí mismo, sus emociones y sus estrategias de aprendizaje. Este “autoconocimiento” es el que le capacita para cuestionar, planificar y evaluar sus acciones

de aprendizaje y su proceso de aprendizaje en sí mismo; en definitiva, le capacita para “gobernarse a sí mismo” (Monereo, 2001, como se citó en García, 2012).

Resumiendo, la autorregulación se constituye como un proceso activo en el que los estudiantes establecen sus objetivos principales de aprendizaje y a lo largo de éste, tratan de conocer, controlar y regular sus cogniciones, motivaciones y comportamientos de cara a alcanzar objetivos (Valle et al., 2008).

1.1.3 Minería de datos

Con el transcurso de los años, las actividades y el desarrollo de nuevas tecnologías han generado de forma considerable el almacenamiento de información, donde todo ese flujo de información recolectado ha permitido satisfacer las necesidades diarias de las organizaciones, pero ha presentado un problema inherente en las capacidades humanas para analizar y transformar la información en conocimiento útil y relevante que apoye a la toma de decisiones (Ballesteros, Sánchez-Guzmán y García, 2013).

Este tipo de problemas comúnmente se han presentado en los grandes sistemas conocidos como bases de datos, en los cuales almacenan una gran cantidad de información; estos datos con frecuencia contienen valioso contenido y pueden ser vistos como una recopilación masiva de información para uso de la mayoría de las organizaciones, aun con el uso de herramientas estadísticas clásicas para la manipulación, el análisis y la extracción de información, esta tarea es casi imposible, por esta razón ha motivado al empleo de técnicas y herramientas de minería de datos (Ballesteros et al., 2013).

Según Ballesteros et al. (2013) Minería de Datos (Por sus siglas en ingles *DM– Data Mining*), posibilita la extracción de conocimiento en forma de reglas y patrones a partir de dichos datos, es decir, es un proceso de análisis de datos que consiste en buscar o encontrar tendencias o variaciones de comportamiento en los datos denominados patrones.

Ribas (2018) describe la minería de datos como un conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, que expliquen el comportamiento de estos datos, para contribuir en la mejorar y crecimiento de las empresas donde son más utilizadas. De igual forma, la minería de datos se ha utilizado también en el sector educativo, en varias disciplinas como las ciencias de la computación, matemáticas, educación, entre otras, para realizar estudios y análisis más detallados y precisos sobre el comportamiento que tienen los alumnos en los diversos sistemas de aprendizaje-enseñanza (Sánchez, 2009).

Ballesteros et al. (2013) hacen mención que la minería de datos forma parte del proceso Descubrimiento de Conocimiento (*KDD* por sus siglas en inglés) considerado como el proceso de acceso a los datos, es decir, la exploración, preparación, métodos de implementación (con base en los algoritmos de DM), modelado y la supervisión de las actividades o métodos de la DM.

1.2 Planteamiento del problema

1.2.1 Definición del problema

La procrastinación o dilación de las actividades es un problema que ha existido desde hace mucho tiempo, pero hoy en día ha crecido de manera significativamente y un sector que afecta es a los estudiantes universitarios, ya que con el uso del Internet los estudiantes pueden dejar a un lado sus actividades escolares y eso puede causar un retraso en sus estudios o está la deserción de estos. Se han realizado estudios sobre la procrastinación, de cuáles son los motivos, factores o relaciones que conlleva a que los estudiantes aplacen o “dejen para mañana” sus actividades.

De igual forma, la procrastinación está relacionada con la autorregulación académica, la cual es el proceso o capacidad que tienen un estudiante para autorregularse en sus estudios y lograr el éxito académico, esta autorregulación, se logra por medio de objetivos, motivación, realización de actividades, entre otras.

El nivel universitario ha visto incrementada su población estudiantil como nunca, reflejo de los cambios demográficos, sociales, económicos y culturales. Ha surgido, así como nuevo reto la respuesta a la diversidad de alumnado que inevitablemente transforma la universidad, y así mismo, también contribuye a cambiar el reflejo de ésta en la sociedad. A su vez, la expansión del uso de las tecnologías de la comunicación y la información marca necesariamente un “antes” y un “después” en el cuestionamiento sobre las formas de enseñanza e, inevitablemente, las formas de aprendizaje (García, 2012).

Algunas razones que pueden originar la procrastinación según Solomon & Rothblum (1984) son: ansiedad ante la evaluación, dificultad para la toma de decisiones, falta de control, miedo a las consecuencias del éxito, percepción de cierto nivel de aversión o rechazo a la tarea y patrones excesivamente perfeccionistas, sobre la propia competencia.

Según García (2012) algunas características de un alumno autorregulado académicamente son: conscientes de la relación entre el proceso de autorregulación y el éxito académico, capaces de ajustar su conducta a las tareas o actividades de aprendizaje que se les proponen por medio de la retroalimentación, llevan a cabo su aprendizaje a través de una serie de estrategias de aprendizaje de todo tipo (cognitivas, metacognitivas, motivacionales y de apoyo) a las que conceden gran importancia, entre otras.

Por lo mencionado anteriormente los investigadores han estado realizando estudios sobre la procrastinación y autorregulación académica para identificar factores, comportamiento o conductas que llevan a que los alumnos procrastinen y a su vez como autorregulen sus estudios. Actualmente en México son pocos los estudios realizados sobre procrastinación y autorregulación académica.

Por lo tanto, en esta investigación se abordó el estudio de procrastinación y autorregulación académica en estudiantes universitarios, con el fin de utilizar el proceso de Descubrimiento de Conocimiento en Base de Datos (*KDD*, por sus siglas en inglés) que contempla el uso de técnicas de minería de datos, para analizar el nivel de relación

entre la procrastinación y autorregulación académica, así como la interacción que mantienen con Internet durante el desarrollo de sus actividades, con la finalidad de identificar patrones de conducta que conllevan a la procrastinación y autorregulación.

Las técnicas de minería de datos utilizadas son las reglas de asociación, árboles de decisión y agrupamiento, las cuales fueron implementadas en tres herramientas para minería de datos. Los datos de trabajo utilizados son los resultados del estudio realizado por García y Silva (2019) a 746 jóvenes universitarios que estudiaron asignaturas en línea en las 12 Divisiones Académicas que integran la Universidad Juárez Autónoma de Tabasco (UJAT).

1.2.2 Delimitación de la investigación

1.2.2.1 Alcances

- El *dataset* utilizado fue el proporcionado por García y Silva (2019) que contiene 746 registros de estudiantes que hubiesen cursado al menos una asignatura en línea en el ciclo escolar febrero-agosto 2019.
- Identificación de patrones de conducta de procrastinación y autorregulación académica en estudiantes universitarios que llevan materias a distancia.
- Implementación de técnicas de minería de datos en la identificación de patrones de conducta de procrastinación y autorregulación académica en estudiantes.
- Utilización de la metodología del proceso *KDD* en la aplicación de la minería de datos.
- Utilización de herramientas para minería de datos.

1.2.2.2 Limitaciones

- La población de estudio fueron los estudiantes de la UJAT, que hubiesen cursado al menos una asignatura en línea en el ciclo escolar febrero-agosto 2019.
- El abordaje teórico es desde la perspectiva de las variables de procrastinación y autorregulación académica.
- Las técnicas de minería de datos que se emplearon son: árboles de decisión, agrupamiento y reglas de asociación.
- Se emplearon las licencias de nivel académico de las herramientas para minería de datos.
- No se realizó ninguna acción para lograr el balanceo de las clases en el dataset.

1.3 Objetivos

1.3.1 Objetivo general

Identificar patrones de conducta, relacionados con las variables de procrastinación y autorregulación académica de estudiantes de licenciatura de la Universidad Juárez Autónoma de Tabasco (UJAT), 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 de procrastinación y autorregulación académica.
- Emplear la técnica de agrupamiento para identificar las similitudes entre los patrones de conducta de procrastinación y autorregulación académica del estudiante.

- Aplicar las reglas de asociación para identificar las relaciones entre procrastinación y autorregulación académica.
- Analizar los patrones de conducta obtenidos de procrastinación y autorregulación académica para la toma de decisiones futuras.

1.4 Justificación

La procrastinación ha sido estudiada desde hace algunos años por diversos autores, los cuales definen la procrastinación como el aplazamiento de actividades, pero a su vez dan diferentes significados los cuales, según Schouwenburg (2004) son retrasar, prolongar, aplazar o posponer la realización de una tarea. La procrastinación se da a menudo en el ámbito académico, en el cual los estudiantes universitarios entre 18 a 22 años, son los más propensos a procrastinar, pero de igual forma encuentran la forma de autorregularse académicamente con sus actividades.

En donde la autorregulación académica, comprende la capacidad o disponibilidad del alumno, para establecerse objetivos y metas, para iniciar y concluir adecuadamente cada una de sus actividades y procrastinar menos.

De acuerdo con Ferrari, O'Callahan, & Newbegin (2004), en América latina aproximadamente el 61% de las personas presentan procrastinación y el 20% presenta procrastinación crónica; esta última implica un patrón regular de comportamientos de aplazamiento frente a las tareas y actividades que son su responsabilidad (Quant y Sánchez, 2012).

Álvarez (2010) afirma que al menos el 25% de estudiantes presentan procrastinación académica y esta se relaciona con resultados académicos desfavorables, es decir, bajo rendimiento académico, aplazamiento de materias y abandono de cursos.

Actualmente el uso de internet en los estudiantes universitarios cobra mayor relevancia, en la utilización para realizar sus tareas y actividades en el ámbito académico y sus tiempos libres, según Lam-Figueroa et al. (2011) el uso de las nuevas tecnologías han impactado en los estilos de vida y desarrollo de las relaciones de socialización, en los últimos años. Al mismo tiempo, el internet se ha posicionado como una herramienta útil, global e inevitable, que hace que las personas la utilicen.

Griffith (1998, como se citó en Navarro-Mancilla y Rueda-Jaimes, 2007), identificó siete áreas específicas de dependencia al Internet: (a) tolerancia, (b) gastar más tiempo del planeado en Internet, (c) gastar la mayor parte del tiempo en actividades que permitan estar en línea, (d) gastar más tiempo en Internet que en actividades sociales o recreativas, (e) persistencia en el uso a pesar de dificultades laborales, académicas, económicas o familiares, (f) intentos fallidos para detener o disminuir el uso de Internet y (g) abstinencia.

En el caso de las investigaciones realizadas sobre procrastinación en idioma español el escenario está dominado especialmente por los investigadores peruanos (Alegre, 2013; Domínguez-Lara, 2017a; Domínguez-Lara, 2017b; Domínguez-Lara y Campos-Uscanga, 2017; Domínguez, Villegas y Centeno, 2014; Matalinares et al., 2017; Olazábal y Zegarra, 2017; Paz, Aranda, Navarro, Delgado y Sayas, 2014; Rodríguez, 2017; y

Yarlequé et al., 2016); siendo esta presencia también evidente en las tesis de las universidades peruanas como sería el caso de la Pontificia Universidad Católica del Perú (Vallejos, 2015), la Universidad Peruana Unión (Rafael y Ramírez, 2016), la Universidad César Vallejo (Morales, 2018), la Universidad Señor de Sipán (Vázquez, 2016), la Universidad Privada Antenor Orrego (Valdiviezo, 2017), y la Universidad San Pedro (Cerqueira, Castro, Tello y Beltrán, 2016), entre otras.

En el caso de México se reportan siete investigaciones realizadas por: Barraza y Barraza (2018); Chávez y Morales (2017); García, Olvera, Jiménez y Ramírez (2014); García et al. (2012); Miranda y Flores (2016); Morales y Chávez (2017) y Vázquez (2016)

De las siete investigaciones, realizadas en México, las primeras seis utilizan la Escala de Procrastinación Académica de Busko (1998) que fue traducida por Álvarez (2010) y adaptada por Domínguez et al., (2014) para utilizarla en investigaciones sobre procrastinación en la población peruana (Barraza y Barraza, 2018).

Actualmente en México son pocas las investigaciones realizadas sobre procrastinación y autorregulación académica, esto se identificó por medio del proceso de revisión de literatura, además de identificar que algunos estudios no validaban el instrumento de investigación. Así mismo, para el estudio en cuestión que se realizó, se revisaron investigaciones sobre uso de internet en estudiantes, ya que el instrumento de investigación fue aplicado a estudiantes que estudian asignaturas a distancia, en la UJAT.

El campo de estudio de esta investigación utilizando minería de datos todavía está comenzando ya que actualmente solo existe un estudio realizado por González, Calet,

Rubio y Dumitrache (2018), en el cual analizan el impacto de la procrastinación en el rendimiento académico de estudiantes universitarios usando técnicas de minería de datos, en este estudio emplean otras técnicas de minería de datos a las que se emplearán en esta investigación.

Con el análisis de datos utilizado por medio de técnicas de minería de datos se puede profundizar más en la identificar patrones de conducta y relaciones entre procrastinación y autorregulación académica que tal vez no se hayan podido identificar por medio de métodos estadísticos. Aunado a esto, la identificación de patrones de conducta de los estudiantes servirá para la toma de decisiones futuras en cuanto a proponer estrategias para contrarrestar la procrastinación y ayudar a la autorregulación académica de los estudiantes; y contribuir con información a esta área de estudio.

1.5 Metodología utilizada

La metodología utilizada para esta investigación consistió en tres etapas, las cuales se describen a continuación:

La primera etapa fue la realización del análisis de la información del *dataset*, que se obtuvo a partir de la investigación realizada por García y Silva (2019), con un total de 746 campos, con el propósito de identificar y visualizar datos en blanco o nulos, así como descartar ítems a no utilizar.

La segunda etapa de la metodología consistió en la búsqueda, evaluación y selección de herramientas para minería de datos. En la búsqueda se obtuvieron un total de 16 herramientas.

La tercera y última etapa fue la realización de la minería de datos, utilizando la metodología *KDD*, la cual consiste en nueve pasos, considerando etapas de aprendizaje antes y después de la minería de datos.

A continuación, se desglosan los aspectos relativos a la metodología utilizada, el enfoque y las fuentes de investigación, los instrumentos y técnicas de recolección de datos, así como el tipo de programación utilizada.

1.5.1 Enfoque de investigación

El enfoque de investigación que se utilizó en la presente tesis fue el cuantitativo basado en el diseño no experimental-transversal. Lo anterior en virtud que de acuerdo con Espinoza (2014) y Kerlinger (2002, como se citó en Ramírez y Zwerg-Villegas, 2012) la investigación cuantitativa busca la objetividad, bajo la perspectiva que la realidad es una y se puede observar sin afectarla, además hace énfasis que los métodos creados y utilizados controlan o eliminan los factores que debilitan la habilidad de descubrir la realidad.

1.5.2 Fuentes de investigación

Las fuentes de información que se emplearon en el desarrollo de esta investigación fueron fuentes primarias y secundarias. Entre las fuentes primarias empleadas se encuentran los artículos científicos, libros, ponencias, tesis de maestría y doctorado, todos ellos de alcance internacional y nacional.

De igual forma las fuentes secundarias utilizadas fueron: análisis e interpretación de información presentada en artículos, resúmenes, sitios o páginas electrónicas,

relacionadas con procrastinación, autorregulación académica y técnicas de minería de datos, las cuales son de investigaciones y libros de alcance internacional y nacional.

1.5.3 Técnicas de recolección de datos

La técnica de recolección de datos consistió en la utilización de un *dataset* para el análisis y aplicación de las técnicas de minería de datos en las respectivas herramientas a seleccionar utilizadas para minería de datos.

Este *dataset* se obtuvo de los resultados derivados de la investigación de García y Silva (2019) titulada “Procrastinación académica entre estudiantes de cursos en línea. Validación de un cuestionario”, en la cual se diseñó, aplicó y validó un cuestionario constituido por 22 *ítems*, el cual es mismo número de variables en el *dataset*.

1.5.4 Modelo de desarrollo

Análisis de la información del dataset

Durante la primera etapa del modelo, se verificó y analizó la información del *dataset* con 746 registros, con el objetivo de conocer cómo están conformados y distribuidos los campos y registros, es decir, la visualización del número de campos o columnas y verificar los nombres de éstas. De igual forma identificar que dentro de los campos no haya datos nulos o vacíos y rectificar que todos los registros estén completos, esto para descartar la pérdida de datos.

Selección de las herramientas para aplicar minería de datos

La metodología de calidad de software FURPS, se empleó para analizar las características y calidad del software de cada herramienta previamente buscada para

minería de datos, con el objetivo de hacer una ponderación de acuerdo con los factores y criterios establecidos por la metodología FURPS y visualizar cuál(es) es(son) la(s) herramienta(s) más adaptable(s) para ser utilizada(s).

Esta metodología establece cinco factores de calidad que deben cumplir las herramientas o softwares los cuales son: Funcionalidad, Usabilidad, Confiabilidad, Rendimiento y Capacidad de Soporte; y proporciona criterios a evaluar como características del programa, tiempo de respuesta, eficiencia, ente otros.

Metodología del proceso KDD

En la última etapa, se utilizó la metodología iterativa e interactiva *KDD*, para aplicar las técnicas de árbol de decisión, agrupamiento y reglas de asociación al *dataset*, siendo empleadas en la(s) herramienta(s) selecciona(s) a utilizar; con el propósito de identificar y descubrir conocimiento y patrones de conducta en relación con procrastinación y autorregulación académica.

A partir de los pasos que contempla la metodología se realizó el aprendizaje y dominio de la herramienta a utilizar, se procedió a la creación, visualización y limpieza de los datos, además de describir el motivo de la elección de las técnicas y algoritmos de minería de datos a utilizadas, posteriormente la realización de minería de datos al *dataset* y por último se realizó la interpretación y presentación de los datos con ayuda de gráficos y tablas.

1.5.5. Tipo de programación

El tipo de programación utilizada principalmente fueron programas o herramientas que se utilizan para minería de datos, dentro de estas herramientas se encuentran herramientas *framework*, lenguajes de programación y servicios en la nube. De igual forma se utilizó Microsoft Excel, ya que los datos se encontraban almacenados en este programa.

Capítulo 2. Marco teórico

Los temas o conceptos principales que fundamentan la investigación son: procrastinación, autorregulación académica y minería de datos. El término procrastinación lo definen como una insuficiencia en los procesos de autorregulación, que provoca la demora voluntaria de actividades planificadas, aunque se anticipe una situación peor como consecuencia de la demora (Steel & Ferrari, 2012).

Según Alvarado, Vega, Cepeda y Del Bosque (2014) la autorregulación académica implica la modulación del pensamiento, motivación, atención y conducta mediante la utilización de mecanismos específicos y estrategias de apoyo; en cambio la minería de datos, según Witten, Frank & Hall (2011) es el proceso de descubrir patrones en los datos, este proceso es automático o semiautomático, además la minería de datos trata de resolver problemas mediante el análisis de datos ya presentes en las bases de datos.

2.1 Marco referencial

Con la revisión literaria realizada, se encontraron investigaciones internacionales y nacional relacionadas con la procrastinación, la autorregulación académica y la minería de datos, de igual forma al ser la población de estudio estudiantes con asignaturas a distancia de esta investigación, se hace referencia al uso de internet en los estudiantes.

Las investigaciones se encuentran organizadas de acuerdo a las variables de estudios, es decir, las primeras siete investigaciones corresponden a estudios sobre procrastinación (de los subtítulos 2.1.1 al 2.1.7), las siguientes cinco investigaciones corresponden a estudios sobre autorregulación académica (los subtítulos 2.1.8 al 2.1.12),

las cuatro investigaciones posteriores corresponden a estudios sobre minería de datos (del 2.1.13 al 2.1.16) y las últimas tres investigaciones corresponden a estudios sobre el uso de internet entre los estudiantes (subtítulos 2.1.17 al 2.1.19); las cuales están ordenadas a partir de investigaciones internacionales a investigaciones nacionales, para tener una mejor apreciación de las investigación que se han realizado a los largo de los años con respecto a estas variables de estudio.

2.1.1 Análisis del impacto de la procrastinación en el rendimiento académico de estudiantes universitarios usando técnicas de minería de datos

González et al. (2018) realizaron esta investigación con una muestra de estudio de 145 estudiantes, entre 18 y 46 años, 29% hombres y 71% mujeres, los cuales cursaban titulaciones de grado en educación primaria, educación infantil y pedagogía, en las Universidades de Granada o Burgos. El instrumento utilizado fue la escala de Evaluación de la Procrastinación para estudiantes con 44 *ítems*; además evaluaron el rendimiento académico y características sociodemográficas sobre la edad, género, curso y estudios que cursan, partiendo los datos en: entrenamiento 80% y test 20%.

Los algoritmos aplicados a la muestra fueron la regresión lineal, *gaussian process*, *random forest* y *multilayer perceptron*; obteniendo como resultado, que la variable curso y tres ítems del cuestionario fueron las variables que demostraron ser más informativas para la predicción clasificatoria de la clase, y a su vez la variable procrastinación podría ser una variable relevante para el rendimiento académico y además los autores sugieren

que la aplicación de minería de datos, en el ámbito de investigaciones educativas puede ser novedosa y de utilidad para mejorar el entendimiento sobre los procesos de enseñanza-aprendizaje y su efecto en el rendimiento académico.

2.1.2 El papel de la procrastinación académica como factor de la deserción universitaria

Garzón y Gil (2017a) realizaron una revisión teórica de las tendencias de estudios de procrastinación académica y comprensión de la deserción universitaria, analizando las variables de autorregulación académica, rendimiento académico, manejo del tiempo, razones para procrastinar e instrumentos para medir la procrastinación.

Los autores resumen algunos resultados sobre las dificultades de autorregulación con la procrastinación (ver tabla 1); estos resultados se obtuvieron según las investigaciones y poblaciones de estudio; los autores sugieren que la prevalencia de procrastinación académica entre los estudiantes es elevada.

Tabla 1
Dificultades de autorregulación con la procrastinación.

Fallo autorregulatorio de los procrastinadores	Autor(es)
Quienes presentan mayor confianza en sus habilidades académicas (niveles elevados de autoeficacia) tienden a procrastinar menos. La autoeficacia está relacionada con el uso de herramientas de planificación e inicio de las tareas de manera oportuna.	Wolters (2003) Klassen & Kuzucu (2009)
Las tareas percibidas como difíciles, poco atractivas, ambiguas y que requieren más esfuerzo, presentan mayor probabilidad de ser objeto de procrastinación.	Ferrari, Mason & Hammer (2006)
La procrastinación parece relacionarse con la impulsividad (la tendencia a ser presa de los deseos inmediatos); el procrastinado tiende a elegir los beneficios a corto plazo sobre las ganancias a largo plazo.	Wilson & Nguyen (2012)

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Fallo autorregulatorio de los procrastinadores	Autor(es)
Los procrastinadores frecuentes subestiman los esfuerzos destinados a las tareas actuales sobre el éxito final en la obtención de una meta. Tienden a reportar menos uso de estrategias de manejo del tiempo que los no procrastinadores.	Lay & Schouwenburg, (1993)

Nota: Garzón y Gil (2017a).

2.1.3 Gestión del tiempo y procrastinación en la educación superior

Esta investigación la realizaron Garzón y Gil (2017b) con una muestra de 494 estudiantes, 367 hombres y 127 mujeres, con edades entre 16 y 46 años, el 55.6% dedicados a estudiar y el 44.4% estudian y trabajan, perteneciendo al primer año, de la Universidad El Bosque y Uniempresarial de la ciudad de Bogotá, utilizando el instrumento de Escala de Evaluación de Procrastinación-Estudiante (PASS) y cuatro variables de Time Management Behavior Questionnaire (TMBQ).

Los resultados obtenidos no muestran correlación entre la edad y procrastinación, pero con el género, los autores identificaron una mayor procrastinación en hombres que en mujeres, además identificaron que los estudiantes que combinan desempeño laboral con estudios corren el riesgo de procrastinar académicamente; los autores concluyen que los aspectos relevantes para prevenir la procrastinación deben de estar relacionadas con las habilidades de autorregulación y la percepción del estudiante acerca del control de su tiempo.

2.1.4 Procrastinación, adicción al internet y rendimiento académico de estudiantes universitarios ecuatorianos

Ramos-Galarza et al. (2017) realizaron esta investigación con una muestra de 250 estudiantes universitarios, la muestra estuvo compuesta por 135 hombres y 115 mujeres,

con edades entre 17 y 31 años; los estudiantes eran de carreras de Psicología, Ingeniería Industrial, Diseño Gráfico, Arquitectura y Administración de Empresas, de dos universidades de Quito, Ecuador.

Los autores identificaron una correlación directa proporcionalmente entre procrastinación y adicción a internet, lo cual sugiere que los estudiantes con mayores niveles de procrastinación, pueden estar conectados por largas horas a sitios web y pueden utilizar ese tiempo en revisar sus bandejas de entrada de sus aplicaciones para estar en comunicación; además los autores concluyen en la invitación a los docentes de educación superior, a considerar las dos variables estudiadas como factores que pueden influir en el rendimiento académico de los estudiantes.

2.1.5 Procrastinación académica como predictor en el rendimiento académico en jóvenes de educación superior

Chan (2011) efectuó esta investigación con una muestra de 200 estudiantes universitarios del área de letras en Lima, entre edades de 19 a 23 años; el instrumento aplicado fue una encuesta de características de los jóvenes y la Escala de Procrastinación Académica Busko (1998) y la escala adaptada de Álvarez (2010). Los primeros hallazgos de la investigación fueron que los hombres tienen a procrastinar mayormente que las mujeres, pero el autor considera que las mujeres estarían en camino a la procrastinación.

El autor identificó que el comportamiento de procrastinación se presenta en estudiantes con notas altas como en estudiantes con bajo rendimiento académico,

concluyendo que los estudiantes participantes del estudio procrastinan en actividades como: dejar las tareas para el último minuto, no estudiar antes para las evaluaciones, no cumplir con entregar sus trabajos, postergar trabajos o lecturas que no les agradan, incumplir con entregar los trabajos en fechas específicas, desmotivaciones y falta de hábitos de estudio.

2.1.6 Evidencias de validez y confiabilidad de la escala de procrastinación académica en una población estudiantil

Barraza y Barraza (2018) efectuaron esta investigación con una muestra de 361 alumnos, 46.6% mujeres y 53.4% hombres de educación media superior del Centro de Bachillerato Tecnológico e Industrial y Servicios de la ciudad de Durango, México; aplicando el instrumento de la Escala de Procrastinación Académica de Busko (1998), que consta de 16 *ítems*, por la validación realizada en otra investigación la escala consta de 12 *ítems*.

Los autores identificaron que la escala presentaba problemas con dos *ítems*, al momento de realizar un análisis factorial exploratorio y confirmatorio, decidieron eliminar esos *ítems* y proponer una nueva versión de la escala con 10 *ítems*. Los autores sugieren que los resultados no son concluyentes, en el entendido que la principal limitación del estudio fue la aplicación a una muestra de estudiantes de una sola institución y un solo nivel educativo, por lo que recomiendan mayores estudios, además proponen realizar comparaciones entre las versiones de 10, 12 y 16 *ítems* de la escala.

2.1.7. Procrastinación académica entre estudiantes de cursos en línea.

Validación de un cuestionario

García y Silva (2019) realizaron esta investigación con una muestra de 746 estudiantes de una universidad pública de Tabasco, México, los cuales cursaron asignaturas en línea como alternativa para avance curricular, durante el periodo febrero-agosto 2019, la muestra la conformaron 275 hombre y 471 mujeres, con edades entre 17 y 46 años; aplicando la Escala de Procrastinación Académica (EPA), la cual está constituida por 16 *ítems* y la agregación de seis *ítems* relacionados con la interacción con internet mientras desarrollan sus actividades académicas.

A los resultados que obtuvieron se les realizó un análisis descriptivo y determinaron la normalidad de los datos y posterior a diversos análisis estadísticos, la escala quedó constituida por siete *ítems* de la escala original, seis de autorregulación y uno de procrastinación. Las autoras concluyen que los hallazgos obtenidos, están en consonancia con los que se encuentran en evidencias en otros estudios en el contexto universitario en términos generales, aun cuando hay una diferencia entre dos factores, la estructura final del cuestionario se ajusta a dos factores: el de procrastinación e internet y el de autorregulación académica.

Las autoras identifican que el internet representa un medio extraordinario para la formación en línea, sin embargo, también puede convertirse en un elemento de procrastinación y expresan la necesidad de realizar más estudios con muestras diferentes, distintos perfiles, enfoques y considerar otras variables.

2.1.8. La autorregulación académica como varianza explicativa de los procesos de aprendizaje universitarios

García (2012) realizó una revisión teórica sobre la autorregulación académica para identificar los procesos de autorregulación y el entendimiento del desarrollo de aprendizaje a lo largo de la vida.

El autor consideró la visión de los procesos de autorregulación deriva de distintos modelos explicativos y resulta importante la consideración de la autorregulación académica como un conjunto multidimensional de procesos cognitivos-metacognitivos, emocionales-motivacionales, conductuales y ambientales, es decir, para aprender se necesita saber cómo, además de sentir la necesidad de aprender, tener el deseo y los motivos.

2.1.9 Teorías de autorregulación educativa: una comparación y reflexión teórica

Panadero y Alonso-Tapia (2014) realizaron una revisión sobre las aportaciones de las principales teorías de autorregulación a partir de cinco procesos para entender cómo los alumnos adquieren competencias de autorregulación; los cinco procesos para examinar las siete teorías autorregulatorias son: 1) origen de la motivación, 2) toma de conciencia, 3) procesos, 4) influencia del entorno social y físico y 5) capacidad de autorregulación.

En la tabla 2, se observan los resultados que los autores obtuvieron sobre las teorías de autorregulación, conforme a las fortalezas y críticas de las teorías; además los autores concluyen que es importante tener una visión comprensiva de las diferentes teorías, para

facilitar la comprensión de los procesos que los alumnos activan en el aula y entender las estrategias que los estudiantes emplean, además que el profesor tiene que considerar las necesidades de los estudiantes que trabajan y desde una visión holística de lo que supone la autorregulación promover que los estudiantes adquieran y activen las estrategias necesarias para las tareas requeridas (ver tabla 2).

Tabla 2
Comparación de teorías de autorregulación.

Teoría	Fortalezas	Criticas
Operante Fenomenológica	Retraso de las graficaciones Papel de las autoidentidades	Origen del auto-refuerzo Definición, evaluación y validación de las autoidentidades
Procesamiento de la información Sociocognitiva	Monitorización a través de ciclos de retroalimentación Metas cognitivas y expectativas Modelado social	Ciclos de retroalimentación negativa vs positivos Auto-eficacia: redundante o de alcance limitado
Volicional	Persistencia y atención	Delimitación de la volición y la motivación.
Vygotskiana	Auto-verbalización y el desarrollo del diálogo social	Auto-verbalizaciones vs. Co-constructivismo social como métodos de enseñanza de la autorregulación.
Constructivista	Teorías personales y estrategias	Papel de la disonancia cognitiva vs. Contexto situado

Nota: Panadero y Alonso-Tapia (2014).

2.1.10 Autorregulación y rendimiento académico en la transición secundaria-universidad

Elvira-Valdés y Pujol (2012) efectuaron esta investigación con una muestra de 172 estudiantes, 95 mujeres y 77 hombres, con edades entre los 16 y 30 años, los cuales cursaban el primer trimestre del curso de iniciación Universitaria de la Universidad Simón

Bolívar en Litoral; el instrumento aplicado fue la Escala de Aprendizaje Autorregulado PCR (Planeación, Control y Reflexión) y una encuesta sobre datos sociodemográficos.

Los resultados que obtuvieron los autores indicaron que los estudiantes presentan un moderado uso de estrategias de autorregulación académica en general, además los resultados sobre uso de estrategias de planificación y control fue moderada y la estrategia que utilizan mayormente es la reflexión. Los autores concluyen que los resultados obtenidos son iniciales y no como indicadores concluyentes, y que podría resultar interesante llevar a cabo investigaciones sobre el uso de estrategias autorregulatorias en estudiantes de bachillerato y estudiantes universitarios a lo largo de sus carreras universitarias contemplando otras variables.

2.1.11 Evaluación de autorregulación académica en estudiantes de psicología en modalidad en línea

Meza, de la Rosa, Rivera y González (2018) utilizaron una muestra de 204 estudiantes, 145 mujeres y 59 hombres, entre 18 y 49 años, de 21 estados de la República Mexicana, de primer semestre de la Licenciatura en Psicología del Sistema de Universidad Abierta y Educación a distancia (SUAYED) del ciclo escolar 2016-2. El instrumento utilizado fue el Inventario de Estilos de Aprendizaje y Orientación Motivacional (EDAOM), la aplicación del instrumento fue a distancia a través de una encuesta en línea.

Los resultados encontrados con el estudio muestran que los participantes cuentan con niveles moderados de autorregulación, además identificaron que los estudiantes del

sistema en línea pertenecen a un tipo de estudiante distinto al de sistemas presenciales, porque el rango de edad es más amplio, que el promedio de edad de los estudiantes de nuevo ingreso de licenciatura presencial.

Los autores concluyeron que la autorregulación en el ámbito académico es una variable individual que permite indagar acerca de cómo los estudiantes planifican, monitorean y evalúan su propio aprendizaje en contextos específicos y es importante continuar indagando cómo es la autorregulación de los estudiantes en línea, porque son distintas las demandas autorregulatorias de un estudiante en línea a un estudiante presencial.

2.1.12 Comparación de estrategias de estudio y autorregulación en universitarios

Alvarado *et al.* (2014) realizaron esta investigación con una muestra de 41 estudiantes de la Licenciatura de Psicología de la Facultad de Estudios Superiores Iztacala de la Universidad Nacional Autónoma de México. La muestra se conformó por 23 estudiantes no regulares, que al menos habían reprobado una materia de las asignaturas de Métodos Cuantitativos o Psicología Experimental Humana III o IV y 18 estudiantes regulares con promedio superior a 8; el instrumento utilizado fue el de Inventario de Estrategias de Estudio y Autorregulación (IEEA).

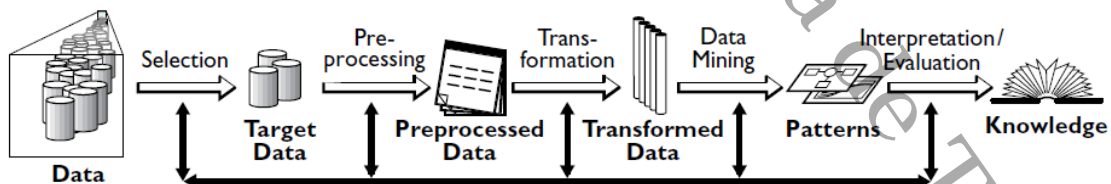
Los autores identificaron que las dos muestras de estudiantes requieren apoyo para desarrollar estrategias de estudio, también los autores hacen mención que los estudiantes regulares contaban con mejores capacidades para planificar, evaluar,

controlar y supervisar su proceso de estudio; los autores concluyen que es necesario que las universidades diseñen e implanten programas para acrecentar la motivación de los estudiantes y la autorregulación durante el proceso de aprendizaje e incrementar la calidad, además de realizar estudios de las características de los alumnos para construir soluciones congruentes con necesidades concretas del estudiante.

2.1.13 The *KDD* Process for extracting useful knowledge from volumes of data

Fayyad, Piatetsky-Shapiro & Smyth (1996) realizaron esta investigación sobre el proceso del *KDD* para la extracción de conocimiento útil a partir de volúmenes de datos; en el cual mencionan el descubrimiento de conocimiento en el proceso de extraer el conocimiento útil. En la figura 1, se describe el proceso de *KDD*, además hacen mención que el proceso *KDD* es un proceso de tipo interactivo e iterativo, el cual implica numerosos pasos (ver figura 1).

Figura 1
Pasos que constituyen el proceso KDD.



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

Los autores mencionan que *KDD* ha evolucionado y sigue evolucionando, con la intersección de investigaciones en campos como base de datos, aprendizaje automático, reconocimiento de patrones, estadística, inteligencia artificial y razonamiento de

incertidumbre, adquisición de conocimiento de sistemas expertos, visualización de datos y descubrimiento automático, descubrimiento científico, recuperación de información y computación de rendimiento; además los sistemas de software *KDD* incorporar teorías, algoritmos y métodos, y un método para realizar análisis en almacenes de datos es el procesamiento analítico en línea (*OLAP*, por sus siglas en inglés).

2.1.14 Análisis comparativo usando minería de datos en la predicción del rendimiento académico de adolescentes basados en emociones

Gamarra y Farfán (2019) utilizaron datos recopilados de una institución de la ciudad de Arequipa, por medio de una encuesta, en la que consideran variables como: edad, género, depresión, estado de ánimo (feliz, triste), grado que cursan, grado académico. La herramienta que utilizaron fue *Weka*, las técnicas de minería de datos empleadas en la investigación fueron árbol de decisión con los algoritmos *J48* y redes bayesianas.

Los resultados obtenidos con el algoritmo *J48*, centrado en las emociones, consideraron un estado de ánimo feliz como momentáneo, además el 85% de los estudiantes sufren emociones negativas, relacionadas con el *bullying* y estar enamorados, generando en ellos una depresión; las redes bayesianas con relación al grado de estudio, los estudiantes presentan mayor interacción en sus emociones y mientras transcurren los años de estudio disminuye la sensación de depresión, haciendo que mejoren su calidad y estabilidad emocional y su promedio se incrementa.

Los autores concluyen que el árbol de decisión ayuda a predecir que los estudiantes con emociones negativas tienden a tener un rendimiento académico menor al resto de

los estudiantes y los resultados del estudio permiten a las autoridades de los centros educativos, tomar medidas para contrarrestarlas o anteponerse a problemas académicos y sociales.

2.1.15 Análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos

Miranda y Guzmán (2017) realizaron esta investigación considerando la deserción y un periodo para retomar sus estudios superiores, equivalente a tres años, como referencia del tiempo que el Sistema de Información de la Educación Superior (SIES) del Ministerio de Educación de Chile utiliza estándares para su investigación, el cual fue aplicado a estudiantes de Universidad Católica del Norte de Chile (UCN) en Antofagasta y Coquimbo, Chile; en el periodo 2000 al 2013, con una muestra de 9,195 estudiantes. Las herramientas que utilizaron fue *SQL Server*, *SPSS Statistics* y *Weka*, las técnicas de minería de datos empleadas en la investigación fueron redes neuronales y árbol de decisión con *SPSS Statistics* y redes bayesianas con *Weka*; la metodología de análisis empleada fue el proceso *KDD*.

Los resultados obtenidos respecto al abandono de estudios de 33.9% (red bayesiana) y 21.7% (árbol de decisión), mientras el no abandono de estudios fue 66.1% (red bayesiana) y 78.3% (árbol de decisión); la clasificación de casos de abandono y no abandono con las redes neuronales fue correcta con 83% de alta exactitud de casos positivos con un 73% de precisión y 88% para ratio de casos negativos; los autores

concluyen que los resultados académicos como la situación económica influye en la decisión de permanencia de los estudiantes en sus estudios.

2.1.16 Análisis de deserción-permanencia de estudiantes universitarios utilizando técnicas de clasificación en minería de datos

Eckert y Suénaga (2015) efectuaron esta investigación con una muestra de 855 estudiantes de Ingeniería en Informática, modalidad presencial a cinco años de duración y tesis de grado, el periodo seleccionado fue del año 2000 al 2009. La herramienta que utilizaron fue *Weka*, las técnicas de minería de datos empleadas en la investigación fueron, los árboles de decisión implementado su algoritmo clasificador C4.5 (J48), el algoritmo Clasificador *Naive Bayes* Aumentando a Árbol (TAN) y por último el Clasificador de Reglas *OneR* y como metodología de análisis empleada fue el proceso *KDD*.

Los autores identificaron que durante el primer año de la carrera es donde adquieren mayor importancia las acciones de contención, apoyo tutorías y todas las actividades que mejoren la situación académica del alumno, además detectaron relaciones entre factores como deserción y permanencia; los autores concluyen que las herramientas de minería de datos brindan resultados que deben ser interpretados y traducidos a diagnósticos y consecuencias del ámbito real, y las etapas de preparación y acondicionamiento de los datos en el proceso *KDD*, son fundamentales porque los resultados dependen de ellas.

2.1.17. Adicción a las nuevas tecnologías y a las redes sociales en jóvenes:

un nuevo reto

Echeburúa y de Corral (2010) realizaron una revisión teórica sobre el problema del uso inadecuado de tecnologías con respecto a los jóvenes. Los autores mencionan un estudio realizado por la Fundación *Pfizer* en el año 2009, donde “el 98% de los jóvenes españoles de 11 a 20 años son usuarios de Internet, del cual siete de cada 10 afirman acceder un tiempo de al menos 1.5 horas, pero solo el 3% o 6%” hace uso abusivo de internet”.

Los autores mencionan que las TIC simplifican considerablemente los quehaceres cotidianos. El atractivo de Internet para los jóvenes es la respuesta rápida, las recompensas inmediatas, la interactividad y las múltiples ventanas con diferentes actividades. El uso es positivo, siempre que no se dejen de lado el resto de las actividades propias de una vida normal (estudiar, hacer deporte, ir al cine, salir con los amigos o relacionarse con la familia). Otra cosa es cuando el abuso de la tecnología provoca aislamiento, induce ansiedad, afecta a la autoestima y le hace perder al sujeto su capacidad de control.

2.1.18 Gestión del tiempo y uso de las TIC en estudiantes universitarios

Garrote, Jiménez y Serna (2018) realizaron esta investigación con una muestra de 108 estudiantes de grado de Maestro en Educación Infantil y Primaria, utilizando el cuestionario de Uso Problemático de Nuevas Tecnologías (UPNT), el cual se compone por 41 *ítems*, los dos primeros se refieren a la frecuencia y problema de uso de nuevas

tecnologías y los otros comprenden las subescalas de Internet, Videojuegos, Teléfono móvil y Televisión.

Los autores identificaron una frecuencia en el uso diarios de Internet y el teléfono móvil es 99.1%, el 50.9% presenta algún problema por dedicación excesiva de tiempo; en relación con características de uso de internet, el 90.7% consideran que dedican más tiempo del que es necesario; los autores concluyen que es conveniente ampliar el estudio sobre el uso las nuevas tecnologías y analizar cómo educar a la población en un uso responsable y correcta gestión de las herramientas que existen.

2.1.19 Adicción a internet: desarrollo y validación de un instrumento en escolares adolescentes de Lima, Perú

Lam-Figueroa *et al.* (2011) efectuaron esta investigación con una muestra de 248 estudiantes de primero a quinto año de secundaria, para evaluar variables de edad, sexo, año de estudio, jornada de estudios, modelo de acceso a internet, tiempo promedio por semana, antecedentes de repetir un año escolar, problemas de disciplina, falta injustificada, plan de futuro, motivos de uso de internet; además realizaron la Escala de Adicción a Internet de Lima (EAIL).

Los autores identificaron que los estudiantes que refieren un futuro incierto, así como antecedentes de faltas injustificadas a la escuela y problemas de disciplina, presentaron mayores índices de problema de adicción a internet, y concluyen que el campo de la adicción a internet, aún resta mucho que investigar y consideran que el aporte de su investigación es significativo en cuanto a desarrollar nuevos estudios de carácter local y

nacional, utilizando el instrumento EAIL, para comprobar su validez y contemplando el conocimiento en torno a la adicción a internet.

2.2 Marco conceptual

2.2.1 Procrastinación

Existen varias concepciones creadas sobre procrastinación por diferentes autores, a continuación, se exponen algunas:

- Según Díaz-Morales (2019) es la tendencia a aplazar el inicio y/o finalización de tareas planificadas para ser realizadas en un tiempo determinado; esto acompañado de malestar subjetivo y no sólo es baja responsabilidad y gestión del tiempo, sino supone un problema de autorregulación.
- Según Quant y Sánchez (2012) es un patrón de comportamiento caracterizado por aplazar voluntariamente la realización de actividades que deben ser entregadas en un momento establecido.
- De acuerdo con Rodríguez y Clariana (2017) es una demora innecesaria e injustificada de las tareas relacionadas con los estudios.
- Para Chan (2011) implica un grado de afectación en la toma de decisiones y en la resolución de conflictos, en el momento de decidir si culminar los deberes académicos o satisfacer los requerimientos que su ambiente.

2.2.2 Autorregulación académica

Algunas concepciones sobre autorregulación académica son:

- Según Elvira-Valdés y Pujol (2014) son las formas de aprendizaje independientes y efectivas y busca explicar como la persona aumenta sus resultados académicos usando métodos de forma sistemática.
- Según Pintrich (2000) es un proceso activo y constructivo mediante el cual los aprendices fijan metas para su aprendizaje e intentan monitorear, regular y controlar su cognición, motivación y conducta, guiados y restringidos por sus metas y las características contextuales de su ambiente.
- De acuerdo con Moreno, Figueroa y Arrieta (2015) es el proceso por el cual los alumnos toman conciencia del propio conocimiento acerca de la realización de una tarea, a vez permite controlar lo que se hace, se dice o se piensa y al autorregularse los estudiantes se sienten motivados hacia la tarea de aprendizaje.

2.2.3 Proceso KDD

KDD se centra en el proceso general de descubrimiento de conocimiento a partir de datos, incluyendo cómo se almacenan los datos y se accede, cómo algoritmos pueden ser escalados a grandes conjuntos de datos y todavía funcionan de manera eficiente, cómo los resultados pueden ser interpretados y visualizados, y cómo la interacción global humano-máquina puede ser modelada y soportada (Fayyad et al., 1996).

Martínez (2012) la describe como una metodología genérica para encontrar información en un gran conjunto de datos y con ello generar conocimiento y se define como un proceso no trivial de extracción de información a partir de los datos, la cual se

encuentra presente de forma implícita, previamente desconocida y potencialmente útil para el usuario o el negocio.

Según Fayyad et al. (1996) el *KDD* hace un énfasis especial en la búsqueda de patrones comprensibles que pueden ser interpretadas como conocimiento útil o interesante. La mayoría de los algoritmos de minería de datos se pueden ver como composiciones de algunas técnicas básicas y principios, las técnicas más comunes en la práctica de minería de datos son:

- **Clasificación:** clasifica un elemento de datos en una de varias clases predefinidas categóricas.
- **Regresión:** los mapas de un elemento de datos a una variable de predicción de valor real.
- **Agrupación:** asigna un elemento de datos en una de varias clases categóricas o conglomerados, en el que las clases deben determinar a partir de la clasificación de datos.
- **Sumarización:** proporciona una descripción compacta para un subconjunto de datos.
- **Modelado de dependencia:** describe las dependencias significativas entre las variables.
- **Análisis de enlaces:** determina las relaciones entre los campos de las bases de datos.

- **Análisis de secuencia:** modelar los estados de proceso de generación de la secuencia o para extraer y reportar desviación y tendencias en el tiempo.

2.2.4 Minería de datos

La minería de datos es un paso en el proceso de *KDD* que consiste en una enumeración de patrones (o modelos) sobre los datos, sujeto a algunas limitaciones de cálculo de eficiencia aceptables (Fayyad et al., 1996).

Según Medina y Gómez (2014) la minería de datos emerge de las áreas de base de datos (*data base*), repositorio de datos (*Data Warehouse*), la estadística, el aprendizaje automático, la visualización de datos, la búsqueda y recuperación de la información y de la computación de alta ejecución, para elaborar procesos esenciales donde se aplican una serie de métodos inteligentes para poder extraer y descubrir patrones de los datos, mediante el uso de algoritmos como regresión lineal, regresión logística, de asociación, lógica difusa, árboles de decisión y redes neuronales entre otros, que permiten obtener un conocimiento histórico y prospectivo para una toma de decisiones en el área de estudio o de conocimiento (Gorbea-Portal, 2013).

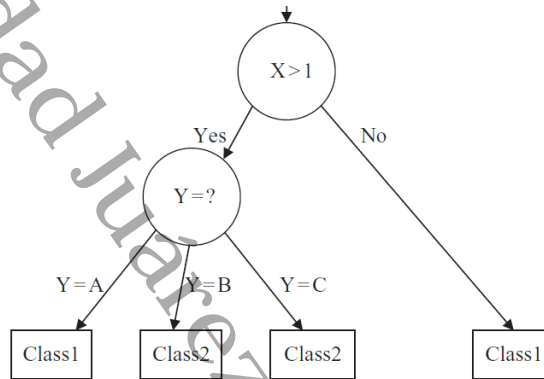
2.2.5 Técnicas de minería de datos

2.2.5.1 Árbol de decisión

Un árbol de decisión es un modelo jerárquico para el aprendizaje supervisado donde la región local se identifica en una secuencia de divisiones recursivas a través de nodos de decisión con función de prueba (ver figura 2). Un árbol de decisión también es un

modelo no paramétrico en el sentido de que no se asume ninguna forma paramétrica para la densidad de clase (Kantardzic, 2011).

Figura 2
Árbol de decisión simple con las pruebas en los atributos X e Y.



Nota: Kantardzic (2011).

2.2.5.2 Agrupamiento

Según Kantardzic (2011) el análisis de conglomerados es el estudio formal de métodos y algoritmos para la agrupación natural o agrupamiento de objetos de acuerdo con características o similitudes intrínsecas medidas o percibidas. Las muestras para la agrupación se representan como un vector de medidas, o más formalmente, como un punto en un espacio multidimensional.

Las muestras dentro de un grupo válido son más similares entre sí que a una muestra que pertenece a un grupo diferente. La metodología de agrupamiento es particularmente apropiada para la exploración de interrelaciones entre muestras para hacer una evaluación preliminar de la estructura de la muestra (Kantardzic, 2011).

2.2.5.3 Reglas de asociación

Según Witten et al. (2011) las reglas de asociación son como las reglas de clasificación, puede encontrar reglas de la misma manera, ejecutando un procedimiento de inducción de reglas de divide y vencerás para cada expresión posible que pueda ocurrir en el lado derecho de la regla. Sin embargo, no solo puede ocurrir un atributo en el lado derecho con cualquier valor posible, sino que una sola regla de asociación a menudo predice el valor de más de un atributo.

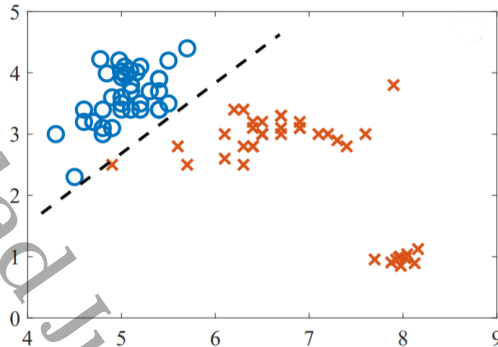
2.2.6 Métodos de aprendizaje en las técnicas de minería de datos

2.2.6.1 Aprendizaje supervisado

El aprendizaje supervisado o aprendizaje con un maestro, se utiliza para estimar una dependencia desconocida a partir de muestras de entrada y salida conocidas, este aprendizaje supone la existencia de una función de aptitud del maestro o algún otro método externo para estimar el modelo propuesto (Kantardzic, 2011).

Según Simeone (2018) el objetivo de este aprendizaje es aprender un mapeo entre los espacios de entrada y salida; en la figura 3, las entradas son puntos en el plano, las salidas con las etiquetas asignadas a cada entrada (círculo o cruces) y el objetivo es aprender un clasificador binario (ver figura 3).

Figura 3
Ejemplo de aprendizaje supervisado.



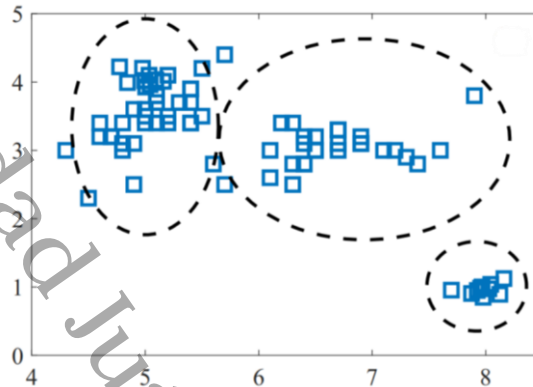
Nota: Simeone (2018).

2.2.6.2 Aprendizaje no supervisado

El aprendizaje no supervisado o aprendizaje sin un maestro, consta solo con valores de entrada a un sistema de aprendizaje y no existe una noción de la salida durante el proceso de aprendizaje; el objetivo del aprendizaje no supervisado es descubrir una estructura “natural” en los datos de entrada (Kantardzic, 2011).

En la figura 4, se muestra un ejemplo del aprendizaje no supervisado, en el cual según Simeone (2018) las entradas son puntos en el plano, pero los datos no proporcionan ninguna indicación sobre la salida deseada y el objetivo es agrupar los puntos de entrada que están cerca uno del otro (ver figura 4).

Figura 4
Ejemplo de aprendizaje no supervisado.



Nota: Simeone (2018).

2.2.7 Uso de internet

El atractivo de Internet para los jóvenes es que se caracteriza por la respuesta rápida, las recompensas inmediatas, la interactividad y las múltiples ventanas con diferentes actividades. El uso es positivo, siempre que no se dejen de lado el resto de las actividades propias de una vida normal (estudiar, hacer deporte, ir al cine, salir con los amigos o relacionarse con la familia); el abuso de la tecnología provoca aislamiento, induce ansiedad, afecta a la autoestima y le hace perder al sujeto su capacidad de control (Echeburúa y de Corral, 2010).

2.2.8 Modelo FURPS

Este modelo fue desarrollado por Hewlett-Packard en 1987, comprende un conjunto de atributos de calidad de software, que forman el acrónimo FURPS: Funcionalidad (*Functionality*), Usabilidad (*Usability*), Confiabilidad (*Reliability*), Desempeño (*Performance*) y Soportabilidad (*Supportability*).

Los atributos de calidad FURPS representan el objetivo de todo diseño de software; la cual se utilizará para analizar y ponderar las herramientas de minería de datos para realizar la selección de ellas y utilizar las que más destaquen este análisis para ser utilizada. Cada atributo del modelo se define de la siguiente manera (Pressman, 2010):

- **Funcionalidad.** Se evalúa de acuerdo con el conjunto de características y capacidades del programa, la generalidad de las funciones que se entregan y la seguridad general del sistema.
- **Usabilidad.** Se evalúa tomando en cuenta factores humanos, la estética general, consistencia y documentación.
- **Confiabilidad.** Se evalúa con la medición de la frecuencia y gravedad de las fallas, la exactitud de los resultados, el tiempo medio para que ocurra una falla, la capacidad de recuperación ante esta y lo predecible del programa.
- **Desempeño.** Se mide con base en la velocidad de procesamiento, tiempo de respuesta, consumo de recursos, rendimiento y eficiencia.
- **Soportabilidad.** Combina la capacidad del programa para ser ampliable (extensibilidad), adaptable y servicial, además que pueda probarse, ser compatible y configurable (capacidad de organizar y controlar los elementos de la configuración del software) y que cuente con la facilidad para instalarse en el sistema y se detecten los problemas.

2.3 Marco tecnológico

Existen una variedad de herramientas para la aplicación de minería de datos, es por ello que se hace mención a cada una de ellas y más adelante se realizó un estudio y selección de las herramientas con las cuales se trabajó en la investigación, para la aplicación de las técnicas de minería de datos, esta selección se realizó con la utilización del modelo FURPS.

2.3.1 Herramientas *framework* para minería de datos

2.3.1.1 *RapidMiner*

Es un programa informático para el análisis y minería de datos. Permite el desarrollo de procesos de análisis de datos mediante el encadenamiento de operadores; esto significa que a través de una Interfaz Gráfica de Usuario (GUI) o de línea de comandos o a través de procesos *batch* (o de lotes) o desde otros programas a través de la Interfaz de programación de aplicaciones (API por sus siglas en inglés) de Java, permite realizar análisis estadístico, minería de datos y análisis predictivo (Vargas et al., 2016).

Es un software de tipo *Open-Source* con licencia *AGPL*, basado en *java*, que trabaja bajo las plataformas Windows y Linux, posee alrededor de 500 operadores que pueden ser combinados, utiliza el lenguaje de *scripting XML* para describir los operadores y su configuración y posee una gran cantidad de extensiones (*plugins*) (Rapidminer, 2016, como se citó en Vargas, et al., 2016).

2.3.1.2 Orange

Según Vargas *et al.* (2016) se trata de una suite para minería de datos y aprendizaje automático, desarrollado en la Facultad de Informática de la Universidad de Ljubljana (Eslovenia). Esta herramienta cuenta con un fácil, potente, rápido y versátil *front end* de programación visual para el análisis exploratorio de datos y visualización, y librerías para *Python* y secuencias de comandos. Contiene un completo juego de componentes desarrollados en C++. Para preprocesamiento de datos, características de puntuación y filtrado, modelado, evaluación del modelo y técnicas de exploración (García, 2013).

A estos componentes se puede acceder de dos formas (García, 2013): por medio de *scripts* desde *Python* y por medio de *widgets* (componentes GUI), desde *CANVAS*. Además, se trata de una aplicación multiplataforma y se distribuye bajo licencia *GPL* y proporciona componentes para: entrada/salida de datos, preprocesamiento de datos, modelado predictivo y métodos de descripción de los datos.

2.3.1.3 Weka

Es de libre distribución (licencia *GPL*) y destaca por la cantidad de algoritmos que presenta, así como por la eficiencia de los mismos, por los generadores de reglas, está desarrollada por miembros de la Universidad de Waikato. Proporciona herramientas para la realización de tareas propias de minería de datos, la visualización y permite la programación en *JAVA* de algoritmos más sofisticados para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades y su sistema operativo es multiplataforma (Rodríguez y Díaz, 2009).

2.3.2 Lenguaje de programación para minería de datos

2.3.2.1 Lenguaje R

Es un lenguaje y entorno de programación, creado en 1993 por Ross Ihaka y Robert Gentleman del Departamento de Estadística de la Universidad de Auckland, cuya característica principal es que forma un entorno de análisis estadístico para la manipulación de datos, cálculos y la creación de gráficos; con la particularidad de que es un software *GNU*, es decir, de uso libre (Contreras, Molina y Arteaga, 2010).

Las características principales son: almacenamiento y manipulación de datos, operadores para cálculo sobre variables indexadas (*arrays*) en particular matrices, herramientas y posibilidades gráficas para análisis de datos (Contreras et al., 2010).

2.3.2.2 Python

Según López, Macías y Sandoval (2019) *Python* es un lenguaje de programación de alto nivel, interpretado y multipropósito; cuenta con facilidades para la programación orientada a objetos, imperativa y funcional, por lo que se considera un lenguaje multiparadigmas.

Python tiene una comunidad activa de soporte de contribuyentes y usuarios que también hacen que su *software* esté disponible para que otros desarrolladores de *Python* lo usen bajo términos de licencia de código abierto. Esto permite a los usuarios de *Python* compartir y colaborar de manera efectiva, beneficiándose de las soluciones que otros ya han creado para problemas comunes, además de contribuir potencialmente con sus propias soluciones al grupo común (Python, s.f.).

2.3.2.3 Matlab

Matlab combina un entorno de escritorio ajustado para análisis iterativo y procesos de diseño con un lenguaje de programación que expresa directamente las matemáticas de matriz y matriz. Incluye el *Live Editor* para crear *scripts* que combinan código, salida y texto formateado en un cuaderno ejecutable (MathWorks, s.f.-a).

Según *MathWorks* (s.f.-b) *Matlab* facilita la ciencia de datos con herramientas para acceder y preprocesar datos, construir modelos predictivos y de aprendizaje automático, e implementar modelos en sistemas de TI empresariales. *Matlab* permite acciones como: acceder a datos almacenados en archivos planos, administrar y limpiar datos utilizando tipos de datos y capacidades de preprocesamiento para la preparación de datos programáticos e interactivos.

2.3.3 Herramientas de Microsoft para minería de datos

2.3.3.1 Microsoft SQL Server

Según *Microsoft* (2014) es un sistema de administración y análisis de base de datos relacionales, para soluciones de comercio electrónico, línea de negocio y almacenamiento de datos, proporciona nuevas soluciones de copia de seguridad y de recuperación ante desastres y contiene una arquitectura híbrida con *Windows Azure*, con lo cual permite utilizar características locales y aprovechar al máximo datos de *Microsoft*.

2.3.3.2 Microsoft Excel

Es la aplicación de *Microsoft Office* que es un tipo de hoja de cálculo, la cual se utiliza para calcular, analizar y gestionar datos. Con este programa se pueden realizar diversas

operaciones desde las más sencillas hasta las más complejas, por medio de funciones y formulas (Anónimo, 2013).

Excel cuenta con un cliente de minería de datos, el cual es un conjunto de herramientas que permiten tareas comunes de minería de datos, desde limpieza de datos hasta generación de modelos y consultas de predicción. Además de utilizar los datos en tablas o rangos en *Excel* y tener acceso a origen de datos externos (Microsoft, 2017).

2.3.3.3 Microsoft Analysis Services

Según *Microsoft* (2021b) *Analysis Services* es un motor de datos analíticos, utilizado en el soporte de decisiones y análisis de negocios. Proporciona modelos de datos semánticos de nivel empresarial para informes comerciales y aplicaciones de clientes como *Power BI*, *Excel*, informes de *Reporting Services* y otras herramientas de visualización de datos.

2.3.4 Servicios en la nube para minería de datos

2.3.4.1 Oracle: Oracle Business Intelligence Server

Según *Oracle* (s.f.) es una herramienta de *suite* de *Enterprise Edición Plus*, que se basa en un verdadero servidor de Business Intelligence (BI) diseñado para ser altamente escalable. Optimiza la simultaneidad y el paralelismo para hacer que el valor de las aplicaciones de BI esté al alcance de tantas personas como sea posible. Proporciona acceso a los datos y cálculos centralizados, creando esencialmente un gran canal a través del cual todos los usuarios pueden usar toda la información, de cualquier forma, dentro de la empresa.

2.3.4.2 IBM: SPSS Modeler

SPSS Modeler es una solución líder en ciencia de datos visual y *machine learning*. Ayuda a las empresas a acelerar el tiempo para obtener valor para la empresa y lograr los resultados deseados acelerando las tareas operativas para los científicos de datos. Las organizaciones líderes de todo el mundo confían en IBM para la preparación y el descubrimiento de datos, la analítica predictiva, la gestión e implementación de modelos y *machine learning* para monetizar activos de datos (IBM, s.f).

Según IBM (s.f.) *SPSS Modeler* ayuda con: aprovechar la innovación basada en código abierto, incluyendo R o *Python*, empoderar a los científicos de datos de todas las habilidades, programáticas y visuales, aplicar un enfoque híbrido - *on premise* y en la nube pública o privada e iniciar con poco y escalar hacia un enfoque gobernado a nivel empresarial.

2.3.4.3 Intel: DAAL

Según Kumar (2015) la Biblioteca de aceleración de análisis de datos Intel® (Intel® DAAL) es la biblioteca de bloques de construcción optimizados de la Arquitectura Intel® que cubre todas las etapas de análisis de datos: adquisición de datos desde una fuente de datos, preprocesamiento, transformación, extracción de datos, modelado, validación y toma de decisiones.

2.3.4.4 Amazon: EMR

Según *Amazon Web Services [AWS]* (s.f.) *Amazon EMR* es la plataforma de *big data* nativa en la nube líder del sector que permite que los equipos procesen grandes

cantidades de datos de forma rápida y rentable a escala. EMR utiliza herramientas de código abierto, que se combinan con la escalabilidad dinámica de *Amazon EC2* y el almacenamiento escalable de *Amazon S3*.

Los beneficios de utilizar *Amazon EMR* son: Ventajas, Facilidad de uso, Bajo costo, Elasticidad, Fiabilidad, Seguridad y Flexibilidad. Además, esta aplicación se puede utilizar en casos como: *Machine Learning*, Extracción, Transformación y Carga (ETL), análisis de secuencias de clics, *streaming* en tiempo real, análisis interactivo y genómica (AWS, s.f.).

2.3.4.5 Google: Google Cloud Datalab

Una herramienta interactiva fácil de usar para el aprendizaje automático y la exploración, el análisis y la visualización de los datos. *Cloud Datalab* es una herramienta interactiva potente creada para explorar, analizar, transformar y visualizar datos, así como para compilar modelos de aprendizaje automático en *Google Cloud Platform* (Google Cloud, s.f.).

Se ejecuta en *Compute Engine* y se conecta con facilidad a varios servicios en la nube, de modo que puedes enfocarte en tus tareas de ciencia de datos; *Cloud Datalab* permite analizar tus datos en *BigQuery*, *Cloud Machine Learning Engine*, *Compute Engine* y *Cloud Storage* con *Python*, *SQL* y *JavaScript* (Google Cloud, s.f.):

2.3.4.6 Microsoft Azure Machine Learning

Azure es un conjunto completo y en expansión constante de servicios de información en la nube que ayudan a las organizaciones a afrontar sus desafíos empresariales. *Azure*

ofrece la flexibilidad de crear, administrar e implementar aplicaciones en una red mundial enorme con las herramientas y las plataformas que prefiera (Microsoft Azure, s.f.-a).

Microsoft Azure Machine Learning es un servicio en la nube que permite acelerar y administrar el ciclo de vida de los proyectos de aprendizaje automático, el cual permite utilizar flujos de trabajo diarios como entrenamiento e implementación de modelos de minería de datos y administración de operaciones de *machine learning* (MLOps por sus siglas en inglés) (Microsoft Azure, 2021b).

2.3.4.7 SAS: SAS Enterprise Miner

Es una solución de minería de datos que permite incorporar patrones inteligentes a los procesos de *marketing*, tanto operativos como estratégicos. El software de SAS es un sistema de entrega de información que provee acceso transparente a cualquier fuente de datos, incluyendo archivos planos, archivos jerárquicos, y los más importantes manejadores de bases de datos relacionales (Rodríguez y Díaz, 2009).

También incluye su propia base de datos de información para almacenar y manejar los datos, es decir, un "*data warehouse*". El sistema soporta un amplio rango de aplicaciones, destacándose el análisis estadístico, análisis gráfico de datos, análisis de datos guiado, mejoramiento de la calidad, entre otros (Rodríguez y Díaz, 2009).

2.4 Marco legal

2.4.1 Licencia de software libre GPL

La Licencia Pública General de GNU (*GPL*, por sus siglas en inglés), es una licencia de *copyleft* para *software* y otros tipos de obras, está destinada a garantizar su libertad

para compartir y cambiar todas las versiones de un programa, para asegurarse de que siga siendo software libre para todos sus usuarios (GNU, 2007a).

Cuando se utiliza el término de *software* libre, se refiere a la libertad, no al precio, licencias públicas generales están diseñadas para garantizar que tenga la libertad de distribuir copias de software gratuito (y cobrar por ellas si lo desea), que reciba el código fuente o que pueda obtenerlo si lo desea, que puede cambiar el software o use partes de él en nuevos programas gratuitos, y que sepa que puede hacer estas cosas (GNU, 2007a).

Según GNU (2007a) los desarrolladores que usan la GNU GPL protegen sus derechos con dos pasos: (1) hacer valer los derechos de autor sobre el software y (2) ofrecerle esta Licencia que le otorga permiso legal para copiarlo, distribuirlo y / o modificarlo.

2.4.2 Licencia de software libre AGPL

La Licencia Pública General *Affero* de GNU (AGPL, sus siglas inglés) es una licencia libre de *copyleft* para *software* y otros tipos de trabajos, diseñada específicamente para garantizar la cooperación con la comunidad en el caso del *software* de servidor de red y el código fuente modificado esté disponible para la comunidad (GNU, 2007b).

Requiere que el operador de un servidor de red proporcione el código fuente de la versión modificada que se ejecuta allí a los usuarios de ese servidor. Por lo tanto, el uso público de una versión modificada, en un servidor de acceso público, le da acceso público al código fuente de la versión modificada (GNU, 2007b).

2.4.3. Licencia de software propietario

Este tipo de licencias son *Microsoft Azure*, *Microsoft SQL Server*, *Analysis Services*, entre otras que en ocasiones se tiene que pagar un costo para poder utilizar el software o tras veces tienen software de tipo académico para estudiantes o periodos de prueba para que los usuarios prueben el software, esta prueba no pasa de 30 días.

Microsoft Azure Machine Learning es un servicio de la plataforma *Microsoft Azure*, el cual es de software propietario y tiene un costo el utilizar los servicios o herramientas que esta ofrece. Por lo anterior *Microsoft Azure Machine Learning* ofrece instancias de trabajo de acuerdo con cinco categorías como: uso general, proceso optimizado, memoria optimizada, unidades de procesamiento de gráficos (GPU por sus siglas en inglés) y proceso de alto rendimiento; cada instancia de trabajo ofrece diferentes funcionalidades y el costo del uso del servicio se calcula por hora (Microsoft Azure, s.f.-b).

Los precios de *Amazon EMR* son simples y predecibles: el costo puede ser una tarifa por cada segundo utilizado, con un cobro mínimo de un minuto y una tarifa por hora dependiendo del tipo de instancia utilizada (p. ej., estándar, uso de unidades de procesamiento central (CPU), memoria, almacenamiento, entre otros) y el costo se calcula por segundo y muestra el tiempo en formato decimal. Los precios por hora oscilan entre 0.011 USD/hora y 0.27 USD/hora (desde 94 USD/año hasta 2367 USD/hora) (AWS, s.f.).

Google Cloud Datalab, tiene costo desde el momento que se utilice para crear algún recurso informático, el software *IBM SPSS Modeler* da una prueba gratis de acceso de 30 días, se tiene que crear una cuenta y contraseña para poder acceder a la prueba.

La licencia del manejador de base de datos de *Microsoft SQL Server*, se puede utilizar la versión de *Developer*, el cual descarga gratuita, esta permite desarrollar, crear, probar y demostrar aplicaciones con software de *SQL Server* de una manera rentable. *Microsoft Excel* tiene una versión de prueba por 30 días, pero igual al igual este software es de costo y tiene un precio de \$2,499.

2.4.4 Ley Federal de Protección de Datos Personales en Posesión de los Particulares

En el año 2010 se emitió la Ley Federal de Protección de Datos Personales en Posesión de los Particulares, la cual es de orden público y tiene el objetivo de dar protección a los datos personales en posesión de los particulares, con la finalidad de proteger, regular, controlar, informar y garantizar la privacidad de los datos personales de las personas.

Esta Ley rige la forma en que se tratarán los datos personales de los estudiantes que participaron en la muestra de información, los cuales fueron recabados con fines académicos, además los datos personales de los estudiantes no serán compartidos con ninguna otra autoridad u organización. En el anexo A, se proporciona más información sobre los artículos que constituyen el capítulo I, sobre las disposiciones generales de esta ley.

Capítulo 3. Aplicación de la metodología y desarrollo

En este capítulo se presenta el modelo de desarrollo que consistió en tres etapas, distribuidas de la siguiente manera: análisis de la información del *dataset*, evaluación y selección de la herramienta tecnológica utilizada y la aplicación de la metodología *KDD* para la ejecución de minería de datos para el descubrimiento de nuevo conocimiento a partir de los datos. A continuación, se describe con más detalle la aplicación de este modelo de desarrollo utilizado.

3.1 Análisis de la información del dataset

El *dataset* analizado y utilizado en esta investigación fue obtenido de la investigación de las doctoras Verónica García Martínez y Martha Patricia Silva Payró, en el que diseñaron, aplicaron, analizaron y validaron el instrumento, el cual en su versión original consta de 22 *ítems*.

En la figura 5, se muestra una parte de la estructura del *dataset*, que está conformada por 746 registros los cuales son los resultados de la aplicación del instrumento a los estudiantes. De igual manera contiene las variables de género, edad, ciclo escolar, División Académica, promedio y los 22 *ítems* que conforman el instrumento de aplicación.

Figura 5
Dataset utilizado en la aplicación de minería de datos.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P										
Marca temporal	Código	Ciclo escolar	División	IC	Asignatura que cursa	Primer	1	Cuando tengo que h	2	Me preparo con antic	3	Cuando me asignan	4	Cuando me asignan	5	Cuando tengo proble	6	Asisto a clase	7	Trato de completar e	8	Postergo los trabaj	9	Postergo las t	10
2/14/2019 13:31:30	2	21	3ro semestre	2019	1	1	2	1	2	3	1	5	5	5	5	5	4	1	1	1	1	1	1	1	1
2/14/2019 13:35:25	2	21	3ro semestre	2019	2	2	2	2	2	2	5	1	4	5	5	5	5	1	1	1	1	1	1	1	1
2/14/2019 13:39:20	2	20	2019-2021	2019	3	2	1	3	3	3	3	3	2	5	5	4	1	1	1	1	1	1	1	1	1
2/14/2019 14:05:23	2	19	Febrero-Agosto 2019	2019	1	3	1	1	4	4	4	3	3	5	4	4	2	2	2	2	2	2	2	2	2
2/14/2019 14:08:13	2	46	6° semestre	2019	4	1	3	3	4	2	2	2	2	5	4	2	2	2	2	2	2	2	2	2	2
2/14/2019 14:09:06	1	20	febrero-agosto 2019	2019	3	4	1	2	4	4	3	2	3	5	4	3	3	3	3	3	3	3	3	3	3
2/14/2019 14:10:57	2	46	6 semestre	2019	6	1	3	2	2	4	2	2	5	4	2	5	5	4	2	2	2	2	2	2	2
2/14/2019 14:27:06	2	20	Febrero-Agosto	2019	1	2	1	3	4	4	4	3	4	4	4	4	3	3	3	3	3	3	3	3	3
2/14/2019 14:30:26	1	20	feb-jun	2017	1	3	1	1	1	5	4	4	4	5	5	5	3	3	3	3	3	3	3	3	3
2/14/2019 14:32:25	2	21	4to semestre	2019	1	3	1	1	1	5	5	1	5	5	5	5	5	1	1	1	1	1	1	1	1
2/14/2019 14:34:16	1	20	feb-jun	2017	4	3	1	2	2	5	3	5	5	5	5	5	5	3	3	3	3	3	3	3	3
2/14/2019 14:36:44	1	19	feb-jun	2017	4	1	1	2	2	5	1	1	4	5	5	5	5	2	2	2	2	2	2	2	2
2/14/2019 15:12:00	2	21	Feb- Ago 2019	2019	3	5	1	4	5	3	3	5	5	5	5	5	2	2	2	2	2	2	2	2	2
2/14/2019 15:17:23	2	21	2015-2019	2019	5	3	2	3	3	4	3	3	5	5	5	5	3	3	3	3	3	3	3	3	3
2/14/2019 15:19:48	1	21	6to semestre	2019	3	1	2	1	3	3	2	3	4	4	4	4	3	3	3	3	3	3	3	3	3
2/14/2019 15:20:37	2	21	6to semestre	2019	3	5	1	1	1	5	1	2	5	5	5	4	1	1	1	1	1	1	1	1	1
2/14/2019 15:27:19	2	21	6to semestre	2019	3	5	2	1	1	5	4	2	5	5	5	4	1	1	1	1	1	1	1	1	1
2/14/2019 15:33:22	2	21	6 semestre	2019	3	5	2	3	3	3	5	4	3	5	5	4	2	2	2	2	2	2	2	2	2
2/14/2019 15:40:43	2	21	6to semestre	2019	3	5	1	2	3	4	2	3	4	5	5	5	3	3	3	3	3	3	3	3	3
2/14/2019 15:51:48	2	19	Febrero-Agosto	2019	3	2	1	1	4	1	4	1	5	5	5	5	1	1	1	1	1	1	1	1	1
2/14/2019 15:55:38	2	18	2018-2019	2019	3	5	2	2	2	5	3	5	5	5	5	5	2	2	2	2	2	2	2	2	2
2/14/2019 16:08:19	1	23	Enero-agosto	2019	5	5	2	2	3	4	3	5	5	5	5	4	3	3	3	3	3	3	3	3	3
2/14/2019 16:22:34	1	19	Febrero-agosto 2019	2019	6	1	2	1	3	4	3	1	4	5	5	5	1	1	1	1	1	1	1	1	1
2/14/2019 16:23:27	2	20	Largo Febrero-Agosto2019	2019	7	5	1	2	2	5	2	5	4	5	5	4	2	2	2	2	2	2	2	2	2
2/14/2019 16:23:27	2	20	Largo Febrero-Agosto2019	2019	7	5	1	2	2	5	2	5	4	5	5	4	2	2	2	2	2	2	2	2	2
2/14/2019 16:23:27	2	20	Largo Febrero-Agosto2019	2019	7	5	1	2	2	5	2	5	4	5	5	4	2	2	2	2	2	2	2	2	2
2/14/2019 16:23:27	2	20	Largo Febrero-Agosto2019	2019	7	5	1	2	2	5	2	5	4	5	5	4	2	2	2	2	2	2	2	2	2
2/14/2019 16:23:38	2	19	Enero-agosto 2019	2019	9	1	2	3	3	3	3	3	4	4	4	4	2	2	2	2	2	2	2	2	2
2/14/2019 16:39:04	2	19	Enero-agosto 2019	2019	9	5	2	3	3	3	3	3	4	4	4	5	3	3	3	3	3	3	3	3	3
2/14/2019 16:52:30	2	19	cuarto semestre	2019	9	5	2	3	3	3	3	3	4	4	4	5	3	3	3	3	3	3	3	3	3
2/14/2019 17:39:49	2	21	6to semestre	2019	3	5	1	1	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2/14/2019 17:42:41	1	18	febrero-agosto 2019	2019	9	5	2	1	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	3
2/14/2019 17:46:42	2	20	febrero-agosto	2019	6	1	2	1	3	4	3	1	4	5	5	4	4	4	4	4	4	4	4	4	4
2/14/2019 17:56:54	2	22	Febrero-Agosto 2019	2019	3	3	2	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2/14/2019 18:22:43	2	19	4to semestre	2019	3	2	1	2	2	3	3	3	4	4	4	5	1	1	1	1	1	1	1	1	1
2/14/2019 18:04:05	2	20	sexto semestre	2019	11	2	2	2	2	3	3	3	4	4	4	4	3	3	3	3	3	3	3	3	3
2/14/2019 18:06:58	2	19	feb-agosto 2019	2019	6	5	0	1	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4
2/14/2019 18:09:28	2	21	Febrero-Agosto 2019	2019	11	2	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2/14/2019 18:11:15	2	21	Febrero-Agosto 2019	2019	11	2	1	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3
2/14/2019 18:20:34	2	23	educación superior	2019	7	3	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2/14/2019 18:22:09	1	19	4to semestre	2019	4	3	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2/14/2019 18:24:23	1	21	6to semestre	2019	4	3	1	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2/14/2019 18:27:56	1	21	6to semestre	2019	4	3	1	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4

Nota: García y Silva (2019).

A lo largo de este apartado se analizará minuciosamente la información del instrumento aplicado y la información del dataset utilizado en esta investigación.

3.1.1 Cuantificación de la población de estudio

La población de estudio de la investigación son jóvenes universitarios matriculados en algunas de las siete asignaturas del área general bajo la modalidad a distancia (en línea) durante el ciclo escolar febrero-agosto 2019 de la UJAT.

Según García y Silva (2019) el motivo de la elección de la población de estudio es medir el nivel de procrastinación académica de los jóvenes y la interacción que mantienen con Internet durante el desarrollo de sus actividades; y a partir de ahí establecer parámetros sobre los hábitos de los estudiantes durante la realización de sus actividades en línea, lo que puede aumentar la conducta procrastinadora.

3.1.2. Depuración de la población

El *dataset* consta de 746 campos, los cuales corresponden a la muestra de 746 estudiantes de las 12 Divisiones Académicas que conforman la UJAT. En este estudio participaron 275 (36.9%) hombres y 471 (63.1%) mujeres, con edades entre 17 y 46 años; la tabla 3, muestra la distribución de participantes por cada División Académica (ver tabla 3).

Tabla 3
Distribución de participantes por División Académica.

División Académica	Frecuencia	Porcentaje
División Académica de Ciencias Agropecuarias (DACA)	10	1.3
División Académica de Ciencias Básicas (DACB)	142	19.0
División Académica de Ciencias Biológicas (DACBIOL)	11	1.5
División Académica de Ciencias Económico Administrativas (DACEA)	44	5.9
División Académica de Ciencias de la Salud (DACS)	209	28.0
División Académica de Ciencias Sociales y Humanidades (DACSYPH)	53	7.1
División Académica de Ciencias y Tecnologías de la Información (DACYTI)	18	2.4
División Académica de Educación y Artes (DAEA)	57	7.6
División Académica de Ingeniería y Arquitectura (DAIA)	107	14.3
División Académica Multidisciplinaria de Comalcalco (DAMC)	40	5.4
División Académica Multidisciplinaria de Jalpa de Méndez (DAMJM)	49	6.6
División Académica Multidisciplinaria de los Ríos (DAMR)	6	0.8
Total	746	100

Nota: García y Silva (2020).

3.1.3 Diseño de instrumento

El instrumento diseñado para la obtención del *dataset*, fue un cuestionario propuesto por García y Silva (2019) utilizando la Escala de Procrastinación Académica (EPA) de Busko (1998), que consta originalmente de 16 *ítems*, los cuales miden el nivel de

procrastinación y agregaron seis preguntas diseñadas para conocer el uso que los estudiantes hacen del Internet durante la realización de las actividades de las asignaturas en línea. En el anexo B se encuentra el instrumento diseñado y aplicado a los estudiantes.

Cada pregunta o *ítem* se responde con una escala tipo *Likert*, que va de (1) nunca, (2) casi nunca, (3) a veces, (4) casi siempre y (5) siempre. El instrumento en un principio estaba compuesto con un total de 22 *ítems*, de los cuales según García y Silva (2019) realizaron adaptaciones a la escala de Busko (1998), en donde se incluyeron ocho *ítems* sobre la autorregulación académica, ochos *ítems* para la procrastinación y seis *ítems*, denominados VGM (ver tabla 4).

Tabla 4

Ítems del instrumento para medir la procrastinación en estudiantes universitarios.

No.	Ítems
1	Cuando tengo que hacer una tarea la dejo para el último minuto.
2	Me preparo con anticipación para los exámenes.
3	Cuando me asignan lecturas las leo la noche anterior.
4	Cuando me asignan lecturas las reviso el mismo día de clase.
5	Cuando tengo problemas para entender algo, inmediatamente trato de buscar ayuda.
6	Asisto a clase.
7	Trato de completar el trabajo asignado lo más pronto posible.
8	Postergo los trabajos de los cursos que no me gustan.
9	Postergo las lecturas de los cursos que me disgustan.
10	Intento mejorar mis hábitos de estudio.
11	Invierto el tiempo necesario en estudiar aun cuando el tema sea aburrido.
12	Trato de motivarme para mantener mi ritmo de estudio.
13	Trato de terminar mis trabajos importantes con tiempo de sobra.
14	Me tomo el tiempo para revisar mis tareas antes de entregarlas.
15	Dejo para mañana lo que puedo hacer hoy.
16	Disfruto la mezcla de desafío con emoción de esperar hasta el último minuto para completar una tarea.
17	Mientras desarrollo mis actividades del curso en línea me conecto a otras páginas que son de mi agrado.
18	El impulso por revisar mis redes sociales mientras estoy revisando mi curso en línea es incontrolable.

No.	Ítems
19	Invierto más tiempo en revisar diversos sitios mientras estoy en la plataforma del curso que en desarrollar otras actividades.
20	Me siento nervioso y hasta desesperado si no puedo navegar
21	He dejado de hacer tareas de mi curso por distraerme con el internet.
22	He trabajado en la plataforma con mi celular, por un lado.

Nota: García y Silva (2019).

3.1.4 Aplicación del instrumento

El instrumento se aplicó de manera virtual, por medio de un formulario electrónico que fue enviado a los alumnos matriculados con asignaturas en línea de la Universidad. Se compartió el formulario durante tres semanas para que fuera respondido y se obtuvo una base de datos en Excel (García y Silva, 2019); la cual será utilizada como *dataset* para la implementación de las técnicas de minería de datos.

3.1.5 Visualización y exploración de los datos del *dataset*

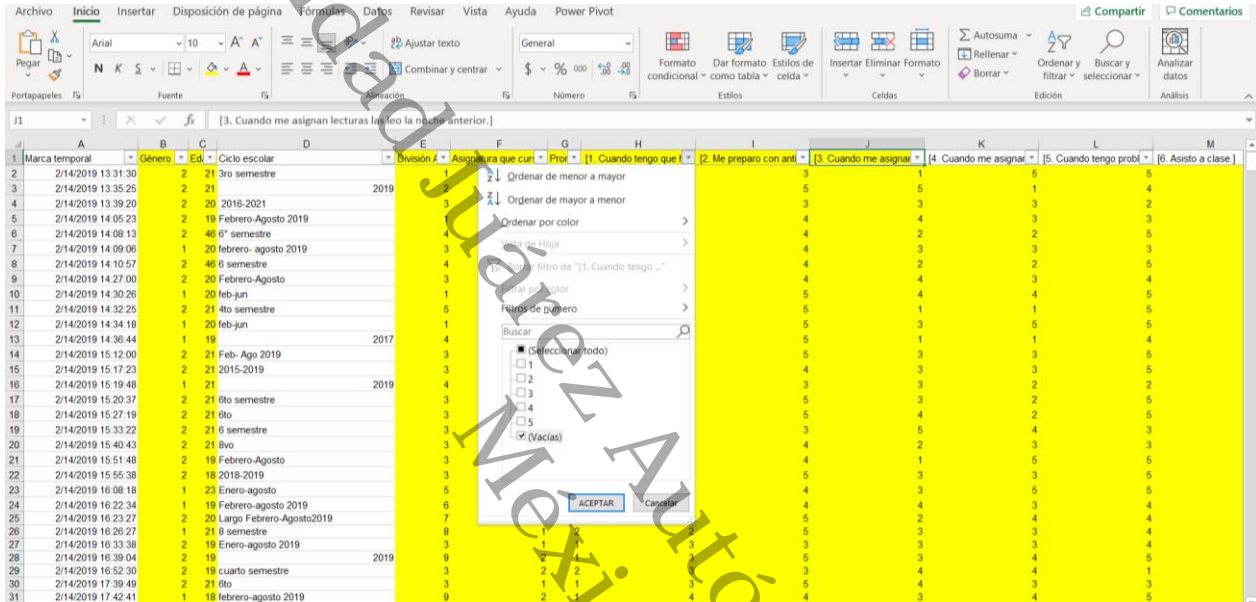
La visualización de los datos que conforman el *dataset*, se realizó por medio de una hoja de cálculo de Excel, la cual fue proporcionada por García y Silva (2019), donde se tuvo un primer acercamiento con los datos recabados y se realizó una exploración de los datos para buscar e identificar datos nulos o vacíos.

Lo anterior se llevó a cabo por medio de un filtrado de datos, es decir, se insertó un filtro en cada una de las variables del que conforman el *dataset*, esto dando clic en la opción Ordenar y filtrar, del cual se despliega una lista de opciones y se da clic en la opción Filtro. Una vez insertado el filtro en las variables se procedió a ir dando clic derecho en cada una de las variables para visualizar los datos que contiene cada variable y dar

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

clic en la casilla de vacías. En la figura 6 se muestra el filtrado de datos realizado a cada una de las variables para identificar datos vacíos o nulos (ver figura 6).

Figura 6
Filtro para identificar datos nulos o vacíos en el dataset.



Nota: Elaboración propia en base en García y Silva (2019).

Con el filtrado de datos se identificó que dentro de los registros en los 22 ítems que conforman el cuestionario se encuentran datos vacíos o nulos y el rango de incidencias es de uno a siete datos vacíos en los ítems. De igual forma se obtuvo que una fila o registro no contiene ningún dato en los 22 ítems, pero sí tiene datos en las variables de edad, género, ciclo escolar, División Académica, asignatura que cursa y promedio. En la figura 7, se muestra esta incidencia que ocurre con la fila o registro 138 (ver figura 7).

Figura 7
Identificación de datos nulos o vacíos en el dataset.

	A	B	C	D	E	F	G	H	I
1	Marca temporal	Género	Edad	Ciclo escolar	División Aca	Asignatura que cursa	Promex	1. Cuando tengo que hac	2. Me preparo con anti
138	2/14/2019 23:26:29	2	21		6	3	1	1	
288	2/15/2019 18:53:47	2	20	6to		3	1	1	1
748									
749									
750									
751									
752									
753									
754									

Nota: Elaboración propia en base en García y Silva (2019).

En la tabla 5, se desglosa el número de campos vacíos o nulos encontrados en los 22 ítems que conforman el cuestionario. Cabe mencionar esta situación no se presenta en variables como lo son edad, género, asignatura que cursa, ciclo escolar, División Académica y promedio (ver tabla 5).

Tabla 5
Datos nulos o vacíos identificados por cada ítem.

Variables o ítems	Campos vacíos o nulos
1. Cuando tengo que hacer una tarea, la dejo para el último minuto.	1
2. Me preparo con anticipación para los exámenes.	2
3. Cuando me asignan lecturas las leo la noche anterior.	3
4. Cuando me asignan lecturas las reviso el mismo día de clase.	3
5. Cuando tengo problemas para entender algo, inmediatamente trato de buscar ayuda.	7
6. Asisto a clase.	4
7. Trato de completar el trabajo asignado lo más pronto posible.	7
8. Postergo los trabajos de los cursos que no me gustan.	6
9. Postergo las lecturas de los cursos que me disgustan.	6
10. Intento mejorar mis hábitos de estudio.	5
11. Invierto el tiempo necesario en estudiar aun cuando el tema sea aburrido.	4
12. Trato de motivarme para mantener mi ritmo de estudio.	3
13. Trato de terminar mis trabajos importantes con tiempo de sobra.	3
14. Me tomo el tiempo de revisar mis tareas antes de entregarlas.	6
15. Dejo para mañana lo que puedo hacer hoy.	4
16. Disfruto la mezcla de desafío con emoción de esperar hasta el último minuto para completar una tarea.	3

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

VARIABLES O ÍTEMS	CAMPOS VACÍOS O NULOS
17. Mientras desarrollo mis actividades del curso en línea me conecto a otras páginas que son de mi agrado (a través de la computadora o mi celular).	3
18. El impulso por revisar mis redes sociales mientras estoy revisando el curso en línea es incontrolable.	7
19. Invierto más tiempo en revisar diversos sitios mientras estoy en la plataforma del curso que en desarrollar las actividades.	2
20. Me siento nervioso y hasta desesperado si no puedo navegar libremente en el Internet mientras hago mis actividades.	5
21. He dejado de hacer tareas de mi curso por distraerme con el Internet.	4
22. Trabajo en la plataforma con mi celular, por un lado.	3

Nota: Elaboración propia.

3.2 Selección de las herramientas para aplicar minería de datos

Actualmente existen diferentes herramientas o *software* que son utilizados para aplicar minería de datos ya sean libres o comerciales, de igual forma se encuentran los servicios en nube que proporcionan módulos para la aplicación de minería de datos. Derivado de lo anterior, resulta indispensable efectuar una selección de herramientas basadas en parámetros de calidad; en ese sentido, la metodología utilizada para realizar el análisis y selección de las herramientas para la aplicación de minería de datos en esta investigación fue la metodología de calidad de *software* bajo el acrónimo de FURPS.

FURPS fue desarrollada por *Hewlett-Packard* en 1987 y consta de cinco indicadores primordiales los cuales son: funcionalidad, usabilidad, confiabilidad, rendimiento y capacidad de soporte. En la tabla 6 se presentan los factores y criterios de calidad que conforman esta metodología (ver tabla 6).

Tabla 6
Factores y criterios de la metodología 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	Frecuencia y severidad de fallos. Exactitud de las salidas. Tiempo medio de fallos. Capacidad de recuperación ante fallos. Capacidad de predicción.
Rendimiento	Velocidad de procesamiento. Tiempo de respuesta. Consumo de recursos. Rendimiento efectivo total. Eficacia.
Capacidad de Soporte.	Extensibilidad. Adaptabilidad. Capacidad de prueba. Capacidad de configuración. Compatibilidad. Requisitos de instalación.

Nota: Conzanzo (2014).

Esta metodología ha sido utilizada en otros estudios para analizar las características y calidad de software de diferentes herramientas y programas; es el caso del estudio realizado por Núñez, Ferreira, Ayala, Ruiz y Cuin (2018) en el cual realizaron una búsqueda de las principales tecnologías utilizadas para el desarrollo web, donde posteriormente ponderaron las herramientas con base en la experiencia de utilización y cumplimiento de cada uno de los factores que contempla la metodología; obteniendo como resultado que las herramientas como *Angular JS*, *React JS* y *Bootstrap* son las que

obtuvieron mayor ponderación y son las destacadas para ser utilizadas para el desarrollo web.

Lo descrito anteriormente es un ejemplo de una investigación y la forma en que se trabaja con la metodología FURPS para lograr el objetivo de selección de las herramientas adecuadas utilizadas. A continuación, se describe paso a paso la utilización de la metodología y las características de las herramientas seleccionadas.

3.2.1 Aplicación de la metodología FURPS

De los cinco indicadores o factores que complementan la metodología, estos contemplan criterios desde características, capacidades del programa, compatibilidad y adaptabilidad en otros *softwares*. Teniendo estos criterios se realizó una sumatoria de ellos donde se obtuvo un total de 23 criterios que deben cumplir y analizarse de las herramientas, las cuales pueden tener una ponderación desde cero hasta 23. En la tabla 7 se muestran los factores, criterios y ponderación que pueden tener cada una de las herramientas (ver tabla 7).

Tabla 7

Factores, criterios y ponderación de la metodología FURPS para la selección de las herramientas para minería de datos.

Factores	Criterios	Ponderación
Funcionalidad	Características y capacidades del programa. Generalidad de las funciones. Seguridad del sistema.	3
Usabilidad	Factores Humanos. Factores estéticos. Consistencia de la interfaz. Documentación.	4
Confiabilidad	Frecuencia y severidad de fallos. Exactitud de las salidas. Tiempo medio de fallos. Capacidad de recuperación ante fallos.	5

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Factores	Criterios	Ponderación
Confiabilidad	Capacidad de predicción.	5
Rendimiento	Velocidad de procesamiento.	
	Tiempo de respuesta.	
	Consumo de recursos.	
	Rendimiento efectivo total.	6
	Eficacia.	
Capacidad de Soporte.	Extensibilidad.	
	Adaptabilidad.	
	Capacidad de prueba.	
	Capacidad de configuración.	
	Compatibilidad.	
	Requisitos de instalación.	

Nota: Elaboración propia con base en Constanzo (2014).

Cada uno de estos indicadores o factores y criterios se aplicaron a las 16 herramientas o servicios en la nube para minería de datos (ver tabla 8). Las que tuvieron mejor ponderación basadas en los criterios de los indicadores, fueron las herramientas seleccionadas para utilizar en el desarrollo del análisis de datos de la investigación.

Tabla 8

Clasificación de las herramientas analizadas con el modelo FURPS.

Clasificación	Herramientas
Herramientas framework	RapidMiner Orange Weka
Lenguajes de programación	Lenguaje R Python Matlab
Herramientas de Microsoft	Microsoft SQL Server Microsoft Excel Microsoft Analysis Services
Servicios en nube	Oracle Business Intelligence Server IBM: SPSS Modeler Intel: DAAL Amazon EMR Google Cloud Datalab Microsoft Azure Machine Learning SAS Enterprise Miner

Nota: Elaboración propia.

En tabla 9 se observa en primera instancia la forma en que se llevó a cabo el análisis de las herramientas para minería de datos, con los factores o indicadores, junto con los criterios que estos contienen de la metodología. De igual manera, se colocó en la parte inferior los nombres de las herramientas que se analizaron y si estas cumplían con alguno(s) de los criterios de la metodología FURPS, se marcaba la casilla con una X y si no era el caso se dejaba en blanco (ver tabla 9).

México.

Autónoma de Tabasco.

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Tabla 9

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 Machine Learning	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	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
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

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica 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 Machine Learning	SAS Enterprise Miner
Rendimiento																
Velocidad de procesamiento	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	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
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
Compatibilidad	x	x		x	x		x	x	x				x	x	x	
Requisitos de instalación	x		x	x			x	x	x	x			x			x

Nota: Elaboración propia.

Una vez realizado el análisis de los criterios que cumplen o no estas herramientas, se procedió a efectuar la sumatoria de los criterios para obtener el total que cumple cada una de las herramientas y observar cuál o cuáles son las que obtuvieron mejor ponderación (ver tabla 10).

Tabla 10
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 Machine Learning	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		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	5	4	4	3	6	3	5	5
Total	16	16	21	21	12	12	20	19	20	17	16	8	18	16	21	18

Nota: Elaboración propia.

De acuerdo con la tabla anterior, se puede observar que las herramientas con mejor ponderación para la aplicación de minería de datos son las herramientas: *Weka*, *Lenguaje R* y *Microsoft Azure Machine Learning*, las cuales serán las herramientas que se utilizarán para desarrollar y realizar la minería de datos.

3.2.2 Características de las herramientas seleccionadas

3.2.2.1 Lenguaje R

El lenguaje R es un entorno de software libre con licencia GNU GLP y lenguaje de programación interpretado, es decir, ejecuta instrucciones directamente, sin una previa

compilación del código; a su vez es utilizado para la computación estadística, gráfica y análisis de datos, ya que dispone de una amplia variedad de técnicas estadísticas y algoritmos, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, agrupamiento, entre otras. Además de tiene compatibilidad con sistemas operativos como *UNIX*, *Windows*, *MacOS* y *Linux* (Unir Revista, 2019).

Esta herramienta es muy utilizada porque es de *software* libre pero también porque tiene compatibilidad con otra herramienta que se utiliza para análisis de los datos como lo es *Python*, dentro de las principales características que es utilizada según Unir Revista (2019) son: el manejo y almacena efectivo de datos, herramientas o modelos para el análisis de datos, visualización de datos de manera gráfica, el lenguaje de programación se encuentra bien desarrollado e incluye saltos condicionales, bucles, funciones, recursivas, etc.

Las características del lenguaje R en el ámbito del análisis de datos, minería de datos o *Big Data* permite crear visualización de datos de alta calidad, dispone de herramientas de análisis estadísticos para profundizar en el conocimiento de los datos, crear *dashboards* para visualizar y analizar datos, crear informes automáticos. Además, según Unir Revista (2019) el usuario no solamente realiza programación o análisis, sino que R permite la interactividad, es decir, ensaya, se equivoca y vuelve a probar, y solo termina este ciclo si el resultado es satisfactorio para el usuario y permite la creación de un resultado final con el que se realiza un informe.

Lenguaje R cuenta con una herramienta o interfaz que permite acceder y visualizar de manera más sencilla los códigos, figuras, tablas o gráficos, esta herramienta es *RStudio*, la cual es un entorno de desarrollo integrado (IDE) para R. Para esta investigación se utiliza el Lenguaje R, pero a su vez utilizando *RStudio*.

3.2.2.2 Microsoft Azure Machine Learning

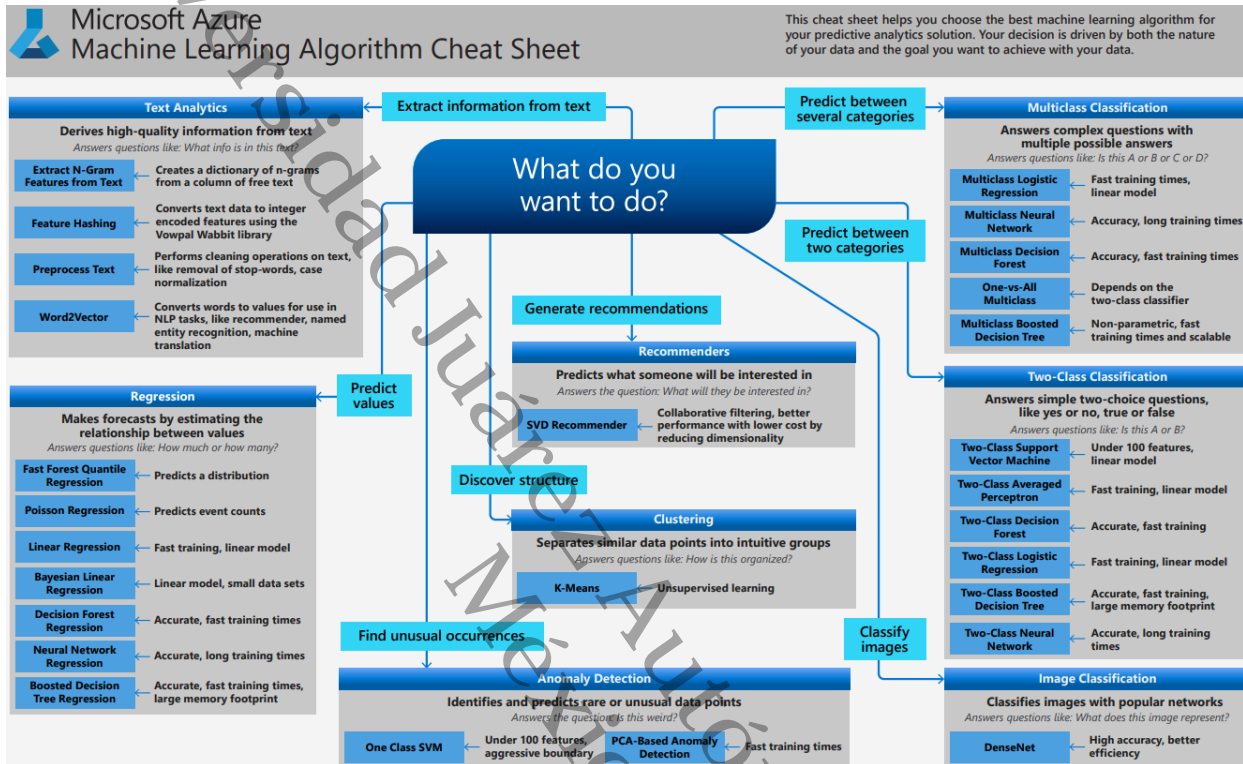
Microsoft Azure Machine Learning es un servicio en la nube que permite acelerar y administrar el ciclo de vida de los proyectos de aprendizaje automático. Esta herramienta permite a los individuos y equipos de trabajo, implementar MLOps dentro de su organización en un entorno de producción seguro y auditable (Microsoft Azure, 2021b).

De igual forma esta herramienta ofrece a los científicos de datos e ingenieros de *Machine Learning* (ML) herramientas para acelerar y automatizar sus flujos de trabajo diarios, para los desarrolladores de aplicaciones herramientas para integrar modelos de aplicaciones o servicios, así como un conjunto de herramientas de respaldo con API duraderas de *Azure Resource Manager* y garantiza seguridad conocida y control de acceso basado en rol (RBAC por sus siglas en inglés) (Microsoft Azure, 2021b).

Microsoft Azure Machine Learning ofrece una gran biblioteca de técnicas de clasificación, sistemas de recomendación, agrupamiento en *clústeres*, detección de anomalías, regresión y análisis de texto, cada uno está diseñado para abordar un tipo diferente de problema de aprendizaje automático (Microsoft Azure, 2021a). La figura 8 se presenta un mapa mental o diagrama referente a las técnicas y algoritmos de minería de datos que ofrece esta herramienta (ver figura 8).

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 8
Técnicas y algoritmos para minería de datos que ofrece Microsoft Machine Learning.



Nota: Microsoft Azure (2021a).

Azure Machine Learning es una herramienta o servicio en la nube comercial, es decir, tiene un costo el uso de la herramienta, una de la ventaja es que solo paga lo que usa, además de poder transformar los datos complejos en datos útiles para trabajar.

Existen cinco categorías que contienen diferentes tipos de instancias de trabajo, estas categorías son: uso general, proceso optimizado, memoria optimizada, GPU y proceso de alto rendimiento; cada una de estas categorías cuenta con características claves que se pueden utilizar dependiendo la instancia con el que se trabaje y el precio varía en

función al núcleo de procesamiento virtual (vCPU) y el tamaño de memoria RAM. La tabla

11 se muestra información de las categorías e instancias de trabajo (ver tabla 11).

Tabla 11

Características de las categorías e instancias de trabajo de Microsoft Azure Machine Learning.

Categorías	Detalles	Grupo de Instancias
Uso general	Las instancias de esta categoría se utilizan para sitios web, bases de datos pequeñas y medianas y otras aplicaciones habituales.	D2-64 v3 D2s-64s v3 D1-5 v2 D1s-5s v2 D2ds-D64ds v4 Serie Fsv2
Proceso optimizado	Son para uso elevado de la CPU en proporción de memoria, para servidores web de tráfico medio, aplicaciones de red, procesos por lotes y servidores de aplicaciones.	E2-64 v3 D-series D11-15 v2 D11S-15S v2 E2a-E96a v4 Serie M
Memoria optimizada	Las instancias esta dadas de acuerdo con el núcleo de memoria alta, excelente para servidores de bases de datos relacionales, memorias caché de capacidad media o grande y análisis en memoria.	Serie NC Serie NCsv2 Serie NCsv3 Serie NV Serie NVv3 Serie NDs Serie NCas_T4_v3 Serie NDv2 Serie ND A100 v4
GPU	Son instancias para máquinas virtuales especializadas específicas para la representación de gráficos pesados y la edición de vídeo disponibles con una o varias GPU.	Serie H Serie Hb Serie HBv2 Serie HC
Proceso de alto rendimiento	Utilizadas para máquinas virtuales de CPU más rápidas y eficaces con interfaces de red de alto rendimiento opcionales (RMDA).	

Nota: Microsoft Azure (s.f.-b).

Microsoft Azure proporciona 30 días de uso gratis donde se pueden utilizar todas las herramientas y servicios de esta plataforma en este caso *Azure Machine Learning*, una vez terminado esos días, la herramienta comienza a tener un costo por su utilización de

acuerdo con lo descrito anteriormente sobre las categorías e instancias de trabajo, se decidió trabajar con la categoría memoria optimizada, porque trabaja con servidores de base de datos relacionales y se está trabajando con datos de una base de datos.

De acuerdo con el grupo de instancias de trabajo que ofrece esta categoría se eligió trabajar con el grupo de instancia D11S-15S v2, ya que el costo de uso por hora y mes no es muy costoso. En la tabla 12 se muestra las instancias pertenecientes a este grupo de instancia de la categoría memoria optimizada, con número de número de vCPU, gigabytes de memoria RAM, precio por uso en hora y mes. Estos precios se obtuvieron a partir de los datos establecidos en la página web de *Microsoft Azure* en relación con la herramienta de *Azure Machine Learning* (ver tabla 12).

Tabla 12
Instancias y costo de las instancias de trabajo de Microsoft Azure Machine Learning.

Instancia	vCPU	Memoria RAM (GB)	Precio por hora en USD*	Precio por mes en USD*
DS11 v2	2	14	\$ 0.166	\$ 121.18
DS12 v2	4	28	\$ 0.332	\$ 242.36
DS13 v2	8	56	\$ 0.665	\$ 485.450
DS14 v2	16	112	\$ 1.33	\$ 970.900
DS15 v2	20	140	\$ 1.662	\$ 1,213.26

*Se aplica las tarifas estándar de transferencia de datos. Para la estimación de precios mensuales, se consideran 730 horas al mes.

Nota: Microsoft Azure (s.f.-b).

Los precios presentados en la tabla 12, representan los precios establecidos por la herramienta de acuerdo con la categoría, región donde se va a trabajar con la herramienta, la divisa de la moneda o dinero y el precio mostrado está expresada en hora y mes. En este caso en particular se seleccionó la región en centro sur de Estados Unidos

y la divisa en dólar estadounidense (USD), porque no se encuentra la divisa en peso mexicano (MXN).

Una vez establecidas y visualizadas las instancias del grupo de instancia seleccionado, se realizó la selección de la instancia de trabajo más apropiada a trabajar, en este caso se eligió trabajar con la instancia DS11 v2, por ser una instancia que se encuentra en un precio considerado y accesible, además de contar con memoria RAM suficiente para realizar los procedimientos y análisis de los datos. En la tabla 13 se muestra la instancia, características y costos de la instancia de trabajo seleccionada anteriormente (ver figura 13).

Tabla 13
Instancia de trabajo seleccionada en Microsoft Azure Machine Learning .

Instancia	vCPU	Memoria RAM (GB)	Precio por hora en USD*	Precio por mes en USD*
DS11 v2	2	14	\$ 0.166	\$ 121.18

*Se aplican las tarifas estándar de transferencia de datos. Para la estimación de precios mensuales, se consideran 730 horas al mes.

Nota: Microsoft Azure (s.f.-b).

3.2.2.3 Weka

Weka (Waikato Environment for Knowledge Analysis) es una plataforma o herramienta de software destinada para el aprendizaje automático y minería de datos, está escrita en lenguaje Java y desarrollada por la Universidad de Waikato, Nueva Zelanda. Además de ser un software de código libre con la licencia pública general GNU, se caracteriza por su adaptabilidad y compatibilidad con cualquier plataforma o sistema

operativo, contiene un gran repertorio de técnicas para preprocesamiento de datos y modelado de datos (Redacción España, 2020).

Dentro de las técnicas o algoritmos que se encuentran para realizar minería de datos son los de clasificación, agrupamiento (*clustering*), regresión y reglas de asociación, también se pueden utilizar técnicas estadísticas, algoritmos matemáticos y algoritmos de aprendizaje automático.

Según Redacción España (2020) *Weka* dispone de cuatro tipo de interfaces de usuario, las cuales el usuario puede utilizar de acuerdo a lo que necesite realizar dentro de la herramienta, estas interfaces son: a) *explorer*, permite procesar, clasificar, asociar y visualizar datos intuitivamente sobre archivos de datos, b) *experimenter*, permite comparación sistemática de ejecución de algoritmos predictivos de *Weka* sobre una colección de datos, c) *knowledgeFlow*, es una interfaz gráfica para el desarrollo de proyectos de flujos de información y d) simple CLI o *Simple Command-Line Interface*, es una consola que permite acceder a todas las opciones de *Weka* desde línea de comando.

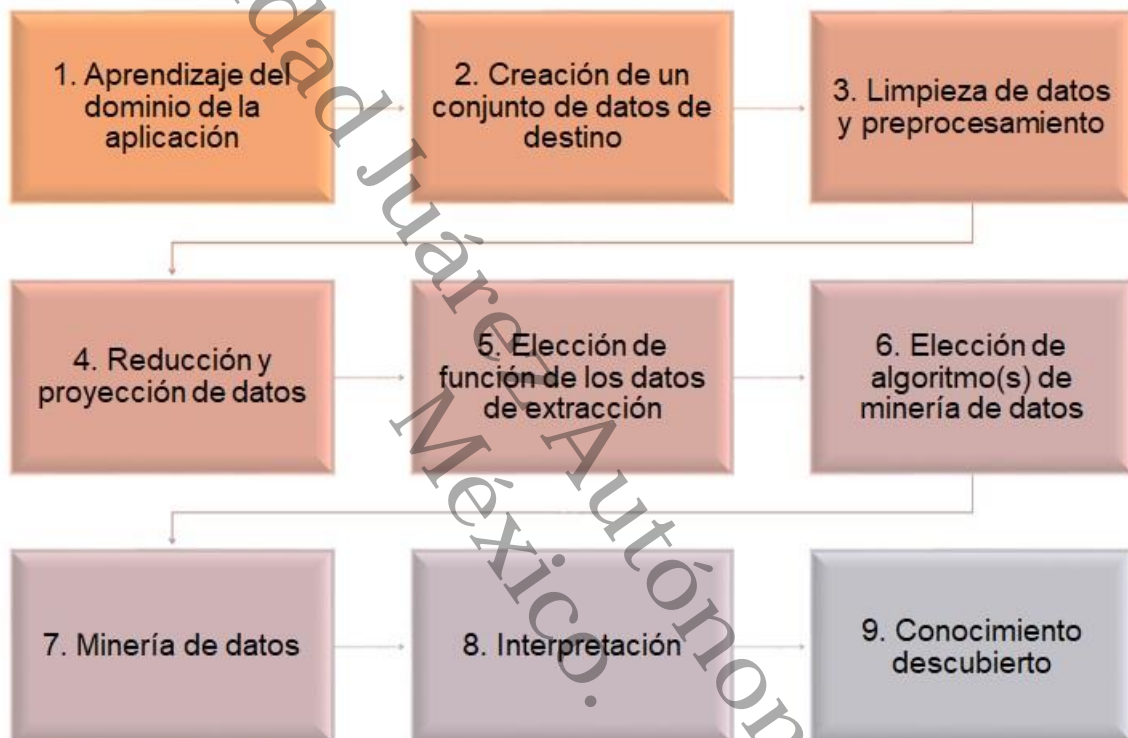
3.3 Metodología del proceso *KDD*

La metodología de Descubrimiento de Conocimiento en Base de Datos (*KDD* por sus siglas en inglés) es una de las primeras metodologías creadas y utilizadas para realizar análisis de datos almacenados en bases de datos, con el pasar de los años esta metodología se ha visto modificada.

La metodología *KDD* cuenta con nueve pasos, pero con las modificaciones que ha sufrido en algunos investigadores o profesionales del área de minería de datos, han

disminuido los pasos a cinco. En la figura 9 se presenta un esquema de los nueve pasos que contempla originalmente la metodología (ver figura 9).

Figura 9
Metodología del proceso KDD.



Nota: Elaboración propia con base en Fayyad, Piatetsky-Sapiro & Smyth (1996).

Por la gran demanda de datos que las organizaciones requieren analizar, se han creado otras metodologías para el descubrimiento de conocimiento de los datos, dentro de los cuales siempre se encuentra un paso o fase que es el de minería de datos. De acuerdo con Fayyad, Piatetsky-Sapiro & Smyth (1996) el valor de almacenar volúmenes de los datos depende de nuestra capacidad para extraer útiles informes de eventos interesantes y tendencias, decisiones de apoyo y basado en políticas en inferencia y

análisis estadístico, y explotar los datos para lograr negocios, operativos o metas científicas.

Por lo anterior descrito se decidió utilizar esta metodología ya que es una de las primeras metodologías utilizadas para el descubrimiento de conocimiento o análisis de datos y la cual contempla dentro de uno de los pasos la minería de datos, a su vez que se utilizarán los pasos nueve pasos que contemplan la metodología y será aplicada a las tres herramientas seleccionadas para minería de datos en el apartado anterior, las cuales son: Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*.

3.3.1 Aprendizaje del dominio de la aplicación

El primer paso de la metodología *KDD* es el aprendizaje del dominio de la aplicación, donde según Fayyad, Piatetsky-Sapiro & Smyth (1996) describen este paso con los conocimientos previos que se tienen hacer de la aplicación o herramienta que se utilizará durante todo el desarrollo del análisis de los datos y de igual forma se deben conocer los objetivos de la utilización de la aplicación.

En este paso se comenzó con la instalación de las tres herramientas seleccionadas a utilizar para minería de datos, a su vez se procedió a identificar las interfaces que en primera instancia las herramientas presentan una vez abierta las herramientas, así como identificar los elementos que estas contienen.

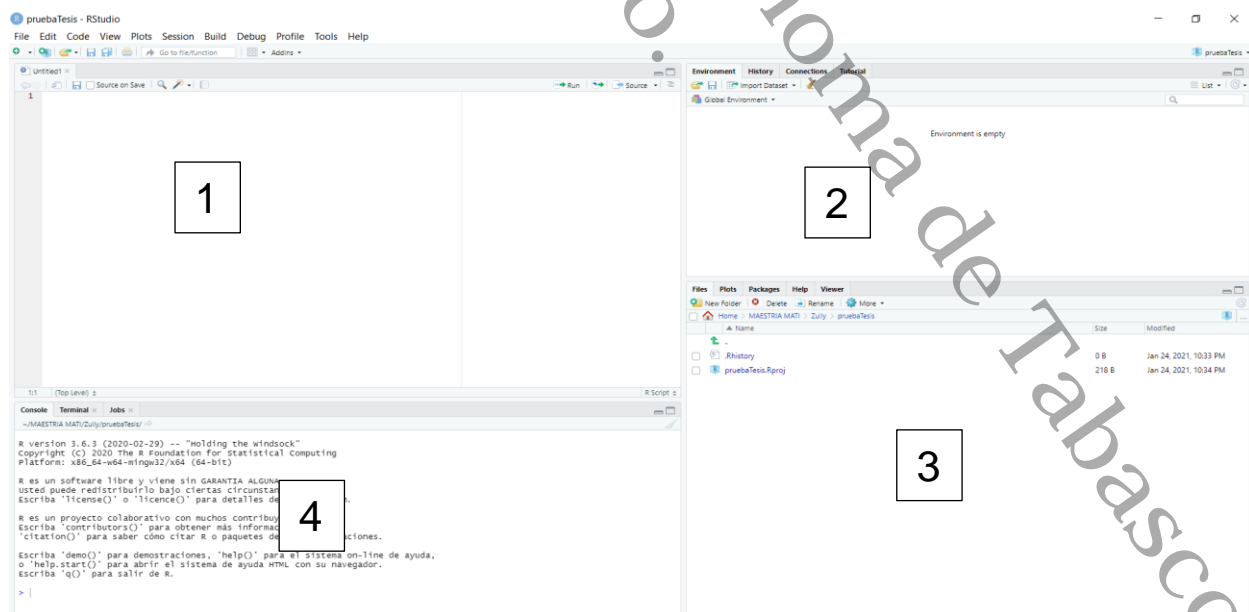
La primera herramienta utilizada fue Lenguaje R, donde se realizó la instalación de la herramienta, de igual forma se instaló *RStudio*, el cual es un entorno de trabajo integrado

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

(IDE) para facilitar el desarrollo y edición del código del algoritmo, y visualización de los datos.

En la figura 10 se muestra la principal interfaz de la herramienta de *RStudio*, que está compuesta por cuatro ventanas, las cuales son: 1) ventana de editor de sintaxis, donde se edita y se ejecuta el código; 2) ventana del entorno de trabajo, es donde se muestra el conjunto de datos y objetos que se almacenan al momento de ejecutar códigos; 3) ventana de subpestañas, se encuentran cinco pestañas que permiten visualizar el historial de archivos trabajados, visualización de gráficos, paquetes descargados y guardados de la herramienta, la pestaña de ayuda y *viewer* que muestra los resultados de reportes del tipo *markdown* y 4) ventana de consola, donde se observa si la ejecución fue correcta o hay errores en el código (ver figura 10).

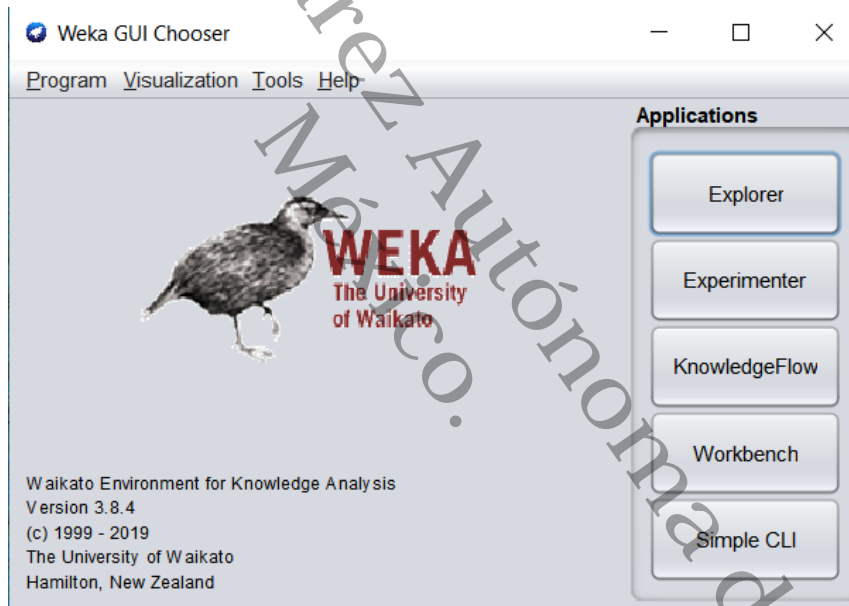
Figura 10
Interfaz principal de la herramienta Lenguaje R.



Nota: Elaboración propia.

La segunda herramienta con la cual se trabajó fue *Weka*; en la figura 11 se observa la pantalla principal de esta herramienta, donde se visualizan las cinco opciones de instancias donde el usuario puede seleccionar para trabajar, estas interfaces son: *explorer*, *experimenter*, *knowledgeFlow*, *workbench* y *simple CLI*, esta última se trabaja por medio de la Interfaz de Línea de Comandos (*CLI* por sus siglas en inglés) (ver figura 11).

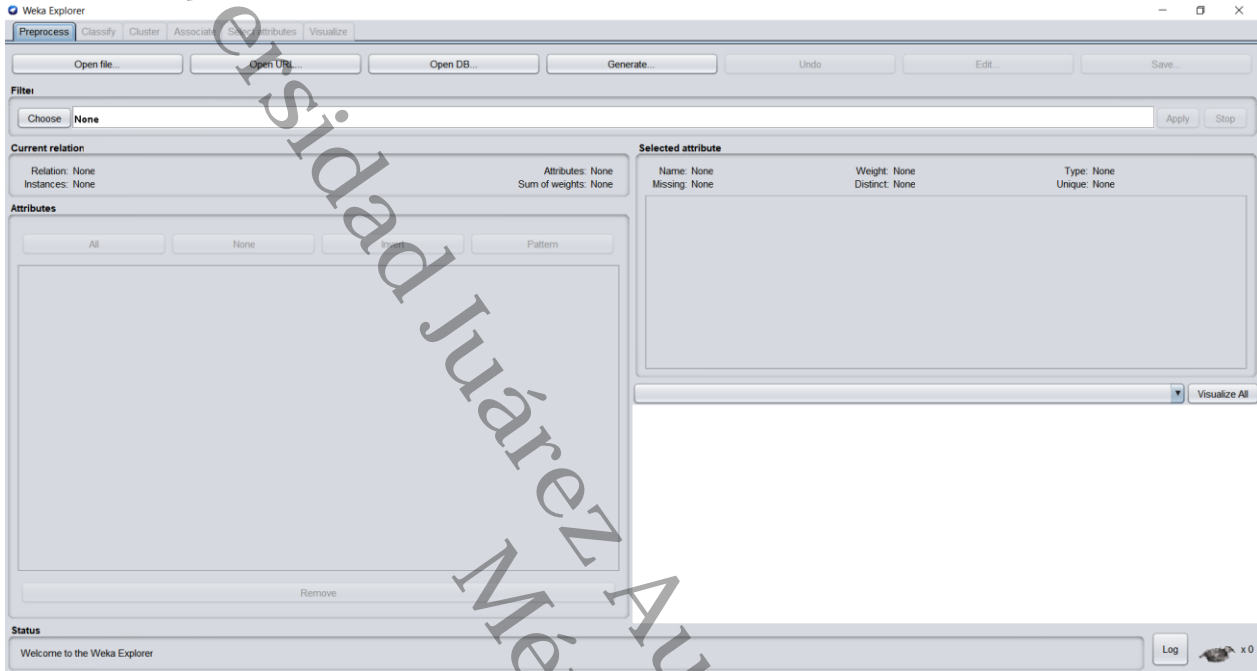
Figura 11
Interfaz principal de la herramienta *Weka*.



Nota: Elaboración propia.

Una vez visualizada y abierta cada interfaz de usuario de *Weka*, se decidió trabajar con la interfaz de *explorer*, la cual permite procesar, clasificar, asociar y visualizar datos intuitivamente sobre archivos de datos, además de contener una gran cantidad de algoritmos que se pueden utilizar para el análisis de los datos (ver figura 12).

Figura 12.
Interfaz principal de Explorer de la herramienta Weka.

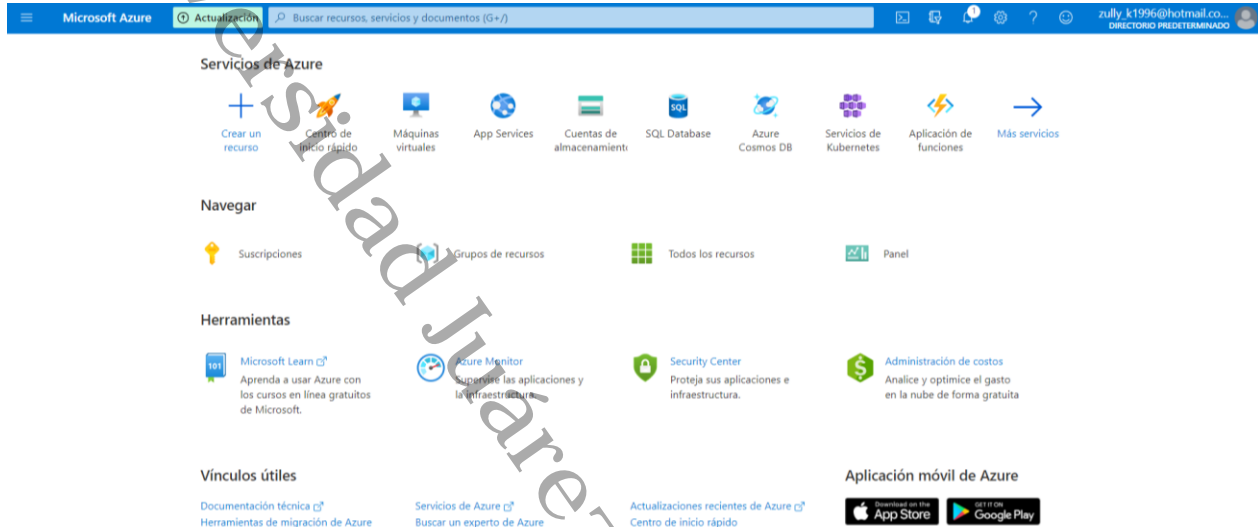


Nota: Elaboración propia.

La última herramienta utilizada fue la herramienta de *Microsoft Azure Machine Learning*, donde se creó una cuenta de *Azure*, para tener acceso a la herramienta, con esta cuenta se tuvo acceso a todos los servicios que ofrece *Azure* sin que se genere algún costo por 30 días, transcurridos estos días se generó un costo a pagar de \$0.166 dólares por hora, es decir, cada hora que se trabajaba en la herramienta generaba un costo.

En la figura 13 se presenta la interfaz principal de *Microsoft Azure*, conformada por cuatro secciones principales, las cuales son: servicios de *Azure*, navegar, herramientas y vínculos útiles.

Figura 13.
Interfaz principal de Microsoft Azure.



Nota: Elaboración propia.

Una vez analizado la interfaz, se procedió a buscar el servicio de *Azure Machine Learning*, que se encuentra en la categoría de Inteligencia artificial (AI) + aprendizaje automático. Donde de la herramienta de *Azure Machine Learning* se tiene que crear un área de trabajo para administrar las soluciones de aprendizaje automático para minería de datos.

En la figura 14 se muestra la interfaz para la creación del área de trabajo de la herramienta, donde se tienen que llenar los siguientes campos: a) grupo de recursos, el cual se puede crear rápidamente en esa interfaz, b) nombre del área de trabajo, c) región, d) cuenta de almacenamiento, e) almacén de claves y f) *application insights* (información de la aplicación) (ver figura 14).

Figura 14. Interfaz para la creación del área de trabajo en Microsoft Azure Machine Learning.

Todos los servicios > Aprendizaje automático >

Aprendizaje automático << >>
Directorio predeterminado

+ Crear ⚙ Administrar vista

Filtrar por cualquier campo...

Nombre ↑↓

Aprendizaje automático ...
Creación de un área de trabajo de Machine Learning

Aspectos básicos Redes Opciones avanzadas Etiquetas Revisión y creación

Detalles del proyecto

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción * ⓘ Suscripción de Azure 1

Grupo de recursos * ⓘ (Nuevo) Minería
[Crear nuevo](#)

Detalles del área de trabajo

Especifique el nombre y la región del área de trabajo.

Nombre del área de trabajo * ⓘ Algoritmos_datos ✓

Región * ⓘ Centro-Sur de EE. UU.

Cuenta de almacenamiento * ⓘ (nuevo) algoritmosdato3313614280
[Crear nuevo](#)

Almacén de claves * ⓘ (nuevo) algoritmosdato7004726063
[Crear nuevo](#)

Application Insights * ⓘ (nuevo) algoritmosdato4979344079
[Crear nuevo](#)

[Revisión y creación](#) < Anterior Siguiente: Redes

Nota: Elaboración propia.

Una vez llenado los campos y dando clic en el botón revisión y creación, la herramienta comienza con la implementación del área de trabajo, una vez completada la creación e implementación del área de trabajo, se visualiza la misma con la información esencial del área de trabajo y aparece la opción de iniciar en *Azure Machine Learning Studio*, el cual es un portal web de *Azure Machine Learning* para la creación, entrenamiento, evaluación e implementación de los modelos de aprendizaje automático (ver figura 15).

Figura 15.
Área de trabajo creada en Microsoft Azure Machine Learning.

^ Información esencial		Vista JSON
Grupo de recursos :	Minería	URL web de Studio : https://ml.azure.com/?tid=41db69c1-3805-42b3-b27d-681d956efe3a&ws...
Ubicación :	Centro-Sur de EE. UU.	Almacenamiento : algoritmosdato3313614280
Suscripción :	Suscripción de Azure	Registro : ...
Id. de suscripción :	42a54610-1206-4523-9995-24d00832001c	Key Vault : algoritmosdato7004726063
		Application Insights : algoritmosdato4979344079

Administración del ciclo de vida del aprendizaje automático

Use Azure Machine Learning Studio para crear, entrenar, evaluar e implementar modelos de aprendizaje automático.

[Más información](#)

[Iniciar Studio](#)

Nota: Elaboración propia.

3.3.2 Creación de un conjunto de datos de destino

El segundo paso de la metodología *KDD* de creación de un conjunto de datos de destino, es donde se comienza a seleccionar el conjunto de datos o un subconjunto de diversas tablas, variables o muestras de datos, los cuales son los que se utilizan para realizar el análisis de los datos (Fayyad, Piatetsky-Sapiro & Smyth, 1996).

El conjunto de datos utilizado fue el *dataset* obtenido de la investigación de García y Silva (2019), que está conformado por solo una tabla con 22 *ítems* y 746 registros, además de seis *ítems* en relación con información sobre el estudiante con lo es el género, edad, ciclo escolar, División Académica, asignatura que cursa, promedio obtenido en el bachillerato.

Además, al conjunto de datos se agregó un *ítem* o variable más, la cual contiene información sobre si el alumno es autorregulador o procrastinador, ya que esta variable es importante al momento de realizar minería de datos con algoritmos supervisados,

como lo son los árboles de decisión. De los 746 registros que contiene el *dataset*, 672 registros pertenecen a la clase de autorregulador y 74 registros pertenecen a la clase procrastinador, es decir la mayoría de los registros pertenecen la clase autorregulador.

En la tabla 14 se muestra el número y nombre de las variables que componen el conjunto de datos de destino, que se utilizó para la realización de minería de datos. Este conjunto de datos contiene datos numéricos y no numéricos, es decir, texto (ver tabla 14).

Tabla 14
Conjunto de datos del dataset.

No.	Variables o ítems
1	Género
2	Edad
3	Ciclo escolar
4	División Académica
5	Asignatura que cursa
6	Promedio obtenido en el bachillerato
7	1. Cuando tengo que hacer una tarea, la dejo para el último minuto.
8	2. Me preparo con anticipación para los exámenes.
9	3. Cuando me asignan lecturas las leo la noche anterior.
10	4. Cuando me asignan lecturas las reviso el mismo día de clase.
11	5. Cuando tengo problemas para entender algo, inmediatamente trato de buscar ayuda.
12	6. Asisto a clase.
13	7. Trato de completar el trabajo asignado lo más pronto posible.
14	8. Postergo los trabajos de los cursos que no me gustan.
15	9. Postergo las lecturas de los cursos que me disgustan.
16	10. Intento mejorar mis hábitos de estudio.
17	11. Invierto el tiempo necesario en estudiar aun cuando el tema sea aburrido.
18	12. Trato de motivarme para mantener mi ritmo de estudio.
19	13. Trato de terminar mis trabajos importantes con tiempo de sobra.
20	14. Me tomo el tiempo de revisar mis tareas antes de entregarlas.
21	15. Dejo para mañana lo que puedo hacer hoy.
22	16. Disfruto la mezcla de desafío con emoción de esperar hasta el último minuto para completar una tarea.
23	17. Mientras desarrollo mis actividades del curso en línea me conecto a otras páginas que son de mi agrado (a través de la computadora o mi celular).
24	18. El impulso por revisar mis redes sociales mientras estoy revisando el curso en línea es incontrolable.

No.	VARIABLES O ÍTEMS
25	19. Invierto más tiempo en revisar diversos sitios mientras estoy en la plataforma del curso que en desarrollar las actividades.
26	20. Me siento nervioso y hasta desesperado si no puedo navegar libremente en el Internet mientras hago mis actividades.
27	21. He dejado de hacer tareas de mi curso por distraerme con el Internet.
28	22. Trabajo en la plataforma con mi celular, por un lado.
29	¿El estudiante es procrastinador o autorregulador?

Nota: Elaboración propia con base en García y Silva (2019).

El *dataset* se encontraba en un formato hoja de cálculo (xlsx) y las herramientas a utilizar no leen este tipo de archivo, por ese motivo se decidió cambiar el formato de archivo por el de valores delimitados por coma (CSV) o mejor conocido como CSV, que es un formato de archivo que la mayor parte de las herramientas lee y este es el caso de Lenguaje R, Microsoft Azure Machine Learning y Weka.

3.3.3 Limpieza de datos y procesamiento previo

Según Fayyad, Piatetsky-Sapiro & Smyth (1996) en este paso se realizan las operaciones básicas para la limpieza de datos y procesamiento previo de los mismos, es decir, operaciones tales como: eliminación de ruido o valores atípicos en su caso, selección de información necesaria para modelar, determinación de estrategias para el manejo de los campos de datos que faltan y contabilidad de la información en la secuencia de tiempo y cambios conocidos, así como problemas con la base de datos, como el tipo de datos, esquemas y mapeo de valores perdidos y desconocidos.

Una vez analizado cada uno de los datos y variables que contienen el *dataset*, se procedió a la limpieza de los datos y preparación para el preprocesamiento previo de los datos para las herramientas. Durante la etapa de limpieza de los datos, se detectaron

datos nulos o vacíos y el campo Ciclo escolar, contenía datos no homogéneos y con alto grado de variabilidad.

Al momento de realizar la lectura del *dataset* en la herramienta *Weka* se obtuvo una tabla con los datos que contenía el campo, esto permitió tener un recuento de la variabilidad del dato y visualizar detenidamente los datos almacenados en este campo. En la figura 16, se puede apreciar la variabilidad de los datos que se puede tener en un solo campo, cuando los datos no están homogéneos y la interpretación que el usuario final (en este caso estudiantes) hacen sobre dicho campo (ver figura 16).

Figura 16

Tabla generada en *Weka* en relación con el campo *Ciclo escolar*.

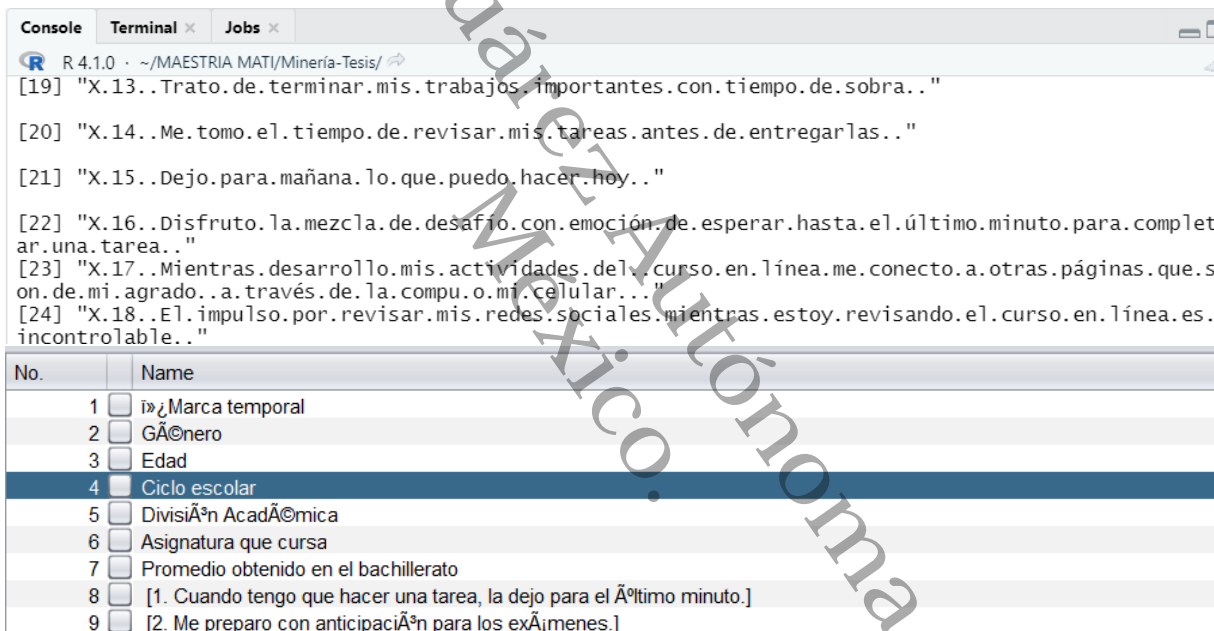
Name: Ciclo escolar		Type: Nominal	
Missing: 0 (0%)		Distinct: 264	
		Unique: 166 (22%)	
No.	Label	Count	Weight
1	3ro semestre	1	1.0
2	2019	91	91.0
3	2016-2021	1	1.0
4	Febrero-Agosto 2019	32	32.0
5	6º semestre	1	1.0
6	febrero- agosto 2019	3	3.0
7	6 semestre	10	10.0
8	Febrero-Agosto	5	5.0
9	feb-jun	2	2.0
10	4to semestre	10	10.0
11	2017	5	5.0
12	Feb- Ago 2019	1	1.0

Nota: Elaboración propia.

De igual forma se consideró que cada uno de los campos, datos y registro que contiene el *dataset* serán utilizados, con el fin de obtener más información, porque mientras más datos y campos se tengan para realizar minería de datos, más información se puede obtener de ellos para después poder interpretarlos.

Aunado a lo anterior se realizó el preprocesamiento de los datos en las herramientas, para verificar que la lectura y visualización de los datos del *dataset* sea correcta. Finalmente se detectó que los nombres de algunos campos se leían de forma correcta, pero por los acentos de las palabras y espacios entre palabras, se apreciaban con algunos símbolos y puntos en el caso de *RStudio* (ver figura 17).

Figura 17
Preprocesamiento del dataset en las herramientas.



Nota: Elaboración propia.

Por lo anterior se modificaron el nombre de cada una de las variables, por nombres más cortos, que se pudieran identificar rápidamente y que no tuvieran espacio entre cada palabra, número o símbolos. En la tabla 15 se muestra la modificación o renombre de cada variable (ver tabla 15).

Tabla 15
Modificación del nombre de las variables del dataset.

No.	Variables o ítems	No.	Variables o ítems
1	Género	16	Ítem10
2	Edad	17	Ítem11
3	CicloEscolar	18	Ítem12
4	DivisiónAcadémica	19	Ítem13
5	AsignaturaCursa	20	Ítem14
6	PromedioBachillerato	21	Ítem15
7	Ítem1	22	Ítem16
8	Ítem2	23	Ítem17
9	Ítem3	24	Ítem18
10	Ítem4	25	Ítem19
11	Ítem5	26	Ítem20
12	Ítem6	27	Ítem21
13	Ítem7	28	Ítem22
14	Ítem8	29	ClaseEstudiante
15	Ítem9		

Nota: Elaboración propia.

Por otro parte, se detectó que era necesario modificar el tipo de dato de la variable Ciclo escolar, ya que al momento de utilizar uno de los algoritmos de minería de datos, este algoritmo mandaba error al momento de visualizar los resultados obtenidos y por ese motivo se homogeneizaron los datos y por consecuente a cada dato se le dio un número, para que el tipo de dato sea numérico igual que los campos y no se presentara este error.

Lo anterior se llevó a cabo gracias al filtrado de datos donde se identificaron que los datos compartían similitudes, referencias y conceptos parecidos, los cuales se podrían agrupar como un solo dato para homogenizar esos datos y así consecutivamente con los demás datos. Al final este campo quedo con una variabilidad o catálogo de 68 datos.

En la tabla 16 se muestra un extracto de lo anteriormente descrito donde se realizó la homogeneización de los datos y numeración para identificarlos (ver tabla 16). Aunado a lo anterior, en el apéndice A se encuentra el catálogo de la descripción de los valores de los datos de cada que contiene cada variable o *ítem* que conforma el *dataset*.

Tabla 16
Extracto de la homogeneización y numeración de los datos del campo Ciclo Escolar.

Datos no homogéneos	Datos homogéneos	Numeración
01/02/2019	01/02/2019	1
2	2	2
3	3	3
...
feb-ago2019	Febrero-Agosto 2019	62
feb-agosto 2019	Febrero-Agosto 2019	62
feb-agostp 2019	Febrero-Agosto 2019	62
FEBERO- AGOSTO 2019	Febrero-Agosto 2019	62
...
Preparatoria	Preparatoria	68

Nota: Elaboración propia.

De igual forma al haber campos nulos o vacíos muy pocos por cada variable o ítem del *dataset* se llenaron estos campos con el número cero, para indicar que estos datos no representaban algún dato relevante dentro de los datos. Esto se realizó con el objetivo de prevenir cualquier error al momento implementar cualquiera de los algoritmos en las tres herramientas.

Aunado a lo anterior, se detectó que era necesario modificar el tipo de dato de las variables, ya que uno de los algoritmos para minería de datos utiliza datos nominales o de carácter, con el propósito de no tener problemas de lectura o análisis de los datos al momento de realizar la implementación de la técnica de minería de datos en las tres herramientas.

Esta modificación se realizó de acuerdo con lo que representan o significan los datos numéricos. Por ejemplo, la variable o columna Género contenía los datos uno y dos, cuales representaban hombres o mujeres, y con la modificación los datos de la variable fueron: valor uno a Hombres y valor dos a Mujeres; así como este ejemplo se realizó la modificación de los datos para cada una de las variables (ver figura 18).

Figura 18
Renombre y modificaciones de los datos del dataset.

	A	B	C	D	E	F	G	H	I	J
1	Género	Edad	CicloEscolar	DivisiónAcadé	AsignaturaCur	PromedioBac	Ítem1	Ítem2	Ítem3	Ítem4
2	Mujeres	21 años	3ro. semestre	DAMC	Derechos Hun	Menos de 9 y	Nunca	A veces	Nunca	Siempre
3	Mujeres	21 años	2019	DACBIOL	Filosofía y Ét	Menos de 9 y	Casi nunca	Siempre	Siempre	Nunca
4	Mujeres	20 años	2016-2021	DACS	Filosofía y Ét	Entre 9 y 10	A veces	A veces	A veces	A veces
5	Mujeres	19 años	Febrero-Agos	DAMC	Comunicación	Entre 9 y 10	Nunca	Casi siempre	Casi siempre	A veces
6	Mujeres	46 años	6to. semestre	DACB	Derechos Hun	Menos de 8 y	A veces	Casi siempre	Casi nunca	Casi nunca
7	Hombres	20 años	Febrero-Agos	DACS	Habilidades de	Entre 9 y 10	Casi nunca	Casi siempre	A veces	A veces
8	Mujeres	46 años	6to. semestre	DACB	Derechos Hun	Menos de 8 y	Casi nunca	Casi siempre	Casi nunca	Casi nunca
9	Mujeres	20 años	Febrero-Agos	DACS	Filosofía y Ét	Entre 9 y 10	A veces	Casi siempre	Casi siempre	A veces
10	Hombres	20 años	Febrero-Junio	DAMC	Comunicación	Entre 9 y 10	Nunca	Siempre	Casi siempre	Casi siempre
11	Mujeres	21 años	4to. semestre	DAIA	Comunicación	Entre 9 y 10	Nunca	Siempre	Nunca	Nunca
12	Hombres	20 años	Febrero-Junio	DAMC	Comunicación	Entre 9 y 10	Casi nunca	Siempre	A veces	Siempre
13	Hombres	19 años	2017	DACB	Derechos Hun	Entre 9 y 10	Casi nunca	Siempre	Nunca	Nunca
14	Mujeres	21 años	Febrero-Agos	DACS	Tecnología de	Entre 9 y 10	Casi siempre	Siempre	A veces	A veces

Nota: Elaboración propia con base en García y Silva (2019).

3.3.4 Reducción y proyección de datos

El cuarto paso de la metodología es la reducción y proyección de datos, donde según Fayyad, Piatetsky-Sapiro & Smyth (1996) se realiza una búsqueda de características útiles para representar los datos. Lo anterior, se efectúa, dependiendo del objetivo de la tarea y el uso de métodos de reducción de dimensionalidad o transformación para reducir el número efectivo de las variables de estudio o encontrar representaciones invariantes en los datos.

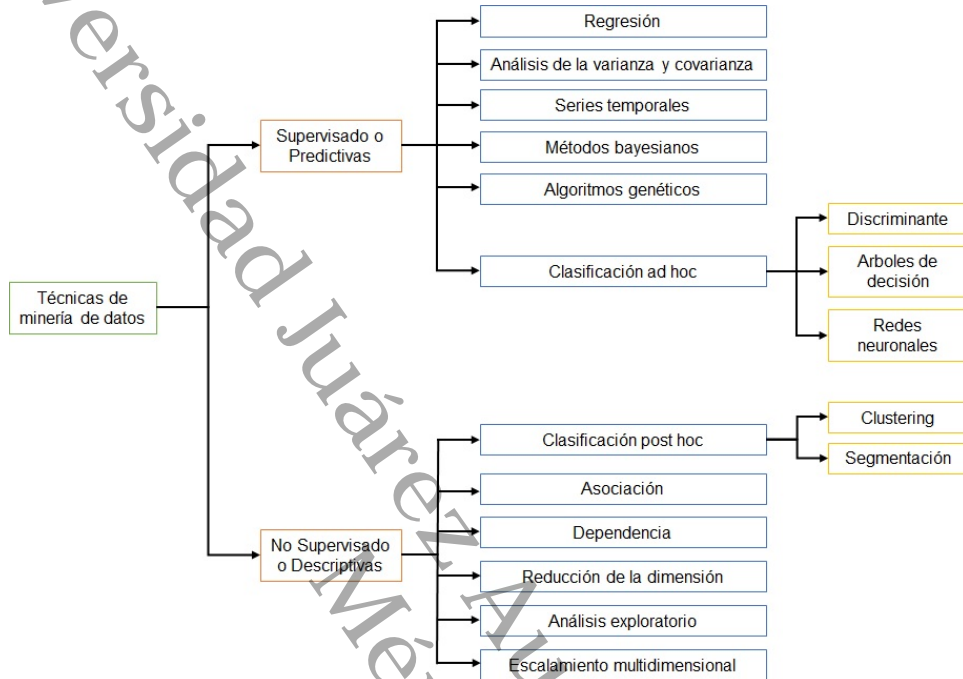
En el caso de la investigación una vez analizado el *dataset* se llegó a la conclusión que no se reducirán los datos almacenados en el *dataset*, sino que se realizará un análisis con cada una de las variables y datos para poder identificar de mejor forma los patrones de conducta de los estudiantes asociados a procrastinación y autorregulación académica.

El motivo del por qué no se reduce el número de datos a utilizar y variables es porque permite tener una mejor interpretación de los datos, ya que, si se reduce desde el principio, podrían existir datos o patrones que tal vez no se analicen y se pierda información y conocimiento importante acerca de la procrastinación y autorregulación académica.

3.3.5 Elección de función de los datos de extracción

El propósito del paso de elección de función de los datos de extracción es seleccionar y describir por qué la utilización de la técnica de minería de datos a utilizar. En la figura 19 se muestra la clasificación de estas técnicas (ver figura 19).

Figura 19.
Clasificación de las técnicas de minería de datos.



Nota: Elaboración propia con base en Pérez y Satín (2008 como se citó en Gutiérrez y Bernardo, 2015).

Durante ese paso de la metodología se analizaron las diferentes técnicas y funciones de extracción de datos o conocimiento, pero desde antes ya se había realizado una selección de las técnicas de minería de datos a utilizar, las cuales son: árbol de decisión, agrupamiento y reglas de asociación.

El motivo de la elección de los árboles de decisión es que permiten la visualización de un modelo de forma jerárquico donde cada elemento representa una o varias clases que se van ampliando a través de nodos. Por otro lado, la técnica de agrupamiento es realizar un análisis de conglomerados y agruparlos a través de varias categorías, clases o grupos que dependen de esa clase.

Por último, la técnica de reglas de asociación establece las reglas de clasificación y asociación dependiendo de los elementos y variables que se estén utilizando. A su vez una sola regla de asociación predice el valor de más de un atributo o variable. Es por estos motivos y porque son las técnicas un poco más utilizadas y populares para el análisis de datos y descubrimiento de conocimiento a partir de los datos que se seleccionaron estas técnicas.

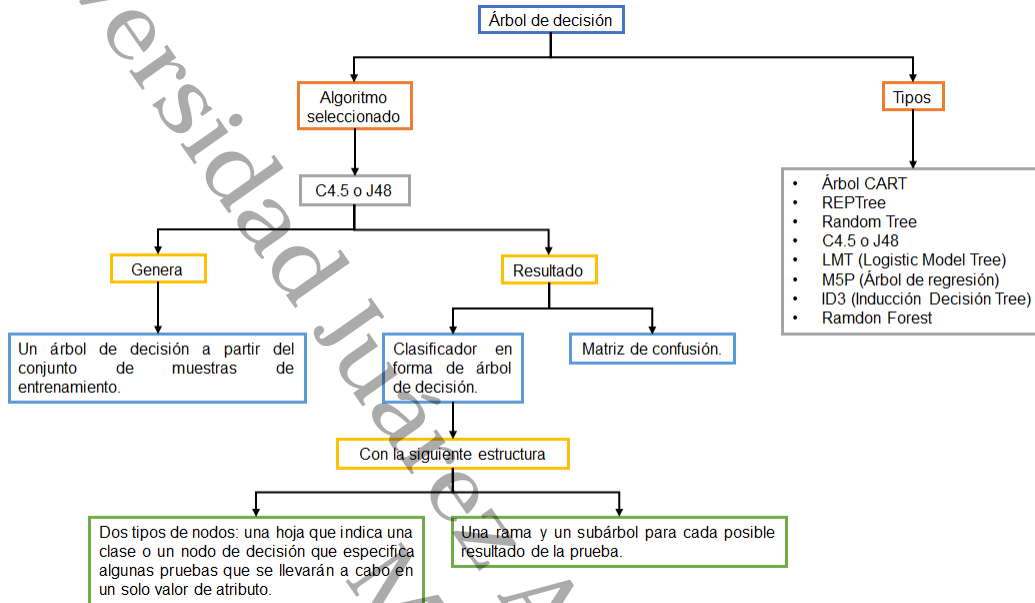
3.3.6 Elección de algoritmo(s) de minería de datos

El sexto paso de la metodología es la elección de algoritmos(s) de minería de datos, de acuerdo con Fayyad, Piatetsky-Sapiro & Smyth (1996) en este paso se seleccionan los algoritmos a implementar en la búsqueda de patrones en los datos, tales como decidir qué modelos y parámetros pueden ser apropiados; búsqueda de un método de minería de datos, en particular con los criterios generales del proceso de KDD.

Como este paso lo indica se procedió a la elección de los algoritmos de minería de datos, de acuerdo con las técnicas seleccionadas anteriormente en el paso cinco de la metodología.

La técnica de árbol de decisión contempla diferentes algoritmos para la generación de árboles de decisión; dentro de estos se encuentra el algoritmo C4.5 o J48, el cual es el algoritmo que se utilizó en esta investigación. En la figura 20 se muestran los tipos de algoritmos de árbol de decisión e información del algoritmo seleccionado (ver figura 20).

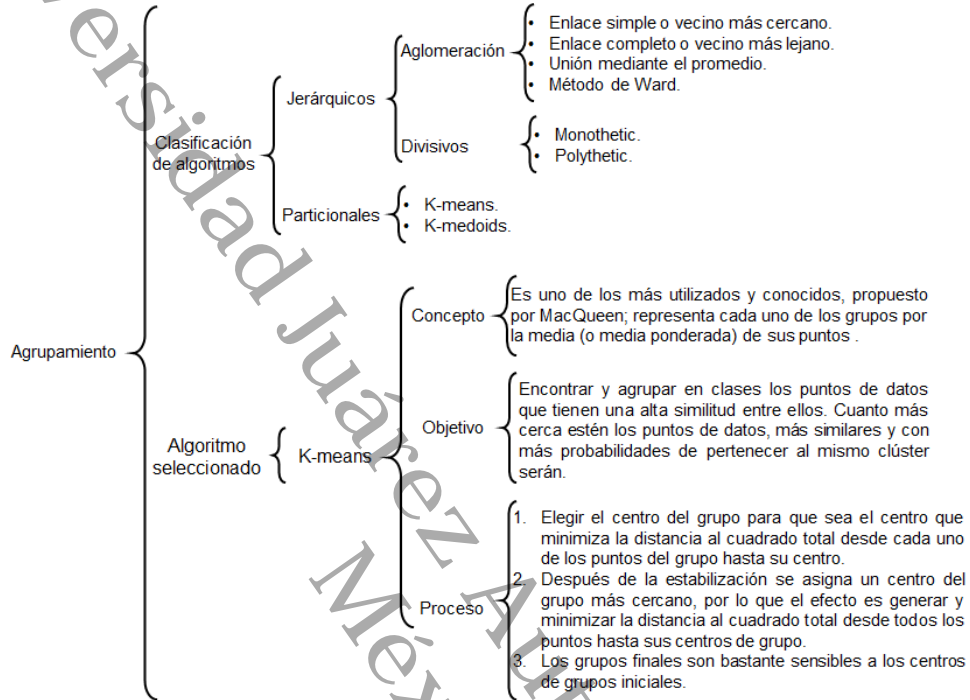
Figura 20.
Tipos de algoritmos y Algoritmo C4.5 o J48 de árbol de decisión.



Nota: Elaboración propia con base en Charris et al. (2018) y Kantardzic (2011).

Dentro de la técnica de agrupamiento seleccionado anteriormente existen diferentes algoritmos a utilizar, de los cuales se seleccionó el algoritmo *K-means*; en la figura 21 se muestra los tipos de algoritmos de agrupamiento e información del algoritmo seleccionado (ver figura 21).

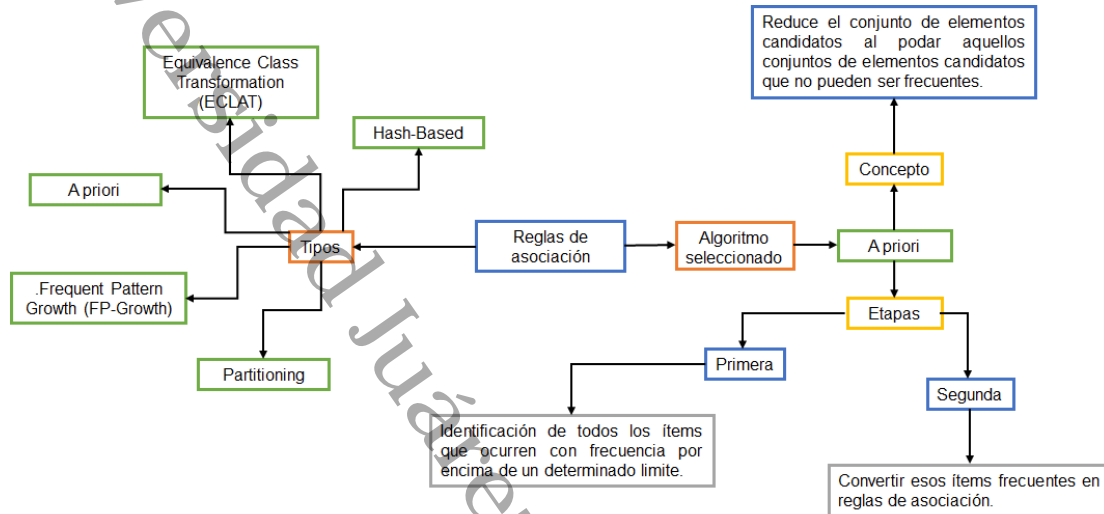
Figura 21
Tipos de algoritmos y Algoritmo K-means de agrupamiento.



Nota: Elaboración propia con base en Román (2019); Schiatti (2017); Vallejo (2016) y Witten, Frank y Hall (2011).

El tercer algoritmo seleccionado a utilizar fue el algoritmo A priori, el cual pertenece a la técnica de reglas de asociación. La figura 22 contienen los tipos de algoritmos que existen de esta técnica e información sobre el algoritmo A priori (ver figura 22).

Figura 22.
Tipos de algoritmos y Algoritmo A priori de reglas de asociación.



Nota: Elaboración propia con base en Amat (2018) y Kantardzic (2011).

Por lo anteriormente descrito y por ser algoritmos utilizados para la minería de datos y análisis de datos, se decidió utilizar estos algoritmos.

3.3.7 Minería de datos

El séptimo paso es el más importante dentro de la metodología ya que en este paso se lleva a cabo el análisis de los datos por medio de la minería de datos, donde se utiliza el *dataset* ya anteriormente preprocesado y realizada la limpieza de los datos a utilizar, y donde se utilizan las técnicas y algoritmos seleccionados en los pasos anteriores.

Según Fayyad, Piatetsky-Sapiro & Smyth (1996) este paso es donde se realiza la búsqueda de patrones de interés en una forma de representación en particular o un conjunto de representaciones, dependiendo del modelo y algoritmo utilizado.

Este paso se dividirá en tres partes porque se describirá el proceso que se llevó a cabo para la implementación de la minería de datos utilizando los tres algoritmos en las tres herramientas.

3.3.7.1 Lenguaje R

La minería de datos se llevó a cabo utilizando la herramienta *RStudio*, la cual permite visualizar y acceder de forma más rápida y gráfica a los códigos, tablas y gráficos. La primera técnica de minería de datos que se implementó en el *dataset*, fueron los árboles de decisión y se utilizó el paquete o librería de C50 el cual es una librería para realizar árbol de decisión y modelos basados en reglas, además que este algoritmo está basado en el algoritmo C4.5 o J48, el cual es el algoritmo que se seleccionó anteriormente.

En primer paso para la realización de los árboles de decisión fue realizar la lectura del *dataset*, esto implementando comandos como *read.csv* para leer el *dataset*, *view()* para la visualización del mismo y *summary()* para resumir los detalles de cada variable, además de renombrar el *dataset* dentro de la herramienta. En la figura 23 se muestra el código utilizado para la lectura del *dataset* y la ejecución de estos (ver figura 23).

Figura 23
Lectura del dataset en RStudio para árbol de decisión.

```
2 ##LECTURA DEL DATASET##
3 procras_autorre1 = read.csv("../Mineria-Tesis/procras_autorre1.csv") #Leer dataset
4 View(procras_autorre1) #Visualización del dataset
5 summary(procras_autorre1) #Resumen de las variables del dataset
6 datos <- procras_autorre1 #Renombre del dataset
```

6:48 (Top Level) ↓

Console Terminal Jobs

R 4.1.0 ~ /MAESTRIA MATI/Mineria-Tesis

```
> ##LECTURA DEL DATASET##
> procras_autorre1 = read.csv("../Mineria-Tesis/procras_autorre1.csv") #Leer dataset
> View(procras_autorre1) #Visualización del dataset
> summary(procras_autorre1) #Resumen de las variables del dataset
```

Género	Edad	CicloEscolar	DivisiónAcadémica	AsignaturaCursa	PromedioBachillerato	Item1
Min. :1.000	Min. : 17.00	Min. : 1.00	Min. : 1.000	Min. :1.000	Min. :1.000	Min. :0.000
1st Qu.:1.000	1st Qu.: 20.00	1st Qu.:23.25	1st Qu.: 3.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000
Median :2.000	Median : 21.00	Median :47.00	Median : 4.000	Median :2.000	Median :2.000	Median :3.000
Mean :1.631	Mean : 21.43	Mean :43.01	Mean : 5.044	Mean :2.476	Mean :1.655	Mean :2.477
3rd Qu.:2.000	3rd Qu.: 22.00	3rd Qu.:62.00	3rd Qu.: 6.000	3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:3.000
Max. :2.000	Max. :189.00	Max. :681.00	Max. :12.000	Max. :5.000	Max. :3.000	Max. :5.000

Item2	Item3	Item4	Item5	Item6	Item7	Item8	Item9
Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000
1st Qu.:3.000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.:3.000	1st Qu.:5.000	1st Qu.:4.000	1st Qu.:2.000	1st Qu.:2.000
Median :4.000	Median :3.000	Median :3.000	Median :4.000	Median :5.000	Median :4.000	Median :3.000	Median :3.000
Mean :3.934	Mean :3.032	Mean :3.021	Mean :3.901	Mean :4.792	Mean :4.229	Mean :2.503	Mean :2.466
3rd Qu.:5.000	3rd Qu.:4.000	3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:3.000	3rd Qu.:3.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000

Item10	Item11	Item12	Item13	Item14	Item15	Item16	Item17
Min. :0.000	Min. :0.00	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000
1st Qu.:4.000	1st Qu.:3.00	1st Qu.:4.000	1st Qu.:4.000	1st Qu.:4.000	1st Qu.:2.000	1st Qu.:1.000	1st Qu.:2.000
Median :4.000	Median :4.00	Median :4.000	Median :4.000	Median :4.000	Median :2.000	Median :2.000	Median :3.000
Mean :4.227	Mean :3.74	Mean :4.282	Mean :4.074	Mean :4.147	Mean :2.395	Mean :2.342	Mean :2.702
3rd Qu.:5.000	3rd Qu.:4.00	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:5.000	3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:4.000
Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000

Item18	Item19	Item20	Item21	Item22
Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000	Min. :0.000
1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:1.000	1st Qu.:2.000
Median :2.000	Median :2.000	Median :1.000	Median :1.000	Median :3.000
Mean :2.001	Mean :1.985	Mean :1.839	Mean :1.716	Mean :2.714
3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.:2.000	3rd Qu.:4.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000

```
> datos <- procras_autorre1 #Renombre del dataset
```

Nota: Elaboración propia.

Una vez realizado la lectura del *dataset* y comprobar el tipo de dato y detalles de las variables o ítems, se procedió a instalar el paquete y librería de C50, utilizando los comandos *install.packages()* y *library()* (ver figura 24).

Figura 24.
Librería cargada C50 para el árbol de decisión en RStudio.

```
10 ##CARGA DE LIBRERIAS DEL PAQUETE DE ARBOL DE CLASIFICACIÓN C5.0##
11 install.packages("C50")
12 library(C50)
13
16:1 (Top Level) ↕

Console Terminal Jobs
R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/ ↗
> ##CARGA DE LIBRERIAS DEL PAQUETE DE ARBOL DE CLASIFICACIÓN C5.0##
> install.packages("C50")
Installing package into 'C:/Users/Zully Guzmán Caraveo/Documents/R/win-library/4.1'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/C50_0.1.5.zip'
Content type 'application/zip' length 476599 bytes (465 KB)
downloaded 465 KB

package 'C50' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Zully Guzmán Caraveo\AppData\Local\Temp\RtmpiOrEiz\downloaded_packages
> library(C50)
Warning message:
package 'C50' was built under R version 4.1.1
> library(C50)
```

Nota: Elaboración propia.

Después de cargar las librerías que se utilizaron, se procedió a realizar un cambio del tipo de dato de la variable ClaseEstudiante de carácter a factor, ya que el dato de la variable categórica en *RStudio* debe ser de tipo factor, además de renombrar la variable ClaseEstudiante a clase (ver figura 25).

Figura 25
Cambio del tipo de dato de la variable categórica ClaseEstudiante para árbol de decisión en RStudio.

```
16 ##CAMBIO DEL TIPO DE DATO DE CARACTER A FACTOR DE LA COLUMNA CLASE##
17 clase <- vector(length = dim(datos)[1])
18 clase [datos$ClaseEstudiante=="Autorregulador"] <-"AUTORREGULADOR"
19 clase [datos$ClaseEstudiante=="Procrastinador"] <-"PROCRASTINADOR"
20 datos$ClaseEstudiante <- factor(clase)
21
22:1 (Top Level) ↕

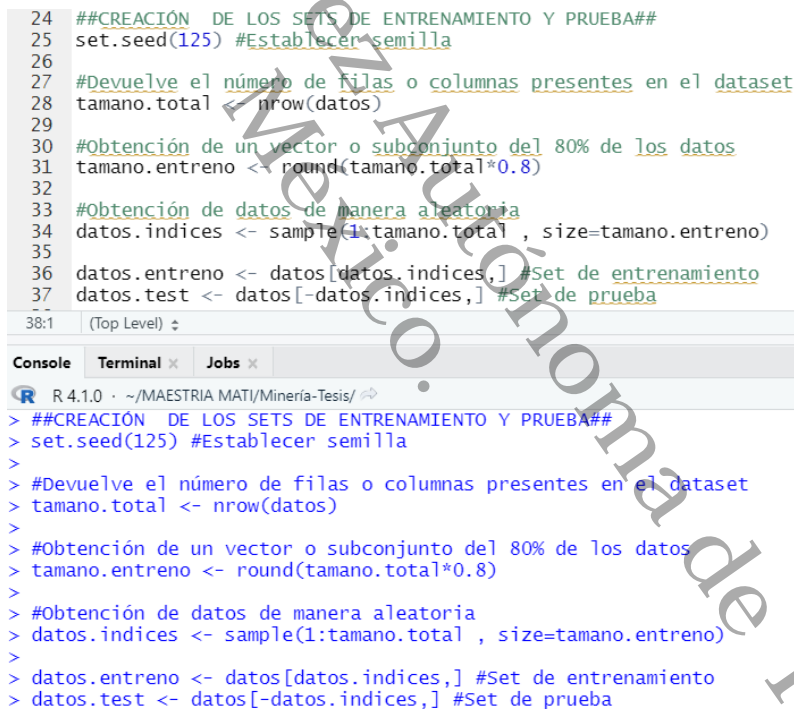
Console Terminal Jobs
R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/ ↗
> ##CAMBIO DEL TIPO DE DATO DE CARACTER A FACTOR DE LA COLUMNA CLASE##
> clase <- vector(length = dim(datos)[1])
> clase [datos$ClaseEstudiante=="Autorregulador"] <-"AUTORREGULADOR"
> clase [datos$ClaseEstudiante=="Procrastinador"] <-"PROCRASTINADOR"
> datos$ClaseEstudiante <- factor(clase)
```

Nota: Elaboración propia.

El siguiente paso para la creación del modelo de árbol de decisión, fue la creación de un set de entrenamiento para generar un modelo predictivo y un set de prueba para comparar la eficacia de este modelo para realizar predicciones correctas en cuanto al *dataset*. En la figura 26 se muestra la creación de estos dos sets, donde el set de entrenamiento representa el 80% de total de los datos y el set de prueba representa el 20% de los datos del *dataset* (ver figura 26).

Figura 26
Creación de sets de entrenamiento y prueba del modelo de árbol de decisión en RStudio.

```
24 ##CREACIÓN DE LOS SETS DE ENTRENAMIENTO Y PRUEBA##
25 set.seed(125) #Establecer semilla
26
27 #Devuelve el número de filas o columnas presentes en el dataset
28 tamaño.total <- nrow(datos)
29
30 #Obtención de un vector o subconjunto del 80% de los datos
31 tamaño.entreno <- round(tamaño.total*0.8)
32
33 #Obtención de datos de manera aleatoria
34 datos.indices <- sample(1:tamaño.total , size=tamaño.entreno)
35
36 datos.entreno <- datos[datos.indices,] #Set de entrenamiento
37 datos.test <- datos[-datos.indices,] #Set de prueba
38:1 (Top Level) ↕
```



The screenshot shows the RStudio interface with a console window. The console displays the execution of the R code provided in the text block above. The code defines variables for the total number of rows, the number of rows for the training set (80%), and the indices for the training set. It then creates the training and test datasets. The console output shows the code being executed line by line, with the final line being the creation of the test set.

Nota: Elaboración propia.

La figura 27 se presenta el código utilizado para la creación del modelo de árbol de decisión, utilizando la librería de C5.0(), en la cual se utiliza la variable que representa la clase categórica a clasificar con las demás variables que se tienen en el *dataset*, en este

caso utilizando el set de entrenamiento, además utilizando el comando `summary()` se obtiene información del árbol de decisión y con `plot()` el gráfico del mismo (ver figura 26).

Figura 27

Código para la creación del árbol de decisión en RStudio.

```
41 ##CREACIÓN DEL MODELO DE ARBOL DE DECISIÓN##
42 modelo <- C5.0(ClaseEstudiante ~ ., data = datos.entreno) #Arbol de decisión
43 summary(modelo) # Información sobre el modelo
44 plot(modelo) # Gráfico
```

38:1 (Top Level) ↓

Console Terminal x Jobs x

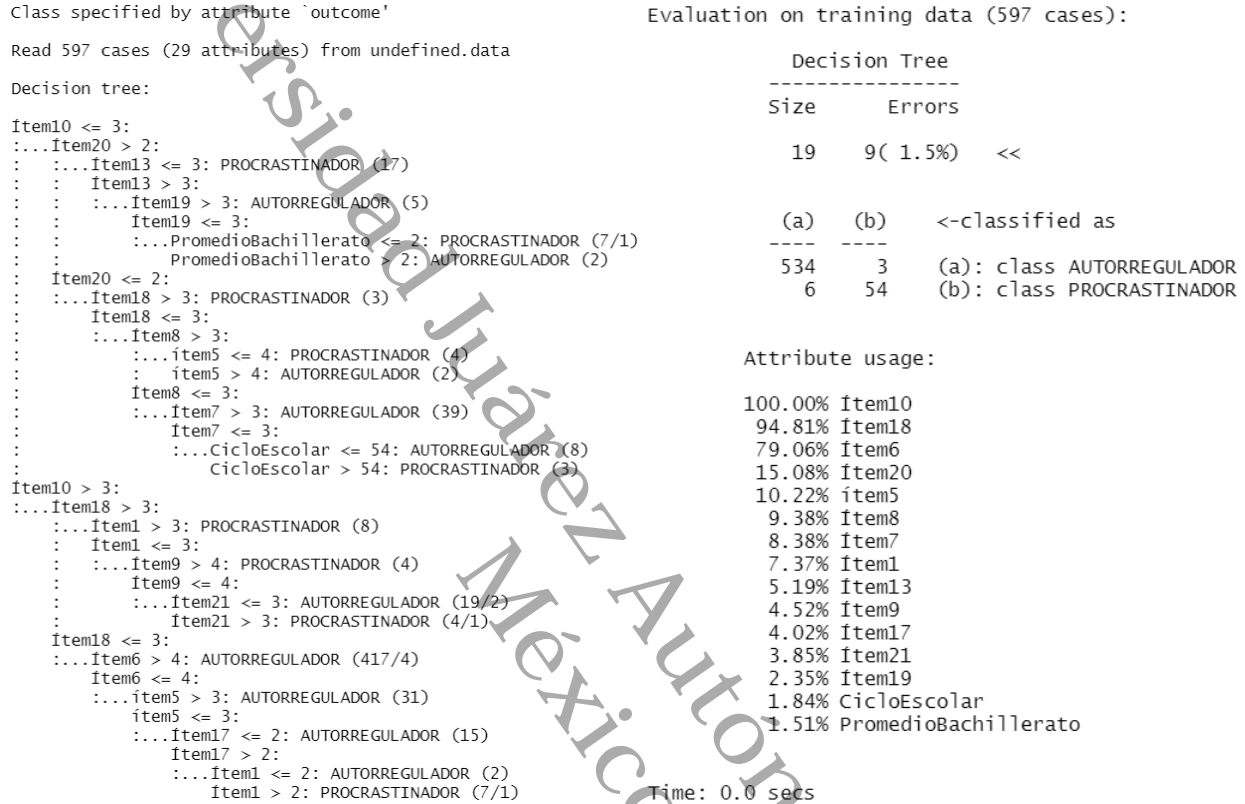
R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/

```
> ##CREACIÓN DEL MODELO DE ARBOL DE DECISIÓN##
> modelo <- C5.0(ClaseEstudiante ~ ., data = datos.entreno) #Arbol de decisión
```

Nota: Elaboración propia.

Después de la creación del modelo de árbol de decisión, se ejecutó el comando `summary()` para visualizar la información del árbol de decisión obtenido, en el cual se aprecia el modelo de árbol de decisión, así como información de los atributos utilizados y clasificación de la variable categórica o clase (ver figura 28).

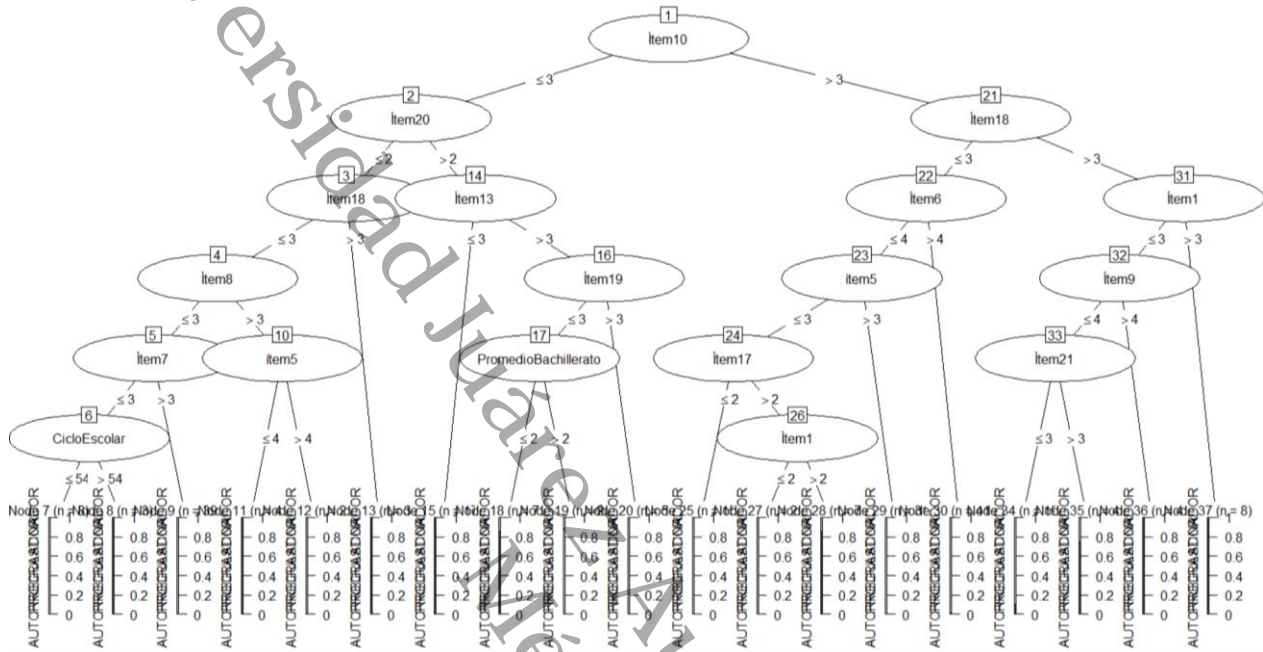
Figura 28.
Modelo e información del Árbol de decisión en RStudio.



Nota: Elaboración propia.

En la figura 29 se muestra el árbol de decisión obtenido de forma gráfica, para que la visualización de la información del árbol de decisión sea más clara (ver figura 29). De igual forma, se puede visualizar un nodo en particular, esto con el código `plot(nombre del modelo, subtree = número del nodo del árbol)` y se subdividirá o extraerá información de ese nodo.

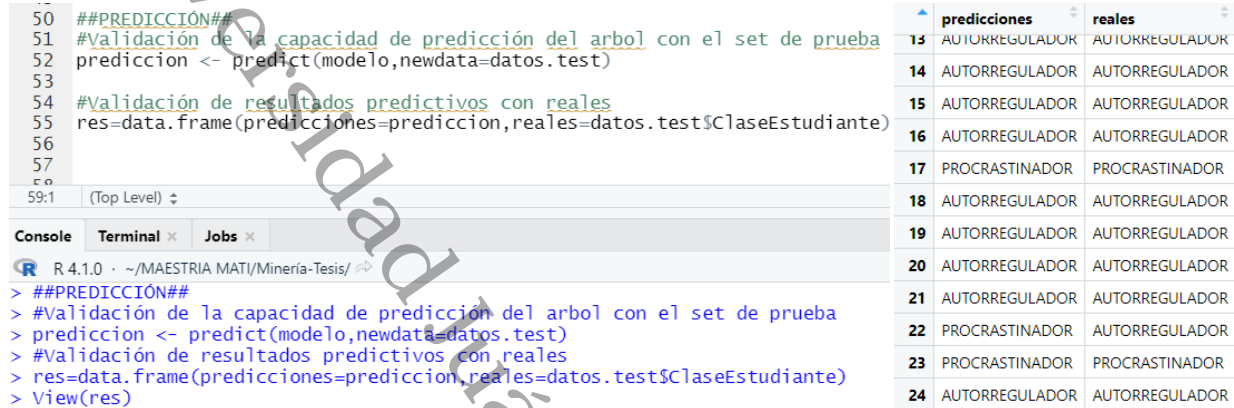
Figura 29.
Árbol de decisión en RStudio.



Nota: Elaboración propia.

La figura 30 se muestra la realización de la predicción realizada al modelo, es decir, verificar si el modelo de árbol de decisión realizó la clasificación de la clase correctamente, esto utilizando el comando o la función *predict()* en el set de prueba, para generar un vector con los valores predichos por el set de entrenamiento. Además de verificar la validación de los resultados predictivos con los reales (ver figura 30).

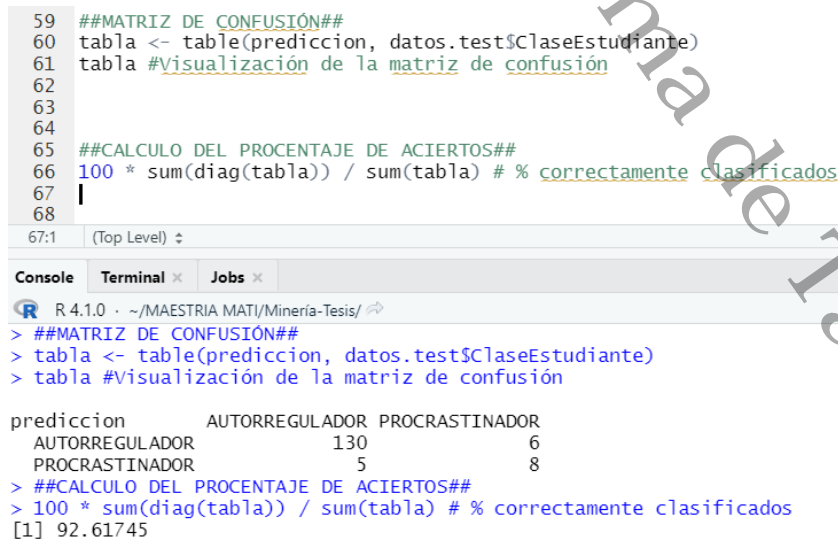
Figura 30.
Verificación de la capacidad de predicción del árbol de decisión en RStudio.



Nota: Elaboración propia.

Por último, se realizó la matriz de confusión del modelo de árbol de decisión, utilizando los datos almacenados en la función predicción y los datos de la clase del set de prueba del modelo, de igual forma se realizó el cálculo de porcentaje de aciertos clasificados correctamente de la clase (ver figura 31).

Figura 31
Matriz de confusión y porcentaje de aciertos de árbol de decisión en RStudio.



Nota: Elaboración propia.

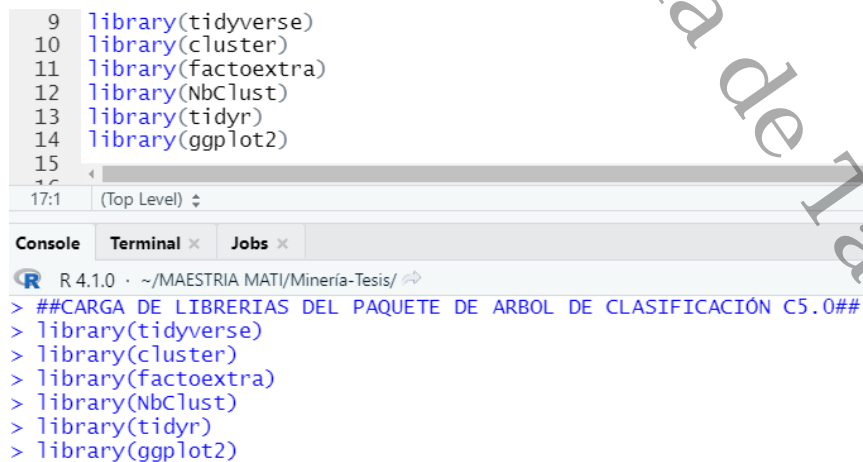
Una vez implementado el algoritmo de árbol de decisión a los datos, se procedió a realizar la minería de datos con el algoritmo *K-means* de la técnica de agrupamiento o conocida también como *clustering*.

De igual forma como en la realización de los árboles de decisión se realizó la lectura del *dataset*, visualización de los datos y detalle o resumen de cada una de las variables, así como renombrar el *dataset* en *RStudio*. Lo anterior, se realizó de manera correcta y el resumen o detalle de las variables fue lo mismo que se obtuvo anteriormente.

El segundo paso fue instalar y cargar las librerías o paquetes para realizar el modelo de agrupamiento, estas librerías fueron: *tidyverse()* para análisis, *tidyr()* para ordenar los datos, *cluster()* se utilizó para el análisis de datos conglomerados o grupos, *NbCluster()* librería para determinar el mejor número de *cluster* para un implementar en un conjunto de datos, *factoextra()* visualización de los resultados de análisis de datos multivariados y *ggplot2()* para la visualización de datos (ver figura 32).

Figura 32

Librerías utilizadas para la creación del modelo de agrupamiento en RStudio.



```
9 library(tidyverse)
10 library(cluster)
11 library(factoextra)
12 library(NbClust)
13 library(tidyr)
14 library(ggplot2)
15
16
17:1 (Top Level) ↓

Console Terminal x Jobs x
R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/ ↗
> ##CARGA DE LIBRERIAS DEL PAQUETE DE ARBOL DE CLASIFICACIÓN C5.0##
> library(tidyverse)
> library(cluster)
> library(factoextra)
> library(NbClust)
> library(tidyr)
> library(ggplot2)
```

Nota: Elaboración propia.

Después de instalar las librerías, se procedió a normalizar o escalar los datos del *dataset*, para que ninguna variable tenga más peso que otras variables, esto utilizando la función o comando *scale()* y solo se normalizan las variables numéricas, en este caso son 28 variables y se utiliza el comando *head()* para visualizar las primeras líneas de las variables escaladas. En la figura 33 se muestra el código completo utilizado para normalizar los datos de las variables y a su vez la visualización de los mismos datos escalados (ver figura 33).

Figura 33
Normalización de los datos para el modelo de agrupamiento en RStudio.

```

17 ##NORMALIZACIÓN DE LOS DATOS##
18 datos.scale <- as.data.frame(scale(datos[1:28])) #Normalizar datos para que ninguna variable tenga más peso que otra
19 head(datos.scale) #Visualizar las primeras filas de datos
20
15:1 (Top Level)

```

```

R 4.1.0 ~ /MAESTRIA MATI/Mineria-Tesis/
> ##NORMALIZACIÓN DE LOS DATOS##
> datos.scale <- as.data.frame(scale(datos[1:28])) #Normalizar datos para que ninguna variable tenga más peso que otra
> head(datos.scale) #Visualizar las primeras filas de datos

```

	Género	Edad	CicloEscolar	DivisiónAcadémica	AsignaturaCursa	PromedioBachillerato	Item1	Item2
1	0.7635977	-0.06199767	0.05131516	-1.5473029	-1.0883183	0.5051779	-1.7097773	-1.12727479
2	0.7635977	-0.06199767	-1.24022036	-1.1647083	+0.3509110	0.5051779	-0.5523418	1.28577254
3	0.7635977	-0.20563201	-0.72360615	-0.7821137	-0.3509110	-0.9612140	0.6050936	-1.12727479
4	0.7635977	-0.34926636	0.98122074	-1.5473029	0.3864963	-0.9612140	-1.7097773	0.07924887
5	0.7635977	3.52886096	0.20629943	-0.3995190	-1.0883183	1.9715697	0.6050936	0.07924887
6	-1.3078346	-0.20563201	0.98122074	-0.7821137	1.1239036	-0.9612140	-0.5523418	0.07924887

	Item3	Item4	Item5	Item6	Item7	Item8	Item9	Item10	Item11	Item12
1	-2.14093517	1.79648625	1.08925536	0.3291871	-0.2798569	-1.4833980	-1.4208191	0.9491565	-0.8220948	-1.6472687
2	2.07314830	-1.83543446	0.09829865	0.3291871	0.9410393	-1.4833980	-1.4208191	0.9491565	0.2889246	0.9235732
3	-0.03389343	-0.01947411	-1.88361476	0.3291871	-0.2798569	-1.4833980	-1.4208191	-0.2780025	0.2889246	-1.6472687
4	1.01962743	-0.01947411	-0.89265805	0.3291871	-0.2798569	-0.4962304	-0.4519607	-0.2780025	-0.8220948	-1.6472687
5	-1.08741430	-0.92745428	1.08925536	0.3291871	0.9410393	-0.4962304	-0.4519607	0.9491565	1.3999441	0.9235732
6	-0.03389343	-0.01947411	-0.89265805	0.3291871	-0.2798569	0.4909373	0.5168976	-1.5051615	0.2889246	0.9235732

	Item13	Item14	Item15	Item16	Item17	Item18	Item19	Item20	Item21	Item22
1	-0.08783044	0.9377425	-1.4413026	-1.1536621	-0.578605	-0.925024259	-0.99475504	-0.7683105	-0.7504897	1.0549802
2	1.10346976	0.9377425	-0.4084383	1.4256544	-0.578605	-0.925024259	-0.99475504	-0.7683105	-0.7504897	-0.5863446
3	-0.08783044	-1.2621188	-0.4084383	-1.1536621	-0.578605	0.922547621	1.02453002	-0.7683105	-0.7504897	0.2343178
4	-0.08783044	-0.1621882	-0.4084383	-0.2938900	-1.402344	-0.925024259	-0.99475504	-0.7683105	-0.7504897	-0.5863446
5	1.10346976	-0.1621882	-0.4084383	-1.1536621	-0.578605	-0.001238319	0.01488749	-0.7683105	-0.7504897	1.0549802
6	-0.08783044	-0.1621882	0.6244260	0.5658822	-0.578605	-0.001238319	0.01488749	0.1472800	1.3463841	0.2343178

Nota: Elaboración propia.

La figura 34 muestra la creación y cálculo de la matriz de distancia de cada uno de los datos escalados, utilizando el método euclídea, que es una de las más utilizadas y esta distancia se calcula entre dos puntos y define la longitud del segmento que une a ambos puntos (ver figura 34).

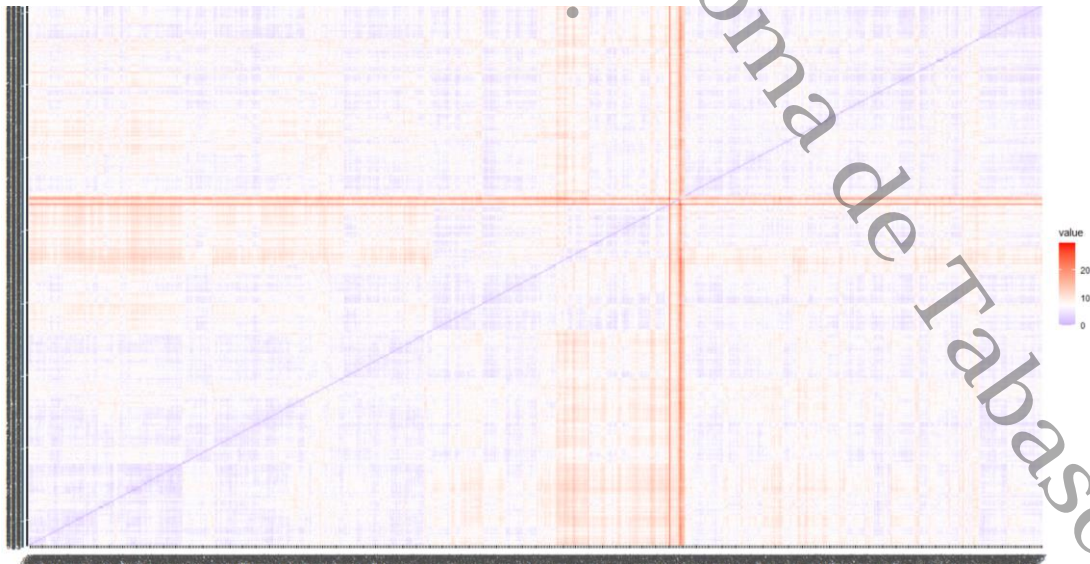
Figura 34
Cálculo de la matriz de distancia utilizando el método euclídeo en RStudio.

```
22 ##CALCULO DE LA MATRIZ DE DISTANCIAS##  
23 m.distancia <- get_dist(datos.scale,method = "euclidean") #El metodo euclidean es el más utilizado  
24 fviz_dist(m.distancia, gradient = list(low = "blue", mid = "white", high = "red")) #Grafico de la matriz  
25  
26  
27  
27:1 (Top Level) R Scri  
Console Terminal Jobs  
R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/  
> ##CALCULO DE LA MATRIZ DE DISTANCIAS##  
> m.distancia <- get_dist(datos.scale,method = "euclidean") #El metodo euclidean es el más utilizado  
> fviz_dist(m.distancia, gradient = list(low = "blue", mid = "white", high = "red")) #Grafico de la matriz de  
distancia
```

Nota: Elaboración propia.

Aunado a lo anterior, la figura 35 se muestra la matriz de distancia obtenida por el método euclídeo, donde las distancias bajas se muestran en azul, las distancias medias en blanco y las distancias más lejanas o altas de color rojo (ver figura 35); al ser un *dataset* con varias variables calculando la distancia entre los datos se dificulta apreciar de forma más clara las distancias.

Figura 35
Matriz de distancias de los datos en RStudio.

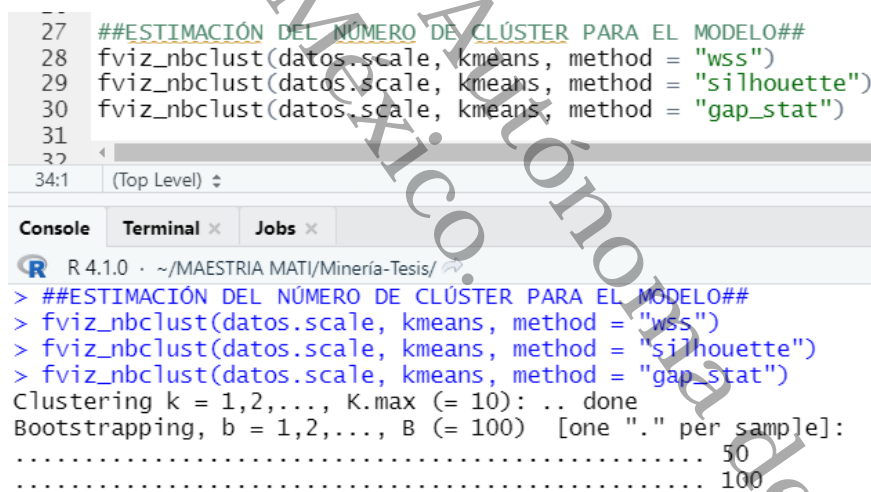


Nota: Elaboración propia.

Subsecuentemente se realizó el cálculo o estimación de números de *cluster* óptimos o favorables para implementarse en el modelo de agrupamiento, por lo cual se realizó la estimación por tres métodos, el método *wss* o método del codo, el cual se tiene que encontrar en el gráfico el punto de corte como la flexión del codo, el segundo método el *silhouette*, donde se maximiza la media de los coeficientes de los datos y por último el método *gap_statistic* en el cual se realiza una comparación entre los valores y varianza de los mismos. La figura 36 se muestra las líneas de código y la ejecución de éstas para encontrar el número de *cluster* óptimo a utilizar en el modelo (ver figura 36).

Figura 36

Estimación de número de cluster para el modelo de agrupamiento en RStudio.



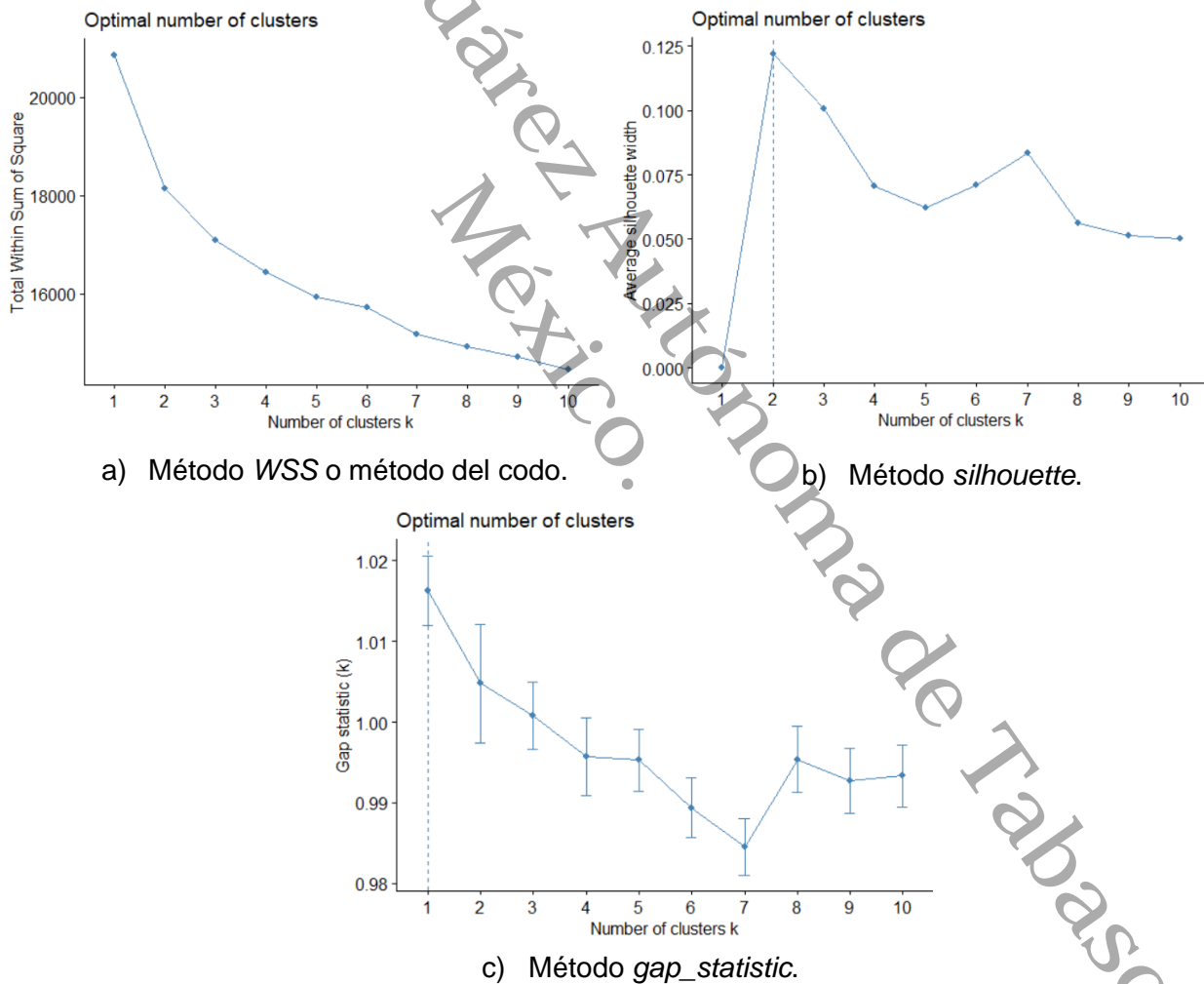
```
--
27 ##ESTIMACIÓN DEL NÚMERO DE CLÚSTER PARA EL MODELO##
28 fviz_nbclust(datos.scale, kmeans, method = "wss")
29 fviz_nbclust(datos.scale, kmeans, method = "silhouette")
30 fviz_nbclust(datos.scale, kmeans, method = "gap_stat")
31
32
33
34:1 (Top Level) ↕
Console Terminal x Jobs x
R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/
> ##ESTIMACIÓN DEL NÚMERO DE CLÚSTER PARA EL MODELO##
> fviz_nbclust(datos.scale, kmeans, method = "wss")
> fviz_nbclust(datos.scale, kmeans, method = "silhouette")
> fviz_nbclust(datos.scale, kmeans, method = "gap_stat")
Clustering k = 1,2,..., K.max (= 10): .. done
Bootstrapping, b = 1,2,..., B (= 100) [one "." per sample]:
..... 50
..... 100
```

Nota: Elaboración propia.

En la figura 37 se muestran los gráficos obtenidos a partir de los tres métodos utilizados para la estimación de los *clústeres* óptimos implementados en el modelo; como se aprecia en el gráfico del método *wss* o del codo, la “flexión del codo” se realiza más notoriamente en el número dos, por lo tanto, el número de *clústeres* óptimos estimados

por este método a implementar en el *dataset* son dos *clústeres*. De igual manera con el método *silhouette* indica que el número óptimo de *clúster* a implementar es de dos *clústeres*, y por último el método *gap_statistic* indica que debe ser uno, pero al ser dos métodos que indican que deberían ser dos *clústeres*, se decidió implementar dos *clústeres* en el modelo de agrupamiento (ver figura 37).

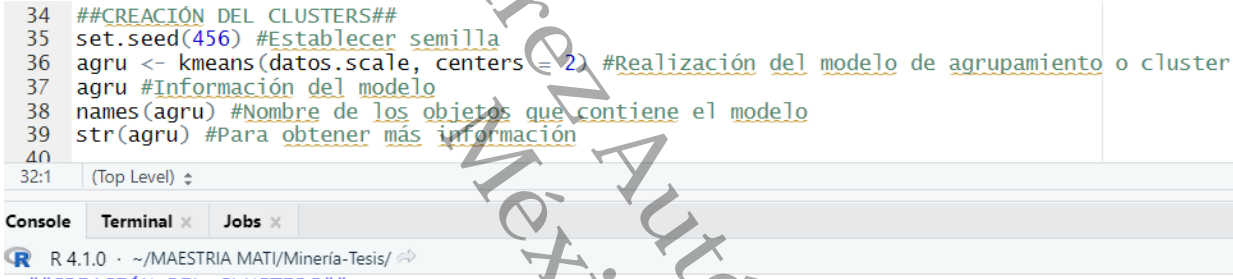
Figura 37
Estimación del número óptimo de *clústeres* a implementar en los datos en RStudio.



Nota: Elaboración propia.

Después de determinar el número óptimo de *clúster* a implementar en el *dataset*, se realizó la creación del modelo de agrupamiento, utilizando la función *kmeans*, los datos escalados o normalizados con anterioridad y especificando que serán dos *clústeres* a implementar en los datos. La figura 38 se muestra el código utilizado para creación del modelo de agrupamiento, así mismo se muestran líneas de código de cómo visualizar los resultados obtenidos por los *clustering* (ver figura 38).

Figura 38
Creación del modelo de agrupamiento en RStudio.



```
34 ##CREACIÓN DEL CLUSTERS##
35 set.seed(456) #Establecer semilla
36 agru <- kmeans(datos.scale, centers = 2) #Realización del modelo de agrupamiento o cluster
37 agru #Información del modelo
38 names(agru) #Nombre de los objetos que contiene el modelo
39 str(agru) #Para obtener más información
40
```

32:1 (Top Level) ↕

Console Terminal x Jobs x

R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/ ↗

```
> ##CREACIÓN DEL CLUSTERS##
> set.seed(456) #Establecer semilla
> agru <- kmeans(datos.scale, centers = 2) #Realización del modelo de agrupamiento o cluster
> agru #Información del modelo
```

Nota: Elaboración propia.

En la figura 39 se muestra el resultado obtenido del modelo de agrupamiento, como lo son los vectores de *cluster*, los nombres de los objetos que contiene el modelo y visualizar el resultado obtenido de ello, esto utilizando la función *str()* (ver figura 39).

Figura 41.
Representación gráfica del modelo de agrupamiento en RStudio.



Nota: Elaboración propia.

Por último se realizó una inspección o revisión de los datos, utilizando dos variables del *dataset* y el modelo de agrupamiento, para visualizar de forma gráfica cómo han sido agrupados los datos en función de los dos *cluster* creados en el modelo, además se utilizó la función *aggregate()* para agrupar las variables y agruparlas de acuerdo a los *cluster* que han sido asignadas y generar la media de los datos.

En la figura 42 se visualiza el código utilizado para realizar lo anteriormente descrito y las variables utilizadas del *dataset* solo se utilizan como ejemplo para visualizar el gráfico, pero se puede utilizar cualquier variable del *dataset* para realizar esta revisión de los datos en función del modelo de agrupamiento, así como el cálculo de las medianas de los datos del *dataset* sin normalizar ordenas de acuerdo con los *cluster* (ver figura 42).

Figura 42
Revisión de los datos en función y medias en función de los cluster del modelo de agrupamiento en RStudio.

```

56 ##INSPECCIONAR LOS RESULTADOS##
57 plot(datos$DivisiónAcadémica, datos$Item16, col=agru$cluster ,xlab = "grupo1", ylab = "grupo2")
58 aggregate(datos[,1:28] ,by = list(agru$cluster),mean)
59 text(datos$DivisiónAcadémica, datos$Item16, labels = datos$ClaseEstudiante, pos = 1, col = agru$cluster,cex =
60
61 ggplot(datos, aes(x = datos$DivisiónAcadémica, y = datos$Item16)) + geom_point()
62

```

```

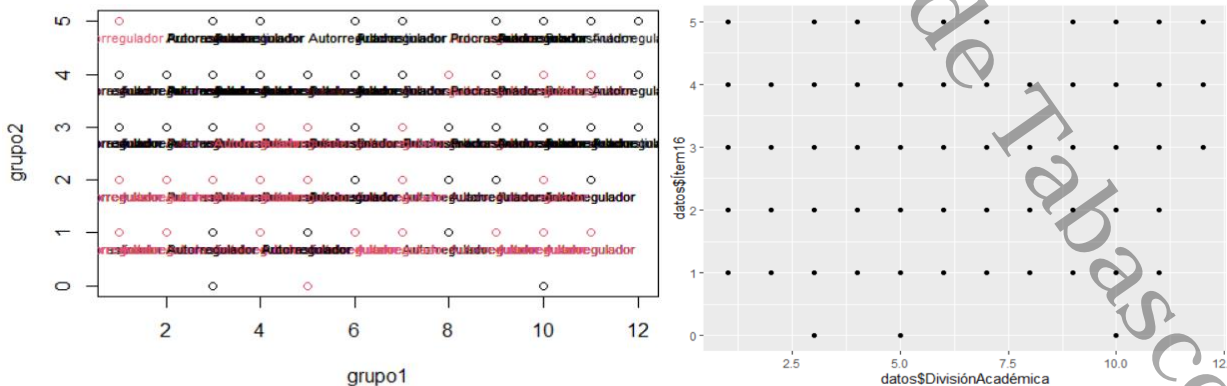
R 4.1.0 · ~/MAESTRIA MATI/Mineria-Tesis/
> ##INSPECCIONAR LOS RESULTADOS##
> plot(datos$DivisiónAcadémica, datos$Item16, col=agru$cluster ,xlab = "grupo1", ylab = "grupo2")
> aggregate(datos[,1:28] ,by = list(agru$cluster),mean)
  Group.1 Género  Edad  CicloEscolar DivisiónAcadémica AsignaturaCursa PromedioBachillerato  Item1
1      1      1 1.556757 20.95946      42.42973      5.508108      2.429730      1.759459 2.864865
2      2      1 1.704787 21.89628      43.57447      4.587766      2.521277      1.553191 2.095745
  Item2  Item3  Item4  Item5  Item6  Item7  Item8  Item9  Item10  Item11  Item12  Item13
1 3.594595 3.132432 3.064865 3.521622 4.659459 3.951351 2.997297 2.981081 3.956757 3.391892 3.951351 3.724324
2 4.268617 2.933511 2.978723 4.273936 4.922872 4.502660 2.015957 1.960106 4.492021 4.082447 4.606383 4.417553
  Item14  Item15  Item16  Item17  Item18  Item19  Item20  Item21  Item22
1 3.813514 2.900000 2.883784 3.162162 2.518919 2.527027 2.340541 2.191892 3.081081
2 4.476064 1.898936 1.808511 2.250000 1.492021 1.452128 1.345745 1.247340 2.353723
> text(datos$DivisiónAcadémica, datos$Item16, labels = datos$ClaseEstudiante, pos = 1, col = agru$cluster,cex = .7)
> ggplot(datos, aes(x = datos$DivisiónAcadémica, y = datos$Item16)) + geom_point()

```

Nota: Elaboración propia.

La figura 43 muestra dos formas de gráficos para visualizar la revisión de agrupamiento de los datos de dos variables en función de los *cluster* creados con la técnica de agrupamiento de minería de datos, esta revisión se puede realizar con cada una de los variables que conforman el *dataset* (ver figura 43).

Figura 43
Ejemplo de revisión de datos en función al modelo de agrupamiento en RStudio.



Nota: Elaboración propia.

Por último se realizó la asignación de los *clústeres* al *dataset*, para poder visualizarlos en conjunto con las variables y datos que contiene el *dataset*, además se realizó una matriz de confusión o tabla de frecuencia con el fin de visualizar la distribución y agrupamiento de los 746 registros que conforman el *dataset* en cada grupo de *cluster* implementados, de acuerdo con la variable clasificatoria, en este caso es la variable ClaseEstudiante, la cual contiene el dato de la clase a la cual pertenece cada registro.

En la figura 44 se muestra lo anteriormente descrito y a su vez el agrupamiento de los registros, donde se agruparon registros de la clase autorregulador en los dos *clústeres* y de la clase procrastinador solo en un *clúster*, ya que el propósito de la técnica de agrupamiento es agrupar los datos o registros en grupos que compartan similitudes o semejanzas entre los datos del mismo grupo y en este caso hay similitudes entre los registros de ambas clases, esto se visualiza en el cluster uno donde el algoritmo agrupó registros de ambas clases (ver figura 44).

Figura 44.
Visualización de la asignación de los cluster al dataset del modelo de agrupamiento en RStudio.

```
65 ##ASIGNACIÓN DE LOS CLUSTERES AL DATASET##
66 tabla <- data.frame (datos) %>%
67   mutate(Cluster = agru$cluster)
68 #tabla #Visualización de la tabla
69
70
71 ##MATRIZ DE CONFUSIÓN DE LOS AGRUPAMIENTOS##
72 cluster=agru$cluster
73 table(tabla$ClaseEstudiante,cluster)
74
```



```
70:1 | (Top Level) ⌵
Console Terminal Jobs
R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/
> ##ASIGNACIÓN DE LOS CLUSTERES AL DATASET##
> tabla <- data.frame (datos) %>%
+   mutate(Cluster = agru$cluster)
> ##MATRIZ DE CONFUSIÓN DE LOS AGRUPAMIENTOS##
> cluster=agru$cluster
> table(tabla$ClaseEstudiante,cluster)
      cluster
      1     2
Autorregulador 296 376
Procrastinador  74   0
```

Nota: Elaboración propia.

La última técnica de minería de datos implementada al *dataset* fue la de reglas de asociación con el algoritmo A priori. Lo primero que se realizó fue la lectura y visualización del *dataset* en RStudio, utilizando las mismas funciones o comandos utilizados anteriormente, de igual forma el tipo de dato se modificó desde el archivo del *dataset* para sea de tipo nominal o carácter y no numérico como anteriormente se había utilizado con las otras dos técnicas de minería.

En la figura 45 se visualizan las características o resumen de las variables que conforman el *dataset*, las cuales indican que el tipo de dato, denominado *class* en RStudio son de tipo carácter y otros datos de las variables (ver figura 45).

Figura 45.
Resumen de las variables del dataset con tipo de dato carácter en RStudio.

```
1 ##LECCURA DEL DATASET DE PRUEBA##
2 procras_autorre = read.csv("../Mineria-Tesis/procras_autorre.csv") #Leer dataset
3 View(procras_autorre) #Visualización del dataset
4 summary(procras_autorre) #Resumen de las variables del dataset
5 datos <- procras_autorre #Renombre del dataset
```

7:1 (Top Level) ↓

Console Terminal Jobs

R 4.1.0 - ~/MAESTRIA MATI/Mineria-Tesis/

```
> ##LECCURA DEL DATASET DE PRUEBA##
> procras_autorre = read.csv("../Mineria-Tesis/procras_autorre.csv") #Leer dataset
> View(procras_autorre) #Visualización del dataset
> summary(procras_autorre) #Resumen de las variables del dataset
  Género      Edad      CicloEscolar      DivisiónAcadémica      AsignaturaCursa      PromedioBachillerato
Length:746    Length:746    Length:746    Length:746    Length:746    Length:746
Class :character  Class :character  Class :character  Class :character  Class :character  Class :character
Mode :character  Mode :character  Mode :character  Mode :character  Mode :character  Mode :character
 Item1      Item2      Item3      Item4      Item5      Item6
Length:746  Length:746  Length:746  Length:746  Length:746  Length:746
Class :character  Class :character  Class :character  Class :character  Class :character  Class :character
Mode :character  Mode :character  Mode :character  Mode :character  Mode :character  Mode :character
 Item7      Item8      Item9      Item10     Item11     Item12
Length:746  Length:746  Length:746  Length:746  Length:746  Length:746
Class :character  Class :character  Class :character  Class :character  Class :character  Class :character
Mode :character  Mode :character  Mode :character  Mode :character  Mode :character  Mode :character
 Item13     Item14     Item15     Item16     Item17     Item18
Length:746  Length:746  Length:746  Length:746  Length:746  Length:746
Class :character  Class :character  Class :character  Class :character  Class :character  Class :character
Mode :character  Mode :character  Mode :character  Mode :character  Mode :character  Mode :character
 Item19     Item20     Item21     Item22
Length:746  Length:746  Length:746  Length:746
Class :character  Class :character  Class :character  Class :character
Mode :character  Mode :character  Mode :character  Mode :character
> datos <- procras_autorre #Renombre del dataset
```

Nota: Elaboración propia.

El segundo paso fue cargar la librería de *arules*, para la identificación de conjunto de elementos frecuentes y creación de reglas de asociación, a través del algoritmo A priori. Para la obtención de las reglas de asociación con el algoritmo antes mencionado, fue necesario trabajar con la función *apriori()*, la cual contempla varios argumentos que el usuario puede utilizar.

Los argumentos que se utilizaron de esta función fueron: *data*, *parameter*, del cual se utilizaron los parámetros soporte igual a 0.1, confianza igual a 0.8 (estos valores son por defecto en el algoritmo, pero se pueden modificar los valores) y *target* con el cual indicamos qué tipo de resultado que debe de generar el algoritmo es *rules* (reglas).

La figura 46 muestra la carga de la librería *arules()* y la realización del modelo de reglas de asociación, además muestra algunos parámetros e información obtenidos de la creación del modelo (ver figura 46).

Figura 46
Creación del modelo *A priori* de reglas de asociación en RStudio.

```
9 ##CARGA DE LIBRERIAS DEL PAQUETE ARULES##
10 library(arules)
11
12
13
14 ##REGLAS DE ASOCIACIÓN##
15 reglas <- apriori(data = datos,
16                   parameter = list(support = 0.1,
17                                   confidence = 0.8,
18                                   target = "rules")) #Especificación que se creen reglas
19
15:32 (Top Level) ↓
```

Console Terminal Jobs

```
R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/
> ##CARGA DE LIBRERIAS DEL PAQUETE ARULES##
> library(arules)
> ##REGLAS DE ASOCIACIÓN##
> reglas <- apriori(data = datos,
+                   parameter = list(support = 0.1,
+                                   confidence = 0.8,
+                                   target = "rules")) #Especificación que se creen reglas
Apriori

Parameter specification:
confidence minval smax arem aval originalSupport maxtime support minlen maxlen target ext
           0.8   0.1   1 none FALSE              TRUE     5   0.1     1    10 rules TRUE

Algorithmic control:
filter tree heap memopt load sort verbose
  0.1 TRUE TRUE  FALSE TRUE   2    TRUE

Absolute minimum support count: 74

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[244 item(s), 746 transaction(s)] done [0.00s].
sorting and recoding items ... [89 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 5 6 7 8 done [0.02s].
writing ... [13215 rule(s)] done [0.01s].
creating S4 object ... done [0.02s].
```

Nota: Elaboración propia.

La figura 47 muestra la ejecución del comando *summary()*, para visualizar información del modelo, con el número de las reglas de asociación identificadas del *dataset* y resumen de las medidas que cuantifica la calidad de las reglas como: *lift*, *support*, *confidence*, *count* y *coverage* (ver figura 47).

Figura 47
 Información del modelo de reglas de asociación en RStudio.

```

21 summary(reglas) #Visualización de la información de modelo
22
23:1 (Top Level) ↓

```

Console	Terminal	Jobs
<pre> R 4.1.0 · ~/MAESTRIA MATI/Minería-Tesis/ ↵ > summary(reglas) #Visualización de la información de modelo set of 13215 rules rule length distribution (lhs + rhs):sizes 1 2 3 4 5 6 7 8 2 168 2187 4412 3904 2044 486 12 Min. 1st Qu. Median Mean 3rd Qu. Max. 1.000 4.000 4.000 4.527 5.000 8.000 summary of quality measures: support confidence coverage lift count Min. :0.1005 Min. :0.8000 Min. :0.1005 Min. :0.8881 Min. : 75.00 1st Qu.:0.1059 1st Qu.:0.8652 1st Qu.:0.1166 1st Qu.:1.0849 1st Qu.: 79.00 Median :0.1180 Median :0.9259 Median :0.1287 Median :1.1176 Median : 88.00 Mean :0.1306 Mean :0.9191 Mean :0.1428 Mean :1.3996 Mean : 97.46 3rd Qu.:0.1381 3rd Qu.:0.9802 3rd Qu.:0.1528 3rd Qu.:1.6955 3rd Qu.:103.00 Max. :0.9008 Max. :1.0000 Max. :1.0000 Max. :4.6548 Max. :672.00 mining info: data ntransactions support confidence datos 746 0.1 0.8 </pre>		

Nota: Elaboración propia.

Para visualizar las reglas de asociación identificadas con el algoritmo A priori, se utilizó la función inspect(), que pertenece a la librería de arules, esta función permite mostrar asociaciones y transacciones más información adicional sobre las ellas. La utilización de esta función se muestra en la figura 48, donde se visualizan todas las reglas de asociación obtenidas por el modelo, pero se indica que éstas sean ordenadas en forma descendente, de acuerdo al número de confianza de las reglas (ver figura 48).

Figura 48
Visualización de las todas las reglas de asociación identificadas por el modelo en RStudio.

```
23 #Visualización de las todas las reglas de asociación obtenidas del modelo ordenas por orden descendente de confianza
24 inspect(sort(x = reglas, decreasing = TRUE, by = "confidence"))
25
27:1 | (Top Level) | R Script
```

Console	Terminal	Jobs	Spellcheck						
[13207]	{ClaseEstudiante=Autorregulador}	=>	{Item18=Nunca}	0.1394102	0.8000000	0.1742627	1.9631579	104	
[13208]	{Item5=Siempre, Item6=Siempre, Item7=Siempre, Item12=Siempre, ClaseEstudiante=Autorregulador}	=>	{Item10=Siempre}	0.1179625	0.8000000	0.1474531	1.9439739	88	
[13209]	{Item13=Siempre, Item14=Siempre, Item19=Nunca, Item20=Nunca, ClaseEstudiante=Autorregulador}	=>	{Item10=Siempre}	0.1018767	0.8000000	0.1273458	1.9439739	76	
[13210]	{Item12=Siempre, Item14=Siempre, Item19=Nunca, Item21=Nunca, ClaseEstudiante=Autorregulador}	=>	{Item13=Siempre}	0.1072386	0.8000000	0.1340483	2.3588933	80	
[13211]	{Item6=Siempre, Item13=Siempre, Item19=Nunca, Item20=Nunca, ClaseEstudiante=Autorregulador}	=>	{Item12=Siempre}	0.1179625	0.8000000	0.1474531	1.8030211	88	
	{Item6=Siempre, Item13=Siempre, Item18=Nunca, Item21=Nunca, ClaseEstudiante=Autorregulador}	=>	{Item14=Siempre}	0.1179625	0.8000000	0.1474531	1.9631579	88	

Nota: Elaboración propia.

Aunado a lo anterior, se filtraron o presentaron las reglas de asociación con mayor confianza y ordenadas por orden descendente, máximo 20 reglas de asociación. En la figura 49 se muestra la ejecución y visualización de estas reglas, que las podemos definir como las asociaciones más relevantes de los datos (ver figura 49).

Figura 49
Reglas con mayor confianza de las reglas de asociación en RStudio.

```

27 #Filtrado de las 20 reglas de asociación con mayor confianza
28 filtrado_reglas <- sort(reglas, decreasing = TRUE, by = "confidence")[1:20]
29 inspect(filtrado_reglas) #Visualización de las reglas
30:1 (Top Level)

```

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Item1=Nunca,Item7=Siempre}	=> {Item6=Siempre}	0.1058981	1	0.1058981	1.172956	79
[2]	{Item8=Nunca,Item13=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1166220	1	0.1166220	1.110119	87
[3]	{Item7=Siempre,Item8=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1193029	1	0.1193029	1.110119	89
[4]	{Item8=Nunca,Item14=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1179625	1	0.1179625	1.110119	88
[5]	{Item8=Nunca,Item10=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1072386	1	0.1072386	1.110119	80
[6]	{Item8=Nunca,Item12=Siempre}	=> {Item6=Siempre}	0.1139410	1	0.1139410	1.172956	85
[7]	{Item8=Nunca,Item12=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1139410	1	0.1139410	1.110119	85
[8]	{Item6=Siempre,Item8=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1729223	1	0.1729223	1.110119	129
[9]	{Item10=Siempre,Item17=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1058981	1	0.1058981	1.110119	79
[10]	{Item12=Siempre,Item17=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1085791	1	0.1085791	1.110119	81
[11]	{Item6=Siempre,Item17=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1595174	1	0.1595174	1.110119	119
[12]	{DivisiónAcadémica=DACB,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1045576	1	0.1045576	1.110119	78
[13]	{Item7=Siempre,Item15=Nunca}	=> {Item6=Siempre}	0.1206434	1	0.1206434	1.172956	90
[14]	{Item5=Siempre,Item9=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1018767	1	0.1018767	1.110119	76
[15]	{Item9=Nunca,Item13=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1246649	1	0.1246649	1.110119	93
[16]	{Item7=Siempre,Item9=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1246649	1	0.1246649	1.110119	93
[17]	{Item9=Nunca,Item14=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1260054	1	0.1260054	1.110119	94
[18]	{Item9=Nunca,Item10=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1206434	1	0.1206434	1.110119	90
[19]	{Item9=Nunca,Item12=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1286863	1	0.1286863	1.110119	96
[20]	{Item6=Siempre,Item9=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1836461	1	0.1836461	1.110119	137

Nota: Elaboración propia.

3.3.7.2 Microsoft Azure Machine Learning

Microsoft Azure Machine Learning es el servicio que Microsoft Azure ofrece para implementar algoritmos de minería de datos, pero esta misma herramienta tiene otra instancia para realizar los esquemas y modelos de los algoritmos para ejecutar en el dataset, esta instancia es Azure Machine Learning Studio.

Lo primero que se realizó para realizar minería de datos con esta herramienta, fue crear una instancia de proceso para utilizar paquetes de Machine Learning (ML), herramientas como JupyterLab, Jupyter y RStudio, esto dando clic en la opción proceso. En la figura 50 se muestra la instancia para la creación de la instancia de proceso, donde se configuran y seleccionan los valores pertinentes para la creación de esta instancia, uno de los parámetros o valores a seleccionar es la instancia de trabajo con la que

trabajar, como ya se había mencionado anteriormente la instancia de trabajo seleccionada fue la instancia DS11_v2 (ver figura 50).

Figura 50
Creación de instancia de proceso para trabajar en Azure Machine Learning Studio.

Creación de instancias de proceso

Valores obligatorios
 Configuración avanzada

Configurar los valores obligatorios

Nombre del proceso *

Ubicación

Tipo de máquina virtual CPU GPU

Tamaño de la máquina virtual Seleccionar de entre las opciones recomendadas Seleccionar de entre todas las opciones

Cuota total disponible: 6 núcleos

Nombre ↑	Categoría	Tipos de carga de trabajo	Cuota disponible	Costo
<input checked="" type="radio"/> Standard_DS11_v2 2 núcleos, 14 GB de RAM, 28 GB de almacenamiento	Optimizada para ...	Desarrollo en cuadernos (u otros IDE) y pruebas atenuadas	6 núcleos	0.17 US...
<input type="radio"/> Standard_DS3_v2	De uso general	Entrenamiento de modelos de ML clásico, ejecuciones de	6 núcleos	0.25 US...

Nota: Elaboración propia.

Aunado a los anterior se pueden crear otros procesos, por ejemplo, *clústeres* de proceso, *clústeres* de inferencia y proceso asociados. Después de la creación de la instancia de proceso, se procedió a la creación de un conjunto de datos para acceder de forma fácil y rápida a los datos durante el entrenamiento del modelo, sin preocuparse de las cadenas de conexión o rutas de acceso a los datos.

La creación del conjunto de datos se realizó seleccionando la opción de archivos locales y se presenta un cuadro de diálogo, donde se configura como primer punto la información básica del conjunto de datos, es decir, nombre, tipo de conjunto de datos (puede ser tabular o archivo) y la descripción correspondiente al conjunto de datos. En la

figura 51 se visualiza el cuadro de diálogo para la creación del conjunto de datos y el llenado de los valores o parámetros descritos anteriormente (ver figura 51).

Figura 51
Creación del conjunto de datos en Azure Machine Learning Studio.

Creación de un conjunto de datos a partir de archivos locales

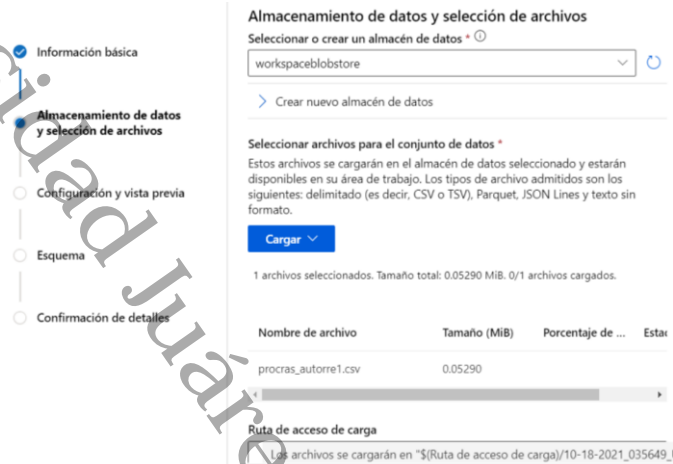
The screenshot shows a multi-step wizard for creating a dataset. The first step, 'Información básica', is active. It includes a sidebar with steps: 'Información básica', 'Almacenamiento de datos y selección de archivos', 'Configuración y vista previa', 'Esquema', and 'Confirmación de detalles'. The main area contains the following fields:

- Nombre ***: A text input field containing 'procras_autorre1'.
- Tipo de conjunto de datos ***: A dropdown menu set to 'Tabular'.
- Descripción**: A text area containing 'dataset con datos sobre procrastinación y autorregulación académica'.

Nota: Elaboración propia.

La figura 52 muestra el almacenamiento de los datos y la selección del archivo, en este caso el archivo seleccionado es de tipo CSV (delimitado por comas), el cual es uno de los tipos de archivo que acepta esta herramienta y el almacenamiento de los datos fue en el almacén de *workspaceblobstore* que es uno de los almacenes que la herramienta da por predeterminado (ver figura 52).

Figura 52
Almacenamiento y selección del archivo para la creación del conjunto de datos en Azure Machine Learning Studio.



Nota: Elaboración propia.

Lo siguiente es realizar la configuración y vista previa, es decir, verificar y visualizar que el archivo seleccionado se haya leído de forma correcta y configurar algunos parámetros de ser necesario (ver figura 53).

Figura 53
Configuración y vista previa del conjunto de datos en Azure Machine Learning Studio .



Nota: Elaboración propia.

De igual forma se puede visualizar el esquema del conjunto de los datos y se muestra el tipo de datos de cada variable. Y por último se visualizó el cuadro de diálogo donde se visualizaron todos los detalles del nuevo conjunto de datos a crear y dar clic en el botón crear (ver figura 54).

Figura 54
Confirmación de los detalles y creación del conjunto de datos en Azure Machine Learning Studio.

Confirmación de detalles

Información básica

Nombre
procras_autorre1

Tipo de conjunto de datos
Tabular

Descripción
dataset con datos sobre procrastinación y autorregulación académica

Almacenamiento de datos y selección ...

Almacén de datos
workspaceblobstore

Archivos seleccionados (1)
procras_autorre1.csv

Ruta de acceso
UI/10-18-2021_035649_UTC/procras_autorre1.csv

Configuración de archivo

Formato de archivo
Delimitada

Delimitador
Coma

Codificación
ISO-8859-1

Encabezados de columna

Atrás Crear

Nota: Elaboración propia.

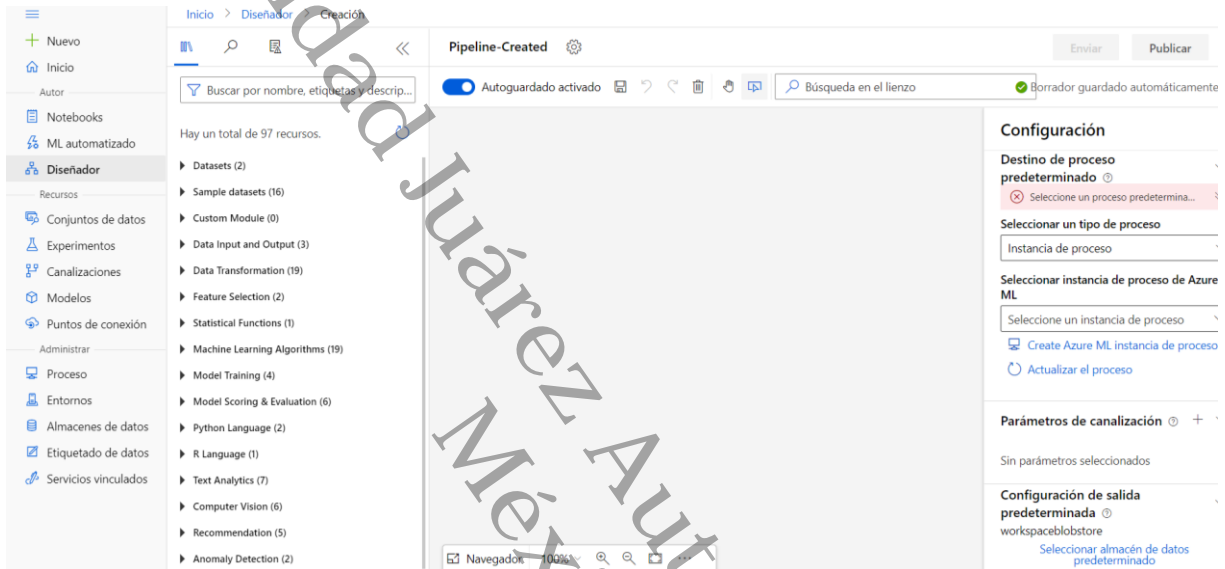
Una vez creado el conjunto de datos, este aparecerá en el área de trabajo Conjunto de datos, donde se indicará el nombre, versión, origen de datos, fecha de creación, fecha de modificación, propiedades (si es tabular o archivo), autor y etiquetas.

Después de crear el proceso y el conjunto de datos, se proceden a realizar los modelos de algoritmos para realizar minería de datos, esto se realiza donde clic en la opción experimentos y se parecerá la ventana para la creación de este, donde escogerá la opción de canalización, la cual nos mandará a la instancia de trabajo de diseñador para

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

crear un modelo. La figura 55 muestra la instancia de trabajo para creación y ejecución de los modelos de algoritmos para realizar minería de datos (ver figura 55).

Figura 55
Instancia de trabajo de Azure Machine Learning Studio.

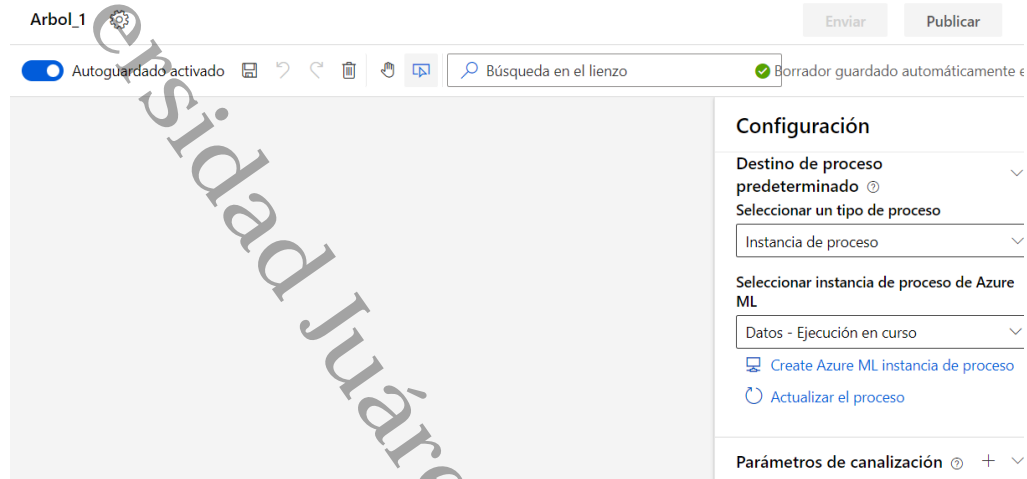


Nota: Elaboración propia.

La primera técnica de minería de datos que se implementará en el *dataset* es la técnica de árbol de decisión, por lo cual se tiene que elegir el destino de proceso, que en este caso es la instancia de proceso creada anteriormente y se selecciona el nombre de esta instancia, a su vez se modifica el nombre del modelo, en este caso el nombre es *Árbol_1* (ver figura 56).

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

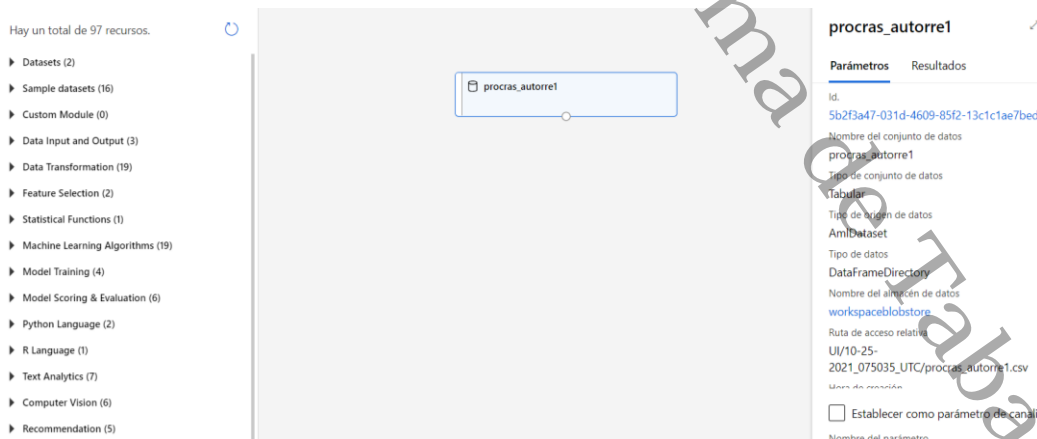
Figura 56
Configuración de la instancia de trabajo en Azure Machine Learning Studio.



Nota: Elaboración propia.

Después de realizar la configuración de la instancia, se procedió a seleccionar el *dataset*, esto dando clic en pestaña *Dataset*, que se encuentra en la parte izquierda de la pantalla y se arrastra el conjunto de datos al área de trabajo (ver figura 57).

Figura 57
Selección del dataset para realizar árbol de decisión en Azure Machine Learning Studio.

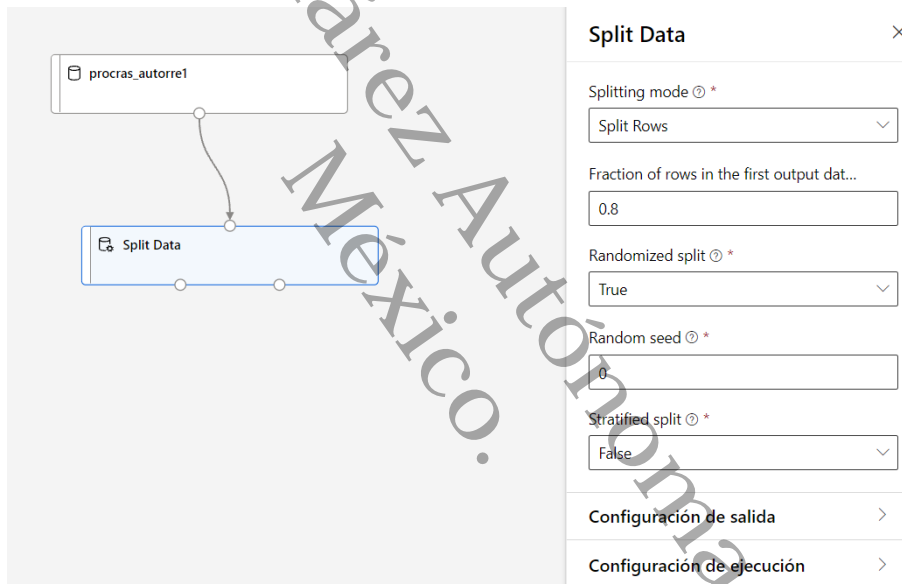


Nota: Elaboración propia.

Lo siguiente a realizar es entrenar los datos, es decir, se dividen los datos en dos *dataset*: uno de entrenamiento y otro de prueba, esto con la opción de *Split data* que se encuentra en la pestaña *Data transformation*. La figura 58 muestra la conexión entre el *dataset* y la división de los datos que crear un *dataset* con 80% de los datos y otro con el 20% de los datos (ver figura 58).

Figura 58

Creación de los sets de entrenamiento y prueba del árbol de decisión en Azure Machine Learning Studio.

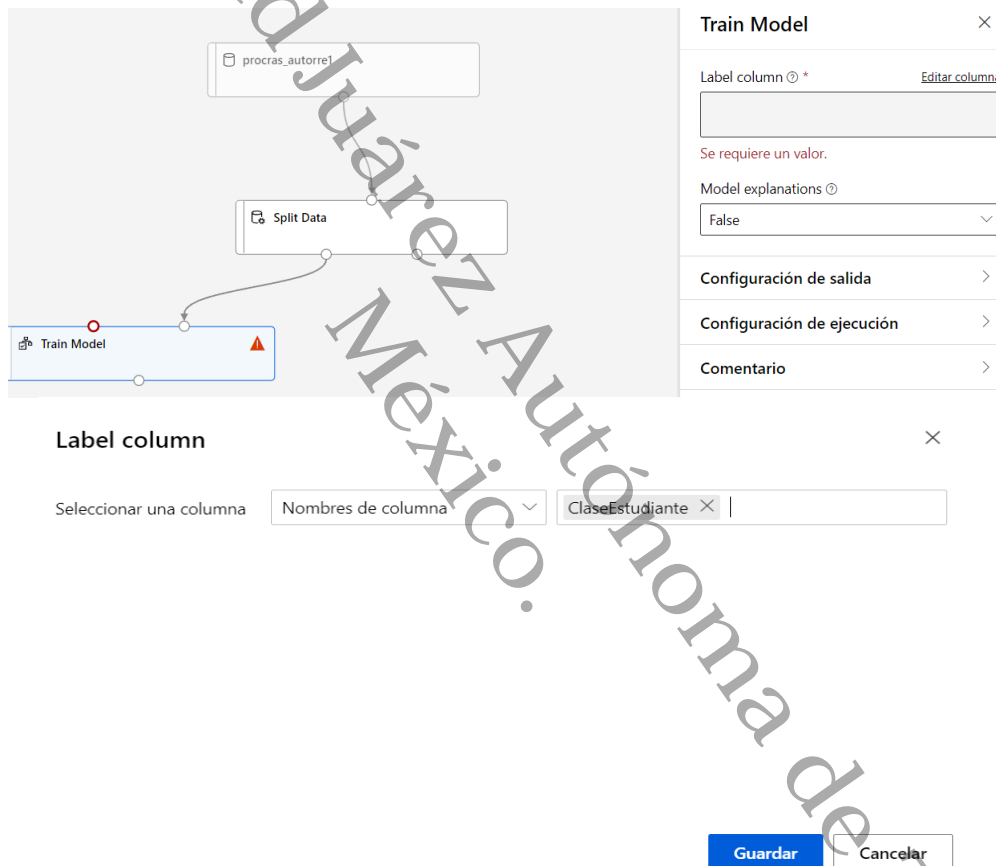


Nota: Elaboración propia.

Se prosigue a realizar el modelo de entrenamiento, es decir, seleccionar la variable que se desea entrenar o clasificar, es este caso es la variable ClaseEstudiante. Para realizar este paso antes se tiene que elegir la opción *train model*, que se encuentra en la pestaña *Model training* y se cual se arrastra al área de trabajo y se conecta con el *Split data*, después en la opción editar columna que aparece en el cuadro de diálogo derecho

de la interfaz de trabajo se seleccionará la columna a entrenar o clasificar. En la figura 59 se presenta la selección de la variable a clasificar y la conexión de los recursos utilizados hasta para realizar el modelo de árbol de decisión (ver figura 59).

Figura 59
Entrenamiento del modelo y selección de la variable a entrenar en el árbol de decisión en Azure Machine Learning Studio.



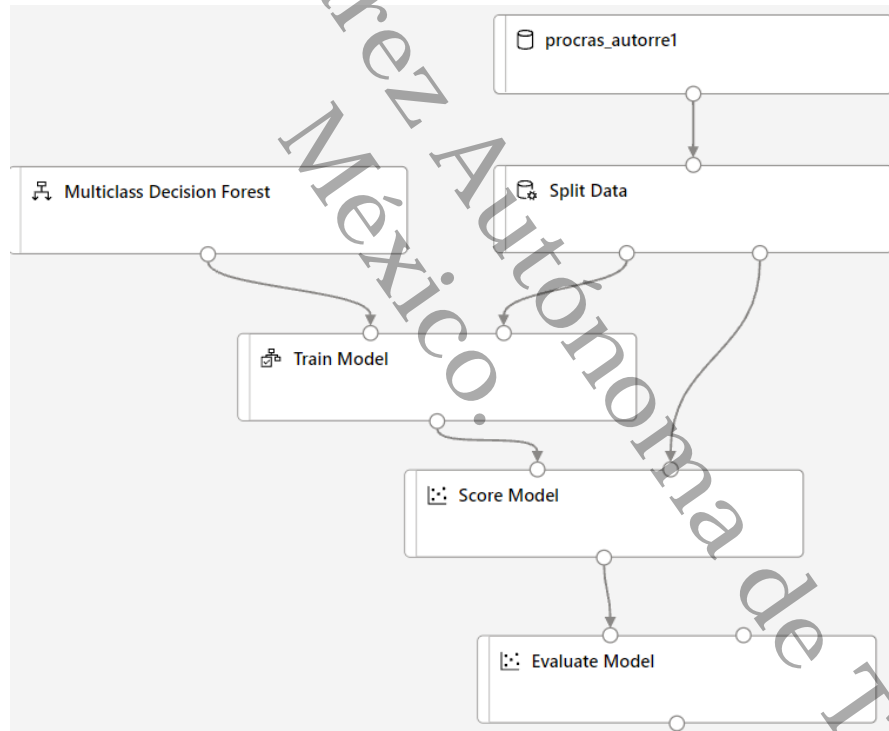
Nota: Elaboración propia.

Aunado a lo anterior, se realizó la selección del algoritmo de clasificación a utilizar, en este caso es el algoritmo *Multiclass Decision Forest*, el cual clasifica un objeto con un

máximo de dos resultados. Este modelo de clasificación se encuentra en la pestaña *Machine Learning Algorithms* en la opción de algoritmos *Classification*.

Por último, se utilizaron los recursos de *Score model* y *Evaluate model* para terminar el modelo de árbol de decisión, estos recursos se encuentran en la pestaña *Model scoring & evaluation*. En la figura 60 se visualiza el modelo de árbol de decisión creado (ver figura 60).

Figura 60
Modelo de árbol de decisión en Azure Machine Learning Studio.



Nota: Elaboración propia.

Una vez terminado el modelo, se procedió a realizar la ejecución del modelo, esto dando clic en el botón enviar, el cual se encuentra en la parte superior de la interfaz de trabajo y aparece un cuadro de diálogo, donde se configurarán el nombre del

experimento. En la figura 61 se muestra el cuadro de diálogo donde se configuró el nombre del experimento y se dio clic en enviar. Finalmente se procedió a realizar la ejecución del modelo de árbol de decisión en el *dataset* (ver figura 61).

Figura 61
Ejecución del modelo de árbol de decisión en Azure Machine Learning Studio.

Configuración de una ejecución de canalización

Experimento

Seleccionar existente Crear nuevo

Experimento existente*

Modelo_arbol

Descripción de la ejecución*

Arbol_1

Continuar con paso de error

Destino de proceso

Predeterminado Datos

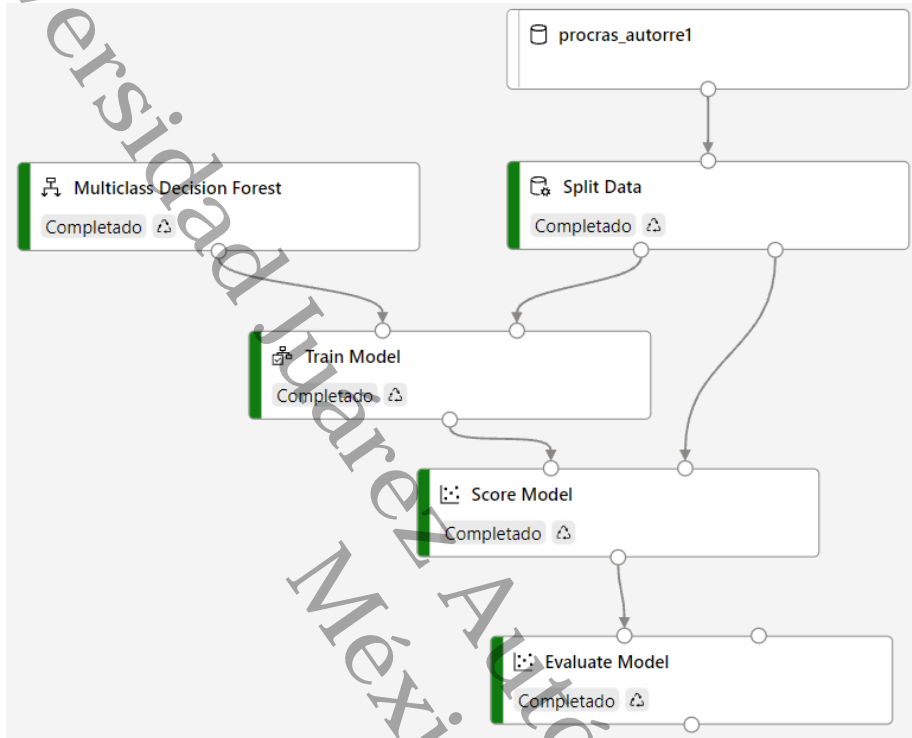
Enviar

Cancelar

Nota: Elaboración propia.

En la figura 62 se muestra que la ejecución del modelo fue completado y no ocurrió ningún problema al momento de realizar la ejecución e implementación del modelo en el *dataset* (ver figura 62).

Figura 62.
Ejecución completa de árbol de decisión en Azure Machine Learning Studio.

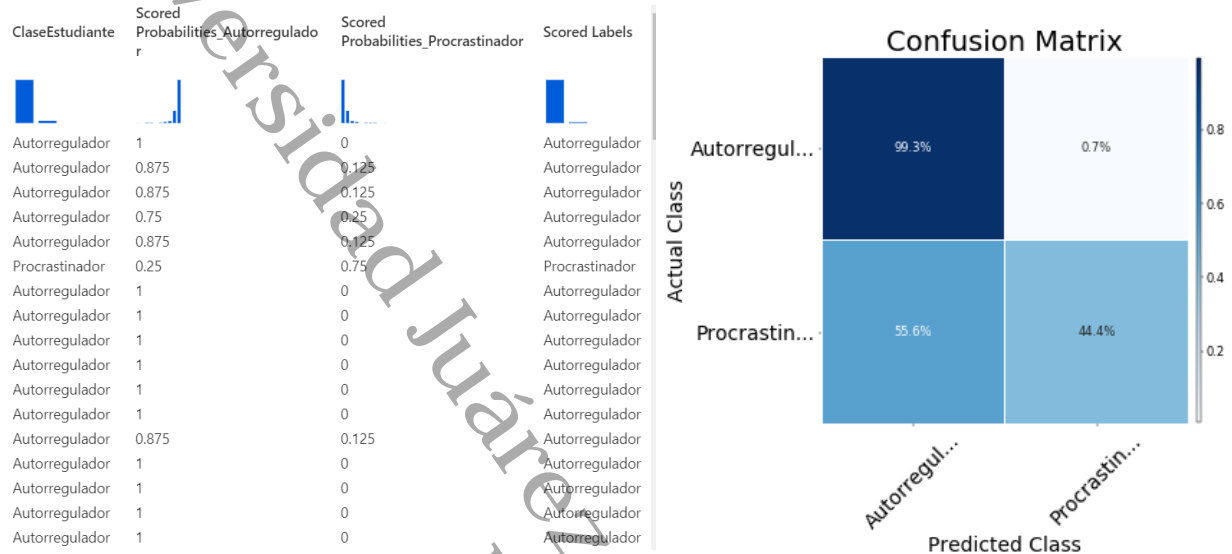


Nota: Elaboración propia.

La visualización de los resultados se puede realizar dando clic sobre la caja de *Evaluate model* y en la parte derecha del interfaz aparece un cuadro de diálogo con las opciones de visualizar parámetros, resultados y registros, etc., donde se dará clic en la segunda opción y por último clic en el ícono de una gráfica de barras.

La figura 63 muestra los resultados obtenidos a partir del modelo de árbol de decisión y la matriz de confusión (ver figura 63), a su vez si da clic en cada uno de los recuadros del modelo, se visualizan los resultados obtenidos en cada proceso o paso del árbol de decisión.

Figura 63.
Árbol de decisión en Azure Machine Learning Studio.



Nota: Elaboración propia.

Al finalizar de la realización del algoritmo de árbol de decisión, se procedió a realizar el modelo para implementar la técnica de agrupamiento en el *dataset*, esto utilizando el algoritmo *K-means*.

Se creó un nuevo proyecto para realización de este modelo, esto a partir de la pestaña diseñador y se procedió a seleccionar el destino de proceso, el cual será la misma instancia de proceso anteriormente creada. Seguidamente se seleccionó y arrastró el conjunto de datos al área de trabajo, esto dando clic en la pestaña *Dataset*, que se encuentra en la parte izquierda de la pantalla.

Posteriormente se normalizaron los datos, con el objetivo de reducir la redundancia que pudieran tener, esta opción se encuentra en la pestaña *Data transformation* con el

nombre de *Normalize data*, la cual realiza la conexión entre éste con el *dataset*. La figura 64 muestra lo anterior descrito (ver figura 64).

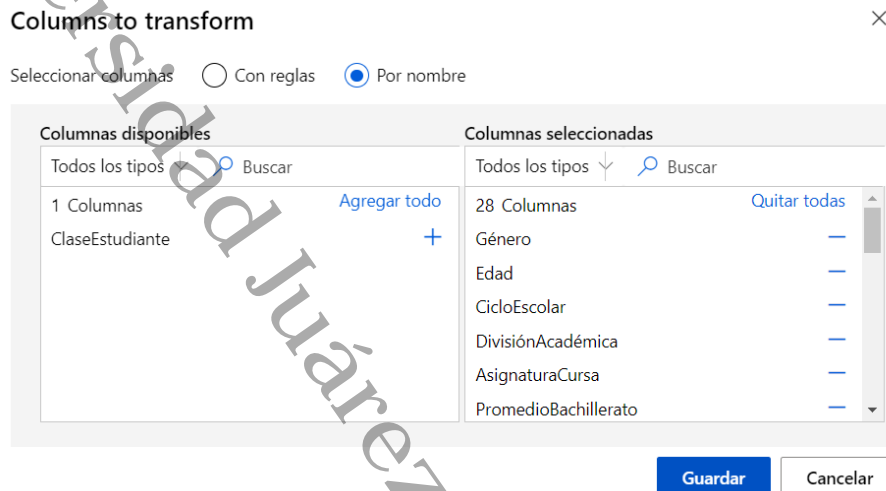
Figura 64
Normalizar datos en el modelo de agrupamiento en Azure Machine Learning Studio.



Nota: Elaboración propia.

En la figura 65 se presenta la selección de las variables a normalizar los datos, en este caso son las variables de tipo de dato numérico, la única variable a no normalizar es la variable ClaseEstudiante. Esto se realiza dando clic al cuadrado de *Normalice Data* y aparecerá en la parte derecha de la interfaz un cuadro de diálogo con opciones referentes a la normalización de los datos, dar clic en la opción editar columna y aparece otro cuadro de diálogo donde se seleccionaron las variables a normalizar (ver figura 65).

Figura 65. Selección de las variables a normalizar para el modelo de agrupamiento en Azure Machine Learning Studio.



Nota: Elaboración propia.

Seguidamente se selecciona el algoritmo de *K-means* que se encuentra en la pestaña *Machine Learning Algorithms*, en el apartado de *Clustering* y se arrastra al área de trabajo, donde aparecerá un cuadro de diálogo en la parte derecha de la interfaz con diferentes parámetros que se pueden configurar de este algoritmo.

La figura 66 muestra los parámetros configurados, donde se indica que el número de *cluster* serán dos y la métrica o método a utilizar para calcular la distancia entre los datos que conformaron el *cluster* será el método euclídean (ver figura 66).

Figura 66.
Configuración de parámetros del algoritmo K-means en Azure Machine Learning Studio.



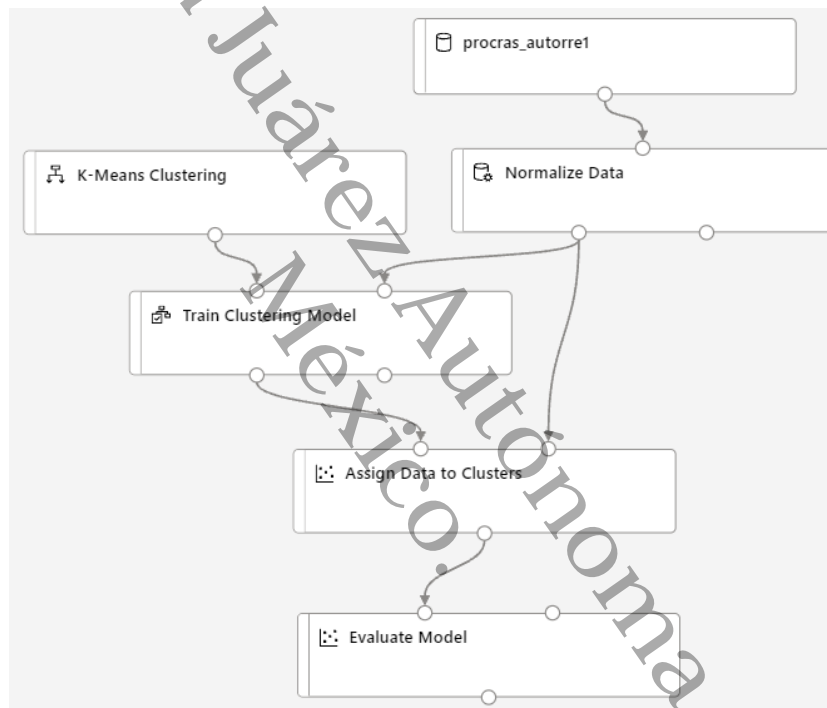
Nota: Elaboración propia.

Subsecuentemente se busca el recurso de *Train clustering model* en la pestaña *Model Training*, la cual cumple con la función de entrenar un modelo de agrupamiento en clústeres y asigna datos al conjunto de datos en entrenamiento. Además de realiza la selección del set de entrenamiento, las cuales son las variables seleccionadas con anterioridad en la normalización de los datos.

El recurso de *Train clustering model* se conecta al recurso de *Normalize data* y al algoritmo *K-means Clustering*. Subsecuentemente se utilizó el recurso de *Assign data to cluster* que se encuentra en la pestaña *Model scoring and evaluation*, el cual devuelve un conjunto de datos que contiene las asignaciones probables de cada registro a los clústeres, además este recurso se conecta con los recursos de *Normalize data* y *Train Clustering Model*.

Por último, se utilizó el recurso de *Evaluate model* que se encuentra en la pestaña *Model scoring & evaluation*, para evaluar y visualizar los resultados obtenidos del *dataset*. En la figura 67 se muestra el modelo de agrupamiento completo y todas las conexiones entre cada recurso para poder realizar la ejecución del modelo (ver figura 67).

Figura 67
Modelo de agrupamiento en Azure Machine Learning Studio.

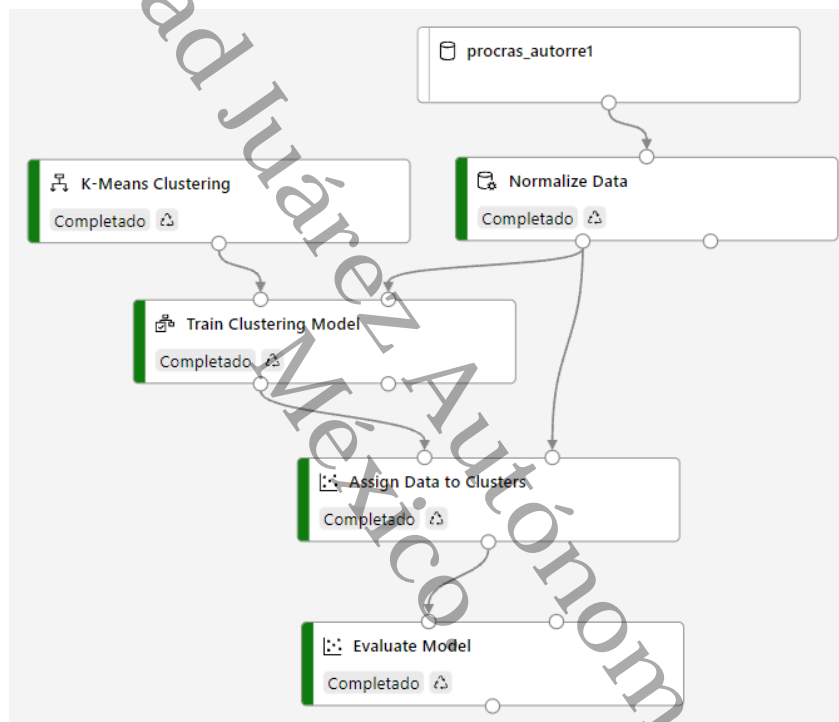


Nota: Elaboración propia.

La ejecución del modelo, se realizó dando clic en el botón enviar, el cual se encuentra en la parte superior de la interfaz de trabajo y aparece un cuadro de diálogo, donde se configurarán el nombre del experimento, en esta ocasión el nombre del experimento fue agrupamiento y por último clic en enviar, donde finalmente se comienza con la ejecución del modelo al *dataset*.

La figura 68 muestra la ejecución del modelo completado y no ocurrió ningún error o problema al momento de realizar la ejecución e implementación del modelo en el *dataset* (ver figura 68).

Figura 68
Ejecución completa de agrupamiento en Azure Machine Learning Studio.



Nota: Elaboración propia.

Una vez terminada la ejecución del modelo de agrupamiento, se procedió a visualizar los resultados de la asignación de los registros a los *clústeres* creados por medio del algoritmo *K-means*, esto dando clic sobre el recurso de *Assign data to clusters* y en la parte derecha de pantalla se visualizará un cuadro de diálogo, en el cual se encuentra la opción resultados y registros, donde se da clic sobre el icono en forma de gráfica de

barras, para poder visualizar los resultados de la asignación de los registros a los clústeres.

En la figura 69 se observa la tabla que la herramienta muestra con los resultados obtenidos de la implementación del algoritmo de *K-means* y la distancia al centro del grupo de agrupamiento (ver figura 69).

Figura 69
Asignación de cluster con la técnica de agrupamiento en Azure Machine Learning Studio.

18	Ítem19	Ítem20	Ítem21	Ítem22	ClaseEstudiante	Assignments	DistancesToClusterCenter no.0	DistancesToClusterCenter no.1
..								
:5645	-0.995422	-0.768826	-0.750993	1.055688	Autorregulador	0	1.185721	1.620564
:5645	-0.995422	-0.768826	-0.750993	-0.586738	Autorregulador	0	1.08998	1.515283
:167	1.025217	-0.768826	-0.750993	0.234475	Autorregulador	0	0.938281	1.423447
:5645	-0.995422	-0.768826	-0.750993	-0.586738	Autorregulador	0	0.893973	1.435208
:1239	0.014897	-0.768826	-0.750993	1.055688	Autorregulador	0	1.085226	1.472722
:1239	0.014897	0.147379	1.347287	0.234475	Autorregulador	1	1.278864	0.822059
:1239	0.014897	0.147379	0.298147	1.055688	Autorregulador	0	1.056759	1.44309
:1239	1.025217	1.063584	0.298147	0.234475	Autorregulador	0	0.82747	1.300951
:1239	0.014897	0.147379	-0.750993	-1.407951	Autorregulador	1	1.389502	1.073134
:5645	-0.995422	-0.768826	-0.750993	0.234475	Autorregulador	0	1.091986	1.577952
:1239	0.014897	0.147379	-0.750993	-1.407951	Autorregulador	1	1.392646	1.069536
:5645	0.014897	-0.768826	-0.750993	-0.586738	Autorregulador	1	1.42518	1.109703
:5645	-0.995422	-0.768826	-0.750993	0.234475	Autorregulador	0	1.063889	1.507115
:167	0.014897	-0.768826	0.298147	1.055688	Autorregulador	0	0.767884	1.227816
:1239	1.025217	1.063584	1.347287	1.055688	Procrastinador	1	1.576961	1.13108
:5645	-2.005742	-1.685031	-1.800134	-2.229164	Autorregulador	0	1.235677	1.712766
:5645	-0.995422	-0.768826	-0.750993	-1.407951	Autorregulador	0	0.984731	1.516951

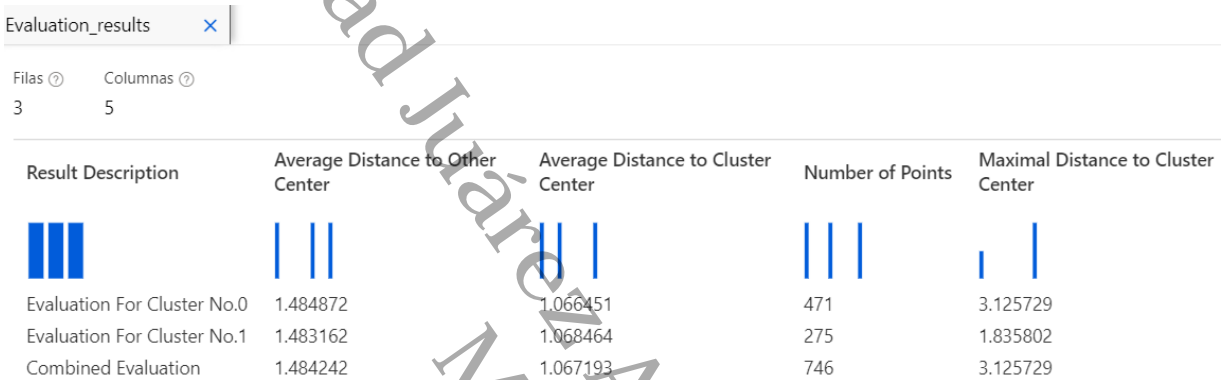
Nota: Elaboración propia.

Además, con el recurso *Evaluate model* se obtienen como resultados algunas métricas de los clústeres creados anteriormente en el modelo. La visualización de estas métricas se da a través de la selección de este recurso y en la parte derecha de la interfaz

se despliega un cuadro de diálogo, donde se da clic sobre el icono de una gráfica de barras para visualizar dichas métricas. En la figura 70 se muestra los resultados y las métricas obtenidos de este proceso del modelo de agrupamiento (ver figura 70).

Figura 70

Métricas obtenidas con la técnica de agrupamiento en Azure Machine Learning Studio.



Nota: Elaboración propia.

La tercera y última técnica de minería de datos implementada en el *dataset* fue la técnica de reglas de asociación, utilizando el algoritmo A priori. Para la creación del modelo de reglas de asociación e implementación de este, se utilizó el servicio de *Machine Learning Studio (clásico)*, el cual es una versión anterior de *Azure Machine Learning*, donde se pueden crear, modelar y ejecutar modelos de aprendizaje automático utilizados para minería de datos.

Esta versión anterior de la herramienta se utilizó porque permite realizar modelos de reglas de asociación, ya que dentro de los modelos o librerías cargados se encuentra el *Discover Association Rules*, el cual se basa en calcular y descubrir las reglas de asociación de un conjunto de datos utilizando el algoritmo A priori.

Para poder utilizar este servicio o herramienta, se tuvo que crear en primera instancia un plan de servicio web de *Machine Learning Studio (clásico)*. La figura 71 muestra la interfaz para la creación de dicha instancia, donde se complementan los campos de: suscripción, grupo de recursos (el cual se crea dentro de la misma instancia), nombre de la instancia, región y plan de tarifa, una vez llenado los campos se da clic en la pestaña revisar y crear, donde por último se crea el plan de servicio web (ver figura 71).

Figura 71
Creación del plan de servicio web de *Machine Learning Studio (clásico)* para la técnica de reglas de asociación.

Todos los servicios > planes de servicio web de Machine Learning Studio (clásico) >

plan de servicio web de Machine Learning Studio (clásico) ...

Basics Revisar y crear

Detalles del proyecto

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción * ⓘ

Grupo de recursos * ⓘ [Crear nuevo](#)

Detalles de la instancia

Nombre *

Región * ⓘ

Plan de tarifa * ⓘ

Nota: Elaboración propia.

Subsecuentemente de la creación del plan de servicio web, se procede a crear el área de trabajo de *Machine Learning Studio (clásico)*, donde se tienen que completar tres formularios para finalmente crear el área de trabajo. La figura 72 muestra la interfaz de la pestaña *Basic* para complementar datos del proyecto y la instancia, estos campos se

complementaron seleccionando la suscripción que se tiene a *Microsoft Azure*, el grupo de recurso creado anteriormente, la región de utilización del servicio y el nombre de la instancia de trabajo (ver figura 72).

Figura 72

Creación del área de trabajo de *Machine Learning Studio (clásico)* para la técnica de reglas de asociación.

Áreas de trabajo de Machine Learning Studio (clásico) ...

Basics Cuenta de almacenamiento Plan de servicio web Revisar y crear

Detalles del proyecto

Seleccione la suscripción para administrar recursos implementados y los costes. Use los grupos de recursos como carpetas para organizar y administrar todos los recursos.

Suscripción * ⓘ

Grupo de recursos * ⓘ [Crear nuevo](#)

Detalles de la instancia

Nombre *

Región * ⓘ

Nota: Elaboración propia.

La figura 73 muestra creación de la cuenta de almacenamiento del modelo de reglas de asociación a crear en *Machine Learning Studio (clásico)*, esto se realiza dando clic en crear nuevo y se despliega una interfaz en la cual se escribe el nombre de la cuenta de almacenamiento, el tipo de cuenta (en este caso es de uso general), rendimiento y replicación (ver figura 73).

Figura 73
Creación de cuenta de almacenamiento para el área de trabajo de Machine Learning Studio (clásico) para la técnica de reglas de asociación.

Todos los servicios > Áreas de trabajo de Machine Learning Studio (clásico) >
Áreas de trabajo de Machine Learning Studio (clásico) ...

Basics Cuenta de almacenamiento Plan de servicio web Revisar y crear

Cuenta de almacenamiento *

Crear cuenta de almacenamie... ×

Nombre *
core.windows.net

Tipo de cuenta

Rendimiento Premium

Replicación

Nota: Elaboración propia.

Una vez creada y seleccionada la cuenta de almacenamiento se procede a seleccionar el plan de servicio web, en este caso este servicio ya fue creado anteriormente, es por lo que se elige la opción de usar existente y se selecciona el plan creado (ver figura 74).

Figura 74
Selección del plan de servicio web para el área de trabajo de Machine Learning Studio (clásico) para la técnica de reglas de asociación.

Áreas de trabajo de Machine Learning Studio (clásico)

Basics Cuenta de almacenamiento Plan de servicio web Revisar y crear

Seleccione un plan de servicio web. * Crear nuevo Usar existente

Seleccione un plan de servicio web.

Nota: Elaboración propia.

Para finalizar la creación del área de trabajo, la herramienta realiza una validación de los campos llenados anteriormente en cada interfaz, si la validación es superada se crea el área de trabajo. La figura 75 muestra que la revisión para la creación del área de trabajo fue correcta y se procede a realizar la creación de ésta, esto dando clic en el botón crear (ver figura 75).

Figura 75

Revisión y creación del área de trabajo de Machine Learning Studio (clásico) para la técnica de reglas de asociación.

Áreas de trabajo de Machine Learning Studio (clásico)

Validación superada

Basics Cuenta de almacenamiento Plan de servicio web Revisar y crear

Basics

Suscripción	Suscripción de Azure 1
Grupo de recursos	Minería_datos
Nombre	pruebaAnálisis
Región	Centro-Sur de EE. UU.

Cuenta de almacenamiento

Cuenta de almacenamiento	pruebademo1
--------------------------	-------------

Plan de servicio web

Seleccione un plan de servicio web.	Usar existente
Seleccione un plan de servicio web.	datasetFinal

[Crear](#) [< Anterior](#) [Siguiente](#) [Descargar una plantilla para la automatizar](#)

Nota: Elaboración propia.

Después de la creación del área de trabajo ya se puede comenzar la utilización de

ésta, para desarrollar el modelo de reglas de asociación. Por lo consiguiente se dio clic en el área de trabajo para seleccionar la opción de compilar e implementar soluciones en la opción de *Machine Learning Studio (clásico)*, donde direcciona a otra página para comenzar a trabajar con los modelos y creación de los mismo.

En primera instancia se tiene que exportar el *dataset* desde los archivos locales y aparece un cuadro de diálogo para seleccionar y subir el conjunto de datos a la herramienta. La figura 76 presenta el cuadro de diálogo para seleccionar el archivo y una vez cargado, se da clic en el botón de marca de verificación o mejor conocido como paloma o revisado (ver figura 76).

Figura 76

Subir el conjunto de datos a Machine Learning Studio (clásico) para la técnica de reglas de asociación.

Upload a new dataset

SELECT THE DATA TO UPLOAD:

Seleccionar archivo procras_autorre2.csv

This is the new version of an existing dataset

ENTER A NAME FOR THE NEW DATASET:

procras_autorre2.csv

SELECT A TYPE FOR THE NEW DATASET:

Generic CSV File with a header (.csv)

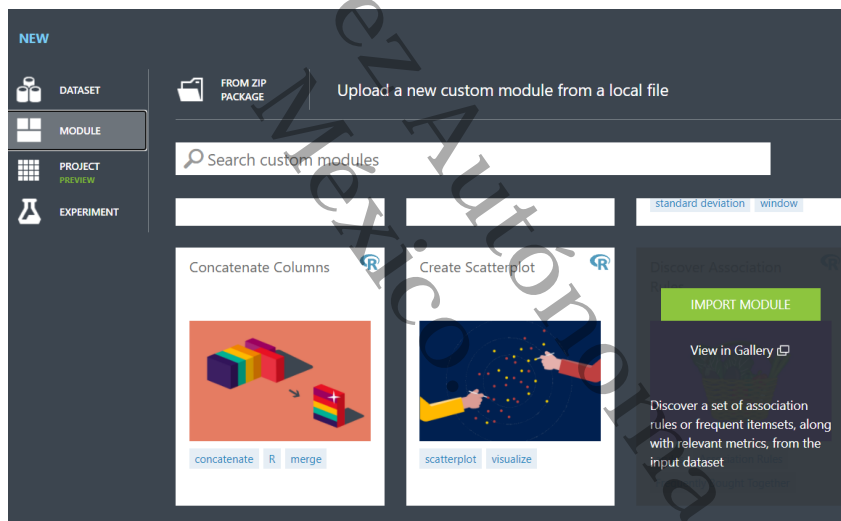
PROVIDE AN OPTIONAL DESCRIPTION:

Nota: Elaboración propia.

Subsecuentemente antes de crear un experimento y diseño del modelo de regla de asociación, se importa el modelo de *Discover association rules*, esto con la opción de *Module*, dando clic en el ícono New, que se encuentra en la parte inferior de la interfaz de la herramienta. En la figura 77 se visualiza lo anteriormente descrito, además de este modelo o paquete se encuentran diversos modelos que se pueden importar para utilizarlos (ver figura 77).

Figura 77

Importar el modelo de Discover association rules en Machine Learning Studio (clásico) para la técnica de reglas de asociación.



Nota: Elaboración propia.

Después de importar el modelo para reglas de asociación, se procedió a realizar un nuevo experimento, esto con la opción *New* en la parte inferior de la interfaz, opción *Experiment*, opción *Blank Experiment*, a partir del cual se presentará el área de trabajo para realizar el modelo.

Dentro del área de trabajo se modifica el nombre del experimento o modelo y posteriormente se busca la pestaña *Saved Dataset – My Dataset*, en esta última se encuentra el *dataset* a utilizar para implementar las reglas de asociación. La figura 78 muestra la interfaz del área de trabajo y el recurso del *dataset* en la misma (ver figura 78).

Figura 78
Selección del dataset para ejecutar la regla de asociación en Machine Learning Studio (clásico).



Nota: Elaboración propia.

Para implementar las reglas de asociación se selecciona el recurso de *Execute R Script*, que se localiza en la pestaña *R Language Modules*, en donde se escriben las líneas de código para la ejecución del algoritmo A priori de las reglas de asociación. La figura 79 muestra el código para la identificación de reglas de asociación del *dataset*, el tipo de dato de las variables es de tipo numérico y se utiliza la librería *arules* (ver figura 79).

Figura 79
Código para identificar reglas de asociación en Machine Learning Studio (clásico).

R Script

```
1 # Map 1-based optional input ports to variables
2 # Lectura del dataset#
3 procras_autorre2 <- mam1.mapInputPort(1) # class: data.frame
4
5 # Cargar librería de arules#
6 library(arules)
7
8 procras_autorre2 <- procras_autorre2[, 1:29]
9
10 # Discretizan las variables para aplicar reglas de asociación#
11
12 procras_autorre2$Genero <- discretize(procras_autorre2$Genero, method="interval", categories=2)
13 procras_autorre2$Edad <- discretize(procras_autorre2$Edad, method="interval", categories=21)
14 procras_autorre2$CicloEscolar <- discretize(procras_autorre2$CicloEscolar, method="interval", categories=68)
15 procras_autorre2$DivisionAcademica <- discretize(procras_autorre2$DivisionAcademica, method="interval", categories=12)
16 procras_autorre2$AsignaturaCursa <- discretize(procras_autorre2$AsignaturaCursa, method="interval", categories=5)
17 procras_autorre2$PromedioBachillerato <- discretize(procras_autorre2$PromedioBachillerato, method="interval", categories=4)
18 procras_autorre2$Item1 <- discretize(procras_autorre2$Item1, method="interval", categories=5)
19 procras_autorre2$Item2 <- discretize(procras_autorre2$Item2, method="interval", categories=5)
20 procras_autorre2$Item3 <- discretize(procras_autorre2$Item3, method="interval", categories=5)
21 procras_autorre2$Item4 <- discretize(procras_autorre2$Item4, method="interval", categories=5)
22 procras_autorre2$Item5 <- discretize(procras_autorre2$Item5, method="interval", categories=5)
23 procras_autorre2$Item6 <- discretize(procras_autorre2$Item6, method="interval", categories=5)
24 procras_autorre2$Item7 <- discretize(procras_autorre2$Item7, method="interval", categories=5)
25 procras_autorre2$Item8 <- discretize(procras_autorre2$Item8, method="interval", categories=5)
26 procras_autorre2$Item9 <- discretize(procras_autorre2$Item9, method="interval", categories=5)
27 procras_autorre2$Item10 <- discretize(procras_autorre2$Item10, method="interval", categories=5)
28 procras_autorre2$Item11 <- discretize(procras_autorre2$Item11, method="interval", categories=5)
29 procras_autorre2$Item12 <- discretize(procras_autorre2$Item12, method="interval", categories=5)
30 procras_autorre2$Item13 <- discretize(procras_autorre2$Item13, method="interval", categories=5)
31 procras_autorre2$Item14 <- discretize(procras_autorre2$Item14, method="interval", categories=5)
32 procras_autorre2$Item15 <- discretize(procras_autorre2$Item15, method="interval", categories=5)
33 procras_autorre2$Item16 <- discretize(procras_autorre2$Item16, method="interval", categories=5)
34 procras_autorre2$Item17 <- discretize(procras_autorre2$Item17, method="interval", categories=5)
35 procras_autorre2$Item18 <- discretize(procras_autorre2$Item18, method="interval", categories=5)
36 procras_autorre2$Item19 <- discretize(procras_autorre2$Item19, method="interval", categories=5)
37 procras_autorre2$Item20 <- discretize(procras_autorre2$Item20, method="interval", categories=5)
38 procras_autorre2$Item21 <- discretize(procras_autorre2$Item21, method="interval", categories=5)
39 procras_autorre2$Item22 <- discretize(procras_autorre2$Item22, method="interval", categories=5)
40 #procras_autorre2$ClaseEstudiante <- discretize(procras_autorre2$ClaseEstudiante, method="frequency", categories=2)
41 procras_autorre2$ClaseEstudiante <- ifelse(procras_autorre2$ClaseEstudiante==1, "PROCRASINADOR", "AUTORREGULADOR")
42
43 # Envía el conjunto de reglas resultante#
44 mam1.mapOutputPort("procras_autorre2");
```

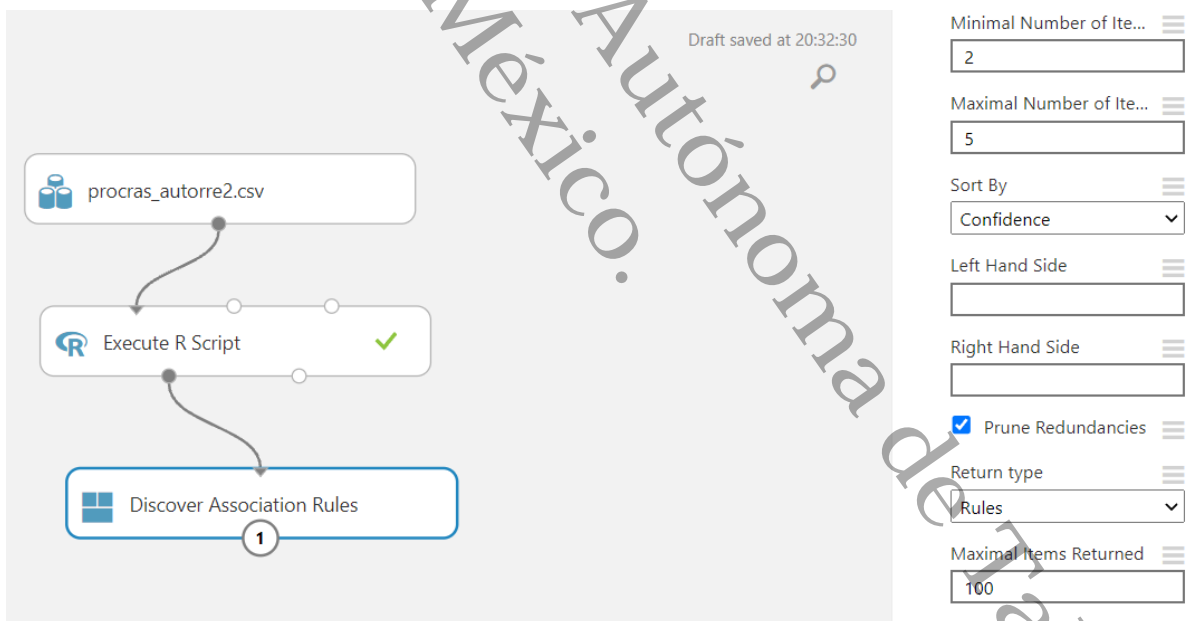
Nota: Elaboración propia.

Lo siguiente a realizar es la ejecución del código esto dando clic derecho sobre el recurso *Execute R Script*, clic en la opción *Run selected*, para comenzar la ejecución y verificar que no haya algún error el código que impida la ejecución del modelo más adelante. Una vez terminada la ejecución del modelo hasta el momento, el cual se ejecutó

de manera correcta el código, se procede a utilizar el recurso de *Discover Association Rules* de la pestaña *Custom*.

La figura 80 muestra el modelo creado para identificar las reglas de asociación en los datos, de igual forma al seleccionar el recurso de *Discover Association Rules*, en la parte derecha de la interfaz o pantalla se visualiza un cuadro de diálogo de los parámetros que contiene dicho recurso, en este caso se decidió que las reglas de asociación encontradas se ordenaran de acuerdo a la confianza de dichas reglas, además que el número máximo de reglas a visualizar es de 100 y se seleccionaron las variables a analizar (ver figura 80).

Figura 80
Modelo de reglas de asociación en Machine Learning Studio (clásico).



Nota: Elaboración propia.

Finalmente se realiza la ejecución del modelo de regla de asociación; una vez terminado la ejecución del modelo se procede a visualizar los resultados, esto dando clic

derecho sobre el recurso de *Discover Associate Rules*, donde se despliega un cuadro de diálogo con diferentes opciones, dentro de estas se encuentra la opción de *rulesDataset*, se da clic en esta opción y clic en *Visualize*.

La figura 81 presenta algunas las reglas de asociación obtenidas con la implementación del modelo diseñado en Machine Learning Studio (clásico) (ver figura 81).

Figura 81
Reglas de asociación obtenidas con el algoritmo A priori en Machine Learning Studio (clásico).

Reglas_1 > Discover Association Rules > Rules

rows 100
columns 6

view as

id	lhs	rhs	support	confidence	lift
1	{Item21=[2.6,3.4]}	{Edad=[17.0, 25.2]}	0.13807	1	1.053672
2	{Item8=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.179625	0.992593	1.101896
3	{Item21=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.536193	0.990099	1.099128
4	{Item19=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.378016	0.989474	1.098434
5	{Item18=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.402145	0.986842	1.095512
6	{Item9=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.191689	0.986207	1.094807
7	{Item17=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.184987	0.985714	1.09426
8	{Item16=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.27748	0.985714	1.09426

Nota: Elaboración propia.

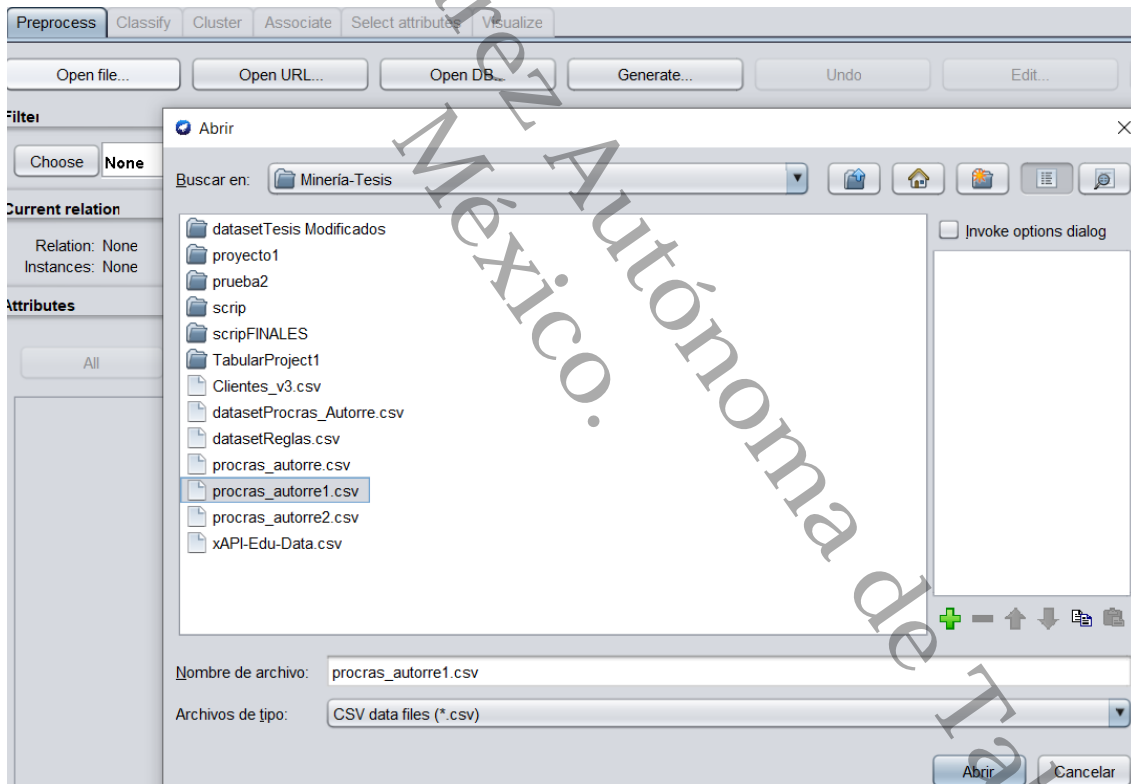
3.3.7.3 Weka

La implementación de minería de datos en esta herramienta se llevó a cabo seleccionando la instancia de trabajo *explorer*, que permite procesar, clasificar, asociar y

visualizar los datos. De igual manera porque permite ver los gráficos obtenidos por cada algoritmo aplicado al *dataset* o conjunto de datos.

Una vez elegida la instancia de trabajo, se muestra la ventana principal de la interfaz, donde se busca y se abre el archivo del *dataset*, esto dando clic en el botón *Open file*, se abre un cuadro de diálogo para buscar y seleccionar el archivo a abrir en el formato de tipo CSV (ver figura 82).

Figura 82
Cargar y abrir dataset en Weka.

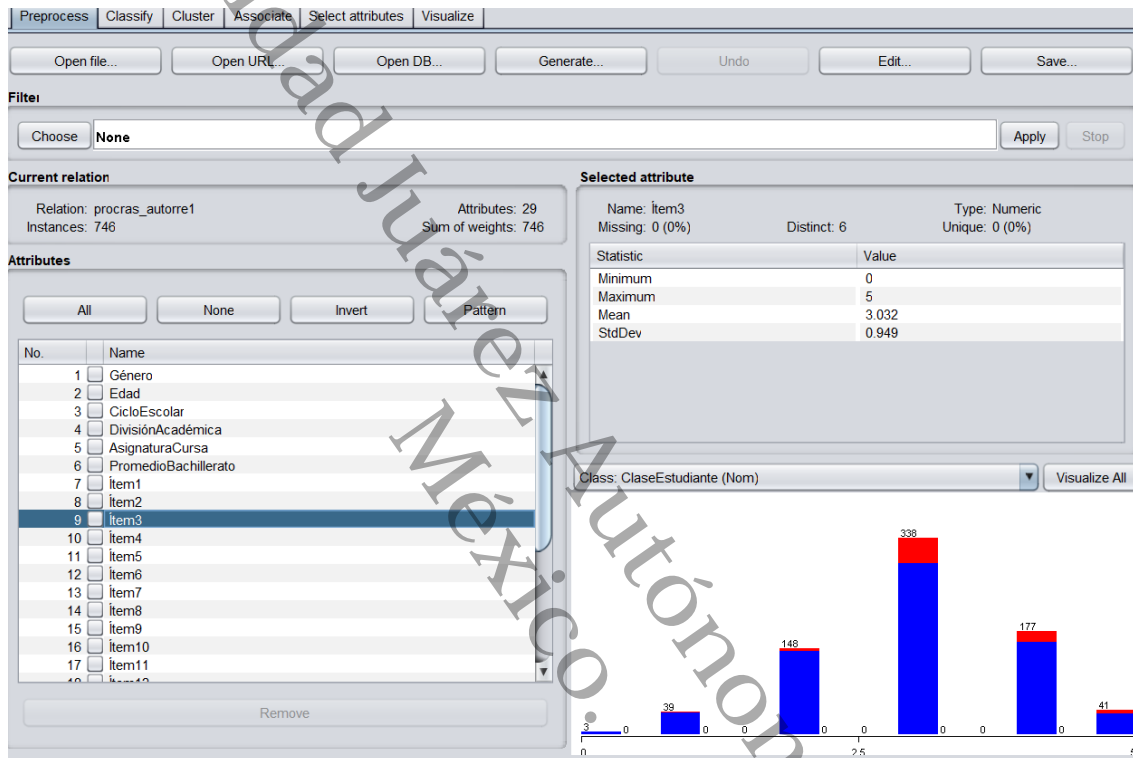


Nota: Elaboración propia.

Una vez abierto el archivo, se muestra en la interfaz recuentos en relación a los datos, las variables que contiene el *dataset*, además si se selecciona una variable o atributo se

presenta una resumen o detalles de la información que contiene ese atributo (ver figura 83).

Figura 83
Visualización del dataset en Weka.



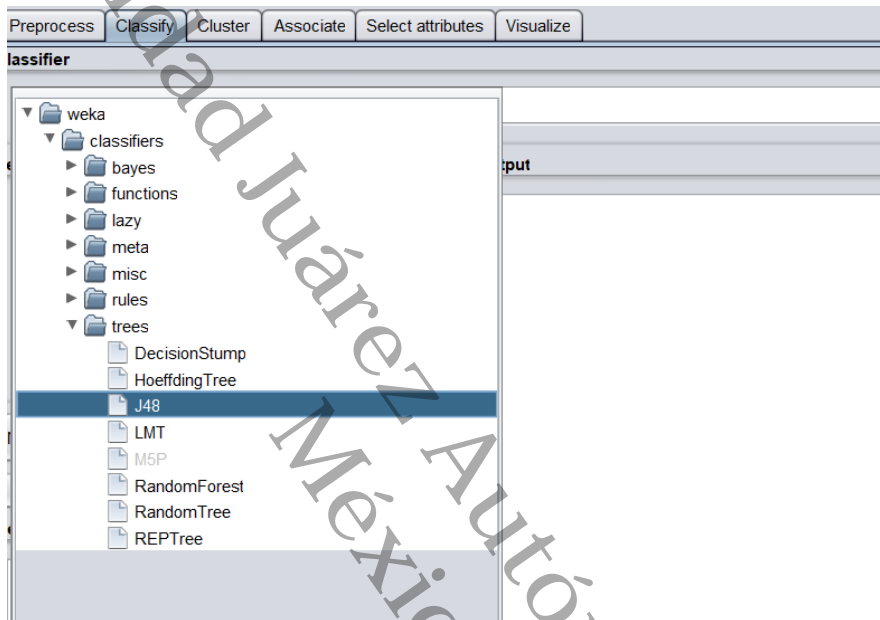
Nota: Elaboración propia.

Después de cargar y verificar que el *dataset* y los datos se hayan cargados correctamente, se procedió a realizar la minería de datos en *Weka*, en primer lugar, se realizó la minería de datos con el algoritmo C4.5 o J48, el cual es un algoritmo de árbol de decisión.

En la figura 84 se muestra la selección de dicho algoritmo, dando clic en la pestaña *Classify*, la cual es la pestaña donde se encuentran algoritmos de clasificación como los

árboles de decisión, luego clic en *Choose* y despegando la carpeta con el nombre *trees* (árboles) y por último damos clic en J48 (ver figura 84).

Figura 84
Selección del algoritmo J48 para árbol de decisión en Weka.



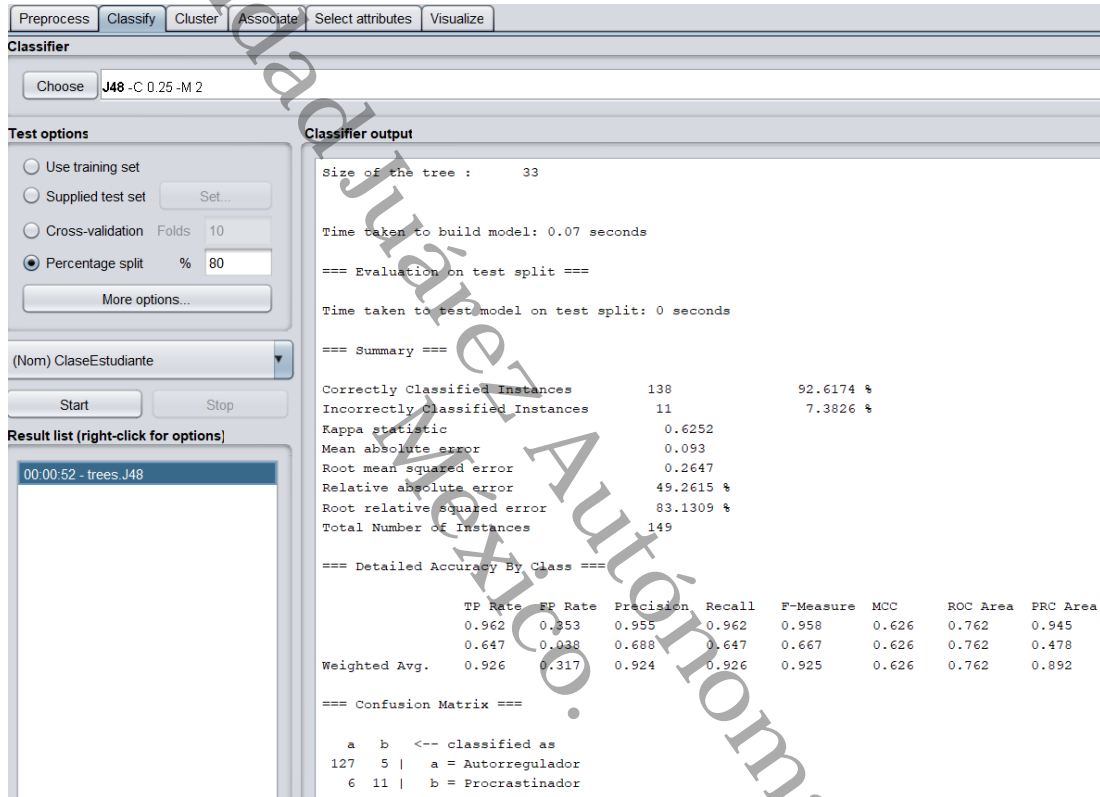
Nota: Elaboración propia.

Una vez seleccionado el algoritmo se pueden elegir cualquiera de las cuatro opciones que la herramienta ofrece para el análisis de los datos, en la parte izquierda de la interfaz, en este caso se seleccionó y utilizó la opción de *percentage split*, de igual forma se verifica que el ítem o variable seleccionada para realizar la minería de datos, sea la variable con el nombre ClaseEstudiante y se da clic en el botón *Start*.

En la figura 85 se muestra la aplicación del algoritmo J48 de árbol de decisión, donde se observa el modelo de clasificación obtenido, número de hoja y tamaño del árbol,

resumen de la información, precisión detallada por clase y matriz de confusión (ver figura 85).

Figura 85
Minería de datos con árbol de decisión en Weka.

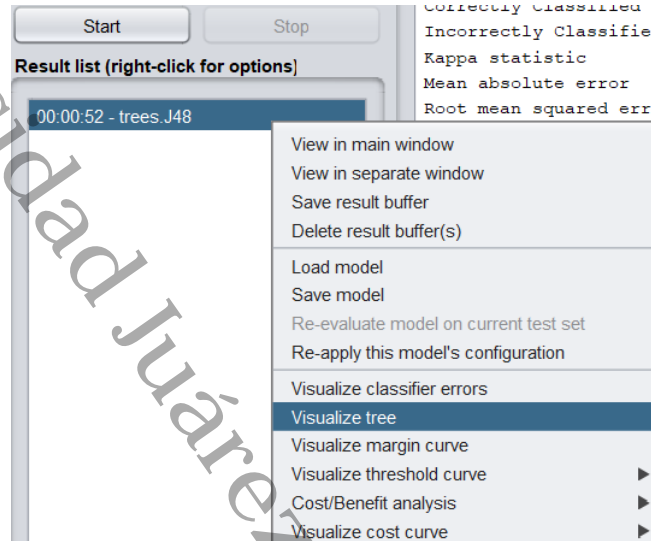


Nota: Elaboración propia.

En *Weka* la forma de visualizar el gráfico del árbol de decisión que se obtuvo, es dando clic derecho en el recuadro *Result list* donde se encuentra el nombre del árbol J48, el cual se ejecutó y se despliega un cuadro de diálogo, donde se dará clic en la opción *Visualize tree* (ver figura 86).

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

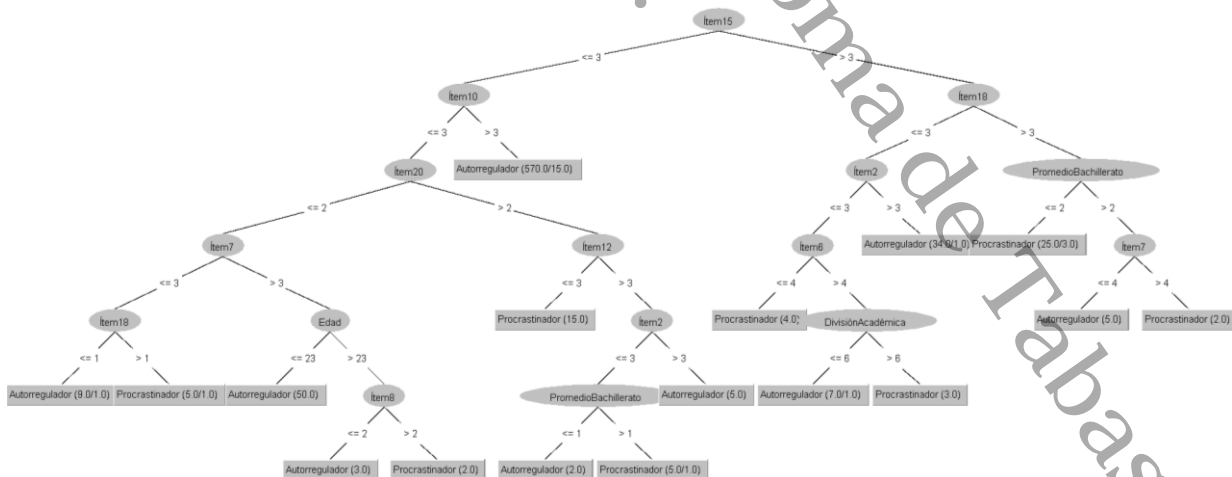
Figura 86.
Visualizar el árbol de decisión en Weka.



Nota: Elaboración propia.

En la figura 87 se muestra de forma gráfica el modelo clasificador del árbol de decisión que obtuvo con el *dataset* (ver figura 87).

Figura 87
Modelo clasificador del árbol de decisión realizado en Weka.

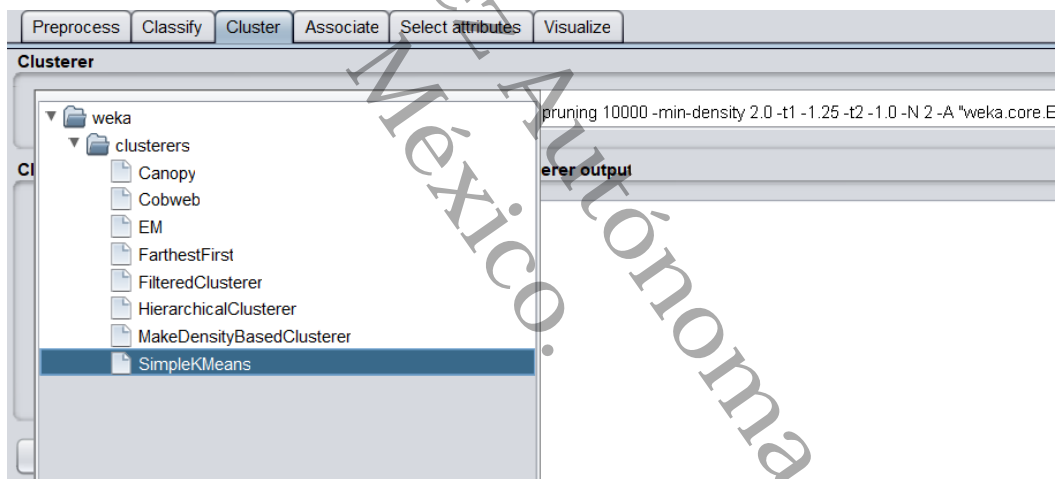


Nota: Elaboración propia.

Después de realizar la minería de datos al *dataset*, utilizando la técnica de árbol de decisión, se procedió a la aplicación de la técnica de agrupamiento con el algoritmo *K-means*. En *Weka* este algoritmo se encuentra con el nombre *SimpleKMeans*; este algoritmo se localiza en la pestaña *Cluster*.

En la figura 88 se muestra la pestaña *Cluster* y selección del algoritmo, esto dando clic en el botón *Choose* donde se despliega la lista de algoritmos y dar clic en *SimpleKMeans*, que es el algoritmo a utilizar para minería de datos (ver figura 88).

Figura 88
Selección del algoritmo *SimpleK-means* en *Weka*.



Nota: Elaboración propia.

Luego de la selección del algoritmo *SimpleKMeans*, se seleccionó la opción de *Classes to cluster evaluation* para el análisis de los datos, de igual forma se verificó que la opción *Store clusters for visualization*, estuviera seleccionada para poder visualizar el agrupamiento y por último se ignoró la variable o ítem *ClaseEstudiante*, la cual es la variable a analizar y se da clic en el botón *Start*.

En la figura 89 se muestra la minería de datos realizada con el algoritmo de *SimpleKMeans* de agrupamiento, donde se observa el modelo de agrupamiento obtenido, número de iteraciones, puntos de partida iniciales aleatorios, grupo de centroides por *cluster* y el modelo y evaluación de agrupamiento (ver figura 89).

Figura 89
Minería de datos con agrupamiento en Weka.

Clusterer

Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A *weka.core.EuclideanDistance -R firs

Cluster mode

Use training set

Supplied test set

Percentage split % 66

Classes to clusters evaluation

(Nom) ClaseEstudiante

Store clusters for visualization

Cluster output

KMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 891.8426053239332

Initial starting points (random):

Cluster 0: 2,20,47,4,2,1,4,3,4,3,3,5,3,4,4,4,4,4,3,4,3,3,1,1,1,1,1,2
Cluster 1: 1,21,62,1,2,1,3,3,1,1,4,5,5,3,3,5,3,4,4,5,3,2,3,3,2,2,4

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster# 0	Cluster# 1
	(746.0)	(471.0)	(275.0)
=====			
Género	1.6314	2	1
Edad	21.4316	21.5541	21.2218
CicloEscolar	43.0067	43.9979	41.3091
DivisiónAcadémica	5.0442	4.8068	5.4509
AsignaturaCursa	2.4755	2.4268	2.56
PromedioBachillerato	1.6555	1.569	1.8036
ítem1	2.4772	2.4268	2.5636
ítem2	3.9343	3.9214	3.9564
ítem3	3.0322	3.0127	3.0655
ítem4	3.0214	2.9766	3.0982

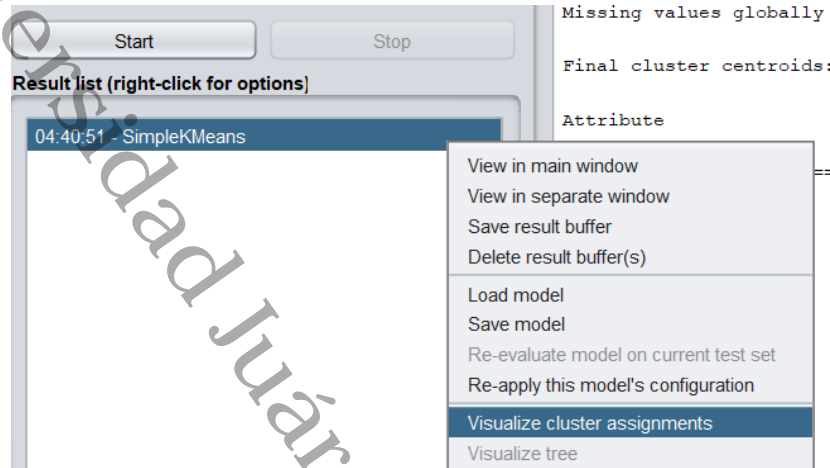
Result list (right-click for options)

04.40:51 - SimpleKMeans

Nota: Elaboración propia.

Además, se puede visualizar la asignación de los datos a los *clusters* o agrupamiento de los datos, dando clic derecho en el recuadro *Result list* el nombre del algoritmo correspondiente y se despliega un cuadro de diálogo, donde se dará clic en la opción *Visualize cluster assignments* (ver figura 90).

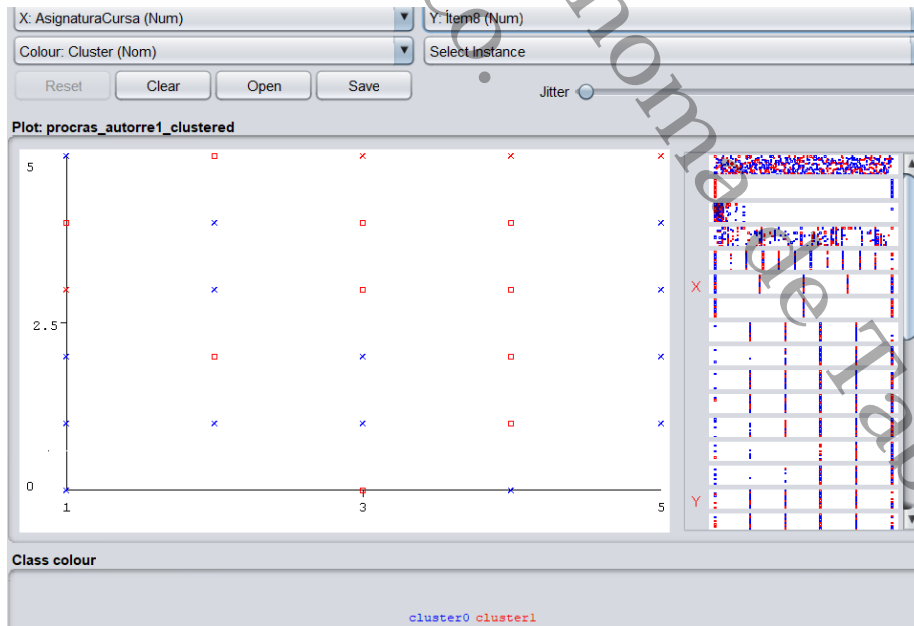
Figura 90
Visualización de agrupamiento en Weka.



Nota: Elaboración propia.

En la figura 91 se visualiza de forma gráfica el agrupamiento que se obtuvo de la minería de datos realizada con la técnica de agrupamiento al *dataset*, (ver figura 91).

Figura 91
Modelo de asignación de agrupamiento de los datos realizados en Weka.

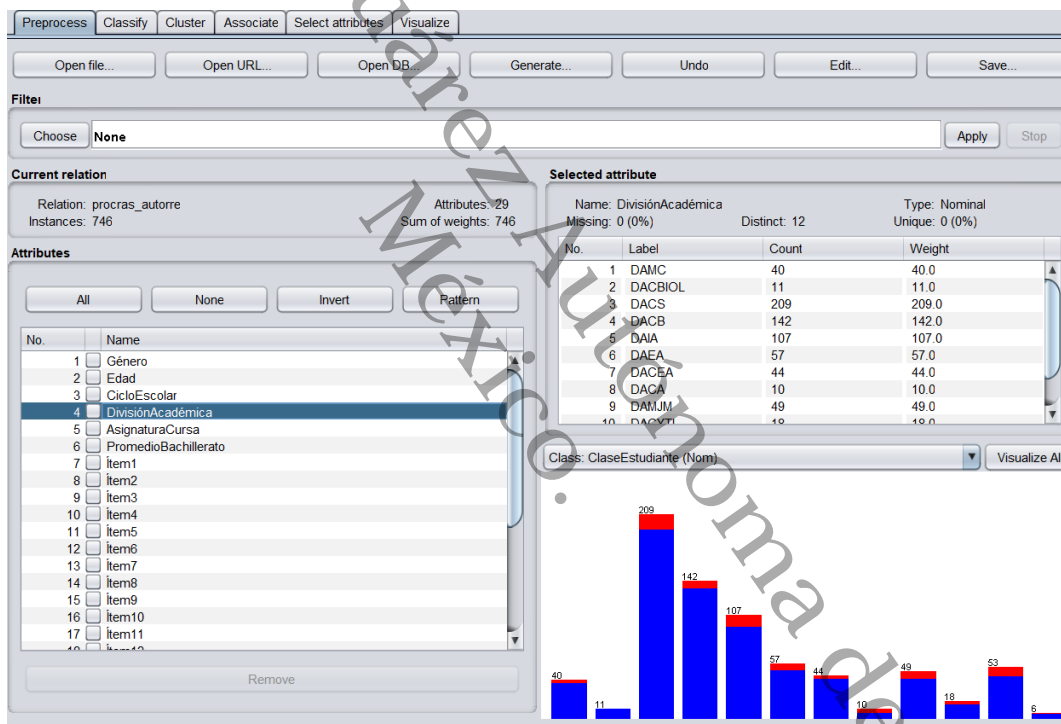


Nota: Elaboración propia.

Por último, se llevó a cabo en *Weka* la implementación de minería de datos con el algoritmo A priori de la técnica de regla de asociación; para realizar esta técnica se volvió a abrir el archivo, pero esta vez el tipo de dato es de tipo nominal, ya que para implementar esta técnica se necesita que los datos no sean numéricos. La figura 92 muestra la visualización del *dataset* con el tipo de datos nominal (ver figura 92).

Figura 92

Carga del dataset con datos nominales para implementar el algoritmo A priori en Weka.



Nota: Elaboración propia.

Este algoritmo se localiza en la pestaña *Associate*, para buscar y seleccionar el algoritmo a utilizar, se dio clic en el botón *Choose* y se despliega una lista de los algoritmos que contiene la herramienta sobre las reglas de asociación.

Una vez seleccionado el algoritmo se pueden realizar las adecuaciones a las propiedades del algoritmo, esto dando clic derecho sobre el nombre del algoritmo y se desglosará un menú con cuatro opciones y dará clic en la opción *Show properties*, donde se muestra un cuadro de diálogo para realizar las modificaciones al algoritmo, en cuanto al número de reglas, tipo de métrica, entre otros (ver figura 93).

Figura 93

Cuadro de diálogo para modificar parámetros del algoritmo Apriori en Weka.

weka.associations.Apriori

About

Class implementing an Apriori-type algorithm.

car False

classIndex -1

delta 0.05

doNotCheckCapabilities False

lowerBoundMinSupport 0.1

metricType Confidence

minMetric 0.8

numRules 20

outputItemSets False

removeAllMissingCols False

significanceLevel -1.0

treatZeroAsMissing False

upperBoundMinSupport 1.0

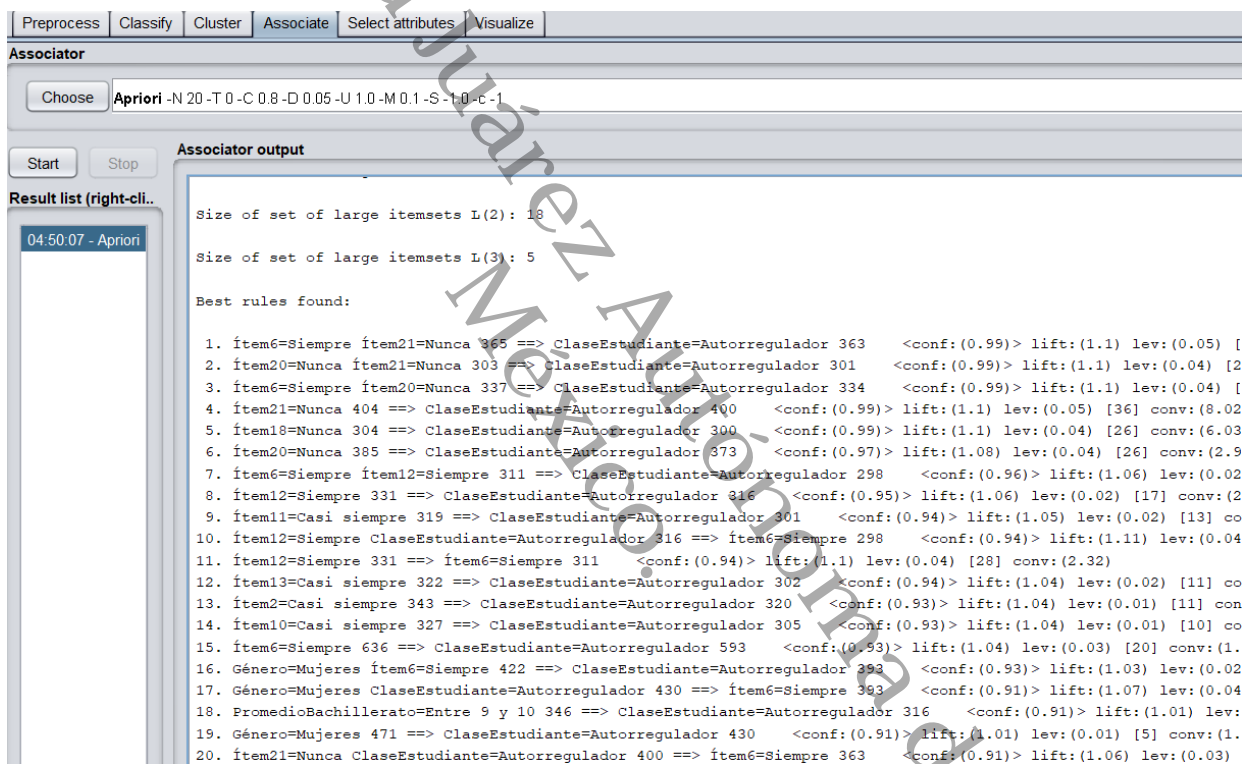
verbose False

Nota: Elaboración propia.

Después de revisar, verificar o en su caso modificar los parámetros o propiedades del algoritmo se procedió a realizar la ejecución del algoritmo, esto dando clic en el botón

Start. En la figura 94 se muestran las reglas de asociación encontradas en el *dataset* e información adicional con respecto a las reglas, el número de reglas de asociación visualizadas se obtiene de acuerdo con el número establecido en las propiedades del algoritmo A priori (ver figura 94).

Figura 94
Reglas de asociación encontradas con el algoritmo A priori en Weka.



Nota: Elaboración propia.

3.3.8 Interpretación

El penúltimo paso de la metodología es la interpretación, donde una vez realizado la minería de datos se procede a analizar la información obtenida. Fayyad, Piatetsky-Sapiro & Smyth (1996) describen este paso como la interpretación de los patrones descubiertos

y la posibilidad de volver a cualquier de los pasos anteriores para realizar una nueva búsqueda de patrones entre el conjunto de datos.

Además, los autores describen que durante este paso se pueden eliminar patrones redundantes o irrelevantes que se hayan obtenido durante el paso anterior y traducir los datos, patrones o relaciones útiles en términos comprensibles para los usuarios.

Por lo anterior se analizaron e interpretaron los resultados obtenidos de la implementación de los modelos creados de las técnicas de minería de datos en las tres herramientas seleccionadas. De igual forma la visualización de los resultados varía de acuerdo con la herramienta, por ejemplo, con Lenguaje R utilizando *RStudio* y *Weka* se obtienen el diagrama del árbol de decisión y la estructura del mismo resultante; en cambio con la herramienta de *Microsoft Azure Machine Learning* se visualiza el diagrama del árbol de decisión.

Con la visualización de los resultados de la técnica de reglas de asociación las tres herramientas muestran de igual forma los resultados. También al realizar un análisis rápido de los resultados, se observaron que algunos resultados pueden variar de acuerdo a la herramienta en que se implementaron las técnicas de minería de datos, es decir, los resultados que se presentan en la herramienta *Weka* de las reglas de asociación pueden ser que en la herramienta *RStudio* se encontraron dos o tres reglas, mientras que en *Weka* no se identificaron y así para las demás técnicas de minería de datos.

Todo lo anterior descrito del análisis, interpretación y visualización de los resultados se describen con mayor amplitud en el capítulo 4 de pruebas y resultados, en el apartado de resultados.

3.3.9 Conocimiento descubierto

El conocimiento descubierto es el último paso de la metodología, donde se realiza la utilización de la información obtenida de la minería de datos y la interpretación de los resultados, para la toma de decisiones futuras y dar a conocer la información encontrada.

Según Fayyad, Piatetsky-Sapiro & Smyth (1996) este paso tiene el propósito de utilizar el conocimiento descubierto de los datos para mejorar rendimiento de sistemas de desempeño de procesos, adopción de medidas basadas en el conocimiento, o simplemente documentar, reportar y presentar informes a las partes interesadas, así como la revisión y resolución de posibles conflictos con los conocimientos previamente extraídos de los datos.

A partir de lo anterior, las pruebas realizadas a los modelos diseñados y creados, así como los resultados y la interpretación realizadas de los mismos, se documentaron para tener conocimiento e identificación de patrones de conducta de los estudiantes que cursan materias en líneas de la UJAT puedan tener ante la procrastinación y la autorregulación académica, con este conocimiento descubierto y la implementación de minería de datos en educación, especialistas del área de educación o interesados puedan tomar decisiones creando o modificando medidas y reglas ya existentes para que el estudiante tenga una nivelación entre estas dos variables.

Capítulo 4. Pruebas y Resultados

En este capítulo se presentan las pruebas realizadas y resultados obtenidos del *dataset*, en relación con la ejecución de la minería de datos utilizando las técnicas y algoritmos empleados en cada una de las tres herramientas.

Lo primero que se realizaron fueron las pruebas en relación con la ejecución de los modelos creados en las herramientas con base en los algoritmos y técnicas de minería de datos seleccionados anteriormente, esto se llevó a cabo realizando cambios o modificaciones con respecto a parámetros y propiedades de los algoritmos.

Una vez realizadas las pruebas y establecidos los parámetros adecuados a utilizar para la obtención de los patrones de conducta respecto a procrastinación y autorregulación académica, fueron ejecutados los algoritmos de las técnicas de minería de datos seleccionadas con anterioridad en las tres herramientas. A continuación, se describen y explican las pruebas realizadas al *dataset* y los resultados obtenidos de la investigación.

4.1 Pruebas de las técnicas de minería de datos

En este apartado se presenta las pruebas realizadas en relación con cada una de las técnicas de minería de datos e implementándolas con las tres herramientas para minería de datos.

El número de pruebas realizadas por cada técnica de minería de datos implementada o desarrollada por cada una de las herramientas de minería de datos seleccionadas anteriormente fueron tres, es decir, que se llevaron a cabo un total de nueve pruebas de

cada una de tres técnicas de minería de datos y de esas nueve, tres fueron realizadas en Lenguaje R, tres en *Microsoft Azure Machine Learning* y por último las otras tres en *Weka*.

El orden de las tres pruebas realizadas en cada una de las herramientas es el siguiente: la primera y segunda prueba se realizaron en el *dataset* de la investigación para encontrar patrones de conducta de procrastinación y autorregulación académica, conformado por 29 variables y 746 registros, estas pruebas se realizaron haciendo modificaciones de parámetros o porcentaje de *sets* de entrenamiento y prueba en caso de los árboles de decisión.

La tercera prueba se realizó en otro *dataset*, en el *dataset* iris, esto para verificar que los modelos creados, modificaciones y parámetros se pueden utilizar en otros datos y sean los adecuados. El *dataset* iris está compuesto por cinco variables o ítems y un total de 150 registros, el tipo de datos de cuatro variables es de tipo numérico y una variable es de tipo carácter.

A continuación, se describen y desarrollan las pruebas realizadas de los árboles de decisión, pruebas de agrupamiento y pruebas de reglas de asociación.

4.1.1 Pruebas de Árboles de decisión

La primera y segunda prueba se realizaron al *dataset* de la investigación, para la clasificación de clase de estudiante si son procrastinadores o autorreguladores, y por último la tercera prueba implementando la técnica de árbol de decisión, se desarrolló en el *dataset* iris, donde se clasificó la especie a la que pertenece la flor.

En cada una de las pruebas se crearon set de entrenamiento y prueba; el porcentaje de datos que contienen cada set de entrenamiento y prueba es el mismo para cada una de las pruebas realizadas en las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*.

Los sets de entrenamiento y prueba están conformados de la siguiente forma para las tres pruebas: la primera prueba está conformada por el 70% y el 30% de los datos respectivamente para cada set, en relación a la segunda prueba el 60% es set de entrenamiento y el 40% set de prueba y la tercera prueba realizada al *dataset* iris es 70% y 30% respectivamente cada set. A continuación, se describen detalladamente las pruebas realizadas con las herramientas.

4.1.1.1 Lenguaje R

La primera prueba realizada al *dataset* de la investigación utilizando el modelo de árbol de decisión, creando el set entrenamiento con el 70% de los datos (522 registros) y el set de prueba con el 30% de los datos (224 registros). Los datos que se encuentran en el set de entrenamiento fueron los utilizados para entrenar el modelo y con la variable clase, en este caso es la variable ClaseEstudiante, que es la variable clasificatoria.

En la figura 95 se muestra el árbol de decisión obtenido a partir de la utilización del set de entrenamiento y la variable clase, de acuerdo con la utilización del algoritmo se obtuvo información del número y porcentaje de errores no clasificados de manera correcta, una clasificación de la clase y los porcentajes de los atributos o variables utilizadas para realizar el árbol de decisión (ver figura 95).

Figura 95.
Árbol de decisión obtenido en la prueba uno con Lenguaje R.

```

Class specified by attribute 'outcome'           Evaluation on training data (522 cases):
Read 522 cases (29 attributes) from undefined.data
Decision tree:
Item21 <= 2:
...Item10 > 3: AUTORREGULADOR (377/6)
: Item10 <= 3:
: ...Item20 <= 2: AUTORREGULADOR (47/4)
: Item20 > 2: PROCRASTINADOR (7/1)
Item21 > 2:
...Item18 <= 3:
...Item10 > 3: AUTORREGULADOR (53/4)
: Item10 <= 3:
: ...Item5 <= 3: PROCRASTINADOR (10/1)
: Item5 > 3: AUTORREGULADOR (3)
Item18 > 3:
...PromedioBachillerato > 2: AUTORREGULADOR (4)
PromedioBachillerato <= 2:
...Item9 > 3: PROCRASTINADOR (14)
Item9 <= 3:
...Item1 <= 3: AUTORREGULADOR (3)
Item1 > 3: PROCRASTINADOR (4)

```

Decision Tree		Evaluation on training data (522 cases):	
Size	Errors		
10	16(3.1%)	<<	
		(a)	(b) <-classified as
473	2	(a): class	AUTORREGULADOR
14	33	(b): class	PROCRASTINADOR

```

Attribute usage:
100.00% Item21
95.21% Item10
17.43% Item18
10.34% Item20
4.79% PromedioBachillerato
4.02% Item9
2.49% Item5
1.34% Item1
Time: 0.0 secs

```

Nota: Elaboración propia.

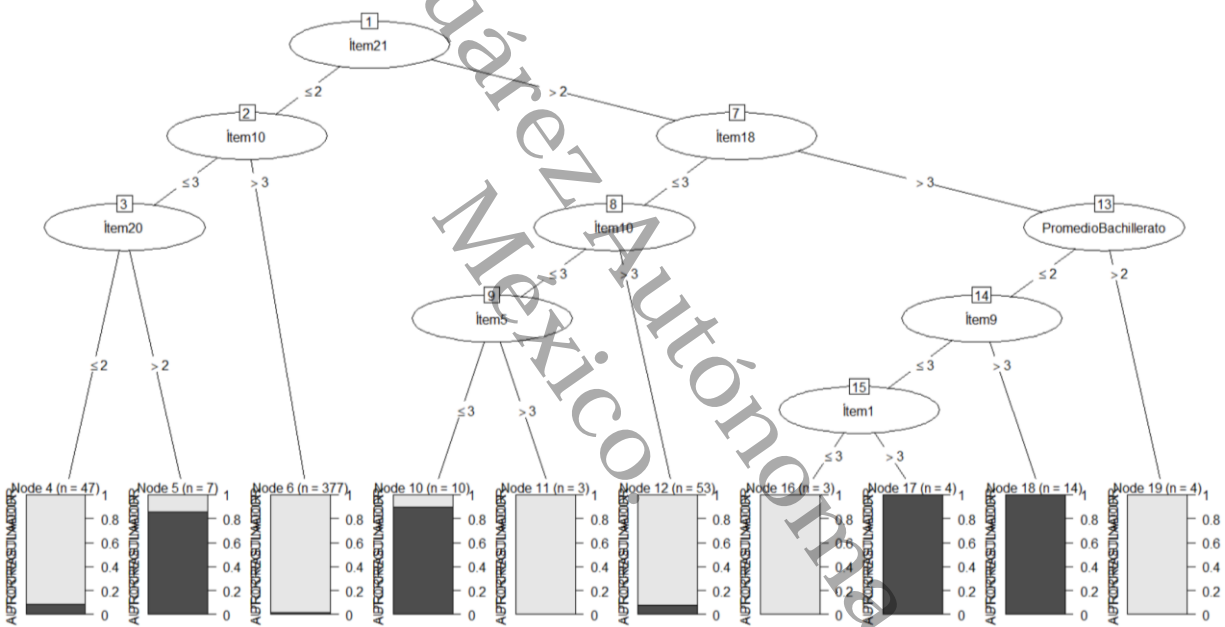
De acuerdo con la información obtenida a partir de la prueba uno realizada al modelo de árbol de decisión, se obtuvo que el tamaño del árbol fue de 10 nodos terminales y hubo nueve errores con un porcentaje del 3.1% al momento de clasificar los datos de la variable clase. De igual forma, para la matriz de confusión obtenida se realizó lo siguiente: se clasificaron correctamente 473 registros de la clase autorregulador y dos clasificados incorrectamente, y de la clase procrastinador se clasificaron correctamente 33 registros y 14 incorrectamente clasificados.

Aunado a lo anterior con el modelo de árbol de decisión se encontraron patrones clasificatorios sobre si el alumno es procrastinador o autorregulador, por ejemplo: cuando el ítem 21 es menor o igual a dos y el ítem 10 mayor que tres el estudiante será autorregulador. Otro ejemplo es que cuando el ítem 21 es mayor que dos, el ítem 18 es

menor o igual que tres y el ítem 10 sea mayor que tres el estudiante es autorregulador; así como estos patrones descritos hay otros más de esta prueba realizada.

La figura 96 se muestra el diagrama de árbol de decisión obtenido de la prueba uno, realizada al *dataset* de la investigación, donde se pueden visualizar de forma más clara y gráfica los patrones clasificatorios y los nodos del árbol (ver figura 96).

Figura 96
Diagrama del árbol de decisión obtenido en la prueba uno con Lenguaje R.



Nota: Elaboración propia.

La figura 97 muestra la matriz de confusión obtenida a partir de la validación de la capacidad de predicción del árbol, utilizando el set de prueba y por último se realizó el cálculo de porcentaje de aciertos clasificados correctamente el cual fue de 89.73214% (ver figura 97).

Figura 97.
Matriz de confusión y porcentaje de acierto de la prueba uno con Lenguaje R.

```
> ##MATRIZ DE CONFUSIÓN##  
> tablaP2 <- table(prediccionP2, datos.testP2$ClaseEstudiante)  
> tablaP2 #Visualización de la matriz de confusión  
  
prediccionP2      AUTORREGULADOR  PROCRASTINADOR  
AUTORREGULADOR      192             18  
PROCRASTINADOR       5              9  
> ##CALCULO DEL PORCENTAJE DE ACIERTOS##  
> 100 * sum(diag(tablaP2)) / sum(tablaP2) # % correctamente clasificados  
[1] 89.73214
```

Nota: Elaboración propia.

Con la segunda prueba se modificaron el porcentaje de datos destinados para cada *set*, en el caso del *set* de entrenamiento estaba conformado por el 60% de los datos (448 registros) y el *set* de prueba tenía el 40% restante de los datos (298 registros).

La figura 98 muestra el árbol de decisión obtenido a partir de los parámetros del *set* de entrenamiento y la variable clasificatoria, al ver modificaciones en el porcentaje de datos que constituyen cada *set*, hubo variaciones en el árbol de decisión y el tamaño de este (ver figura 98).

Figura 98.
Árbol de decisión obtenido en la prueba dos con Lenguaje R.

```

Class specified by attribute 'outcome'
Read 448 cases (29 attributes) from undefined.data
Decision tree:
Item21 > 2:
...Item18 <= 3:
: ...Item10 > 3: AUTORREGULADOR (47/3)
: : Item10 <= 3:
: : ...Item19 <= 3: PROCRASTINADOR (8/1)
: : : Item19 > 3: AUTORREGULADOR (2)
: : Item18 > 3:
: ...PromedioBachillerato > 2: AUTORREGULADOR (4)
: : PromedioBachillerato <= 2:
: : ...Item9 > 3: PROCRASTINADOR (14)
: : : Item9 <= 3:
: : : ...Item1 <= 3: AUTORREGULADOR (3)
: : : : Item1 > 3: PROCRASTINADOR (3)
Item21 <= 2:
...Item14 > 3: AUTORREGULADOR (307/2)
Item14 <= 3:
...Item21 <= 1:
: ...Item6 <= 3: PROCRASTINADOR (2)
: : Item6 > 3: AUTORREGULADOR (38)
Item21 > 1:
...item5 <= 2: PROCRASTINADOR (3)
item5 > 2:
: ...Item13 > 3: AUTORREGULADOR (7)
: : Item13 <= 3:
: : ...Item18 <= 1: AUTORREGULADOR (4)
: : : Item18 > 1:
: : : ...item4 <= 2: AUTORREGULADOR (2)
: : : : item4 > 2: PROCRASTINADOR (4)

```

Evaluation on training data (448 cases):

Decision Tree		
Size	Errors	
15	6 (1.3%)	<<
(a)	(b)	<-classified as
409	1	(a): class AUTORREGULADOR
5	33	(b): class PROCRASTINADOR

Attribute usage:

100.00%	Item21
81.92%	Item14
20.31%	Item18
12.72%	Item10
8.93%	Item6
5.36%	PromedioBachillerato
4.46%	item5
4.46%	Item9
3.79%	Item13
2.23%	Item19
1.34%	Item1
1.34%	item4

Time: 0.0 secs

Nota: Elaboración propia.

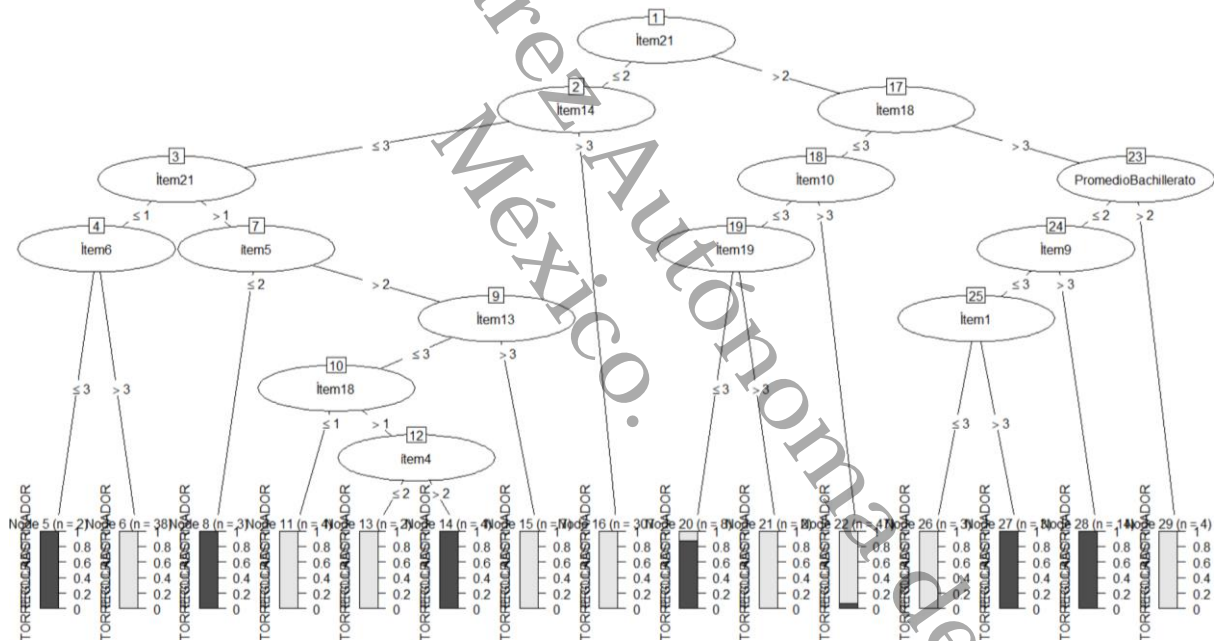
De acuerdo con la información obtenida a partir de esta prueba, el tamaño del árbol fue de 15 nodos terminales y hubo seis errores con un porcentaje del 1.3% al momento de clasificar los datos de la variable clase. De igual forma la matriz de confusión obtenida fue: 409 registros clasificados correctamente de la clase autorregulador y uno clasificado incorrectamente, y de la clase procrastinador se clasificaron correctamente 33 registros y cinco incorrectamente clasificados.

Aunado a lo anterior, con el modelo de árbol de decisión se encontraron patrones clasificatorios si el alumno es procrastinador o autorregulador, por ejemplo: cuando el ítem 21 es mayor que dos, el ítem 18 menor o igual a tres y el ítem 10 mayor que tres el estudiante será autorregulador. Otro patrón clasificatorio obtenido es si el ítem 21 es

mayor a dos, el ítem 18 es menor o igual que tres, el ítem 10 menor o igual a tres y el ítem 19 sea menor e igual a tres el estudiante es procrastinador; así como estos patrones descritos, hay otros más de esta prueba realizada.

La figura 99 muestra el diagrama de árbol de decisión obtenido de la prueba dos, realizada al *dataset* de la investigación, donde se puede visualizar de forma más clara y gráfica los patrones clasificatorios y los nodos del árbol (ver figura 99).

Figura 99
Diagrama del árbol de decisión obtenido en la prueba dos con Lenguaje R.



Nota: Elaboración propia.

La figura 100 muestra la matriz de confusión obtenida a partir de la validación de la capacidad de predicción del árbol, utilizando el set de prueba y por último se realizó el cálculo de porcentaje de aciertos clasificados correctamente el cual fue de 90.6% (ver figura 100).

Figura 100

Matriz de confusión y porcentaje de acierto de la prueba dos con Lenguaje R.

```
> ##MATRIZ DE CONFUSION##
> tablaP3 <- table(prediccionP3, datos.testP3$ClaseEstudiante)
> tablaP3 #visualización de la matriz de confusión

prediccionP3 • AUTORREGULADOR PROCRASTINADOR
AUTORREGULADOR      256          22
PROCRASTINADOR       6          14
> ##CALCULO DEL PORCENTAJE DE ACIERTOS##
> 100 * sum(diag(tablaP3)) / sum(tablaP3) # % correctamente clasificados
[1] 90.60403
```

Nota: Elaboración propia.

De igual forma en la tabla 19 del apéndice A, se encuentran las descripciones de las variables, es decir, que representa cada ítem y dato que se encuentra en el *dataset*, para mayor comprensión de los resultados obtenidos en las pruebas realizadas.

Por último, la tercera prueba realizada con el modelo de árbol de decisión desarrollado y creado en R utilizando *RStudio*, fue implementado en el *dataset* iris, para la verificación de que este modelo desarrollado podría ser utilizado con otros conjuntos de datos.

Esta prueba en el *dataset* iris se realizó utilizando los sets de entrenamiento y prueba con el 70% de los datos (105 registros) y el set de prueba contiene el 30% restante de los datos (45 registros) respectivamente.

En la creación del árbol de decisión se utilizó el set de entrenamiento y la variable *Especie*, la cual es la variable para clasificar o predictora. La figura 101 muestra el árbol de decisión obtenido, al ser un *dataset* con un número de registro más chico, el árbol de decisión obtenido fue más pequeño (ver figura 101).

Figura 101
Árbol de decisión obtenido en la prueba tres con Lenguaje R.

```
Class specified by attribute 'outcome'
Read 105 cases (5 attributes) from undefined.data
Decision tree:
Petal.Length <= 1.9: SETOSA (33)
Petal.Length > 1.9:
...Petal.Width > 1.6: VIRGINICA (34)
  Petal.Width <= 1.6:
  ...Petal.Length <= 4.9: VERSICOLOR (34)
    Petal.Length > 4.9: VIRGINICA (4/1)

Evaluation on training data (105 cases):

-----
Decision Tree
-----
Size      Errors
-----
      4      1( 1.0%)  <<

(a) (b) (c)  <-classified as
-----
      33      34      1
      37      37      1
      (a): class SETOSA
      (b): class VERSICOLOR
      (c): class VIRGINICA

Attribute usage:
100.00% Petal.Length
 68.57% Petal.Width

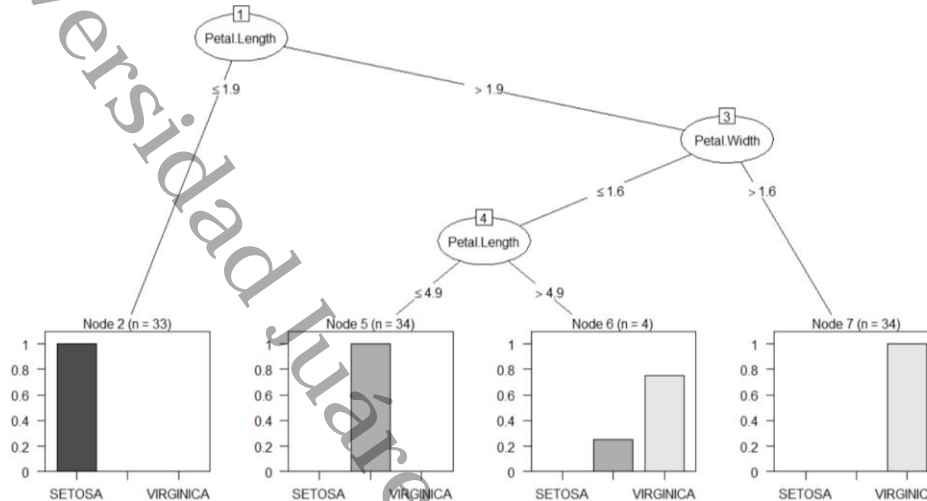
Time: 0.0 secs
```

Nota: Elaboración propia.

De acuerdo con la información obtenida a partir del modelo presentado en la figura 101, el tamaño del árbol fue de cuatro nodos y solo hubo un error al momento de clasificar los datos de la variable clase. El árbol de decisión obtenido del *dataset iris*, muestra cuatro caminos clasificatorios; por ejemplo, si la longitud del pétalo es menor o igual a 1.9 entonces la especie es setosa, o por lo contrario si la longitud del pétalo es mayor que 1.9 y el ancho del pétalo es mayor que 1.6 entonces la especie de la flor es virgínica.

La figura 102 muestra el diagrama de árbol de decisión, donde se pueden visualizar de forma más clara los nodos del árbol y las clasificaciones de los nodos terminales (ver figura 102).

Figura 102
Diagrama del árbol de decisión obtenido en la prueba tres con Lenguaje R.



Nota: Elaboración propia.

Lo último realizado de esta prueba fue validar la capacidad de predicción del árbol utilizando el set de prueba y obtener la matriz de confusión de estos datos, a su vez realizar el cálculo de porcentaje de aciertos clasificados correctamente. La figura 103 muestra la matriz de confusión obtenida de las tres especies, donde dos datos de la variable virgínica fueron clasificados de forma incorrecta y los demás se clasificaron correctamente, y el porcentaje de aciertos fue del 95.5% (ver figura 103).

Figura 103
Matriz de confusión y porcentaje de acierto de la prueba tres con Lenguaje R.

```
> tablaP1 <- table(prediccionP1, datos.testP1$Species)
> tablaP1 #Visualización de la matriz de confusión

prediccionP1 SETOSA VERSICOLOR VIRGINICA
SETOSA       17          0          0
VERSICOLOR   0          13          0
VIRGINICA    0           2          13
> ##CALCULO DEL PORCENTAJE DE ACIERTOS##
> 100 * sum(diag(tablaP1)) / sum(tablaP1) # % correctamente clasificados
[1] 95.55556
```

Nota: Elaboración propia.

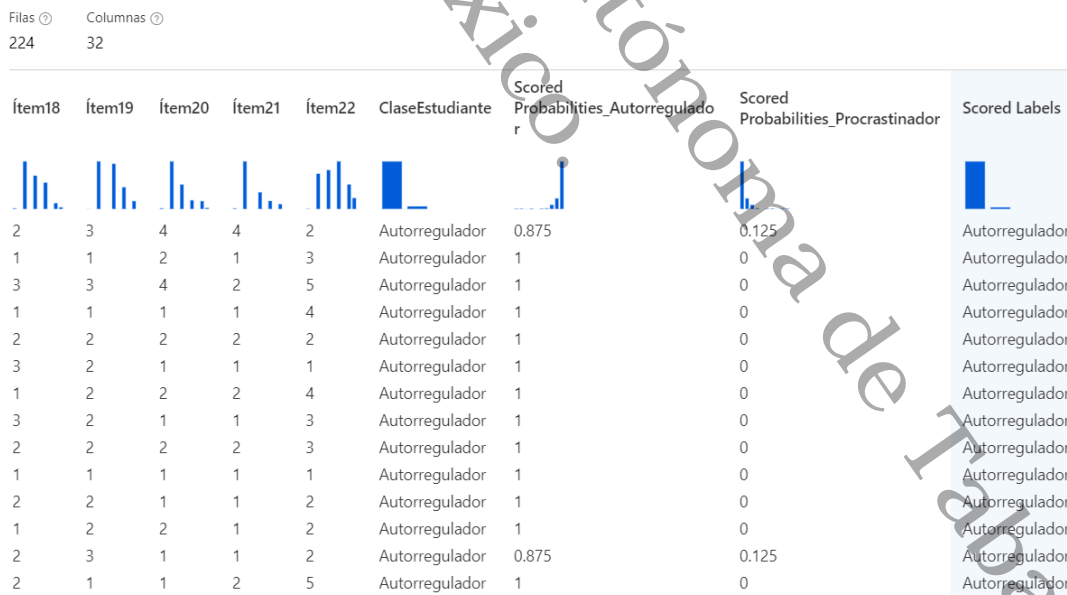
4.1.1.2 Microsoft Azure Machine Learning

La primera prueba realizada al modelo construido y creado en esta herramienta se llevó a cabo dividiendo el *dataset* en los *sets* de entrenamiento y prueba, donde al primero se asignaron el 70% de los registros y al segundo el 30% restante de los registros.

Una vez establecido lo anterior y la variable clasificatoria se realizó la ejecución, se obtuvieron las predicciones de las variables clasificatorias, así como un conjunto de métricas que evalúan la precisión de los resultados y por último la matriz de confusión. Es por lo anterior que en la figura 104 se muestra la predicción realizada y la probabilidad de éstas de acuerdo con las dos clases que contiene el *dataset* (ver figura 104).

Figura 104

Predicción de la clase de acuerdo con el árbol de decisión obtenido en la prueba uno con Microsoft Azure Machine Learning.



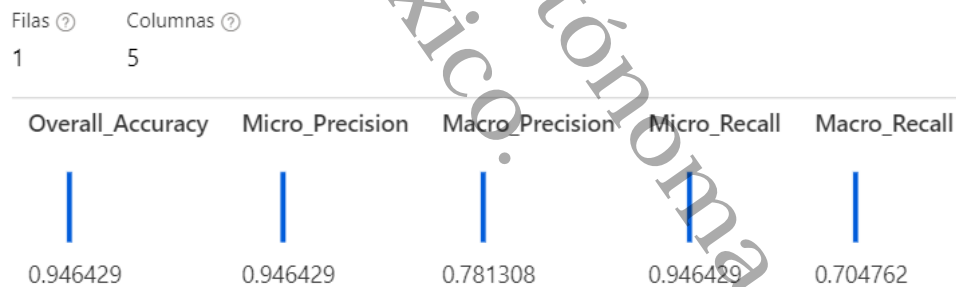
Nota: Elaboración propia.

De acuerdo con la figura anterior, se aprecia que la predicción realizada sobre la variable ClaseEstudiante fue correcta, ya que la probabilidad o puntuación en algunas tienen más valor puntaje en este caso hacia la clase donde se encontraban antes de analizarlas, pero también se encuentran datos con errores clasificados para las dos clases.

En la figura 105 se muestra el conjunto de métricas obtenidas de la evaluación de la precisión (rendimiento) del modelo realizado con la técnica de árbol de decisión en esta herramienta. Estas métricas son: precisión general, micro precisión, macro precisión, micro recuperación y macro recuperación (ver figura 105).

Figura 105

Conjunto de métricas obtenidas de la prueba uno del árbol de decisión con Microsoft Azure Machine Learning.



Nota: Elaboración propia.

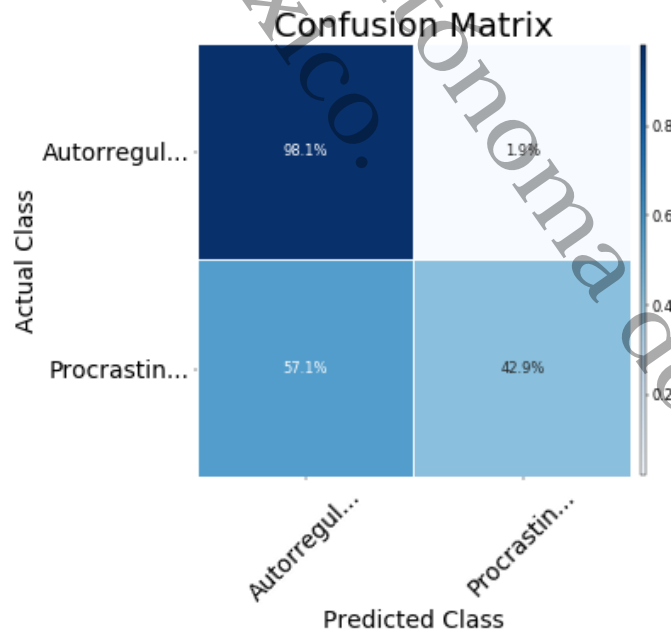
Con la figura anterior se visualiza que la precisión general de clasificación de los resultados fue del 94.64% lo que significa que la mayoría de los datos de la variable clase fueron clasificadas correctamente. También se obtuvo la precisión micro y macro en relación con los resultados verdaderos y resultados positivos, así como la recuperación

de las instancias del total relevantes que se recuperaron las cuales fueron micro del 94.64% y macro 70.47%.

Por último, en relación a la matriz de confusión de la prueba uno, se obtuvieron los siguientes datos de las instancias clasificadas correcta e incorrectamente: para la clase autorregulador se clasificaron correctamente el 98.1% (206 instancias) y 1.9% (4 instancias) clasificadas incorrectamente; con la clase procrastinador se clasificaron 42.9% (6 instancias) correctamente e incorrectas el 57.1% (8 instancias). Lo anteriormente descrito se puede visualizar en la figura 106, donde las instancias clasificadas se representan en porcentaje (ver figura 106).

Figura 106

Matriz de confusión obtenida de la prueba uno del árbol de decisión con Microsoft Azure Machine Learning.



Nota: Elaboración propia.

Por otro lado, la segunda prueba se realizó con un set de entrenamiento del 60% de los registros (448 registros) y el set de prueba con el 40% restante (298 registros), donde se obtuvo la predicción de las variables clasificatorias del set de prueba, en las cuales algunas instancias que pertenecían a la clase procrastinación, después de realizar minería de datos con la técnica de árbol de decisión, fueron clasificadas en la clase autorregulador y de igual forma sucedió con la clase autorregulador.

Lo anteriormente descrito se muestra figura 107, en la que se pueden visualizar algunas de las variables que conforman en *dataset*, la clase, las probabilidades de que dichas instancias o registros pertenezcan a la clase y la predicción realizada por el algoritmo (ver figura 107).

Figura 107

Predicción de la clase de acuerdo con el árbol de decisión obtenido en la prueba dos con Microsoft Azure Machine Learning.

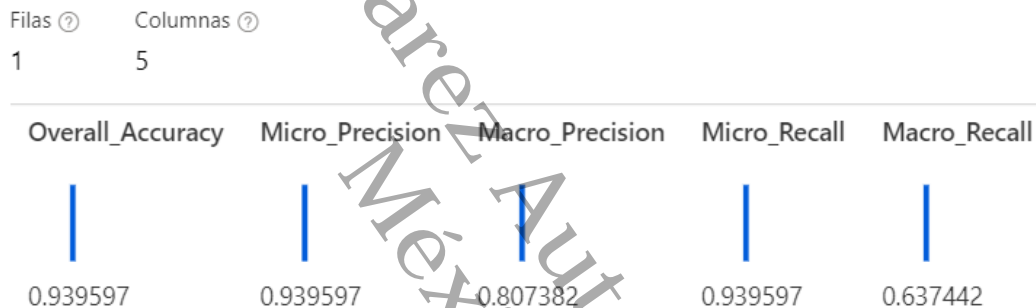
Filas 298 Columnas 32

Ítem18	Ítem19	Ítem20	Ítem21	Ítem22	ClaseEstudiante	Scored Probabilities_Autorregulador	Scored Probabilities_Procrastinador	Scored Labels
2	3	1	1	3	Autorregulador	1	0	Autorregulador
1	1	1	1	2	Autorregulador	0.875	0.125	Autorregulador
2	2	3	1	2	Autorregulador	0.875	0.125	Autorregulador
2	1	1	3	3	Procrastinador	0.5	0.5	Autorregulador
3	2	2	1	1	Autorregulador	1	0	Autorregulador
2	2	2	2	4	Autorregulador	0.875	0.125	Autorregulador
1	1	1	1	2	Autorregulador	1	0	Autorregulador
2	3	2	3	3	Autorregulador	1	0	Autorregulador
1	1	1	1	1	Autorregulador	1	0	Autorregulador
1	1	1	1	2	Autorregulador	1	0	Autorregulador
3	3	3	3	1	Procrastinador	0	1	Procrastinador
2	2	2	1	4	Autorregulador	0.875	0.125	Autorregulador

Nota: Elaboración propia.

En la figura 108 se muestra el conjunto de métricas obtenidas a partir del algoritmo, donde la precisión general de clasificación fue del 93.95% lo que significa que la mayoría de los datos de la variable clase fueron clasificadas correctamente, de igual forma la recuperación de las instancias del total relevantes que se recuperaron las cuales fueron micro del 93.95% y macro 63.74% (ver figura 108).

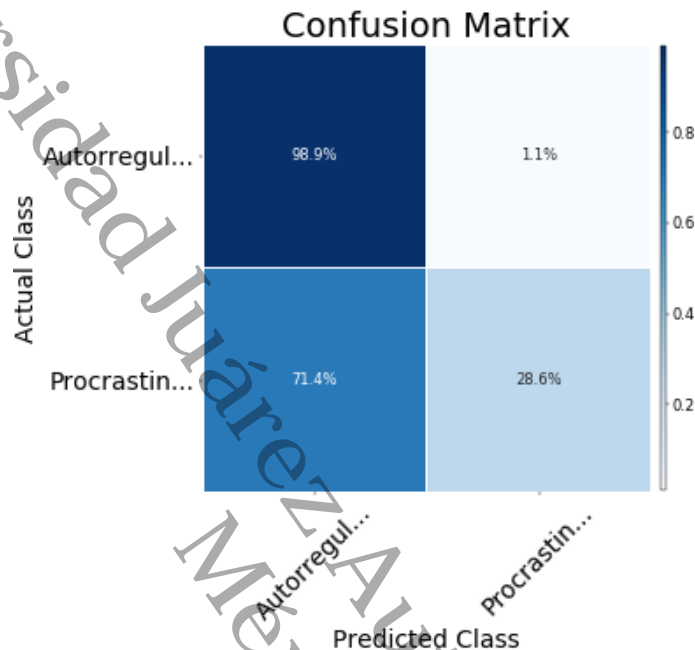
Figura 108
Conjunto de métricas obtenidas de la prueba dos del árbol de decisión con Microsoft Azure Machine Learning.



Nota: Elaboración propia.

Así mismo se obtuvo la matriz de confusión, en la figura 109, se visualiza que para la clase autorregulador se clasificaron el 98.9% (274 instancias) correctamente y el 1.1% (3 instancias) clasificadas incorrectamente y con la clase procrastinador se clasificaron el 71.4% (15 instancias) incorrectas y el 28.6% (6 instancias) clasificadas correctamente (ver figura 109).

Figura 109
Matriz de confusión obtenida de la prueba dos del árbol de decisión con Microsoft Azure Machine Learning.



Nota: Elaboración propia.

Por último, la tercera prueba realizada con el modelo de árbol de decisión desarrollado en *Microsoft Azure Machine Learning*, fue en el *dataset iris*; donde el set de entrenamiento contenía el 70% de los datos (105 registros) y el set de prueba el 30% restante de los datos (45 registros).

En la figura 110 se presentan las tres clases que contiene este *dataset*, las probabilidades de las instancias o registros de pertenecer a la clase y la predicción realizada por el algoritmo; por ejemplo, en la figura se puede visualizar que la probabilidad que una instancia versicolor tiene una mínima probabilidad del 12.5% de ser de la clase virgínica y tiene una mayor probabilidad del 87.5% de pertenecer a la clase versicolor y nula probabilidad en la clase setosa, es por eso que el algoritmo la clasificó como

versicolor, lo cual está en coincidencia con la forma en que se encontraba clasificada antes de la minería de datos, así como este ejemplo hay otros (ver figura 110).

Figura 110
Predicción de la clase de acuerdo con el árbol de decisión obtenido en la prueba tres con Microsoft Azure Machine Learning.

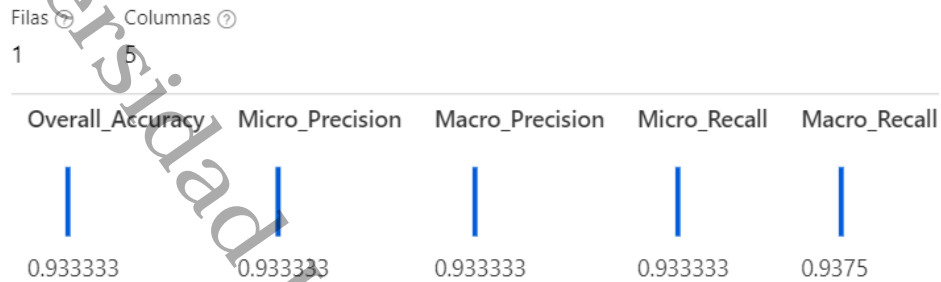
Filas 45 Columnas 9

gth	sepal_width	petal_length	petal_width	variety	Scored Probabilities_Setosa	Scored Probabilities_Versicolor	Scored Probabilities_Virginica	Scored Labels
3.8	6.7	2.2	Virginica	0	0	1	Virginica	
3	4.4	1.4	Versicolor	0	0.875	0.125	Versicolor	
3.2	1.4	0.2	Setosa	1	0	0	Setosa	
4.1	1.5	0.1	Setosa	1	0	0	Setosa	
3.5	1.4	0.2	Setosa	1	0	0	Setosa	
3	4.8	1.8	Virginica	0	0.375	0.625	Virginica	
2.7	3.9	1.2	Versicolor	0	1	0	Versicolor	
2.4	3.3	1	Versicolor	0	0.75	0.25	Versicolor	
3	5.2	2.3	Virginica	0	0	1	Virginica	
2.6	6.9	2.3	Virginica	0	0.125	0.875	Virginica	
2.7	4.1	1	Versicolor	0	1	0	Versicolor	
3.5	1.3	0.2	Setosa	0.875	0.125	0	Setosa	

Nota: Elaboración propia.

La figura 111 muestra el conjunto de métricas obtenidas a partir del *dataset* iris, donde la precisión general de clasificación fue del 93.33% lo que significa que la mayoría de los datos de la variable clase fueron clasificadas correctamente, de igual forma la recuperación de las instancias del total relevantes que se recuperaron las cuales fueron micro del 93.33% y macro 93.75% y por último la precisión micro y macro tienen el mismo porcentaje del 93.33% (ver figura 111).

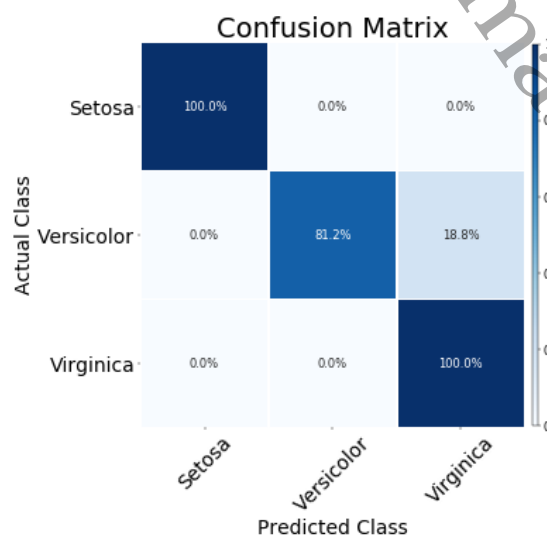
Figura 111
Conjunto de métricas obtenidas de la prueba tres del árbol de decisión con Microsoft Azure Machine Learning.



Nota: Elaboración propia.

Para finalizar esta prueba y las pruebas realizadas con esta herramienta, se obtuvo la matriz de confusión. En la figura 112 se visualiza que la clase setosa y virgínica se clasificaron correctamente todas las instancias que pertenecen a estas clases. Para el caso de las instancias pertenecientes a la clase versicolor se clasificaron correctamente el 81.2% (13 instancias) y el 18.8% (3 instancias) clasificadas incorrectamente.

Figura 112
Matriz de confusión obtenida de la prueba tres del árbol de decisión con Microsoft Azure Machine Learning.



Nota: Elaboración propia.

4.1.1.3 Weka

La primera prueba realizada en *Weka* con la técnica de árbol de decisión, fue realizada utilizando el *dataset* de la investigación. Para ello se dividieron los registros en el *set* de entrenamiento y prueba, con el 70% de los datos (522 registros) y 30% de los datos (224 registros) respectivamente cada uno, esto con la opción de *Percentage Split*.

En la figura 113 se presenta el árbol de decisión obtenido a partir de los parámetros anteriormente descritos, donde el árbol está constituido por 17 hojas o nodos terminales y el tamaño del árbol es de 33 nodos (ver figura 113).

Figura 113

Árbol de decisión obtenido en la prueba uno con *Weka*.

```
item15 <= 3
|  item10 <= 3
|  |  item20 <= 2
|  |  |  item7 <= 3
|  |  |  |  ítem18 <= 1: Autorregulador (9.0/1.0)
|  |  |  |  ítem18 > 1: Procrastinador (5.0/1.0)
|  |  |  |  item7 > 3
|  |  |  |  |  Edad <= 23: Autorregulador (50.0)
|  |  |  |  |  Edad > 23
|  |  |  |  |  |  ítem8 <= 2: Autorregulador (3.0)
|  |  |  |  |  |  ítem8 > 2: Procrastinador (2.0)
|  |  |  |  |  item20 > 2
|  |  |  |  |  |  ítem12 <= 3: Procrastinador (15.0)
|  |  |  |  |  |  ítem12 > 3
|  |  |  |  |  |  |  ítem2 <= 3
|  |  |  |  |  |  |  |  PromedioBachillerato <= 1: Autorregulador (2.0)
|  |  |  |  |  |  |  |  PromedioBachillerato > 1: Procrastinador (5.0/1.0)
|  |  |  |  |  |  |  |  ítem2 > 3: Autorregulador (5.0)
|  |  |  |  |  |  |  ítem10 > 3: Autorregulador (570.0/15.0)
|  item10 > 3: Autorregulador (570.0/15.0)
item15 > 3
|  item18 <= 3
|  |  item2 <= 3
|  |  |  ítem6 <= 4: Procrastinador (4.0)
|  |  |  ítem6 > 4
|  |  |  |  DivisiónAcadémica <= 6: Autorregulador (7.0/1.0)
|  |  |  |  DivisiónAcadémica > 6: Procrastinador (3.0)
|  |  |  |  ítem2 > 3: Autorregulador (34.0/1.0)
|  |  item18 > 3
|  |  |  PromedioBachillerato <= 2: Procrastinador (25.0/3.0)
|  |  |  PromedioBachillerato > 2
|  |  |  |  ítem7 <= 4: Autorregulador (5.0)
|  |  |  |  ítem7 > 4: Procrastinador (2.0)
```

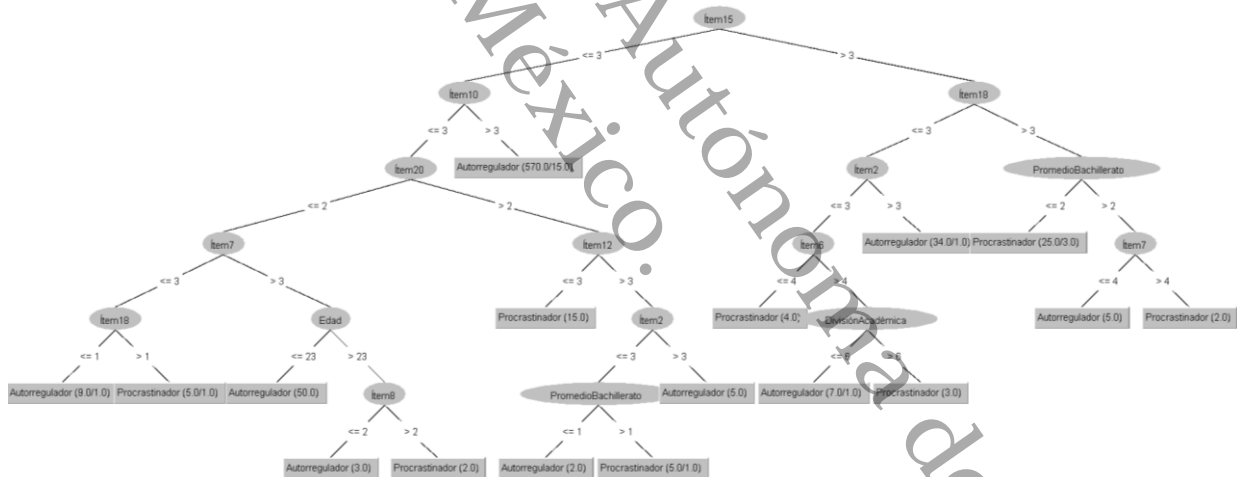
Nota: Elaboración propia.

Aunado a la figura anterior, se encontraron patrones clasificatorios sobre si el alumno es procrastinador o autorregulador. Algunas variables relevantes para identificar estos patrones fueron: ítem 15, ítem 10, promedio del bachillerato, ítem 12, entre otros. Además, se obtuvo una la representación gráfica (diagrama) del árbol de decisión de la prueba uno.

En la figura 114 se muestra el diagrama de árbol de decisión obtenido en esta prueba, donde se pueden visualizar de forma más clara y gráfica los patrones clasificatorios y los nodos del árbol (ver figura 114).

Figura 114

Diagrama del árbol de decisión obtenido en la prueba uno con Weka.



Nota: Elaboración propia.

De igual forma se obtuvo que el porcentaje de instancias clasificadas correctamente fue de 88.8393% (199 instancias) y a su vez el 11.1607% (25 instancias) fueron clasificados de forma incorrecta, esto se obtuvo a partir del uso del set de prueba. A continuación, en la figura 115 se muestra la matriz de confusión dada por el algoritmo,

donde se determina que 184 instancias de la clase autorregulador y 15 instancias de la clase procrastinador fueron clasificadas correctamente (ver figura 115).

Figura 115

Matriz de confusión de la prueba uno con Weka.

```
=== Confusion Matrix ===
      a    b  <-- classified as
184  11 |   a = Autorregulador
  14  15 |   b = Procrastinador
```

Nota: Elaboración propia.

Para la segunda prueba realizada con el algoritmo al *dataset* se modificó el porcentaje de datos destinados para cada *set*, en el caso del *set* de entrenamiento el 60% de los datos (448 registros) y el *set* de prueba tenía el 40% restante de los datos (298 registros).

Una vez ejecutado el algoritmo, se obtuvo el siguiente árbol de decisión, el cual se muestra en la figura 116, donde se identificó que el árbol se encuentra constituido por el mismo número de hojas o nodos terminales y tamaño de árbol, es decir, 17 hojas o nodos terminales y el tamaño del árbol de 33 nodos (ver figura 116).

Figura 116
Árbol de decisión obtenido en la prueba dos con Weka.

```

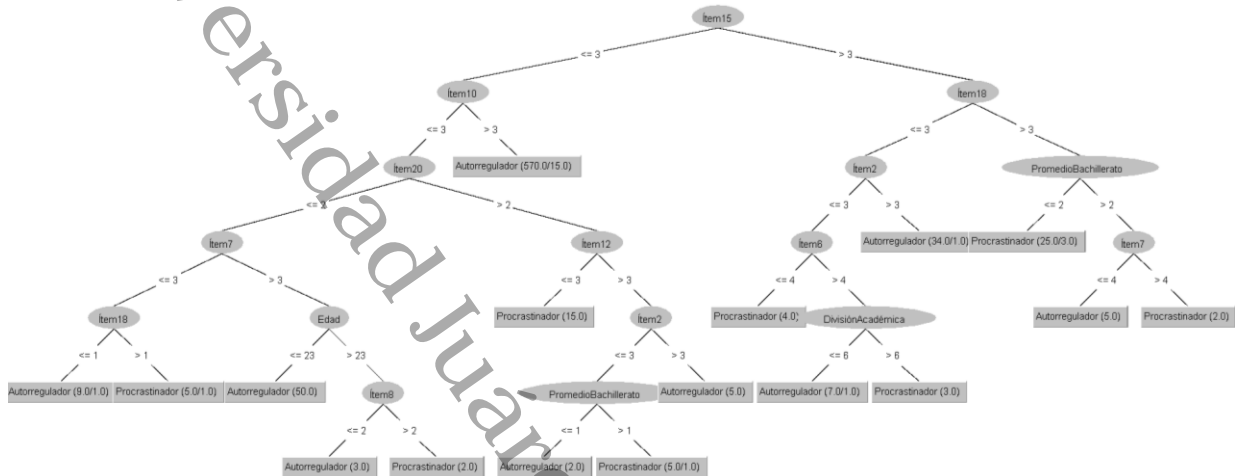
  ítem15 <= 3
  |
  |   ítem10 <= 3
  |   |
  |   |   ítem20 <= 2
  |   |   |
  |   |   |   ítem7 <= 3
  |   |   |   |
  |   |   |   |   ítem18 <= 1: Autorregulador (9.0/1.0)
  |   |   |   |   ítem18 > 1: Procrastinador (5.0/1.0)
  |   |   |   |
  |   |   |   |   ítem7 > 3
  |   |   |   |   |
  |   |   |   |   |   Edad <= 23: Autorregulador (50.0)
  |   |   |   |   |   Edad > 23
  |   |   |   |   |   |
  |   |   |   |   |   |   ítem8 <= 2: Autorregulador (3.0)
  |   |   |   |   |   |   ítem8 > 2: Procrastinador (2.0)
  |   |   |   |   |
  |   |   |   |   |   ítem20 > 2
  |   |   |   |   |   |
  |   |   |   |   |   |   ítem12 <= 3: Procrastinador (15.0)
  |   |   |   |   |   |   ítem12 > 3
  |   |   |   |   |   |   |
  |   |   |   |   |   |   |   ítem2 <= 3
  |   |   |   |   |   |   |   |
  |   |   |   |   |   |   |   |   PromedioBachillerato <= 1: Autorregulador (2.0)
  |   |   |   |   |   |   |   |   PromedioBachillerato > 1: Procrastinador (5.0/1.0)
  |   |   |   |   |   |   |   |
  |   |   |   |   |   |   |   |   ítem2 > 3: Autorregulador (5.0)
  |   |   |   |   |   |   |
  |   |   |   |   |   |   |   ítem10 > 3: Autorregulador (570.0/15.0)
  |   |   |   |   |
  |   |   |   |   |   ítem15 > 3
  |   |   |   |   |   |
  |   |   |   |   |   |   ítem18 <= 3
  |   |   |   |   |   |   |
  |   |   |   |   |   |   |   ítem2 <= 3
  |   |   |   |   |   |   |   |
  |   |   |   |   |   |   |   |   ítem6 <= 4: Procrastinador (4.0)
  |   |   |   |   |   |   |   |   ítem6 > 4
  |   |   |   |   |   |   |   |   |
  |   |   |   |   |   |   |   |   |   DivisiónAcadémica <= 6: Autorregulador (7.0/1.0)
  |   |   |   |   |   |   |   |   |   DivisiónAcadémica > 6: Procrastinador (3.0)
  |   |   |   |   |   |   |   |   |
  |   |   |   |   |   |   |   |   |   ítem2 > 3: Autorregulador (34.0/1.0)
  |   |   |   |   |   |   |   |
  |   |   |   |   |   |   |   |   |   ítem18 > 3
  |   |   |   |   |   |   |   |   |   |
  |   |   |   |   |   |   |   |   |   |   PromedioBachillerato <= 2: Procrastinador (25.0/3.0)
  |   |   |   |   |   |   |   |   |   |   PromedioBachillerato > 2
  |   |   |   |   |   |   |   |   |   |   |
  |   |   |   |   |   |   |   |   |   |   |   ítem7 <= 4: Autorregulador (5.0)
  |   |   |   |   |   |   |   |   |   |   |   ítem7 > 4: Procrastinador (2.0)

```

Nota: Elaboración propia.

Aunado a la figura anterior se encontraron patrones clasificatorios sobre si el alumno es procrastinador o autorregulador, se identificó que el árbol de decisión obtenido es similar al obtenido durante la realización de la prueba uno; de igual forma se muestra en la figura 117 la representación gráfica (diagrama) del árbol de decisión de la prueba dos, para visualizar claramente los patrones clasificatorios (ver figura 117).

Figura 117
Diagrama del árbol de decisión obtenido en la prueba dos con Weka.



Nota: Elaboración propia.

El porcentaje de instancias clasificadas correctamente fue de 90.2685% (269 instancias de un total de 298 instancias) y el otro 9.7315% (29 instancias) restante fueron clasificados de forma incorrecta, esto a partir del set de prueba. A continuación, en la figura 118 se muestra la matriz de confusión, donde se determina que 247 instancias de la clase autorregulador y 22 instancias de la clase procrastinador fueron clasificadas correctamente (ver figura 115).

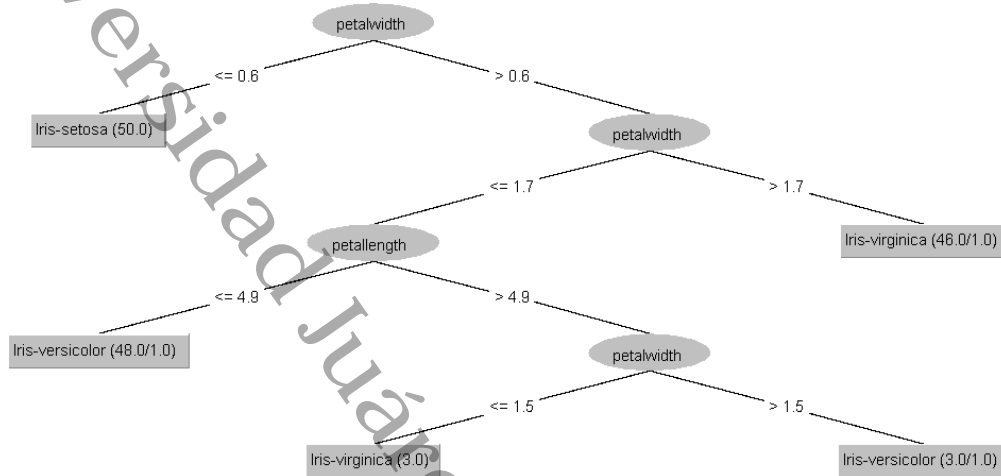
Figura 118
Matriz de confusión de la prueba dos con Weka.

```

=== Confusion Matrix ===
      a    b  <-- classified as
 247  14 |   a = Autorregulador
  15  22 |   b = Procrastinador
    
```

Nota: Elaboración propia.

Figura 120
Diagrama del árbol de decisión obtenido en la prueba uno con Weka.



Nota: Elaboración propia.

Finalmente, el porcentaje de instancias clasificadas correctamente en esta prueba fue del 95.5556% (43 instancias de un total de 45 instancias) y el otro 4.4444% (2 instancias) restante fueron clasificados de forma incorrecta, esto a partir del set de prueba. A su vez, la figura 121 muestra la matriz de confusión, donde se determina que 14 instancias de la clase setosa, 16 instancias de la clase versicolor y 13 instancias de la clase virgínica fueron clasificadas correctamente (ver figura 121).

Figura 121
Matriz de confusión de la prueba tres con Weka.

```

=== Confusion Matrix ===
      a  b  c  <-- classified as
14  0  0 |  a = Iris-setosa
 0 16  0 |  b = Iris-versicolor
 0  2 13 |  c = Iris-virginica
    
```

Nota: Elaboración propia.

Estas tres pruebas realizadas por cada una de las herramientas y la técnica de minería de datos de árbol de decisión sirvieron para poder entender y verificar de una mejor manera la forma en que se trabaja esta técnica desarrollada en cada herramienta y de igual forma a partir de las pruebas se pudieron establecer los parámetros a utilizar para la obtención de los resultados de la investigación.

4.1.2 Pruebas de Agrupamiento

Las pruebas que se realizaron con esta técnica de minería de datos son las siguientes: la primera y segunda prueba se realizaron al *dataset* de la investigación, para el agrupamiento de los datos y por último la tercera prueba se realizó implementando la técnica de agrupamiento en el *dataset* iris, donde se clasificó la especie a la que pertenece la flor.

Estas tres pruebas se realizaron en cada una de las herramientas seleccionadas anteriormente, las cuales son: Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*. En caso de esta técnica se normalizaron los datos que contiene el *dataset* de investigación y el de prueba, así como la selección del número de *clúster* a utilizar en cada prueba.

El número de *clúster* en la primera prueba fue de tres *clústeres*, para la segunda prueba fue de cuatro *clústeres* y por último en la prueba tres el número de *clúster* fue dos *clústeres*. A continuación, se describen detalladamente estas tres pruebas realizadas.

4.1.2.1 Lenguaje R

La primera prueba consistió en visualizar cómo es el comportamiento de los datos al construir tres agrupamientos o *clúster*, pero antes de crear los *clúster*, se normalizaron los datos para que ningún data tenga mayor peso que otro al momento de realizar el agrupamiento.

Una vez normalizado los datos, se realizó la creación e implementación del modelo de agrupamiento con tres *clústeres*. En la figura 122 se visualiza en la parte superior información sobre el número de *clustering* o agrupamientos creados e implementados en los datos, así como el tamaño o número de registros que contiene cada agrupamiento (ver figura 122).

Figura 122
Modelo de agrupamiento con tres *clústeres* de la prueba uno en Lenguaje R.

```
K-means clustering with 3 clusters of sizes 363, 32, 351

Cluster means:
  Género      Edad  CicloEscolar  DivisiónAcadémica  AsignaturaCursa  PromedioBachillerato  Item1  Item2  Item3
1 -0.18366887 -0.069911404  0.01687422      0.1496044         -0.04416580         0.1537286  0.44885575 -0.3328970  0.1721671
2  0.05154285  0.714525511      -0.37973482      0.3537141         0.34048837         0.1385799  0.02637587 -0.8653477 -0.3631187
3  0.18524908  0.007159611      0.01716858      -0.1869665         0.01464136         0.1716183 -0.46660588  0.4229878 -0.1449483

  Item4  Item5  Item6  Item7  Item8  Item9  Item10  Item11  Item12  Item13  Item14
1  0.09058410 -0.2866184  0.04548884 -0.2024998  0.51813196  0.5382499 -0.2171516 -0.2926007  0.2662379 -0.2880214 -0.2227904
2 -0.38834105 -0.8926581  -3.18607966 -1.3862940 -0.06434454 -0.1189157 -1.1216743 -1.3081658 -1.5267605 -1.2046744 -1.2964916
3 -0.05827668  0.3777993  0.24342479  0.3358087 -0.52997971 -0.5458103  0.3268365  0.4218671  0.4145319  0.4076961  0.3486058

  Item15  Item16  Item17  Item18  Item19  Item20  Item21  Item22
1  0.5219932  0.50903779  0.4675210  0.5357282  0.5656016  0.45752137  0.50878990  0.28857648
2 -0.1502222 -0.05207903 -0.6558305 -0.4631313 -0.1113178  0.06144336 -0.06245297  0.08044361
3 -0.5261436 -0.52169284 -0.4237138 -0.5118209 -0.5747898 -0.52049071 -0.30577623

Clustering vector:
[1] 3 3 3 3 3 1 3 1 3 3 3 3 3 1 1 3 3 1 3 3 3 3 1 3 3 3 3 1 1 1 3 1 1 3 1 1 1 3 3 3 1 3 3 1 3 1 1 3 3 3 1 3 1 1 3 3 3 1 3 1 1 3 3 3 1 3 1 2 1 3 3 3 3 1 3 1
[64] 1 1 1 1 3 1 1 1 1 3 3 1 1 1 1 1 3 3 3 3 3 3 3 1 3 1 1 3 3 3 1 1 1 1 3 3 2 1 3 1 3 1 1 3 3 3 1 3 3 3 3 3 1 1 1 1 1 1 2 3
[127] 1 3 1 3 3 3 3 3 1 1 2 2 3 2 3 1 3 1 3 2 1 3 1 3 3 1 3 3 1 1 1 2 1 3 3 3 3 3 3 1 3 3 3 1 1 3 3 1 3 3 3 3 3 3 3 3 3 3 3 1 3 3 1 3 3 1 3 3 1 3 1
[190] 1 3 1 3 1 1 1 1 3 2 1 1 3 3 1 3 1 1 3 3 3 3 3 1 3 1 1 3 3 1 3 3 3 1 3 3 1 1 1 3 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1
[253] 1 1 1 1 1 2 1 3 1 1 1 1 3 3 1 3 1 1 3 3 1 3 3 3 1 3 3 3 1 1 3 3 3 3 1 3 1 3 3 1 3 1 1 3 1 1 3 1 1 1 3 3 1 3 3 3 3 3 3 1 3 3 1 3 3 1 3 3 1 3 3 1 3 3 1 3
[316] 1 1 1 3 3 3 1 3 3 1 3 1 3 1 3 1 3 1 3 3 3 1 3 1 3 3 3 3 1 3 3 3 3 1 1 3 1 3 3 1 1 1 3 3 1 1 3 3 1 1 3 3 1 1 3 3 1 1 3 3 1 1 3 3 1 1 3 3 1 1 1 1
[379] 3 1 1 3 1 1 1 3 1 3 3 3 1 1 3 1 1 1 1 3 3 1 3 2 1 3 1 3 3 3 3 1 3 3 3 1 3 1 2 1 1 3 1 3 2 2 2 3 3 3 1 3 3 1 3 3 3 1 3 3 1 3 3 3 1 3 3 1 3 3 1 2 1
[442] 3 1 3 3 1 3 1 1 3 1 3 1 1 3 3 1 1 3 3 1 1 3 3 3 3 1 2 1 1 3 1 1 3 2 1 3 3 1 3 1 1 3 3 1 3 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 3 3 1 3 3 1
[505] 1 3 1 1 1 1 3 1 1 3 1 1 3 1 1 1 1 3 3 1 3 2 3 1 1 1 3 3 3 1 1 1 1 1 1 1 1 1 1 1 3 3 1 3 1 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 1 1 3 3 1 3 1 2
[568] 3 2 3 2 1 3 1 3 1 1 1 1 3 1 1 1 3 3 1 1 1 3 3 1 1 3 3 3 3 3 1 1 1 3 1 1 3 3 3 3 1 1 1 3 1 1 3 3 3 3 1 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 3 1 1
[631] 1 1 3 1 1 1 1 1 3 3 3 3 1 3 1 1 3 3 2 1 3 3 3 3 1 3 1 1 1 1 2 1 1 3 1 3 3 1 2 1 1 3 3 1 2 1 1 3 3 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 3 3 1 3 1 1
[694] 3 3 3 1 3 2 1 1 1 3 1 3 1 2 1 1 1 1 3 2 3 3 3 3 1 3 3 3 2 3 1 3 3 1 1 3 2 3 1 3 1 1 1 1 1 3 3 3 3 1 3 1 1 1 1 1 3 3 3 3 1 3 1 1 1 1 3 3 3 3 1 3 1 1
```

Nota: Elaboración propia.

Para esta prueba con tres *clústeres*, se obtuvieron los siguientes resultados: el primer agrupamiento está conformado por 363 registros, el segundo por 32 registros y el tercer

agrupamiento contiene 351 registros, y a su vez se muestra el vector de agrupamiento de los 746 registros de acuerdo con el número de *clúster* al que pertenecen.

Con el modelo de agrupamiento ya creado y con los datos clasificados de acuerdo con los *clústeres*, se creó un gráfico para visualizar de una forma más clara los agrupamientos de los datos.

En la figura 123 se muestra el gráfico obtenido a partir del modelo de agrupamiento con tres *cluster*, los cuales son representados en colores rojo, azul y verde, por medio del gráfico se aprecia que la mayoría de los registros se encuentran aglomerados en el centro del diagrama, en su mayoría son datos pertenecientes al *cluster* uno y dos, en comparación con el *cluster* tres que los datos pertenecientes a él, hay algunos datos aglomerados en el centro del gráfico, pero también hay algunos datos con mayor distancia al centro del modelo (ver figura 123).

Figura 123

Gráfico de los tres agrupamientos de la prueba uno en Lenguaje R.



Nota: Elaboración propia.

Lo siguiente que se realizó fue visualizar la asignación de cada registro o instancia a los *clústeres* creados, esto se presenta en la figura 124 y se aprecian los tres grupos de *clústeres* que se encuentran en la clase de estudiante autorregulador, por el contrario, para la clase procrastinador solo se encuentra en dos *clústeres* (ver figura 124).

Figura 124

Visualización de la asignación de las instancias en los clústeres de acuerdo con la variable ClaseEstudiante con la prueba uno en Lenguaje R.

	cluster		
	1	2	3
Autorregulador	298	23	351
Procrastinador	65	9	0

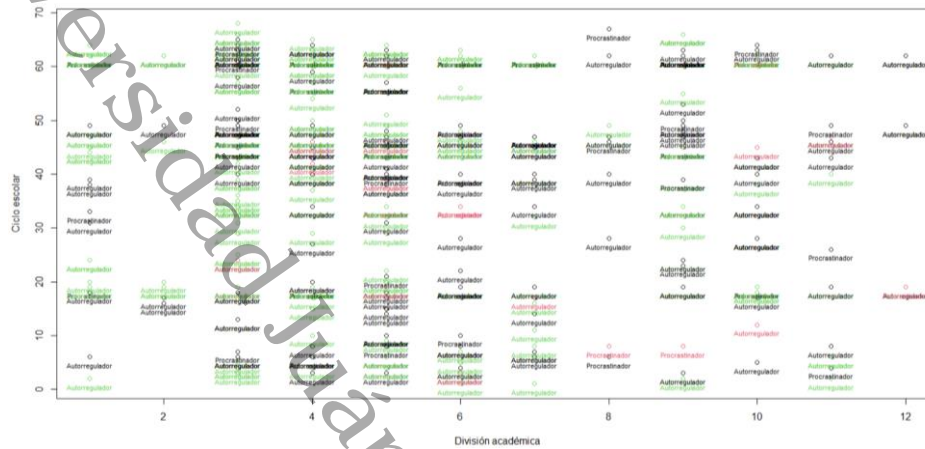
Nota: Elaboración propia.

Finalmente se realizó otra gráfica para identificar patrones o similitudes de conducta procrastinadora o autorreguladora de los estudiantes, esta gráfica se realizó de acuerdo con las variables de División Académica en el eje X y ciclo escolar en el eje Y, y los datos graficados son los agrupamientos obtenidos con las etiquetas de si el alumno es procrastinador o autorregulador.

Lo anteriormente descrito se muestra en la figura 125, en la cual se observa que la mayoría de los estudiantes son autorreguladores y se encuentran muy pocos que son procrastinadores, pero aun así se encuentra que hay divisiones en que todos son autorreguladores y no procrastinadores; así como esta gráfica se pueden realizar otras (ver figura 125).

Figura 125

Visualización de los grupos de agrupamiento de la prueba uno en Lenguaje R.



Nota: Elaboración propia.

La segunda prueba fue realizada implementando cuatro *clústeres* al *dataset*. La ejecución de los cuatro *clústeres* dio como resultado para el primer *clúster* 169 registros, segundo *clúster* 21 registros, tercer *clúster* 257 registros y el cuarto *clúster* 299 registros. Lo anterior descrito se muestra en la figura 126, en la cual también se muestra el vector de agrupamiento de los 746 registros, pero representados con el número del *clúster* al que pertenece cada registro (ver figura 126).

Figura 126
Modelo de agrupamiento con cuatro clústeres de la prueba dos en Lenguaje R.

```

K-means clustering with 4 clusters of sizes 169, 21, 257, 299

Cluster means:
  Género      Edad  CicloEscolar  DivisiónAcadémica  AsignaturaCursa  PromedioBachillerato  Item1  Item2  Item3
1 -0.29050395 -0.06454739 -0.05842714  0.15286609  0.03306442  0.31428664  0.58454739 -0.2277364  0.29650069
2 -0.02551937  1.03235925 -0.37919668  0.42032657  0.35138169  0.01638061 -0.05629808 -1.1272748 -0.33489939
3  0.22357449  0.01233450  0.01653919 -0.21045478  0.05078950 -0.18522450 -0.65592555  0.6191330 -0.20606416
4 -0.02617917 -0.04662543  0.04544062  0.06496874 -0.08702275 -0.01958441  0.23734654 -0.3242707  0.03305271

  Item4  Item5  Item6  Item7  Item8  Item9  Item10  Item11  Item12  Item13  Item14  Item15
1  0.24915908 -0.1831565  0.01981783 -0.0703540  0.7070627  0.7462132 -0.1690831 -0.1186683 -0.1564846 -0.2076654 -0.2533067  0.8199980
2 -0.36537132 -1.3645422 -4.27295958 -1.7333047  0.3029053  0.0555365 -1.2714169 -1.3511517 -1.4024266 -1.2791307 -1.6287623 -0.2608863
3 -0.06540306  0.5340111  0.25520986  0.5372409 -0.6844452 -0.6932329  0.5241870  0.5828909  0.5884635  0.6399288  0.5097539 -0.6616307
4 -0.05895150 -0.2596389  0.06954517 -0.3002732  0.1673840  0.1701824 -0.2656899 -0.3390429 -0.3188571 -0.3428245 -0.1805816  0.1235386

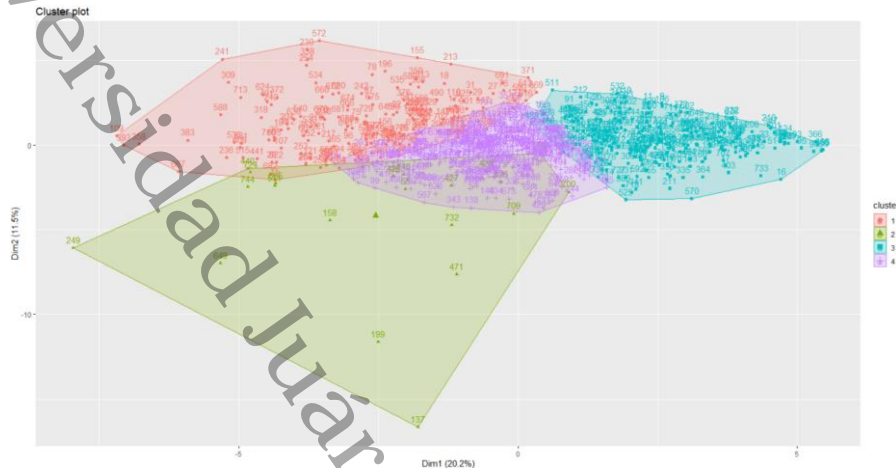
  Item16  Item17  Item18  Item19  Item20  Item21  Item22
1  0.91691348  0.86903125  1.1029319  1.14401434  1.0682881  1.1726786  0.43341341
2 -0.13012383 -0.42170227 -0.3091670 -0.08126894  0.2344791  0.3478728  0.23431781
3 -0.56152332 -0.48565385 -0.5440075 -0.62939801 -0.4939896 -0.5383546 -0.38836376
4 -0.02646918 -0.04413877 -0.1340905 -0.09992136 -0.1956837 -0.2245180  0.07238109

Clustering vector:
[1] 3 3 4 4 3 4 3 1 3 3 3 3 3 4 1 3 3 1 4 3 3 1 3 3 4 4 1 1 1 4 1 4 3 4 4 1 4 3 4 4 3 4 3 1 1 3 1 3 3 3 4 3 4 2 4 3 3 3 3 4 3 4 4 4 1 4
[68] 4 4 4 4 4 4 4 1 1 4 1 1 3 3 4 3 3 4 4 1 4 4 4 3 1 3 3 3 1 1 4 4 3 4 4 4 3 3 1 4 4 3 4 4 4 3 3 1 4 1 1 3 3 4 1 4 3 3 4 3 4 1 4 1 4 2 3 1 3 4 3 3 3 3
[135] 4 4 2 4 4 4 3 4 3 4 3 2 1 4 4 3 3 1 3 3 1 1 4 2 1 4 3 4 4 3 3 1 3 3 3 4 1 4 3 1 3 3 3 3 3 3 4 1 3 3 4 3 1 1 4 4 3 4 1 1 4 3 2 4 4
[202] 3 3 1 3 1 4 3 3 3 3 1 4 1 1 1 3 4 4 3 3 3 4 3 4 4 1 1 1 3 1 3 3 4 1 4 3 3 1 1 1 3 4 4 4 4 4 2 4 4 4 1 1 1 1 1 2 4 3 1 1 1 1 4 3 1
[269] 3 1 4 3 4 3 4 3 4 3 4 1 4 4 3 3 3 4 4 3 1 4 4 4 4 3 4 4 4 1 4 4 1 4 3 4 4 4 1 4 4 3 1 4 3 4 4 4 1 3 3 4 4 3 3 4 3 4 3 1 4 4 3 3 4 3 3 4
[336] 4 3 1 4 4 3 3 4 3 4 4 3 3 3 4 1 4 3 4 3 3 4 1 1 4 4 1 3 1 3 3 1 3 3 1 1 3 3 1 1 4 4 3 1 4 3 1 4 4 1 3 4 3 3 1 1 4 4 4 4 1 3 3 4
[403] 3 4 4 3 4 3 3 3 3 4 3 4 4 4 4 4 4 4 4 4 4 4 3 4 4 2 2 3 3 4 4 4 3 3 3 4 2 1 3 1 3 4 4 3 4 4 3 4 4 1 4 3 3 4 4 4 3 1 4 4 4 3 1 4 4 4 3 4 3
[470] 4 2 4 4 3 1 4 3 4 4 3 4 4 4 4 1 4 4 3 3 1 1 4 3 3 4 1 4 4 3 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
[537] 1 4 1 1 1 4 1 3 3 1 3 4 3 4 1 1 4 4 4 4 4 3 4 4 3 4 4 4 3 1 4 4 3 4 2 1 3 4 3 4 1 4 1 3 1 1 4 4 3 3 4 1 1 1 4 3 1 1 3 3 3 4 3 3 3 1
[604] 4 4 3 4 1 3 4 4 3 4 3 1 1 4 4 4 3 1 1 1 1 1 4 4 4 1 4 4 3 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
[671] 4 1 4 1 1 3 4 4 1 1 1 1 1 4 4 3 3 4 3 1 3 4 3 4 3 4 4 3 2 1 1 1 4 3 4 3 4 2 1 1 1 1 3 4 3 3 3 1 3 4 3 2 4 1 3 4 1 4 3 2 3 4 3 4 1
[738] 1 4 3 4 3 2 3 4
    
```

Nota: Elaboración propia.

De acuerdo con los agrupamientos creados y distribuidos en cada uno de los cuatro clústeres, se creó un gráfico para la visualización más grafica los agrupamientos de los datos; la figura 127 se presenta el gráfico obtenido a partir del modelo de agrupamiento con cuatro clústeres, los cuales se representan por los colores rojo, azul, verde y morado (ver figura 127).

Figura 127
Gráfico de los cuatro agrupamientos de la prueba dos en Lenguaje R.



Nota: Elaboración propia.

Por medio del gráfico se aprecia que la mayoría de los registros se encuentran aglomerados en el centro del diagrama, en su mayoría los datos del *clúster* cuatro se encuentran agrupados en el centro y cercanos entre sí, a diferencia de los *clústeres* uno y tres (rojo y azul respectivamente) están conformados con datos con distancias cercanas, pero con el *clúster* uno se comienzan a visualizar datos con mayor distancia que al resto que conforman este agrupamiento y por último en el *clúster* dos se encuentran datos donde la distancia entre cada dato es mucho más lejana.

En la figura 128 se visualizan los cuatro grupos de *clústeres*, en donde se aprecia el número de instancias o registros asignados de cada clase en los *clústeres*; la clase autorregulador se distribuye en los cuatro *clústeres* y la clase procrastinador se distribuye solo en los *clústeres* uno, dos y cuatro (ver figura 128).

Figura 128
Visualización de la asignación de las instancias en los clústeres de acuerdo con la variable ClaseEstudiante con la prueba tres en Lenguaje R.

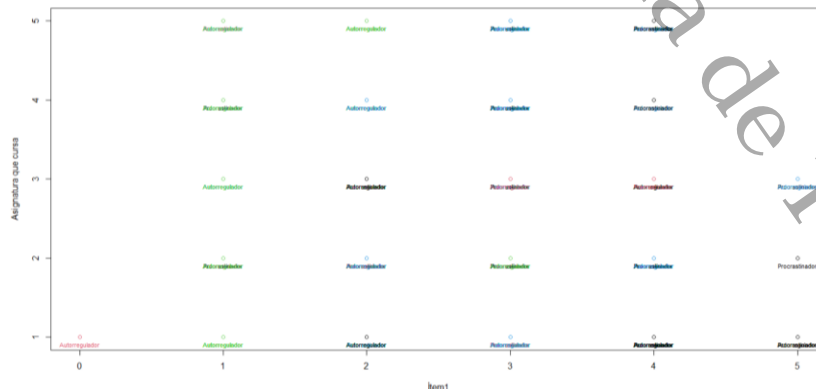
	cluster			
	1	2	3	4
Autorregulador	115	11	257	289
Procrastinador	54	10	0	10

Nota: Elaboración propia.

Por último, con la prueba dos se realizó una exploración de los datos de las variables en relación con los grupos de agrupamientos creados en el *dataset*, además se pueden identificar patrones entre los datos de acuerdo con la variable de ClaseEstudiante, por ejemplo, ya que hay datos que pueden coincidir cuando los dos son autorreguladores o procrastinadores.

La figura 129 muestra un ejemplo de la exploración de los datos en relación con los agrupamientos de los datos, esto utilizando las variables de Ítem 1 en eje X y AsignaturaCursa eje Y, y el etiquetado de la clase (ver figura 129).

Figura 129
Visualización de los grupos de agrupamiento de la prueba dos en Lenguaje R.

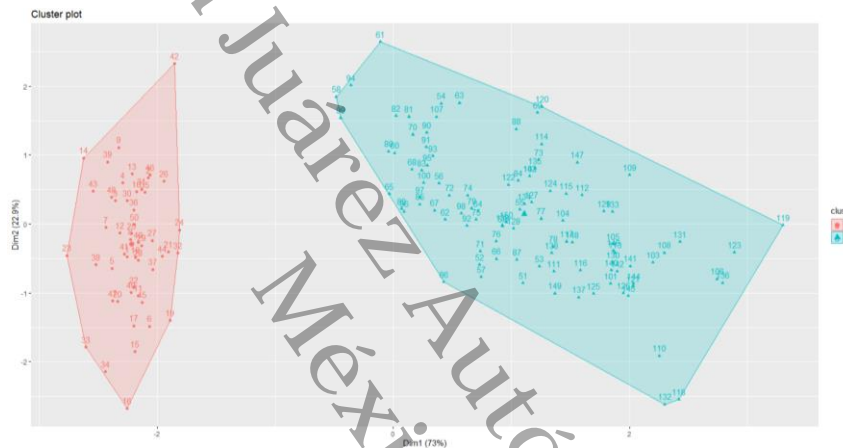


Nota: Elaboración propia.

Después de la implementación del algoritmo en el *dataset* iris, se realizó un gráfico para la visualización más clara de los agrupamientos de los datos; la figura 131 presenta el gráfico obtenido a partir del modelo de agrupamiento con dos *clústeres* en el *dataset* iris, los cuales se representan por los colores rojo y azul respectivamente (ver figura 131).

Figura 131

Gráfico de los dos agrupamientos de la prueba tres en Lenguaje R.



Nota: Elaboración propia.

Por medio del gráfico se aprecia que entre cada agrupamiento se encuentran con una distancia claramente considerable, los registros del grupo uno se encuentra del lado izquierdo del gráfico y hay registros aglomerados en el centro del grupo con una distancia cercana, pero también hay registros con distancias lejanas. Dentro del grupo dos hay más registros con distancias más lejanas entre datos y unos cuantos están con distancias cercanas

Posteriormente se procedió a visualizar la asignación de cada registro de los *clústeres* creados, la figura 132 muestra lo anteriormente descrito y se visualizaron los dos grupos

de *clústeres* donde el *clúster* uno solo contiene registros de la clase setosa y el *clúster* dos contiene registros de las clases versicolor y virgínica (ver figura 132).

Figura 132

Visualización de la asignación de las instancias en los *clústeres* de acuerdo con la variable *Especie* con la prueba tres en Lenguaje R.

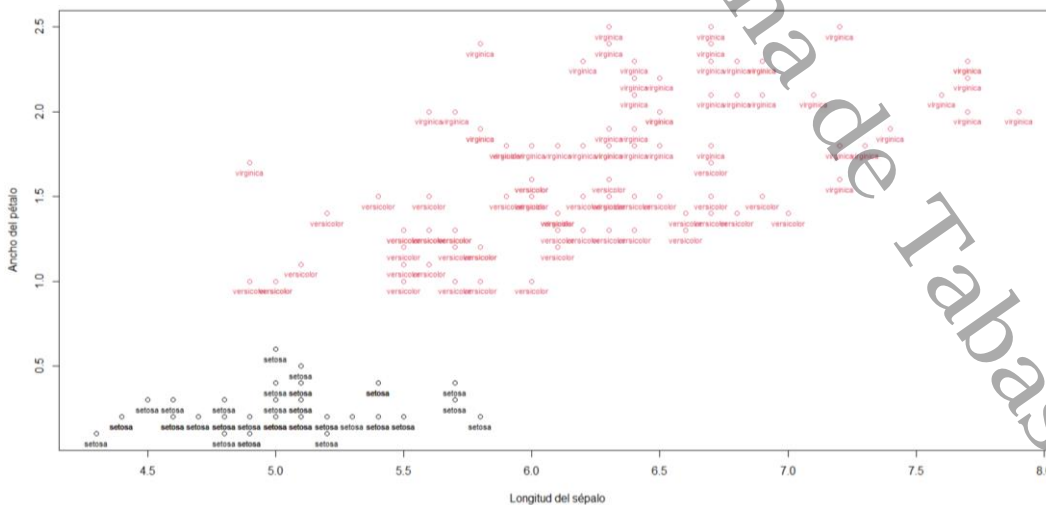
	cluster	
	1	2
setosa	50	0
versicolor	0	50
virginica	0	50

Nota: Elaboración propia.

Para finalizar con la prueba tres, se realizó la exploración de los datos utilizando el modelo *K-means* de agrupamiento. La figura 133 muestra un ejemplo de la exploración de los datos en relación con los agrupamientos de los datos, esto utilizando las variables de sepal length (longitud del sépalo) en eje X y petal width (ancho del pétalo) eje Y, y el etiquetado de la clase especie (ver figura 133).

Figura 133

Visualización de los grupos de agrupamiento de la prueba tres en Lenguaje R.



Nota: Elaboración propia.

En la gráfica se muestra que los datos de estas dos variables se encuentran dispersos y se visualizan los dos agrupamientos por separado y hay datos con distancias cercanas y otros con distancias lejanas con los demás datos.

4.1.2.2 Microsoft Azure Machine Learning

Para la implementación de la técnica de agrupamiento en el *dataset*, se realizó la normalización de los datos; para la primera prueba realizada del modelo creado con el algoritmo *K-means* fue la ejecución de tres *clúster* o grupo de agrupamiento.

En la figura 134 se visualiza el resultado obtenido de la prueba uno realizada, donde se aprecia que al *dataset* se le asignaron cuatro columnas más, una de ellas contiene la asignación de los *clústeres* y las otras tres la distancia en que se encuentran los *clústeres* con relación al centro del grupo o centroide (ver figura 134).

Figura 134

Agrupamiento de los tres clústeres de la prueba uno en Microsoft Azure Machine Learning.

Ítem21	Ítem22	ClaseEstudiante	Assignments	DistancesToClusterCenter no.0	DistancesToClusterCenter no.1	DistancesToClusterCenter no.2	
26	-0.750993	1.055688	Autorregulador	2	1.470125	1.620564	1.086136
26	-0.750993	-0.586738	Autorregulador	2	1.37478	1.515283	0.997113
26	-0.750993	0.234475	Autorregulador	2	1.086881	1.423447	0.964456
26	-0.750993	-0.586738	Autorregulador	2	1.197835	1.435208	0.803535
26	-0.750993	1.055688	Autorregulador	2	1.324743	1.472722	1.029404
9	1.347287	0.234475	Autorregulador	1	1.293514	0.822059	1.35839
9	0.298147	1.055688	Autorregulador	2	1.253086	1.44309	1.036277
4	0.298147	0.234475	Autorregulador	0	0.613361	1.300951	1.051069
9	-0.750993	-1.407951	Autorregulador	1	1.573757	1.073134	1.353383
26	-0.750993	0.234475	Autorregulador	2	1.457835	1.577952	0.92647
9	-0.750993	-1.407951	Autorregulador	1	1.565453	1.069536	1.364384
26	-0.750993	-0.586738	Autorregulador	1	1.632057	1.109703	1.370959

Nota: Elaboración propia.

Como se aprecia en la figura anterior, la herramienta asignó el agrupamiento o *clúster* del rango 0 al 2, pero son tres grupos de *clústeres* que realizaron. De igual forma se visualiza que mientras más cerca se encuentra la instancia o los datos del registro, centroide o centro del grupo del *clúster*, es al grupo que pertenecerá o asignará el algoritmo.

Por ejemplo, en el caso de la instancia dos, está asignada al *clúster* dos (grupo tres), ya que la distancia que tiene al centro del grupo es 0.997113, la cual es menor que de los otros dos *clústeres*, ya que la distancia del *clúster* 0 es del 1.37478 y la del *clúster* 1 es de 1.515283.

Aunado a lo anterior, se extrajo una evaluación de los resultados de los grupos de agrupamiento, las métricas que componen esta evaluación son: distancia media a otro centro, distancia promedio al centro del grupo, número de instancias agrupadas a cada *clúster* y la distancia máxima al centro del grupo.

En la figura 135 se muestran los resultados de dicha evaluación a los grupos de *clústeres*, donde se aprecia que el *clúster* 0 (grupo uno) contiene 178 instancias agrupadas, el *clúster* 1 (grupos dos) con 275 instancias y el *clúster* 2 (grupo tres) 293 instancias (ver figura 135).

Figura 135
Evaluación de la asignación de clúster con la prueba uno con Microsoft Azure Machine Learning.

Result Description	Average Distance to Other Center	Average Distance to Cluster Center	Number of Points	Maximal Distance to Cluster Center
Evaluation For Cluster No.0	1.268614	1.016187	178	1.723338
Evaluation For Cluster No.1	1.433021	1.068464	275	1.835802
Evaluation For Cluster No.2	1.262182	0.99223	293	3.068032
Combined Evaluation	1.326694	1.026048	746	3.068032

Nota: Elaboración propia.

La segunda prueba fue realizada implementando cuatro *clústeres* o grupo de agrupamiento al *dataset*, después de la normalización de los datos. En la figura 136 se visualiza el resultado obtenido de la prueba dos realizada con cuatro *clústeres*, donde se aprecia que al *dataset* se le asignaron cinco columnas más, una de ellas contiene la asignación de los *clústeres* y las otras cuatro la distancia en que se encuentran los *clústeres* con relación al centro del grupo o centroide (ver figura 136).

Figura 136
Agrupamiento de los cuatro *clústeres* de la prueba dos en Microsoft Azure Machine Learning.

ClaseEstudiante	Assignments	DistancesToClusterCenter no.0	DistancesToClusterCenter no.1	DistancesToClusterCenter no.2	DistancesToClusterCenter no.3
Autorregulador	2	1.498236	1.622122	1.052045	1.365276
Autorregulador	2	1.397616	1.515993	0.989865	1.247296
Autorregulador	2	1.11753	1.423592	0.942614	1.170634
Autorregulador	2	1.259213	1.436982	0.846942	0.911411
Autorregulador	2	1.332733	1.473388	0.988542	1.337539
Autorregulador	1	1.332548	0.823287	1.424056	1.247962
Autorregulador	2	1.250915	1.443661	0.999274	1.331294
Autorregulador	0	0.618739	1.299646	1.060137	1.108308
Autorregulador	1	1.612568	1.076145	1.407789	1.344951
Autorregulador	2	1.507909	1.579985	0.990076	1.030549
Autorregulador	1	1.604911	1.072748	1.4218	1.340296
Autorregulador	1	1.64778	1.109753	1.327396	1.629527
Autorregulador	3	1.340847	1.509067	1.182236	0.730963

Nota: Elaboración propia.

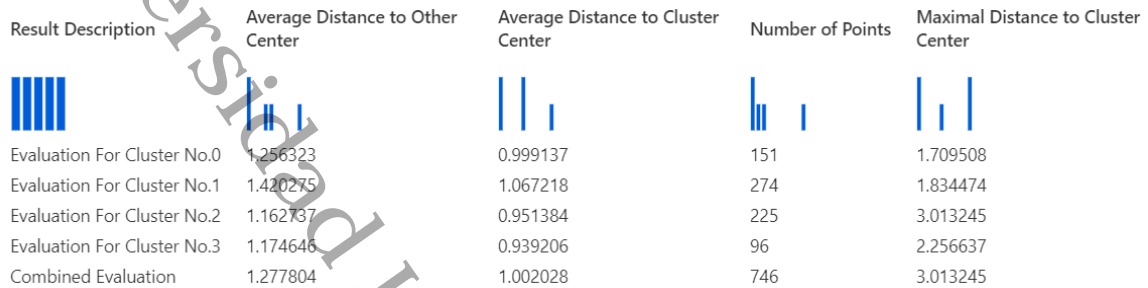
Como se aprecia en la figura anterior, la herramienta asignó el agrupamiento o *clúster* del rango 0 al 3, los cuales son los cuatro grupos de *clústeres* que realizaron. De igual forma se visualiza, que mientras más cerca se encuentra la instancia del centroide o centro del grupo del *clúster*, es al grupo que pertenecerá o asignará el algoritmo.

Por ejemplo, en el caso de la última instancia que se aprecia en la figura anterior, el algoritmo la asignó el *clúster* tres (grupo cuatro), ya que la distancia que tiene al centro del grupo es 0.730963, la cual es menor que de los otros tres *clústeres*, ya que la distancia del *clúster* 0 es de 1.340847, *clúster* 1 es de 1.509067 y *clúster* 2 es de 1.182236.

Aunado a lo anterior, el conjunto de métricas obtenidas por el algoritmo, se pueden apreciar las distancias medias de un centro del grupo a otro, así como las distancia promedio del centro de grupo que hay en cada *clúster*, y se aprecia que las distancias pueden ser cercanas y otras más lejanas.

En la figura 137 se muestran los resultados de dichas métricas de los grupos de *clústeres* y el número de instancias o registros que contiene cada *clúster*, en el caso del *clúster* 0 (grupo uno) contiene 151 instancias, el *clúster* 1 (grupos dos) con 274 instancias, el *clúster* 2 (grupos tres) con 225 instancias y el *clúster* 3 (grupo cuatro) 96 instancias ver figura 137).

Figura 137
Evaluación de la asignación de clúster con la prueba dos con Microsoft Azure Machine Learning.

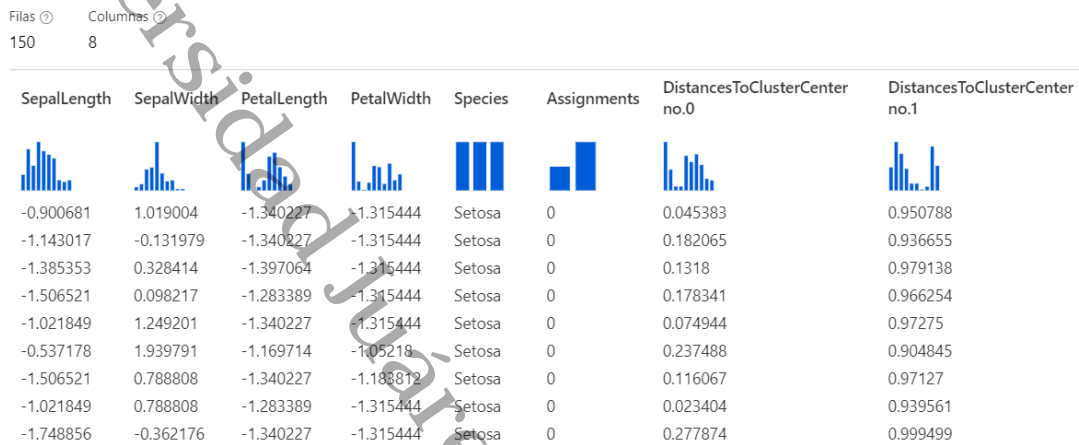


Nota: Elaboración propia.

La tercera y última prueba realiza con el modelo desarrollado en esta herramienta se realizó en el *dataset* iris, el cual anteriormente ya ha sido utilizado; de igual forma que como en los anteriores casos se realizó la normalización de los datos y se ejecutaron dos clústeres, como lo realizado en la prueba tres realizada con el lenguaje R con la técnica de agrupamiento.

En la figura 138 se visualiza el resultado obtenido de la prueba tres realizada con dos *clústeres*, donde se aprecia que al *dataset* se agregaron tres columnas más, una de ellas contiene la asignación de los clústeres y las otras dos la distancia en que se encuentran los *clústeres* con relación al centro del grupo o centroide (ver figura 138).

Figura 138
Agrupamiento de los dos clústeres de la prueba tres en Microsoft Azure Machine Learning.



Nota: Elaboración propia.

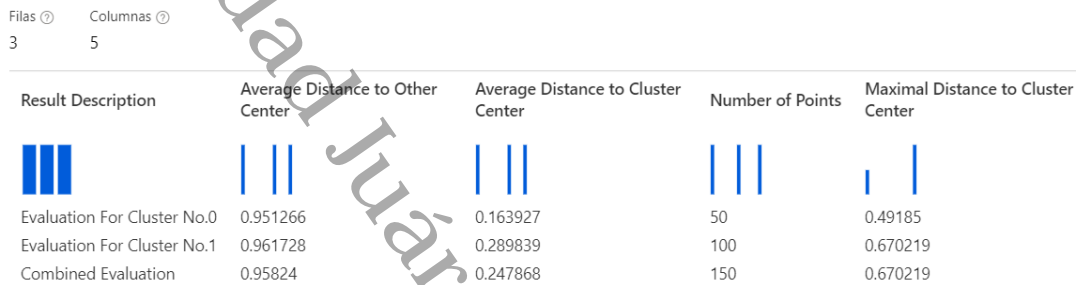
De acuerdo con la figura anterior, los dos clústeres implementados al dataset, corresponden al clúster 0 y clúster 1; además que mientras más cerca se encuentra la instancia del centroide o centro del grupo del clúster, es al grupo que pertenecerá o asignará el algoritmo.

Por ejemplo, en el caso de las instancias que se muestran en la figura 138, todas ellas están agrupadas en el clúster 0 (grupo uno), ya que la distancia que tienen al centroide del clúster 0, es menor que la distancia al centro o centroide del clúster 1.

Aunado a lo anterior, en la figura 139 se visualiza el conjunto de métricas sobre la evaluación de los resultados de los clústeres, donde se indica la distancia media existente entre un centroide del grupo a otro centroide de otro grupo, la distancia promedio existente, la distancia máxima y el número de instancias asignadas a cada clúster, en

este caso el número de instancias agrupadas al *clúster* 0 fueron de 50 y al *clúster* 1 de 100 instancias (ver figura 139).

Figura 139
Evaluación de la asignación de clúster con la prueba tres con Microsoft Azure Machine Learning.



The image shows a screenshot of a data table with 3 rows and 5 columns. The columns are: Result Description, Average Distance to Other Center, Average Distance to Cluster Center, Number of Points, and Maximal Distance to Cluster Center. The rows are: Evaluation For Cluster No.0, Evaluation For Cluster No.1, and Combined Evaluation. Each cell contains a numerical value and a small bar chart visualization.

Result Description	Average Distance to Other Center	Average Distance to Cluster Center	Number of Points	Maximal Distance to Cluster Center
Evaluation For Cluster No.0	0.951266	0.163927	50	0.49185
Evaluation For Cluster No.1	0.961728	0.289839	100	0.670219
Combined Evaluation	0.95824	0.247868	150	0.670219

Nota: Elaboración propia.

4.1.2.3 Weka

Las últimas tres pruebas sobre la técnica de agrupamiento se realizaron en *Weka*; las cuales consistieron y se llevaron a cabo con el mismo número de *clústeres* implementados con anterioridad en las herramientas de Lenguaje R y *Microsoft Azure Machine Learning*, por lo tanto, la prueba uno consistió en la ejecución de tres *clústeres*.

En la figura 140 se muestran los resultados obtenidos a partir de la creación de tres *clústeres* al *dataset*; en dicha figura se visualiza el número de iteraciones que realizó el algoritmo, las cuales fueron 16 iteraciones, también se obtienen los datos aleatorios con los cuales el algoritmo fue creado cada uno de los *clústeres* y la distancia media al centroide (centro) de los tres *clústeres* y la distancia total por cada atributo o variable que conforman el *dataset* (ver figura 140).

Figura 140
Modelo de agrupamiento con tres clústeres de la prueba uno en Weka.

```

KMeans
=====

Number of iterations: 16
Within cluster sum of squared errors: 823.6830625222834

Initial starting points (random):

Cluster 0: 2,20,47,4,2,1,4,3,4,3,3,5,3,4,4,4,4,4,3,4,3,3,1,1,1,1,1,2
Cluster 1: 1,21,62,1,2,1,3,3,1,1,4,5,5,3,3,5,3,4,4,5,3,2,3,3,3,2,2,4
Cluster 2: 2,21,24,3,1,2,3,4,3,3,5,4,3,3,3,2,3,2,5,4,4,1,1,1,1,1,3

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute          Full Data          Cluster#
                   (746.0)           (293.0)           (275.0)           (178.0)
=====
Género              1.6314             2                 1                 2
Edad                21.4316           22.0922           21.2218           20.6685
CicloEscolar        43.0067           44.0068           41.3091           43.9831
DivisiónAcadémica   5.0442            4.4334           5.4509           5.4213
AsignaturaCursa     2.4759            2.4266           2.56              2.427
PromedioBachillerato 1.6555           1.5666           1.8036           1.573
Ítem1               2.4772            2.1775           2.5636           2.8371
Ítem2               3.9343            4.1229           3.9564           3.5899
Ítem3               3.0322            2.8908           3.0655           3.2135
Ítem4               3.0214            2.843           3.0982           3.1966
Ítem5               3.9008            4.1811           3.8255           3.5393
Ítem6               4.7922            4.843           4.7273           4.809
Ítem7               4.2292            4.3447           4.2036           4.0787
Ítem8               2.5027            2.0922           2.6364           2.9719
Ítem9               2.4665            2.0375           2.5891           2.9831
Ítem10              4.2265            4.3618           4.2073           4.0337
Ítem11              3.7359            3.9215           3.7236           3.4663
Ítem12              4.2815            4.4369           4.2873           4.0169
Ítem13              4.0737            4.2491           4.0509           3.8202
Ítem14              4.1475            4.3208           4.0182           4.0618
Ítem15              2.3954            2.0171           2.5164           2.8315
Ítem16              2.3418            1.7679           2.5055           3.0337
Ítem17              2.7024            2.0887           2.84              3.5
Ítem18              2.0013            1.4403           2.1018           2.7697
Ítem19              1.9853            1.43            2.1127           2.7022
Ítem20              1.8391            1.3003           1.8982           2.6348
Ítem21              1.7158            1.2253           1.88              2.2697
Ítem22              2.7145            2.3276           2.7927           3.2303
    
```

Nota: Elaboración propia.

En la figura 141 se visualiza el número de instancias asignadas o agrupadas a los tres clústeres implementados en el dataset, donde el clúster 0 (grupo uno) contiene 293 instancias, el clúster 1 (grupo dos) 275 instancias y el clúster 2 (grupo tres) 178 instancias; además se obtuvo la evaluación o clasificación de los clústeres de acuerdo con la variable clase (ver figura 141).

Figura 141

Visualización de la asignación de las instancias en los clústeres y asignación, de acuerdo con la variable ClaseEstudiante con la prueba uno en Weka.

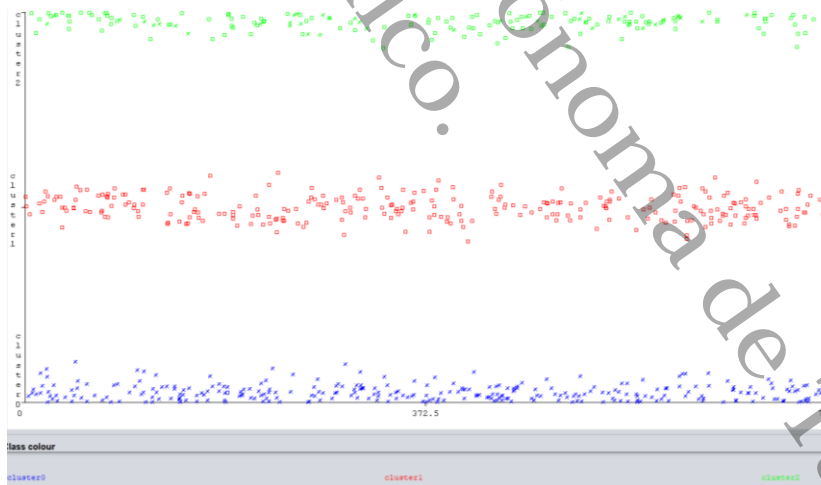
```
=== Model and evaluation on training set === Class attribute: ClaseEstudiante
Classes to Clusters:
Clustered Instances
0 293 ( 39%)
1 275 ( 37%)
2 178 ( 24%)
0 1 2 <-- assigned to cluster
292 242 138 | Autorregulador
1 33 40 | Procrastinador
Cluster 0 <-- Autorregulador
Cluster 1 <-- No class
Cluster 2 <-- Procrastinador
```

Nota: Elaboración propia.

En la figura 142 se muestra la gráfica de asignación de clústeres, donde en el eje X contiene los datos de número de instancias y el eje Y el número de clústeres implementados (ver figura 142).

Figura 142

Gráfico de los tres agrupamientos de la prueba uno en Weka.



Nota: Elaboración propia.

La segunda prueba se realizó con la implementación de cuatro clústeres; con esta prueba el número de iteraciones dadas por el algoritmo fue de 19 iteraciones. En la figura

143 se muestran los resultados obtenidos a partir de la creación de cuatro *clústeres* al *dataset*, en dicha figura se visualiza el número de iteraciones, los datos aleatorios con los cuales el algoritmo fue creado cada uno de los *clústeres* y la distancia media al centroide (centro) de los cuatro *clústeres* y la distancia total por cada atributo o variable que conforman el *dataset* (ver figura 143).

Figura 143
Modelo de agrupamiento con cuatro *clústeres* de la prueba dos en Weka.

```

Number of iterations: 19
Within cluster sum of squared errors: 796.0397218346325

Initial starting points (random):

Cluster 0: 2,20,47,4,2,1,4,3,4,3,3,5,3,4,4,4,4,3,4,3,3,1,1,1,1,1,2
Cluster 1: 1,21,62,1,2,1,3,3,1,1,4,5,5,3,3,5,3,4,4,5,3,2,3,3,2,2,4
Cluster 2: 2,21,24,3,1,2,3,4,3,3,5,4,3,3,3,3,2,3,2,5,4,4,1,1,1,1,3
Cluster 3: 2,18,62,10,1,1,3,3,2,1,5,4,4,2,2,5,0,3,3,3,3,2,2,1,1,4,1,2

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute          Full Data      Cluster#
                   (746.0)      (191.0)      (274.0)      (109.0)      (172.0)
=====
Género             1.6314         1.9948         1           2           2
Edad               21.4316        22.733         21.2226        20.8716        20.6744
CicloEscolar      43.0067        43.5497        41.2409        42.2661        45.686
DivisiónAcadémica 5.0442         4.3979         5.4562         5.2661         4.9651
AsignaturaCursa   2.4759         2.9058         2.5584         2.7431         1.6977
PromedioBachillerato 1.6555        1.6126         1.8066         1.8716         1.3256
ítem1             2.4772         1.9476         2.5693         2.8807         2.6628
ítem2             3.9343         4.288          3.9526         3.5963         3.7267
ítem3             3.0322         2.7853         3.073          3.3303         3.0523
ítem4             3.0214         2.8168         3.1058         3.2018         3
ítem5             3.9008         4.4031         3.8212         3.6239         3.6453
ítem6             4.7922         4.8901         4.7263         4.8165         4.7733
ítem7             4.2292         4.5236         4.2007         4.1743         3.9826
ítem8             2.5027         1.7801         2.646          3.0734         2.7151
ítem9             2.4665         1.7277         2.5949         3.0642         2.7035
ítem10            4.2265         4.5288         4.2044         4.1468         3.9767
ítem11            3.7399         4.0942         3.719          3.578         3.4826
ítem12            4.2815         4.6702         4.2847         4.1284         3.9419
ítem13            4.0737         4.5183         4.0474         3.9266         3.7151
ítem14            4.1475         4.5131         4.0146         4.1468         3.9535
ítem15            2.3954         1.822          2.5219         2.9725         2.4651
ítem16            2.3418         1.6492         2.5109         3.2202         2.2849
ítem17            2.7024         2.0838         2.8467         3.7156         2.5174
ítem18            2.0013         1.4555         2.1058         3.1927         1.686
ítem19            1.9853         1.4346         2.1168         3.0275         1.7267
ítem20            1.8391         1.2984         1.9015         2.9725         1.6221
ítem21            1.7158         1.2304         1.8832         2.578         1.4419
ítem22            2.7145         2.2723         2.7993         3.367         2.657
    
```

Nota: Elaboración propia.

La figura 144 contiene información sobre el número de instancias asignadas o agrupadas en los cuatro *clústeres* implementados en el *dataset*, donde el *clúster* 0 (grupo uno) contiene 191 instancias, el *clúster* 1 (grupo dos) 274 instancias, el *clúster* 2 (grupo tres) 109 y el *clúster* 3 (grupo cuatro) 172 instancias; además se obtuvo la evaluación o clasificación de los *clústeres* de acuerdo con la variable clase (ver figura 144).

Figura 144

Visualización de la asignación de las instancias en los clústeres y asignación, de acuerdo con la variable ClaseEstudiante con la prueba dos en Weka.

```
=== Model and evaluation on training set === Class attribute: ClaseEstudiante
Classes to Clusters:

Clustered Instances
0      191 ( 26%)
1      274 ( 37%)
2      109 ( 15%)
3      172 ( 23%)

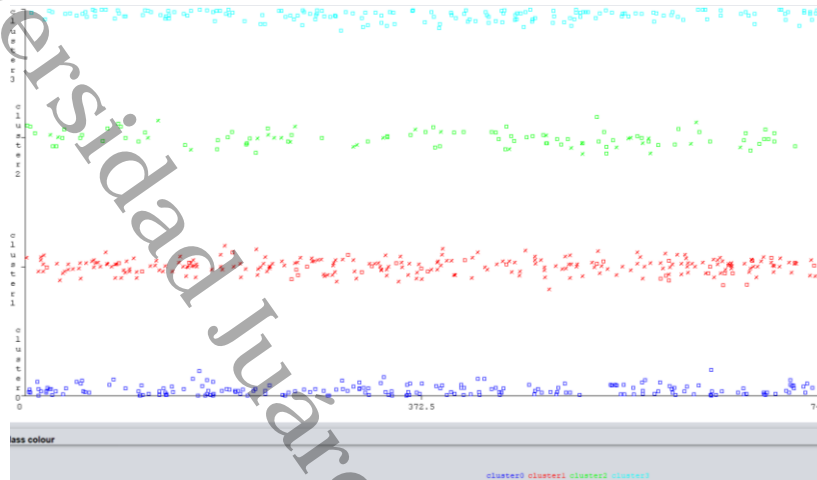
0  1  2  3  <-- assigned to cluster
191 241 79 161 | Autorregulador
0  33 30 11 | Procrastinador

Cluster 0 <-- No class
Cluster 1 <-- Autorregulador
Cluster 2 <-- Procrastinador
Cluster 3 <-- No class
```

Nota: Elaboración propia.

Aunado a lo anterior, en la figura 145 se muestra la gráfica resultante de la técnica de agrupamiento, donde en el eje X contiene los datos de número de instancias y el eje Y el número de clústeres implementados (ver figura 145).

Figura 145
Gráfico de los cuatro agrupamientos de la prueba dos en Weka.



Nota: Elaboración propia.

La tercera prueba, consistió en implementar dos *clústeres* en el *dataset* iris, el cual se ha estado utilizando también con la técnica de árbol de decisión y anteriormente con la técnica de agrupamiento en las otras dos herramientas.

En la figura 146 se visualizan los resultados obtenidos a partir de la implementación el algoritmo de *K-means*, donde el número de iteraciones realizadas por el algoritmo fue de siete iteraciones, de igual forma se visualizan los datos aleatorios con los cuales fueron creados los dos *clústeres* y la distancia media al centroide (centro) en relación con los dos *clústeres* y la distancia total por cada atributo o variable que conforman el *dataset* (ver figura 146).

Figura 146
 Modelo de agrupamiento con dos clústeres de la prueba tres en Weka.

```

KMeans
=====

Number of iterations: 7
Within cluster sum of squared errors: 12.143688281579722

Initial starting points (random):

Cluster 0: 6.1,2.9,4.7,1.4
Cluster 1: 6.2,2.9,4.3,1.3

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
                (150.0)      (100.0)      (50.0)
=====
sepalength     5.8433         6.262         5.006
sepalwidth     3.054          2.872         3.418
petalength     3.7587         4.906         1.464
petalwidth     1.1987         1.676         0.244
    
```

Nota: Elaboración propia.

Así mismo, la figura 147 contiene información sobre el número de instancias asignadas o agrupadas en los dos *clústeres* implementados en el *dataset*, donde el *clúster* 0 (grupo uno) contiene 100 instancias y el *clúster* 1 (grupo dos) 50 instancias; además se obtuvo la evaluación o clasificación de los *clústeres* de acuerdo con la variable clase (ver figura 147).

Figura 147
Visualización de la asignación de las instancias en los clústeres y asignación, de acuerdo con la variable clase con la prueba tres en Weka.

```
=== Model and evaluation on training set === Class attribute: class
Classes to Clusters:

Clustered Instances
0 100 ( 67%)
1  50 ( 33%)

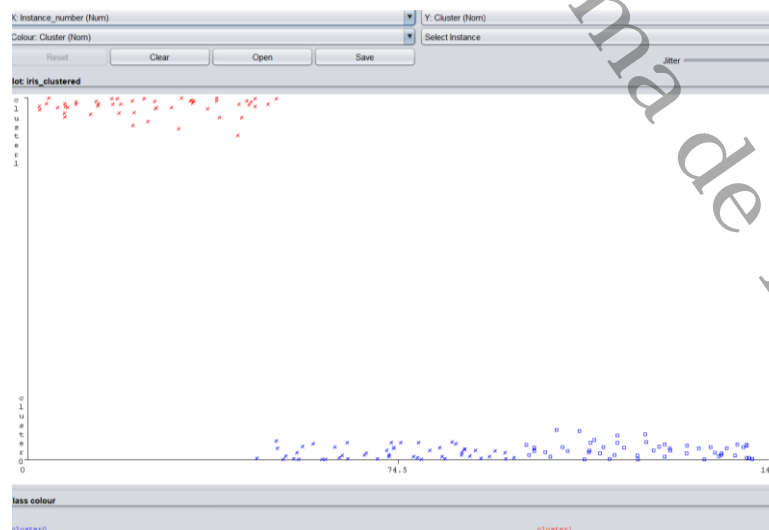
0 1 <-- assigned to cluster
0 50 | Iris-setosa
50 0 | Iris-versicolor
50 0 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-setosa
```

Nota: Elaboración propia.

De igual forma, con esta herramienta se obtuvo como resultado un gráfico, con el cual se pueden visualizar los grupos de agrupamiento resultantes con esta técnica, pero también se pueden graficar las otras variables y junto con los agrupamientos. Lo anterior descrito se puede visualizar en la figura 148, donde en el eje X contiene los datos de número de instancias y el eje Y el número de clústeres implementados (ver figura 148).

Figura 148
Gráfico de los cuatro agrupamientos de la prueba dos en Weka.



Nota: Elaboración propia.

Las pruebas realizadas a la creación del modelo y aplicación del algoritmo de la técnica de agrupamiento en las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*, sirvieron para probar la ejecución de los modelos desarrollados y utilizados, para verificar el número de clústeres adecuados a implementar en el *dataset* de la investigación y se pudo visualizar la forma en que dependiendo la herramienta depende el resultado que se obtendrá, los cuales pueden ser diferentes o similares.

4.1.3 Pruebas de Reglas de asociación

Las pruebas realizadas con esta técnica de minería de datos fueron tres efectuadas con las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*. La primera y segunda prueba se realizaron al *dataset* de la investigación, y la tercera prueba se realizaron en el *dataset* iris.

En cada prueba se realizaron con diferentes parámetros; como en el caso de la prueba uno se llevó a cabo con un soporte del 0.2 y confianza del 0.7 para la obtención de las reglas de asociación del *dataset*.

De igual forma la segunda prueba se llevó a cabo con los parámetros distintos a los anteriores de la prueba uno, esta prueba tenía que el soporte debía ser del 0.3 y la confianza del 0.6; y por último la prueba tres se realizó con un soporte del 0.1 y la confianza del 0.8; además que el número máximo de reglas a obtener son 20 reglas por cada prueba realizada en las herramientas.

A continuación, se describen detalladamente estas tres pruebas realizadas con la técnica de reglas de asociación.

4.1.3.1 Lenguaje R

En la prueba uno realizada con el algoritmo A priori en Lenguaje R, los parámetros de soporte fueron con el mínimo del 0.2 y confianza mínima del 0.7; con estos parámetros el algoritmo A priori extrajo un conjunto de 805 reglas, de las cuales el tamaño o la longitud de cada regla variaban.

Por lo anterior se obtuvo la siguiente distribución de la longitud de las reglas: dos reglas estuvieron conformadas por solo una variable, 125 reglas conformadas por dos variables, 359 reglas construidas a partir de tres variables, 247 reglas con cuatro variables, 66 reglas con una longitud de cinco variables y por último seis reglas conformadas por seis variables.

En la figura 149 se visualiza lo anteriormente descrito y se presentan los resultados obtenidos para los argumentos o parámetros que componen a este algoritmo, como lo son: *support* (soporte), *confidence* (confianza), *coverage* (cobertura), *lift* y *count* (ver figura 149).

Figura 149
Modelo de reglas de asociación obtenido en la prueba uno en Lenguaje R.

```
set of 805 rules

rule length distribution (lhs + rhs):sizes
 1  2  3  4  5  6
 2 125 359 247 66 6

  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 1.000  3.000  3.000  3.333  4.000  6.000

summary of quality measures:
  support      confidence      coverage      lift      count
Min.   :0.2011   Min.   :0.7000   Min.   :0.2024   Min.   :0.9128   Min.   :150.0
1st Qu.:0.2145   1st Qu.:0.8427   1st Qu.:0.2413   1st Qu.:1.0526   1st Qu.:160.0
Median :0.2453   Median :0.9121   Median :0.2735   Median :1.1025   Median :183.0
Mean   :0.2632   Mean   :0.8949   Mean   :0.2965   Mean   :1.2474   Mean   :196.3
3rd Qu.:0.2842   3rd Qu.:0.9585   3rd Qu.:0.3271   3rd Qu.:1.4008   3rd Qu.:212.0
Max.   :0.9008   Max.   :1.0000   Max.   :1.0000   Max.   :2.6276   Max.   :672.0

mining info:
 data ntransactions support confidence
datos1          746      0.2      0.7
```

Nota: Elaboración propia.

Al obtener demasiadas reglas, se realizó un filtrado en las mismas para obtener 20 reglas de asociación con mayor confianza, de las 805 reglas encontradas inicialmente por el algoritmo. Esto se presenta en la figura 150, en la cual se visualizan las 20 reglas con mayor confianza ordenadas en forma decreciente (ver figura 150).

Figura 150
Reglas con mayor confianza de las reglas de asociación obtenidas en la prueba uno con Lenguaje R.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Item1=Casi nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2077748	1	0.2077748	1.110119	155
[2]	{Item7=Siempre,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2037534	1	0.2037534	1.110119	152
[3]	{Item14=Siempre,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2117962	1	0.2117962	1.110119	158
[4]	{Item10=Siempre,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2144772	1	0.2144772	1.110119	160
[5]	{Item6=Siempre,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3404826	1	0.3404826	1.110119	254
[6]	{Item14=Siempre,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2774799	1	0.2774799	1.110119	207
[7]	{Item10=Siempre,Item18=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2064343	1	0.2064343	1.110119	154
[8]	{Item12=Siempre,Item18=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2117962	1	0.2117962	1.110119	158
[9]	{Item10=Casi siempre,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2131367	1	0.2131367	1.110119	159
[10]	{Item7=Casi siempre,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2077748	1	0.2077748	1.110119	155
[11]	{Item6=Siempre,Item14=Nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2064343	1	0.2064343	1.110119	154
[12]	{Item6=Siempre,Item14=Siempre,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2037534	1	0.2037534	1.110119	152
[13]	{Item6=Siempre,Item18=Nunca,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2734584	1	0.2734584	1.110119	204
[14]	{Item6=Siempre,Item12=Siempre,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2064343	1	0.2064343	1.110119	154
[15]	{Item6=Siempre,Item19=Nunca,Item20=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2868633	1	0.2868633	1.110119	214
[16]	{Item6=Siempre,Item19=Nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2989276	1	0.2989276	1.110119	223
[17]	{Género=Mujeres,Item6=Siempre,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2386059	1	0.2386059	1.110119	178
[18]	{Item14=Siempre,Item20=Nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2131367	1	0.2131367	1.110119	159
[19]	{Género=Mujeres,Item14=Siempre,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2064343	1	0.2064343	1.110119	154
[20]	{Item6=Siempre,Item14=Siempre,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.2640751	1	0.2640751	1.110119	197

Nota: Elaboración propia.

De acuerdo con la figura anterior, se puede apreciar que la confianza de las 20 reglas filtradas es de uno, además que con las reglas se pueden visualizar diferentes patrones que ocurren para que la clase del estudiante sea autorregulador. Por ejemplo, con la regla uno ocurrirá u ocurre cada vez que el ítem 1 sea “casi nunca” y el ítem 21 sea “nunca”, entonces la clase es autorregulador; otra regla sería que cuando el ítem 7 es “siempre” y el ítem 19 sea “nunca”, ocurrirá que la clase es autorregulador; así como estos patrones descritos hay otros más encontrados en esta prueba.

La segunda prueba se realizó con un soporte fuera con el mínimo del 0.3 y confianza mínima del 0.6; con estos parámetros el algoritmo A priori extrajo un conjunto de 179 reglas, de las cuales el tamaño o la longitud de algunas reglas son diferentes, algunas solo están conformados por una variable y otras con cuatro variables.

En la figura 151 se visualiza la información sobre las reglas de asociación encontradas en el *dataset*, con los parámetros anteriormente mencionados y otros parámetros, donde tres reglas están conformadas por una variable, 74 reglas con un tamaño de dos variables, 82 reglas con tres variables y 20 reglas con cuatro variables (ver figura 151).

Figura 151
Modelo de reglas de asociación obtenido en la prueba dos en Lenguaje R.

```
set of 179 rules

rule length distribution (lhs + rhs):sizes
 1  2  3  4
 3 74 82 20

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  2.000  3.000  2.665  3.000  4.000

summary of quality measures:
  support      confidence      coverage      lift      count
Min.   :0.3016   Min.   :0.6000   Min.   :0.3016   Min.   :0.9474   Min.   :225.0
1st Qu.:0.3224   1st Qu.:0.8166   1st Qu.:0.3713   1st Qu.:1.0343   1st Qu.:240.5
Median :0.3472   Median :0.8907   Median :0.4075   Median :1.0728   Median :259.0
Mean   :0.3749   Mean   :0.8587   Mean   :0.4444   Mean   :1.1658   Mean   :279.7
3rd Qu.:0.3867   3rd Qu.:0.9447   3rd Qu.:0.4866   3rd Qu.:1.1189   3rd Qu.:288.5
Max.   :0.9008   Max.   :1.0000   Max.   :1.0000   Max.   :1.9806   Max.   :672.0

mining info:
 data ntransactions support confidence
datos2          746      0.3      0.6
```

Nota: Elaboración propia.

Al conjunto de 179 reglas encontradas, se realizó un filtrado para obtener 20 reglas de asociación con mayor confianza. Esto se presenta en la figura 152, en la cual se visualizan las 20 reglas con mayor confianza ordenadas en forma decreciente (ver figura 152).

Figura 152
Reglas con mayor confianza de las reglas de asociación obtenidas en la prueba dos con Lenguaje R.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Item6=Siempre,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3404826	1.0000000	0.3404826	1.110119	254
[2]	{Item6=Siempre,Item18=Nunca,Item20=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3016086	1.0000000	0.3016086	1.110119	225
[3]	{Item6=Siempre,Item18=Nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3016086	1.0000000	0.3016086	1.110119	225
[4]	{Item6=Siempre,Item20=Nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3659517	0.9963504	0.3672922	1.106068	273
[5]	{Género=Mujeres,Item6=Siempre,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3458445	0.9961390	0.3471850	1.105833	258
[6]	{Item18=Nunca,Item20=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3364611	0.9960317	0.3378016	1.105714	251
[7]	{Item19=Nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3243968	0.9958848	0.3257373	1.105551	242
[8]	{Género=Mujeres,Item6=Siempre,Item20=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3163539	0.9957806	0.3176944	1.105435	236
[9]	{Item18=Nunca,Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3042895	0.9956140	0.3056300	1.105250	227
[10]	{Item6=Siempre,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.4865952	0.9945205	0.4892761	1.104036	363
[11]	{Item20=Nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.4034853	0.9933993	0.4061662	1.102792	301
[12]	{Item6=Siempre,Item18=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3592493	0.9925926	0.3619303	1.101896	268
[13]	{Item18=Nunca,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3297587	0.9919355	0.3324397	1.101166	246
[14]	{Item19=Nunca,Item20=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3176944	0.9916318	0.3203763	1.100829	237
[15]	{Item6=Siempre,Item20=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.4477212	0.9910979	0.4517426	1.100237	334
[16]	{Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.5361930	0.9900990	0.5415550	1.099128	400
[17]	{Item19=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3780161	0.9894737	0.3820375	1.098434	282
[18]	{Género=Mujeres,Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3739946	0.9893617	0.3780161	1.098309	279
[19]	{Item18=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.4021448	0.9868421	0.4075067	1.095512	300
[20]	{Género=Mujeres,Item20=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.3458445	0.9809886	0.3525469	1.089014	258

Nota: Elaboración propia.

De acuerdo con la figura anterior, se puede apreciar que la confianza de las 20 reglas filtradas varía, pero todas ellas se encuentran en el rango mínimo establecido anteriormente; además que con las reglas se pueden visualizar diferentes patrones que ocurren para que la clase del estudiante sea autorregulador.

Por ejemplo, con la regla uno ocurrirá u ocurre cada vez que el ítem 6 sea “siempre” y el ítem 19 sea “nunca”, entonces la clase es autorregulador; otra regla sería que cuando el ítem 6 es “siempre”, el ítem 18 “nunca” y el ítem 20 sea “nunca”, ocurrirá que la clase es autorregulador; así como estas reglas descritos hay otras más encontrados en esta prueba.

Por último, la tercera prueba consistió en realizarla en el *dataset* iris, con un soporte mínimo de 0.1 y confianza de 0.8, con estos parámetros el algoritmo extrajo un conjunto de 144 reglas; donde el tamaño de longitud o tamaño de la regla es de dos variables a cinco variables.

En la figura 153 se visualiza la información sobre las reglas de asociación encontradas en el *dataset*, el resumen de las medidas de los parámetros que conforman esta técnica de minería de datos; a su vez muestra que 24 reglas están conformadas por dos variables, 66 reglas por 3 variables, 44 reglas conformadas por cuatro variables y por último 10 reglas con cinco variables (ver figura 153).

Figura 153
Modelo de reglas de asociación obtenido en la prueba tres en Lenguaje R.

```
set of 144 rules

rule length distribution (lhs + rhs):sizes
 2 3 4 5
24 66 44 10

  Min. 1st Qu. Median Mean 3rd Qu. Max.
 2.000 3.000 3.000 3.278 4.000 5.000

summary of quality measures:
  support confidence coverage lift count
Min. :0.1000 Min. :0.8000 Min. :0.1000 Min. :2.366 Min. :15.00
1st Qu.:0.1133 1st Qu.:0.9029 1st Qu.:0.1200 1st Qu.:2.730 1st Qu.:17.00
Median :0.1867 Median :0.9459 Median :0.2000 Median :2.883 Median :28.00
Mean :0.2020 Mean :0.9451 Mean :0.2144 Mean :2.845 Mean :30.31
3rd Qu.:0.2667 3rd Qu.:1.0000 3rd Qu.:0.3017 3rd Qu.:3.000 3rd Qu.:40.00
Max. :0.3333 Max. :1.0000 Max. :0.3467 Max. :3.125 Max. :50.00

mining info:
data ntransactions support confidence
iris 150 0.1 0.8
```

Nota: Elaboración propia.

En la figura 154 se muestran las 20 reglas de asociación con mayor confianza obtenidas del conjunto de 144 reglas, esto se realizó filtrando este conjunto de reglas para obtener las reglas con mayor confianza ordenadas en forma decreciente (ver figura 154).

Figura 154
Reglas con mayor confianza de las reglas de asociación obtenidas en la prueba tres con Lenguaje R.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Petal.Length=[1,2.63]}	=> {Petal.width=[0.1,0.867]}	0.3333333	1	0.3333333	3.000000	50
[2]	{Petal.width=[0.1,0.867]}	=> {Petal.Length=[1,2.63]}	0.3333333	1	0.3333333	3.000000	50
[3]	{Petal.Length=[1,2.63]}	=> {Species=setosa}	0.3333333	1	0.3333333	3.000000	50
[4]	{Species=setosa}	=> {Petal.Length=[1,2.63]}	0.3333333	1	0.3333333	3.000000	50
[5]	{Petal.width=[0.1,0.867]}	=> {Species=setosa}	0.3333333	1	0.3333333	3.000000	50
[6]	{Species=setosa}	=> {Petal.width=[0.1,0.867]}	0.3333333	1	0.3333333	3.000000	50
[7]	{Sepal.Length=[4.3,5.4],Petal.Length=[1,2.63]}	=> {Petal.width=[0.1,0.867]}	0.2666667	1	0.2666667	3.000000	40
[8]	{Sepal.Length=[4.3,5.4],Petal.width=[0.1,0.867]}	=> {Petal.Length=[1,2.63]}	0.2666667	1	0.2666667	3.000000	40
[9]	{Sepal.Length=[4.3,5.4],Petal.Length=[1,2.63]}	=> {Species=setosa}	0.2666667	1	0.2666667	3.000000	40
[10]	{Sepal.Length=[4.3,5.4],Species=setosa}	=> {Petal.Length=[1,2.63]}	0.2666667	1	0.2666667	3.000000	40
[11]	{Sepal.Length=[4.3,5.4],Sepal.Width=[3.2,4.4]}	=> {Petal.Length=[1,2.63]}	0.1866667	1	0.1866667	3.000000	28
[12]	{Sepal.Length=[4.3,5.4],Petal.width=[0.1,0.867]}	=> {Species=setosa}	0.2666667	1	0.2666667	3.000000	40
[13]	{Sepal.Length=[4.3,5.4],Species=setosa}	=> {Petal.width=[0.1,0.867]}	0.2666667	1	0.2666667	3.000000	40
[14]	{Sepal.Length=[4.3,5.4],Sepal.Width=[3.2,4.4]}	=> {Petal.width=[0.1,0.867]}	0.1866667	1	0.1866667	3.000000	28
[15]	{Sepal.Length=[4.3,5.4],Sepal.Width=[3.2,4.4]}	=> {Species=setosa}	0.1866667	1	0.1866667	3.000000	28
[16]	{Sepal.width=[2.9,3.2],Petal.Width=[0.867,1.6]}	=> {Species=versicolor}	0.1133333	1	0.1133333	3.000000	17
[17]	{Sepal.Width=[2.9,3.2],Species=virginica}	=> {Petal.Width=[1.6,2.5]}	0.1200000	1	0.1200000	2.884615	18
[18]	{Petal.Length=[2.63,4.9],Petal.Width=[0.867,1.6]}	=> {Species=versicolor}	0.2866667	1	0.2866667	3.000000	43
[19]	{Petal.Length=[1,2.63],Petal.Width=[0.1,0.867]}	=> {Species=setosa}	0.3333333	1	0.3333333	3.000000	50
[20]	{Petal.Length=[1,2.63],Species=setosa}	=> {Petal.width=[0.1,0.867]}	0.3333333	1	0.3333333	3.000000	50

Nota: Elaboración propia.

De acuerdo con la figura anterior, una de las reglas de asociación identificadas fue que la longitud del pétalo sea 1 y 2.63 entonces la especie de la flor es setosa, otra regla es la especie de la flor es setosa ocurre que el ancho del pétalo debe ser 0.1 y 0.867, así como estas reglas hay otras 18 reglas que contienen más información sobre este *dataset*.

4.1.3.2 Microsoft Azure Machine Learning

Los resultados obtenidos de la ejecución de la técnica de reglas de asociación fueron diferentes a comparación, con lo realizado en la herramienta Lenguaje R, ya que con la herramienta de *Microsoft Azure Machine Learning*, solo se obtienen el número de reglas que el usuario desee, las cuales corresponden a las más relevantes que se pueden obtener con los parámetros establecidos con anterioridad.

Por lo anterior, con la primera prueba se estableció que lo parámetros de soporte fueran con el mínimo del 0.2, confianza mínima del 0.7 y ordenarlos de acuerdo con el parámetro de confianza (*confidence*), pero se tuvieron que transformar los datos a datos categóricos o factor, para ejecutar esta técnica de minería de datos.

En la figura 155 se visualizan algunas de las 20 reglas obtenidas en la prueba uno, donde también se presentan los resultados de los parámetros de *support* (soporte), *confidence* (confianza) y *lift* (ver figura 155).

Figura 155
Modelo de reglas de asociación obtenido en la prueba uno en Machine Azure Machine Learning.

id	lhs	rhs	support	confidence	lift
1	{Item6=[4.2,5.0],Item16=[1.0,1.8],Item21=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.206434	1	1.110119
2	{Item6=[4.2,5.0],Item14=[4.2,5.0],Item19=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.203753	1	1.110119
3	{Item6=[4.2,5.0],Item18=[1.0,1.8],Item19=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.273458	1	1.110119
4	{Item6=[4.2,5.0],Item12=[4.2,5.0],Item19=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.206434	1	1.110119
5	{Item6=[4.2,5.0],Item19=[1.0,1.8],Item20=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.286863	1	1.110119

Nota: Elaboración propia.

De acuerdo con las reglas obtenidas, se tiene que si el ítem 6 está en un rango de 4.2,5.0, además el ítem 16 es 1.0,1.8, y el ítem 21 igual a 1.0,1.8 entonces la clase del estudiante es autorregulador, así como esta regla de asociación se encontraron otras reglas, las cuales se encuentran en el apéndice B.

La segunda prueba se realizó con un soporte mínimo de 0.3 y confianza mínima del 0.6 y las reglas se ordenaron de acuerdo con el parámetro de confianza (*confidence*), pero se realizó la transformación de los datos a datos categóricos o factor.

En la figura 156 se visualizan las primeras cinco reglas de las 20 reglas obtenidas en la prueba dos, donde también se presentan los resultados de los parámetros de *support* (soporte), *confidence* (confianza) y *lift* (ver figura 156).

Figura 156
Modelo de reglas de asociación obtenido en la prueba dos en Machine Azure Machine Learning.

id	lhs	rhs	support	confidence	lift
1	{Item6=[4.2,5.0],Item18=[1.0,1.8],Item20=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.301609	1	1.110119
2	{Item6=[4.2,5.0],Item18=[1.0,1.8],Item21=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.301609	1	1.110119
3	{Item6=[4.2,5.0],Item20=[1.0,1.8],Item21=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.365952	0.99635	1.106068
4	{Genero=[1.5,2.0],Item6=[4.2,5.0],Item21=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.345845	0.996139	1.105833
5	{Genero=[1.5,2.0],Item6=[4.2,5.0],Item20=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.316354	0.995781	1.105435
6	{PromedioBachillerato=[1.0,1.5],Item6=[4.2,5.0],ClaseEstudiante=AUTORREGULADOR}	{Edad=[17.0, 25.2]}	0.376676	0.975694	1.028062

Nota: Elaboración propia.

De acuerdo con las reglas obtenidas, se visualiza que una de las reglas es: cuando el ítem 6 tenga valores de 4.2,5.0; además el ítem 18 sea 1.0,1.8 y el ítem 20 igual a 1.0,1.8 entonces la clase del estudiante es autorregulador; otra de las reglas encontradas es, si el promedio del bachillerato es 1.0,1.5, el ítem 6 sea 4.2,5.0 y la variable ClaseEstudiante sea autorregulador entonces la edad será 17.0,25.2; en el apéndice C se encuentran las reglas de asociación obtenidas de la prueba dos.

En el apéndice A, se encuentra el catálogo de las descripciones de las variables, valores del *dataset* y los valores categóricos que conforman el *dataset*, para una mejor comprensión de los resultados de las pruebas uno y dos.

Por último, la tercera prueba se realizó en el *dataset* iris, con un soporte mínimo de 0.1 y confianza mínima de 0.8, los resultados obtenidos están ordenados de acuerdo con el parámetro de confianza (*confidence*), al igual que en las pruebas uno y dos, se realizó la transformación de los datos a datos categóricos o factor.

En la figura 157 se visualizan las tres reglas de asociación obtenidas del *dataset* iris. La primera regla de asociación es si la longitud del pétalo es 1.27,1.41 entonces la especie de la flor es setosa, la segunda regla es cuando el ancho del pétalo es 0.100,0.209 entonces la especie de la flor es setosa, y la tercera regla encontrada es si la longitud del pétalo es 1.300,1.409 será la especie de la flor versicolor (ver figura 157).

Figura 157

Modelo de reglas de asociación obtenido en la prueba tres en Machine Azure Machine Learning.

id	lhs	rhs	support	confidence	lift
1	{PetalLength=[1.27,1.41]}	{Species=Setosa}	0.133333	1	3
2	{PetalWidth=[0.100,0.209]}	{Species=Setosa}	0.226667	1	3
3	{PetalWidth=[1.300,1.409]}	{Species=Versicolor}	0.133333	0.952381	2.857143

Nota: Elaboración propia.

4.1.3.3 Weka

La primera prueba realizada en *Weka*, se llevó a cabo con los mismos parámetros utilizados en las otras dos herramientas para esta prueba, es decir, con un soporte

mínimo del 0.2 y confianza del 0.7, a su vez se indicó al algoritmo que el número de reglas de salida fueran 20 reglas, las cuales son las más relevantes.

A partir de lo anterior, el algoritmo generó un conjunto de elementos, el cual incluye reglas o patrones conformados por una variable hasta reglas conformadas por seis variables, esto se presenta en la figura 158, donde se visualizan los parámetros utilizados para la obtención de las reglas de asociación (ver figura 158).

Figura 158

Modelo de reglas de asociación obtenido en la prueba uno en Weka.

```
Apriori
=====

Minimum support: 0.2 (149 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 16

Generated sets of large itemsets:

Size of set of large itemsets L(1): 61
Size of set of large itemsets L(2): 247
Size of set of large itemsets L(3): 253
Size of set of large itemsets L(4): 95
Size of set of large itemsets L(5): 15
Size of set of large itemsets L(6): 1
```

Nota: Elaboración propia.

Por lo anterior, se identificó que el tamaño de reglas conformadas por una variable son 61. 247 reglas están conformadas por dos variables. 253 reglas conformados por tres variables. 95 reglas conformadas por cuatro variables. El número de reglas conformadas por cinco variables son 15 reglas y una regla solamente está conformada por seis variables.

Aunado a lo anterior, en la figura 159, se presentan las reglas de asociación obtenidas a partir del algoritmo A priori y de igual manera, se muestran los resultados de los parámetros de *confidence* (confianza), *lift*, *leverage* y *conviction* (convicción) (ver figura 159).

Figura 159

Reglas con mayor confianza de las reglas de asociación obtenido en la prueba uno con Weka.

Best rules found:

1. ítem6=Siempre ítem16=Nunca ítem20=Nunca 149 ==> ClaseEstudiante=Autorregulador 149 <conf:(1)> lift:(1.11) lev:(0.02) [14] conv:(14.78)
2. ítem6=Siempre ítem10=Casi siempre ítem13=Casi siempre 152 ==> ClaseEstudiante=Autorregulador 149 <conf:(0.98)> lift:(1.09) lev:(0.02) [12] conv:(3.77)
3. Género=Mujeres ítem1=Casi nunca ítem6=Siempre 153 ==> ClaseEstudiante=Autorregulador 149 <conf:(0.97)> lift:(1.08) lev:(0.01) [11] conv:(3.04)
4. PromedioBachillerato=Entre 9 y 10 ítem7=Siempre 155 ==> ítem6=Siempre 149 <conf:(0.96)> lift:(1.13) lev:(0.02) [16] conv:(3.27)
5. PromedioBachillerato=Entre 9 y 10 ítem7=Siempre 155 ==> ClaseEstudiante=Autorregulador 149 <conf:(0.96)> lift:(1.07) lev:(0.01) [9] conv:(2.2)
6. ítem10=Siempre ítem12=Siempre ítem14=Siempre 155 ==> ClaseEstudiante=Autorregulador 149 <conf:(0.96)> lift:(1.07) lev:(0.01) [9] conv:(2.2)
7. ítem16=Nunca ítem20=Nunca ClaseEstudiante=Autorregulador 161 ==> ítem6=Siempre 149 <conf:(0.93)> lift:(1.09) lev:(0.02) [11] conv:(1.83)
8. Género=Mujeres ítem1=Casi nunca ClaseEstudiante=Autorregulador 162 ==> ítem6=Siempre 149 <conf:(0.92)> lift:(1.08) lev:(0.01) [10] conv:(1.71)
9. ítem16=Nunca ítem20=Nunca 163 ==> ítem6=Siempre 149 <conf:(0.91)> lift:(1.07) lev:(0.01) [10] conv:(1.6)
10. ítem16=Nunca ítem20=Nunca 163 ==> ítem6=Siempre ClaseEstudiante=Autorregulador 149 <conf:(0.91)> lift:(1.15) lev:(0.03) [19] conv:(2.23)
11. ítem4=Casi siempre ítem6=Siempre 165 ==> ClaseEstudiante=Autorregulador 149 <conf:(0.9)> lift:(1) lev:(0) [0] conv:(0.96)
12. Género=Mujeres ítem1=Casi nunca 168 ==> ítem6=Siempre ClaseEstudiante=Autorregulador 149 <conf:(0.89)> lift:(1.12) lev:(0.02) [15] conv:(1.72)
13. ítem4=Casi siempre ClaseEstudiante=Autorregulador 171 ==> ítem6=Siempre 149 <conf:(0.87)> lift:(1.02) lev:(0) [3] conv:(1.1)
14. ítem7=Siempre ítem13=Siempre 174 ==> ítem14=Siempre 149 <conf:(0.86)> lift:(2.1) lev:(0.1) [78] conv:(3.97)
15. ítem10=Casi siempre ítem13=Casi siempre ClaseEstudiante=Autorregulador 174 ==> ítem6=Siempre 149 <conf:(0.86)> lift:(1) lev:(0) [0] conv:(0.99)
16. ítem10=Casi siempre ítem13=Casi siempre 182 ==> ítem6=Siempre ClaseEstudiante=Autorregulador 149 <conf:(0.82)> lift:(1.03) lev:(0.01) [4] conv:(1.1)
17. ítem10=Siempre ítem20=Nunca 184 ==> ítem21=Nunca 149 <conf:(0.81)> lift:(1.5) lev:(0.07) [49] conv:(2.34)
18. ítem10=Siempre ítem14=Siempre ClaseEstudiante=Autorregulador 184 ==> ítem12=Siempre 149 <conf:(0.81)> lift:(1.83) lev:(0.09) [67] conv:(2.84)
19. ítem13=Siempre ítem14=Siempre 189 ==> ítem7=Siempre 149 <conf:(0.79)> lift:(1.94) lev:(0.1) [72] conv:(2.74)
20. ítem6=Siempre ítem16=Nunca ClaseEstudiante=Autorregulador 191 ==> ítem20=Nunca 149 <conf:(0.78)> lift:(1.51) lev:(0.07) [50] conv:(2.15)

Nota: Elaboración propia.

De acuerdo con la figura anterior, se puede apreciar que la confianza de las 20 reglas se encuentra en un rango de 1 a 0.78, además se observa que se establecen diferentes patrones. Por ejemplo, la regla uno ocurrirá u ocurre cada vez que el ítem 6 sea “siempre”, a su vez el ítem 16 y ítem 20 sea “nunca”, entonces la clase del estudiante será autorregulador, esto ocurre tanto en el antecedente como en la regla completa en 149 instancias.

La segunda prueba se realizó con un soporte mínimo del 0.3 y confianza mínima del 0.6, de igual manera se indicó que el número de reglas de salida fueran 20 reglas, las cuales el algoritmo las toma como las más relevantes.

En la figura 160 se visualiza información sobre los parámetros utilizados en la generación de las reglas de asociación, pero también el algoritmo generó un conjunto de elementos que conforman las reglas de acuerdo con el número de variables que intervienen en cada una (ver figura 160).

Figura 160

Modelo de reglas de asociación obtenido en la prueba dos en Weka.

```
Minimum metric <confidence>: 0.6
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 50
Size of set of large itemsets L(2): 124
Size of set of large itemsets L(3): 98
Size of set of large itemsets L(4): 27
Size of set of large itemsets L(5): 4
```

Nota: Elaboración propia.

De acuerdo con lo anterior, el algoritmo identificó 50 reglas conformadas por una variable, 124 reglas con dos variables, 99 reglas con tres variables, 27 reglas conformadas por cuatro variables, y cuatro reglas conformadas por cinco variables.

Aunado a lo anterior, en la figura 161, se presentan las reglas de asociación obtenidas a partir del algoritmo A priori y los resultados de los parámetros de *confidence* (confianza), *lift*, *leverage* y *conviction* (convicción) (ver figura 161).

Figura 161

Reglas con mayor confianza de las reglas de asociación obtenido en la prueba dos con Weka.

Best rules found:

```
1. ítem6=Siempre ítem19=Nunca ítem21=Nunca 223 ==> ClaseEstudiante=Autorregulador 223 <conf:(1)> lift:(1.11) lev:(0.03) [22] conv:(22.12)
2. ítem6=Siempre ítem19=Nunca ítem20=Nunca 214 ==> ClaseEstudiante=Autorregulador 214 <conf:(1)> lift:(1.11) lev:(0.03) [21] conv:(21.23)
3. ítem14=Siempre ítem21=Nunca 207 ==> ClaseEstudiante=Autorregulador 207 <conf:(1)> lift:(1.11) lev:(0.03) [20] conv:(20.53)
4. ítem6=Siempre ítem18=Nunca ítem19=Nunca 204 ==> ClaseEstudiante=Autorregulador 204 <conf:(1)> lift:(1.11) lev:(0.03) [20] conv:(20.24)
5. ítem6=Siempre ítem18=Nunca ítem20=Nunca ítem21=Nunca 199 ==> ClaseEstudiante=Autorregulador 199 <conf:(1)> lift:(1.11) lev:(0.03) [19] conv:(19.74)
6. ítem6=Siempre ítem14=Siempre ítem21=Nunca 197 ==> ClaseEstudiante=Autorregulador 197 <conf:(1)> lift:(1.11) lev:(0.03) [19] conv:(19.54)
7. Género=Mujeres ítem6=Siempre ítem20=Nunca ítem21=Nunca 197 ==> ClaseEstudiante=Autorregulador 197 <conf:(1)> lift:(1.11) lev:(0.03) [19] conv:(19.54)
8. ítem6=Siempre ítem19=Nunca ítem20=Nunca ítem21=Nunca 193 ==> ClaseEstudiante=Autorregulador 193 <conf:(1)> lift:(1.11) lev:(0.03) [19] conv:(19.14)
9. PromedioBachillerato=Entre 9 y 10 ítem6=Siempre ítem21=Nunca 191 ==> ClaseEstudiante=Autorregulador 191 <conf:(1)> lift:(1.11) lev:(0.03) [18] conv:(18.95)
10. ítem6=Siempre ítem18=Nunca ítem21=Nunca 187 ==> ClaseEstudiante=Autorregulador 187 <conf:(1)> lift:(1.11) lev:(0.02) [18] conv:(18.55)
11. ítem18=Nunca ítem20=Nunca ítem21=Nunca 220 ==> ClaseEstudiante=Autorregulador 219 <conf:(1)> lift:(1.11) lev:(0.03) [20] conv:(10.91)
12. Género=Mujeres ítem20=Nunca ítem21=Nunca 214 ==> ClaseEstudiante=Autorregulador 213 <conf:(1)> lift:(1.1) lev:(0.03) [20] conv:(10.61)
13. ítem19=Nunca ítem20=Nunca ítem21=Nunca 211 ==> ClaseEstudiante=Autorregulador 210 <conf:(1)> lift:(1.1) lev:(0.03) [19] conv:(10.47)
14. ítem7=Siempre ítem21=Nunca 206 ==> ClaseEstudiante=Autorregulador 205 <conf:(1)> lift:(1.1) lev:(0.03) [19] conv:(10.22)
15. ítem12=Siempre ítem21=Nunca 205 ==> ClaseEstudiante=Autorregulador 204 <conf:(1)> lift:(1.1) lev:(0.03) [19] conv:(10.17)
16. ítem18=Nunca ítem19=Nunca ítem21=Nunca 205 ==> ClaseEstudiante=Autorregulador 204 <conf:(1)> lift:(1.1) lev:(0.03) [19] conv:(10.17)
17. PromedioBachillerato=Entre 9 y 10 ítem21=Nunca 203 ==> ClaseEstudiante=Autorregulador 202 <conf:(1)> lift:(1.1) lev:(0.03) [19] conv:(10.07)
18. ítem6=Siempre ítem12=Siempre ítem21=Nunca 201 ==> ClaseEstudiante=Autorregulador 200 <conf:(1)> lift:(1.1) lev:(0.03) [18] conv:(9.97)
19. ítem18=Nunca ítem19=Nunca ítem20=Nunca 201 ==> ClaseEstudiante=Autorregulador 200 <conf:(1)> lift:(1.1) lev:(0.03) [18] conv:(9.97)
20. ítem6=Siempre ítem7=Siempre ítem21=Nunca 199 ==> ClaseEstudiante=Autorregulador 198 <conf:(0.99)> lift:(1.1) lev:(0.03) [18] conv:(9.87)
```

Nota: Elaboración propia.

Una de las primeras reglas obtenidas en la prueba dos, fue si el ítem 14 es “siempre” y el ítem 21 es “nunca” entonces la clase del estudiante será autorregulador, esto ocurre tanto en el antecedente como en la regla completa en 207 instancias. Otra regla obtenida fue cuando el ítem 7 es “siempre” y el ítem 21 “nunca”, entonces la clase del estudiante es autorregulador, esta regla se cumple en el antecedente 206 instancias y con la regla completa 205 instancias.

La tercera y última prueba realizada con el algoritmo A priori en Weka, consistió en realizar esta prueba en otro *dataset*, el cual fue el *dataset* iris. Los parámetros utilizados fueron un soporte mínimo de 0.1 y confianza de 0.8 y se indicó que el número de reglas de salida fueran 20 reglas.

En la figura 162 se visualiza el resultado obtenido, donde se obtuvo solo una regla de asociación, la cual se cumple cuando el ancho del pétalo sea 0.2, entonces el tipo de

especie de la forma es setosa, esto se cumple en la regla completa en 28 instancias (ver figura 162).

Figura 162

Modelo de reglas de asociación obtenido en la prueba tres en Weka.

```
Apriori
=====

Minimum support: 0.1 (15 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 5

Size of set of large itemsets L(2): 1

Best rules found:

1. petalwidth=0.2 28 ==> class=Iris-setosa 28 <conf: (1)> lift: (3) lev: (0.12) [18] conv: (18.67)
```

Nota: Elaboración propia.

Las pruebas realizadas sobre el algoritmo A priori de la técnica de reglas de asociación, sirvieron para establecer los parámetros a utilizar para la obtención del resultado de esta técnica, utilizando las tres herramientas para minería de datos, donde el soporte mínimo por defecto es 0.1 y la confianza es 0.8.

También se pudo entender la manera en que trabajan cada una de las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka* en relación con la técnica de reglas de asociación, la cual se trabaja de forma diferente y los resultados son obtenidos de forma diferente en cada herramienta.

En conclusión, la realización de estas tres pruebas a cada modelo creado, desarrollado e implementado en cada una de las tres técnicas de minería de dato (árbol

de decisión, agrupamiento y reglas de asociación) en las tres herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka* se puede verificar que el funcionamiento de estos modelos es bueno y al momento de modificar algún parámetro y especificación que se requiera se puede realizar sin ningún problema y estos modelos creados pueden ser utilizados en cualquier otro *dataset* o base de datos para el análisis y descubrimiento de conocimiento a partir de los datos.

4.2 Resultados

En este apartado se presentan y describen los resultados obtenidos en esta investigación, sobre los patrones de conducta asociadas con las variables de procrastinación y autorregulación académica de estudiantes de educación superior, a partir de la implementación de las técnicas de minería de datos de árbol de decisión, agrupamiento y reglas de asociación, utilizando las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*.

El orden en que se presentan los resultados es a partir de las técnicas de minería de datos antes mencionadas y se va describiendo el resultado obtenido de la implementación del algoritmo al *dataset* de acuerdo con las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*.

4.2.1 Resultados de Árboles de decisión

Se realizaron y ejecutaron los algoritmos de árboles de decisión con las tres herramientas seleccionadas anteriormente para minería de datos, y se dividieron los 746 registros en dos subconjuntos de datos, es decir, en los sets de entrenamiento y prueba,

donde el primero contiene el 80% de los datos y el segundo el 20% restante. A continuación, se describen los resultados obtenidos con esta técnica de minería de datos.

4.2.1.1 Lenguaje R

La ejecución del árbol de decisión, para esta herramienta se realizó utilizando el set de entrenamiento conformado por el 80% de los datos (597 instancias) y la variable clasificatoria ClaseEstudiante, la cual contiene la clase de procrastinador y autorregulador.

El resultado obtenido a partir de la ejecución de esta técnica de minería de datos en la herramienta de *RStudio*, fue el siguiente: el tamaño del árbol fue de 19 nodos terminales, con un porcentaje de errores de clasificación de la clase del 1.5% equivalentes a 9 instancias.

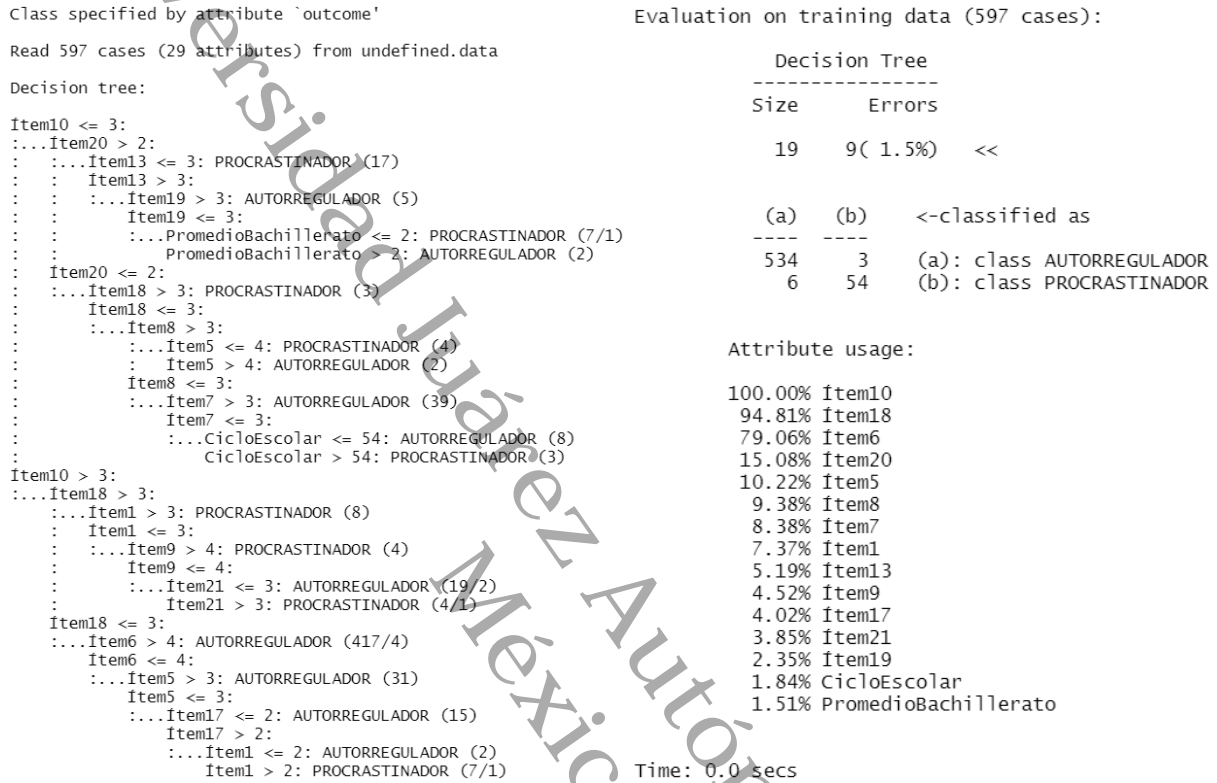
Aunado a lo anterior, la ejecución del algoritmo en esta herramienta se obtuvo una matriz de confusión de la clasificación de las 597 instancias del set de entrenamiento, donde se clasificaron correctamente 534 instancias pertenecientes a la clase autorregulador y solamente tres instancias pertenecientes a esta clase fueron erróneamente clasificadas. De igual forma 54 instancias pertenecientes a la clase procrastinador fueron clasificadas correctamente y seis instancias incorrectamente.

La serie de patrones o decisiones relacionadas con procrastinación y autorregulación académica, se encuentran relacionadas con 15 variables de las 29 variables que conforman el *dataset*, donde 13 de ellas son algunas de las preguntas que conforman el cuestionario, estas variables representan lo siguiente:

- Ítem 1. Cuando tengo que hacer una tarea la dejo para el último minuto.
- Ítem 5. Cuando tengo problemas para entender algo, inmediatamente trato de buscar ayuda.
- Ítem 6. Asisto a clase.
- Ítem 7. Trato de completar el trabajo asignado lo más pronto posible.
- Ítem 8. Postergo los trabajos de los cursos que no me gustan.
- Ítem 9. Postergo las lecturas de los cursos que me disgustan.
- Ítem 10. Intento mejorar mis hábitos de estudio.
- Ítem 13. Trato de terminar mis trabajos importantes con tiempo de sobra.
- Ítem 17. Mientras desarrollo mis actividades del curso en línea me conecto a otras páginas que son de mi agrado.
- Ítem 18. El impulso por revisar mis redes sociales mientras estoy revisando mi curso en línea es incontrolable.
- Ítem 19. Invierto más tiempo en revisar diversos sitios mientras estoy en la plataforma del curso que en desarrollar otras actividades.
- Ítem 20. Me siento nervioso y hasta desesperado si no puedo navegar
- Ítem 21. He dejado de hacer tareas de mi curso por distraerme con el internet.

Lo anterior descrito se muestra en la figura 163 y a su vez se visualiza el árbol de decisión obtenido en esta herramienta, donde se observan los patrones de conducta asociados con procrastinación y autorregulación académica (ver figura 163).

Figura 163
 Resultado del árbol de decisión obtenido en Lenguaje R.



Nota: Elaboración propia.

A continuación, se describen los patrones de conducta visualizados en la figura anterior:

- Cuando el ítem 10 sea menor o igual a tres “a veces”, el ítem 20 mayor que dos “casi nunca” y el ítem 13 menor o igual a tres “a veces” el estudiante es procrastinador.
- Cuando el ítem 10 sea menor o igual a tres “a veces”, el ítem 20 mayor que dos “casi nunca”, el ítem 13 mayor que tres “a veces” y el ítem 19 es mayor que tres “a veces” el estudiante es autorregulador.

- Cuando el ítem 10 sea menor o igual a tres “a veces”, el ítem 20 mayor que dos “casi nunca”, el ítem 13 mayor que tres “a veces”, el ítem 19 menor o igual que tres “a veces” y el promedio del bachillerato es menor o igual que dos (menos de 9 y más de 8) el estudiante es procrastinador.
- Cuando el ítem 10 sea menor o igual a tres “a veces”, el ítem 20 mayor que dos “casi nunca”, el ítem 13 mayor que tres “a veces”, el ítem 19 menor o igual que tres “a veces” y el promedio del bachillerato es mayor que dos (menos de 9 y más de 8) el estudiante es autorregulador.
- Cuando el ítem 10 sea menor o igual a tres “a veces”, el ítem 20 menor o igual que dos “casi nunca” y el ítem 18 mayor que tres “a veces”, el estudiante es procrastinador.
- Cuando el ítem 10 sea menor o igual a tres “a veces”, el ítem 20 menor o igual que dos “casi nunca”, ítem 18 menor o igual que tres “a veces”, ítem 8 mayor a tres “a veces” y el ítem 5 menor o igual que cuatro “casi siempre”, el estudiante es procrastinador.
- Cuando el ítem 10 sea menor o igual a tres “a veces”, el ítem 20 menor o igual que dos “casi nunca”, ítem 18 menor o igual que tres “a veces”, ítem 8 mayor a tres “a veces” y el ítem 5 mayor que cuatro “casi siempre”, el estudiante es autorregulador.
- Cuando el ítem 10 sea menor o igual a tres “a veces”, el ítem 20 menor o igual que dos “casi nunca”, ítem 18 menor o igual que tres “a veces”, ítem 8 menor

- o igual que tres “a veces” y el ítem 7 mayor que tres “a veces”, el estudiante es autorregulador.
- Cuando el ítem 10 sea menor o igual a tres (a veces), el ítem 20 menor o igual que dos (casi nunca), ítem 18 menor o igual que tres (a veces), ítem 8 menor o igual que tres (a veces), ítem 7 menor o igual que tres (a veces) y el ciclo escolar menor o igual que 54 (ciclo largo 2018-2019), el estudiante es autorregulador.
 - Cuando el ítem 10 sea menor o igual a tres (a veces), el ítem 20 menor o igual que dos (casi nunca), ítem 18 menor o igual que tres (a veces), ítem 8 menor o igual que tres (a veces), ítem 7 menor o igual que tres (a veces) y el ciclo escolar mayor que 54 (ciclo largo 2018-2019), el estudiante es procrastinador.
 - Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 mayor que tres (a veces) y el ítem 1 mayor que tres (a veces), el estudiante es procrastinador.
 - Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 mayor que tres (a veces) ítem 1 menor o igual que tres (a veces) y el ítem 9 mayor que cuatro (casi siempre), el estudiante es procrastinador.
 - Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 mayor que tres (a veces) ítem 1 menor o igual que tres (a veces), ítem 9 menor o igual que cuatro (casi siempre) y el ítem 21 menor o igual que tres (a veces), el estudiante es autorregulador.

- Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 mayor que tres (a veces), ítem 1 menor o igual que tres (a veces), ítem 9 menor o igual que cuatro (casi siempre) y el ítem 21 mayor que tres (a veces), el estudiante es procrastinador.
- Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 menor o igual que tres (a veces) y el ítem 6 mayor que cuatro (casi siempre), el estudiante es autorregulador.
- Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 menor o igual que tres (a veces), ítem 6 menor o igual que cuatro (casi siempre) y el ítem 5 mayor que tres (a veces), el estudiante es autorregulador.
- Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 menor o igual que tres (a veces), ítem 6 menor o igual que cuatro (casi siempre), ítem 5 menor o igual que tres (a veces) y el ítem 17 menor o igual que dos (casi nunca), el estudiante es autorregulador.
- Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 menor o igual que tres (a veces), ítem 6 menor o igual que cuatro (casi siempre), ítem 5 menor o igual que tres (a veces), ítem 17 mayor que dos (casi nunca) y el ítem 1 menor o igual que dos (casi nunca), el estudiante es autorregulador.
- Cuando el ítem 10 sea mayor a tres (a veces), ítem 18 menor o igual que tres (a veces), ítem 6 menor o igual que cuatro (casi siempre), ítem 5 menor o igual

que tres (a veces), ítem 17 mayor que dos (casi nunca) y el ítem 1 mayor que dos (casi nunca), el estudiante es procrastinador.

Dentro de los resultados obtenidos se desarrolló de igual forma una matriz de confusión a partir del set de prueba, para la validación de la capacidad de predicción del árbol. En la figura 164 se visualiza la matriz de confusión, en la que 130 instancias de la clase autorregulador fueron clasificadas correctamente y otras seis instancias clasificadas incorrectamente, y de la clase procrastinador se clasificaron correctamente ocho instancias y cinco instancias incorrectamente (ver figura 164).

Figura 164

Matriz de confusión de los resultados del árbol de decisión obtenido en Lenguaje R.

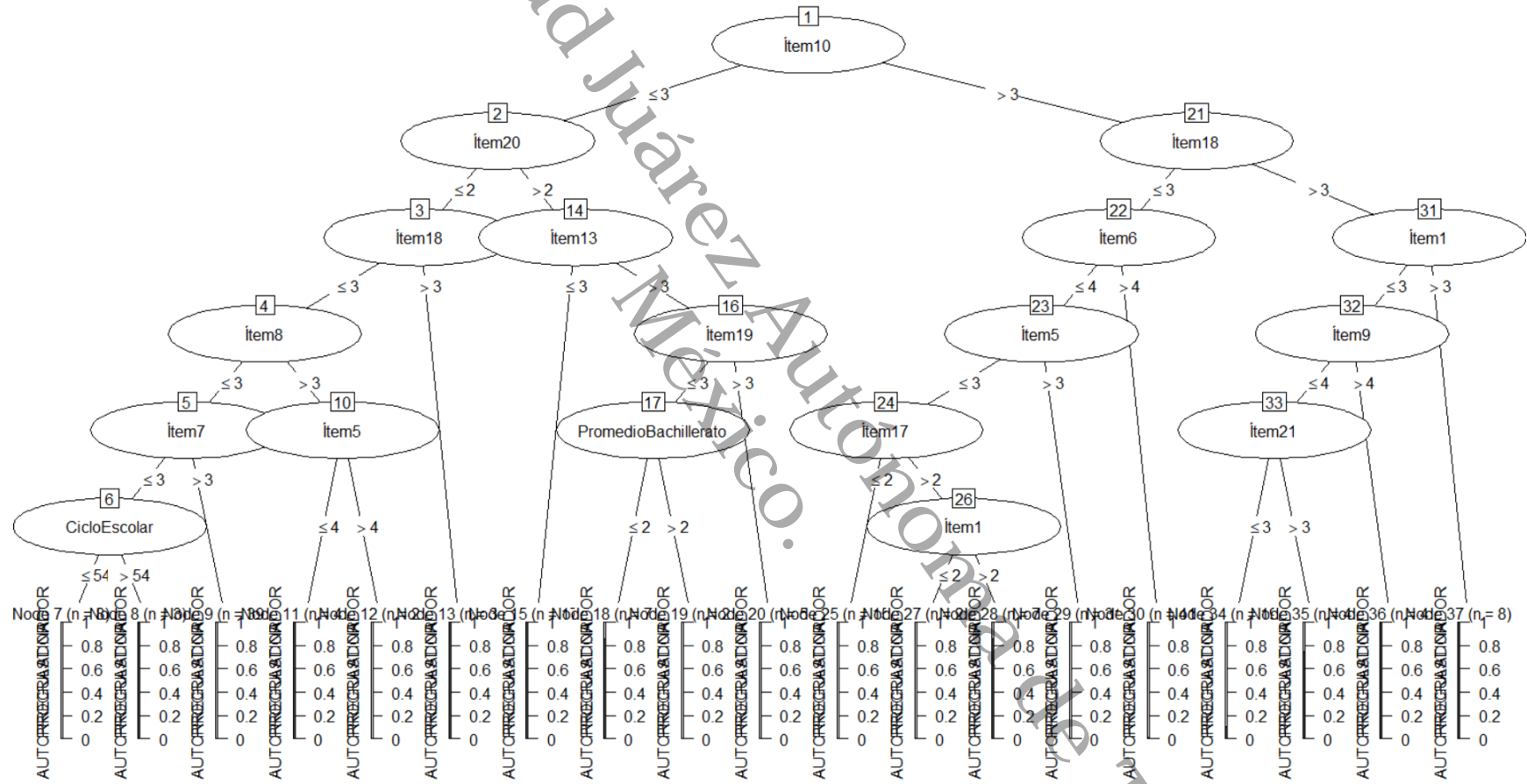
prediccion	AUTORREGULADOR	PROCRASTINADOR
AUTORREGULADOR	130	6
PROCRASTINADOR	5	8

Nota: Elaboración propia.

Por último, se realizó el cálculo de porcentaje de aciertos clasificados correctamente donde el resultado obtenido fue de 92.61745%. También en la figura 165 se muestra el diagrama de árbol de decisión obtenido, donde se pueden visualizar de forma más clara, las ramificaciones y la conexión entre cada nodo del árbol de decisión los patrones clasificatorios (ver figura 165).

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 165
Diagrama del árbol de decisión de los resultados obtenido con Lenguaje R.



Nota: Elaboración propia.

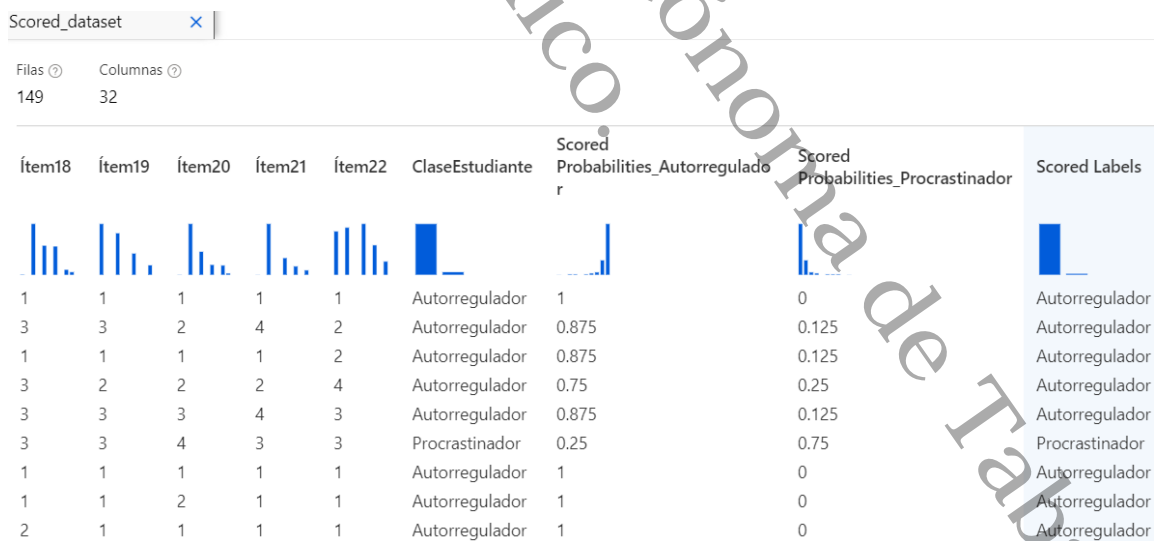
4.2.1.2 Microsoft Azure Machine Learning

En esta herramienta el resultado de la ejecución de la técnica de árbol de decisión es un poco diferente, ya que no se genera un diagrama del árbol de decisión como tal, sino únicamente un subconjunto de los datos con las predicciones y la probabilidad de pertenecer a la clase.

En la figura 166 se observa una muestra de las predicciones extraídas en cuanto a la clase clasificadas, estas predicciones se realizaron con el set de prueba, conformado por 149 registros (20% de los datos); en la figura también se visualizan las probabilidades que una instancia o registro se clasifique si el estudiante es procrastinador o autorregulador (ver figura 166).

Figura 166

Resultado del árbol de decisión obtenido en Microsoft Azure Machine Learning.



Nota: Elaboración propia.

De acuerdo con la figura anterior y revisando detenidamente este subconjunto de datos, se visualizan datos que según la predicción dada por el algoritmo en esta

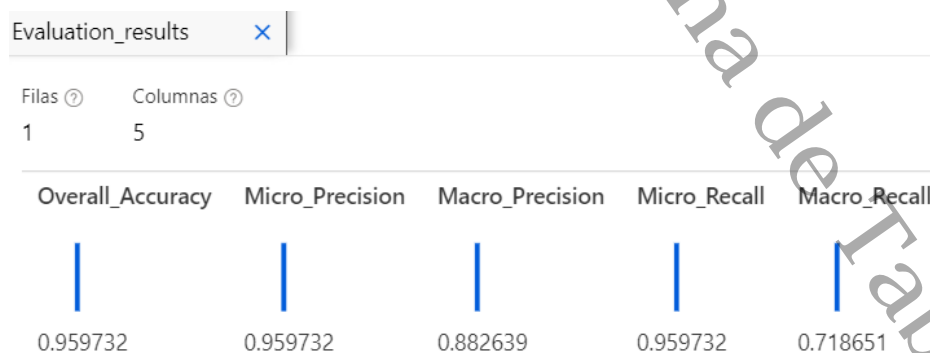
herramienta la clasificaron en otra clase a la que no pertenecía anteriormente. A su vez se pueden visualizar instancias clasificadas de acuerdo con la clasificación que tenían anteriormente.

Por ejemplo, en la figura anterior se visualiza la instancia tres donde la clasificación dada es la clase autorregulador y en la predicción, se tuvo como resultado que la probabilidad de pertenecer a esta clase es del 0.875 (87.5%), pero tiene una mínima probabilidad del 0.125 (12.5%) de ser la clase procrastinador; por lo consiguiente la clasificación con la predicción fue la misma para esta instancia; así como este ejemplo hay resultados donde sí cambia la clasificación dada con relación a la predicción.

En la figura 167 se encuentra el conjunto de métricas que representan la precisión (rendimiento) del algoritmo; estas métricas son: precisión general, micro precisión, macro precisión, micro recuperación y macro recuperación (ver figura 167).

Figura 167

Conjunto de métricas del árbol de decisión obtenido en Microsoft Azure Machine Learning.



Nota: Elaboración propia.

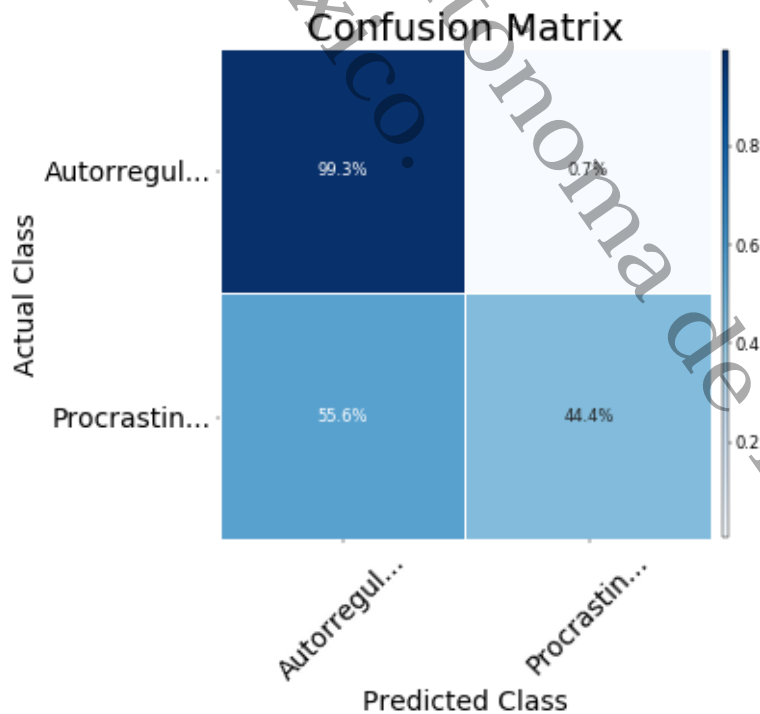
Con la figura anterior se visualiza que la precisión general de clasificación es del 95.9732% lo que significa que la mayoría de los datos de la variable clase fueron

clasificadas correctamente. También se obtuvieron la precisión micro y macro en relación con los resultados verdaderos y resultados positivos, así como la recuperación de las instancias del total relevantes que se recuperaron las cuales fueron micro del 95.9732% y macro 88.2639%.

Por último, la matriz de confusión resultante se muestra en la figura 168, en la que el 99.3% (139 instancias) de la clase autorregulador fueron clasificadas correctamente y el 0.7% (una instancia) clasificada incorrectamente, y para la clase procrastinador se clasificaron 44.4% (cuatro instancias) correctamente e incorrectas el 55.6% (cinco instancias) (ver figura 168).

Figura 168

Matriz de confusión de los resultados del árbol de decisión obtenido en Microsoft Azure Machine Learning.



Nota: Elaboración propia.

4.2.1.3 Weka

La ejecución del algoritmo de árbol de decisión en esta herramienta, se llevó a cabo con las mismas características que en las otras dos herramientas. La implementación del algoritmo se realizó al *set* de entrenamiento conformado por el 80% de los datos (597 instancias) y la variable clasificatoria ClaseEstudiante, y el *set* de prueba con el 20% de los datos (149 instancias) fueron utilizadas para la creación de la matriz de confusión y el porcentaje de aciertos de instancias clasificadas.

De acuerdo con lo anterior, el resultado obtenido con la técnica de minería de datos en la herramienta de *Weka*, fue el siguiente: el número de hojas fue de 17 y el tamaño del árbol 33 nodos, además, el porcentaje de instancias clasificadas correctamente fue del 92.6174% (138 instancias) e incorrectamente clasificadas el 7.3826% (11 instancias), esto quiere decir que la mayor parte de las instancias pertenecen a la clase correcta.

Con el resultado del árbol de decisión obtenido en esta herramienta, se identificaron una serie de patrones o decisiones relacionadas con 12 de las 29 variables que conforman el *dataset*, donde nueve de ellas son algunas de las preguntas que conforman el cuestionario, estas variables representan lo siguiente:

- Ítem 2. Me preparo con anticipación para los exámenes.
- Ítem 6. Asisto a clase.
- Ítem 7. Trato de completar el trabajo asignado lo más pronto posible.
- Ítem 8. Postergo los trabajos de los cursos que no me gustan.
- Ítem 10. Intento mejorar mis hábitos de estudio.

- Ítem 12. Trato de motivarme para mantener mi ritmo de estudio.
- Ítem 15. Dejo para mañana lo que puedo hacer hoy.
- Ítem 18. El impulso por revisar mis redes sociales mientras estoy revisando mi curso en línea es incontrolable.
- Ítem 20. Me siento nervioso y hasta desesperado si no puedo navegar.

En la figura 169 se visualiza el resultado del árbol de decisión implementado el algoritmo C4.5 o mejor conocido como J48, donde se visualizan los patrones de conducta asociados con procrastinación y autorregulación académica (ver figura 169).

Figura 169

Resultado del árbol de decisión obtenido en Weka.

```
ítem15 <= 3
|   ítem10 <= 3
|   |   ítem20 <= 2
|   |   |   ítem7 <= 3
|   |   |   |   ítem18 <= 1: Autorregulador (9.0/1.0)
|   |   |   |   ítem18 > 1: Procrastinador (5.0/1.0)
|   |   |   |   ítem7 > 3
|   |   |   |   |   Edad <= 23: Autorregulador (5.0)
|   |   |   |   |   Edad > 23
|   |   |   |   |   |   ítem8 <= 2: Autorregulador (3.0)
|   |   |   |   |   |   ítem8 > 2: Procrastinador (2.0)
|   |   |   |   |   ítem20 > 2
|   |   |   |   |   |   ítem12 <= 3: Procrastinador (15.0)
|   |   |   |   |   |   ítem12 > 3
|   |   |   |   |   |   |   ítem2 <= 3
|   |   |   |   |   |   |   |   PromedioBachillerato <= 1: Autorregulador (2.0)
|   |   |   |   |   |   |   |   PromedioBachillerato > 1: Procrastinador (5.0/1.0)
|   |   |   |   |   |   |   |   ítem2 > 3: Autorregulador (5.0)
|   |   |   |   |   |   |   ítem10 > 3: Autorregulador (570.0/15.0)
|   |   |   |   |   ítem15 > 3
|   |   |   |   |   |   ítem18 <= 3
|   |   |   |   |   |   |   ítem2 <= 3
|   |   |   |   |   |   |   |   ítem6 <= 4: Procrastinador (4.0)
|   |   |   |   |   |   |   |   ítem6 > 4
|   |   |   |   |   |   |   |   |   DivisiónAcadémica <= 6: Autorregulador (7.0/1.0)
|   |   |   |   |   |   |   |   |   DivisiónAcadémica > 6: Procrastinador (3.0)
|   |   |   |   |   |   |   |   |   ítem2 > 3: Autorregulador (34.0/1.0)
|   |   |   |   |   |   |   |   ítem18 > 3
|   |   |   |   |   |   |   |   |   PromedioBachillerato <= 2: Procrastinador (25.0/3.0)
|   |   |   |   |   |   |   |   |   PromedioBachillerato > 2
|   |   |   |   |   |   |   |   |   |   ítem7 <= 4: Autorregulador (5.0)
|   |   |   |   |   |   |   |   |   |   ítem7 > 4: Procrastinador (2.0)
```

Nota: Elaboración propia.

A continuación, se describen los patrones de conducta visualizados en la figura anterior:

- Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 menor o igual que dos (casi nunca), el ítem 7 menor o igual que tres (a veces) y el ítem 18 menor o igual que uno (nunca), el estudiante es autorregulador.
- Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 menor o igual que dos (casi nunca), el ítem 7 menor o igual que tres (a veces) y el ítem 18 mayor que uno (nunca), el estudiante es procrastinador.
- Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 menor o igual que dos (casi nunca), el ítem 7 mayor que tres (a veces) y edad menor o igual que 23 (23 años), el estudiante es autorregulador.
- Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 menor o igual que dos (casi nunca), el ítem 7 mayor que tres (a veces), la variable edad mayor que 23 (23 años) y el ítem 8 menor o igual que dos (casi nunca), el estudiante es autorregulador.
- Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 menor o igual que dos (casi nunca), el ítem 7 mayor

- que tres (a veces), la variable edad mayor que 23 (23 años) y el ítem 8 mayor que dos (casi nunca), el estudiante es procrastinador.
- Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 mayor que dos (casi nunca) y el ítem 12 menor o igual que tres (a veces), el estudiante es procrastinador.
 - Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 mayor que dos (casi nunca), el ítem 12 mayor que tres (a veces), ítem 2 menor o igual que tres (a veces) y la variable promedio del bachillerato sea menor o igual que uno (entre 9 y 10), el estudiante es autorregulador.
 - Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 mayor que dos (casi nunca), el ítem 12 mayor que tres (a veces), ítem 2 menor o igual que tres (a veces) y la variable promedio del bachillerato sea mayor que uno (entre 9 y 10), el estudiante es procrastinador.
 - Cuando el ítem 15 sea menor o igual a tres (a veces), el ítem 10 menor o igual que tres (a veces), ítem 20 mayor que dos (casi nunca), el ítem 12 mayor que tres (a veces) y el ítem 2 mayor que tres (a veces), el estudiante es autorregulador.
 - Cuando el ítem 15 sea menor o igual a tres (a veces) y el ítem 10 mayor que tres (a veces), el estudiante es autorregulador.

- Cuando el ítem 15 sea mayor a tres (a veces), el ítem 18 menor o igual que tres (a veces), ítem 2 menor o igual que tres (a veces) y el ítem 6 menor o igual que cuatro (casi siempre), el estudiante es procrastinador.
- Cuando el ítem 15 sea mayor a tres (a veces), el ítem 18 menor o igual que tres (a veces), ítem 2 menor o igual que tres (a veces), ítem 6 mayor que cuatro (casi siempre) y la variable División Académica sea menor o igual que seis (DAEA), el estudiante es autorregulador.
- Cuando el ítem 15 sea mayor a tres (a veces), el ítem 18 menor o igual que tres (a veces), ítem 2 menor o igual que tres (a veces), ítem 6 mayor que cuatro (casi siempre) y la variable División Académica sea mayor que seis (DAEA), el estudiante es procrastinador.
- Cuando el ítem 15 sea mayor a tres (a veces), el ítem 18 menor o igual que tres (a veces) y el ítem 2 mayor que tres (a veces), el estudiante es autorregulador.
- Cuando el ítem 15 sea mayor a tres (a veces), el ítem 18 mayor que tres (a veces) y la variable promedio del bachillerato sea menor o igual que dos (menos de 9 y más de 8), el estudiante es procrastinador.
- Cuando el ítem 15 sea mayor a tres (a veces), el ítem 18 mayor que tres (a veces), la variable promedio del bachillerato sea mayor que dos (menos de 9 y más de 8) y el ítem 7 sea menor o igual que cuatro (casi siempre), el estudiante es autorregulador.

- Cuando el ítem 15 sea mayor a tres (a veces), el ítem 18 mayor que tres (a veces), la variable promedio del bachillerato sea mayor que dos (menos de 9 y más de 8) y el ítem 7 sea mayor que cuatro (casi siempre), el estudiante es procrastinador.

En la figura 170 se visualiza la matriz de confusión, con 127 instancias de la clase autorregulador las cuales fueron clasificadas correctamente y cinco instancias pertenecientes a esta clase fueron clasificadas incorrectamente. Para la clase procrastinador se clasificaron correctamente 11 instancias y seis instancias incorrectamente (ver figura 170).

Figura 170

Matriz de confusión de los resultados del árbol de decisión obtenido en Weka.

```
=== Confusion Matrix ===
      a    b  <-- classified as
127   5 |  a = Autorregulador
  6  11 |  b = Procrastinador
```

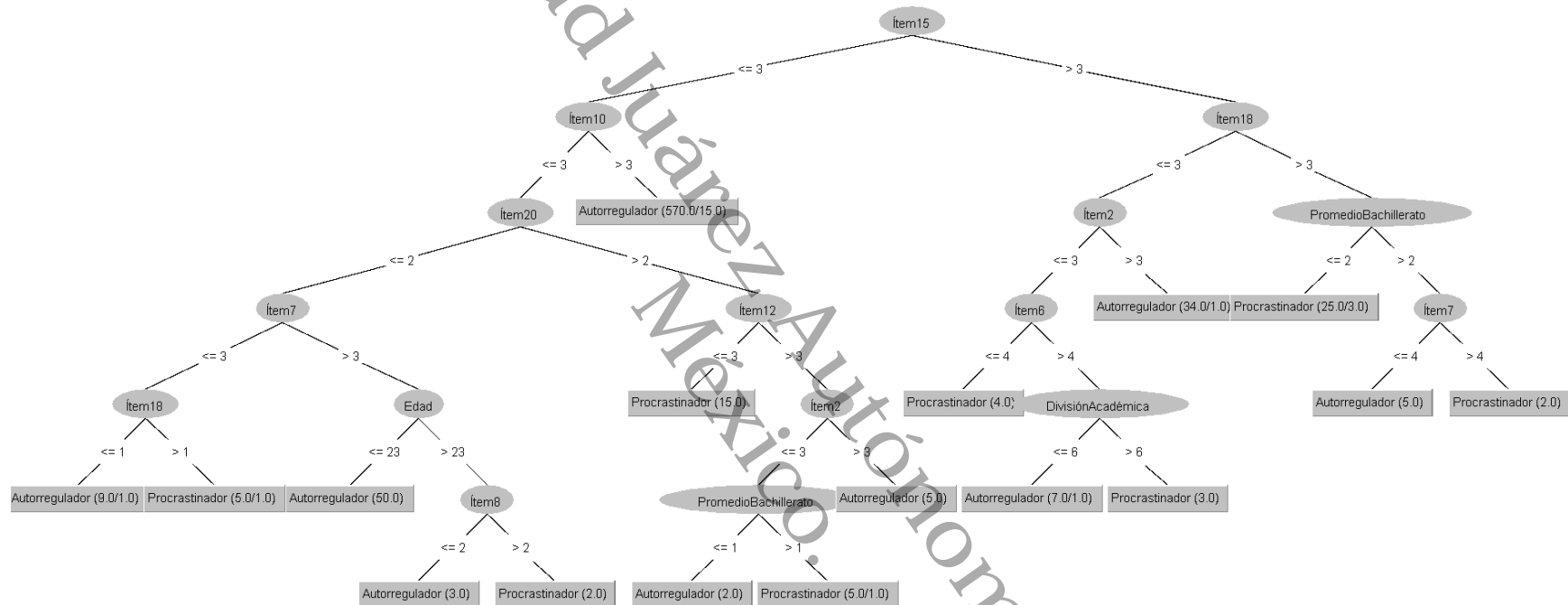
Nota: Elaboración propia.

De igual forma, se realizó el cálculo de porcentaje de aciertos clasificados correctamente donde el resultado obtenido fue de 92.6174% correspondientes a 138 instancias, además que el 7.3826% (11 instancias) fueron clasificadas incorrectamente.

En la figura 171 se muestra el diagrama de árbol de decisión obtenido, donde se pueden visualizar de forma más clara y gráfica las ramificaciones y la conexión entre cada nodo del árbol de decisión los patrones encontrados (ver figura 171).

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Figura 171
 Diagrama del árbol de decisión de los resultados obtenido con Weka.



Nota: Elaboración propia.

En conclusión, con la técnica de árbol de decisión utilizando las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*, se obtuvieron resultados, similares, pero también diferentes entre sí.

En el caso del árbol de decisión, obtenidos en las herramientas Lenguaje R y *Weka*, no son iguales, en cada una de las herramientas se obtuvo un árbol de decisión diferente, en donde intervienen *ítems* o variables y datos, que no se presentan en el otro árbol y viceversa; por lo anterior, se encontraron patrones de conducta de procrastinación y autorregulación académica diferentes y que ayudan a tener más comprensión sobre como actuará o actúa un estudiante que es procrastinador o en un caso si es autorregulador.

En el caso del porcentaje de instancias clasificadas correctamente, se obtuvieron el mismo resultado en el caso de Lenguaje R y *Weka*, en el caso de *Microsoft Azure Machine Learning* el resultado solo fue mayor por 3.3558% que el de las otras herramientas, pero se encuentran en un rango similar.

Lo anteriormente descrito, sucede también con la matriz de confusión, obtenida en cada herramienta, las cuales no presentan diferencias significativas, en la clasificación de las clases, sino que se encuentran en un rango medio.

4.2.2 Resultados de Agrupamiento

En este apartado se describen los resultados obtenidos de la investigación, utilizando el algoritmo *K-means* de la técnica de minería de datos, ejecutado en las herramientas de Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*, donde se implementaron dos

clústeres al *dataset* y los resultados obtenidos variaron, en cuestión de la ejecución de esta técnica en las tres herramientas para minería de datos.

Para establecer que serían dos *clústeres* para implementar en el *dataset*, se realizó la estimación para definir el número de éste, para ello se utilizaron los métodos de *wss*, *silhouette* y *gap_stat*. Con los dos primeros métodos se obtuvo que el número óptimo de *clúster* a utilizar fuera dos y con el último método fuera uno, pero al obtener en dos métodos el mismo resultado, se decidió realizar dos *clústeres*. A continuación, se describen los resultados obtenidos con esta técnica de minería de datos.

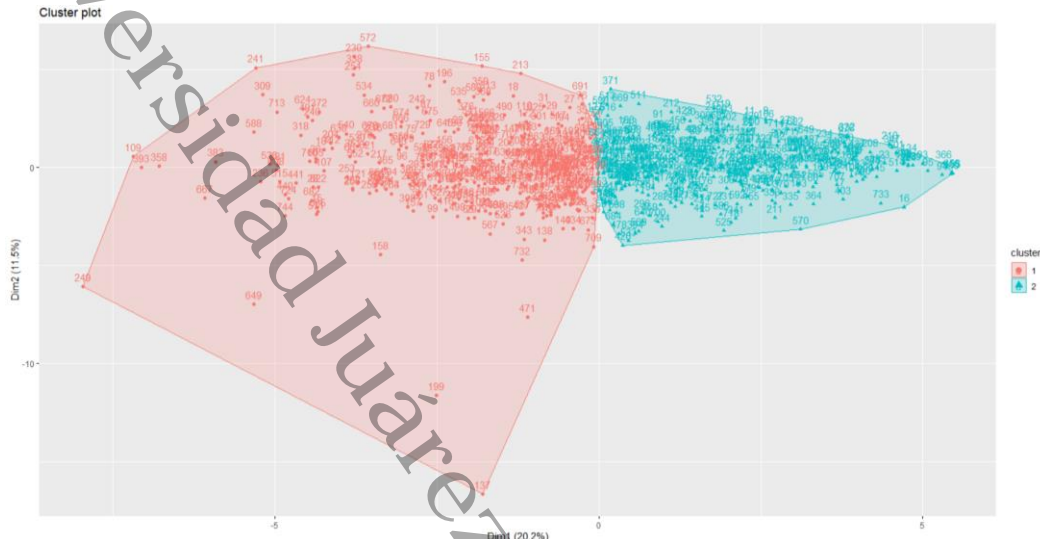
4.2.2.1 Lenguaje R

Para la obtención de los resultados se realizó la ejecución de dos *clústeres* al *dataset*, donde el algoritmo ejecutado dio como resultado la obtención del primer *clúster* con 370 instancias agrupadas y el segundo *clúster* con 376 instancias agrupadas, esto equivalente a las 746 instancias que contiene el *dataset* de la investigación.

Lo anteriormente descrito se muestra en la figura 172, donde aparte de mostrar el vector que contiene la agrupación de las instancias, se muestran las medias obtenidas de cada una de las variables que conforman el conjunto de datos y en la parte superior de la figura, está el número de *clústeres* ejecutados y el tamaño de cada uno (ver figura 172).

Figura 173

Gráfico del resultado obtenido con la técnica de agrupamiento en Lenguaje R.



Nota: Elaboración propia.

El diagrama o gráfico, permite visualizar cómo están agrupadas las instancias y la distancia que existe entre las instancias. En el caso particular de este resultado, las instancias de los dos grupos se encuentran aglomeradas en el centro del gráfico, pero en caso del *clúster* dos, no se encuentra mucha distancia entre cada instancia agrupada en este *clúster*. Caso contrario al *clúster* uno, donde se detecta rápidamente que existe una mayor distancia entre algunas instancias sobre el centro del aglomerado de este *clúster*.

En la figura 174 se muestra una tabla o matriz de confusión, sobre la asignación o distribución de las instancias en los *clústeres*, de acuerdo con la clase que pertenecen las instancias; los resultados obtenidos fueron: para el *clúster* 1 se encuentran agrupadas instancias de las dos clases, es decir, clase autorregulador y clase procrastinador,

mientras que el clúster 2, únicamente contiene instancias de la clase autorregulador (ver figura 174).

Figura 174

Visualización de la asignación de las instancias en los clústeres de acuerdo con la variable ClaseEstudiante en Lenguaje R.

	cluster	
	1	2
Autorregulador	296	376
Procrastinador	74	0

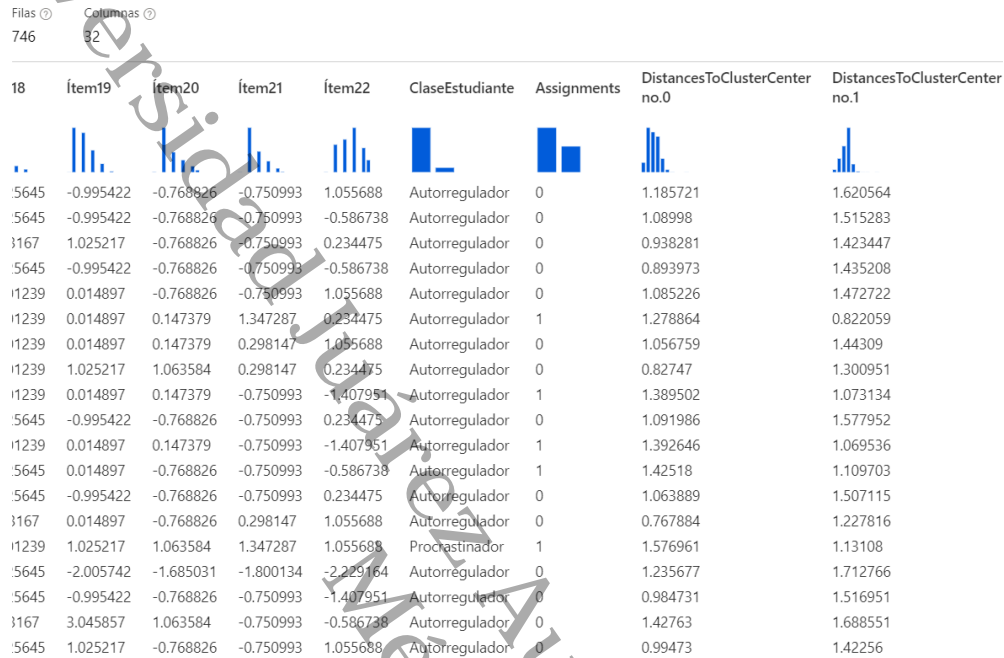
Nota: Elaboración propia.

4.2.2.2 Microsoft Azure Machine Learning

Los resultados obtenidos con esta herramienta fueron diferentes al que se obtuvo anteriormente con la herramienta Lenguaje R. Además, no se obtuvo un diagrama para visualizar las instancias agrupadas en los *clústeres*.

Con la ejecución de dos *clústeres*, se generó una columna con la asignación por cada instancia al grupo al que pertenecen, además de la distancia en que se encuentra la instancia del centro o centroide de los dos *clústeres* ejecutados, esto se presenta en la figura 175 (ver figura 175).

Figura 175
Resultados de agrupamiento obtenido en Microsoft Azure Machine Learning.



Nota: Elaboración propia.

Como se aprecia en la figura anterior, los dos *clústeres* corresponden al 0 y 1, pero son dos grupos de agrupamiento. De igual forma se visualiza que mientras más cerca se encuentre la instancia al centro o centroide del grupo del *clúster*, es al grupo que pertenecerá o asignará el algoritmo.

Por ejemplo, en el caso de las primeras cinco instancias, está asignada al *clúster* 0 (grupo uno), ya que la distancia que tienen hacia al centro del grupo es menor que la distancia que tienen al centro o centroide del *clúster* 1 (grupo dos), y así sucesivamente va ocurriendo con las demás instancias que conforman el *dataset*.

De acuerdo con lo anterior, se extrajo un conjunto de métricas sobre la evaluación de la ejecución de los *clústeres*, donde se tienen datos sobre la distancia media existente

del centro de un *clúster* al otro, la distancia promedio del centro del grupo, el número de instancias agrupadas en los *clústeres* y la distancia máxima al centro del grupo.

Por lo anterior, en la figura 176 se muestra el conjunto de métricas resultantes, en la que se aprecia que el *clúster* 0 está conformado por 471 instancias (más de la mitad de las instancias que conforman el *dataset*) y el *clúster* 1 por las 275 instancias restantes (ver figura 176).

Figura 176

Evaluación de la asignación de clúster con Microsoft Azure Machine Learning.

Filas 3 Columnas 5

Result Description	Average Distance to Other Center	Average Distance to Cluster Center	Number of Points	Maximal Distance to Cluster Center
Evaluation For Cluster No.0	1.484872	1.066451	471	3.125729
Evaluation For Cluster No.1	1.483162	1.068464	275	1.835802
Combined Evaluation	1.484242	1.067193	746	3.125729

Nota: Elaboración propia.

De igual forma la distancia existente entre el centro o centroide un *clúster* al otro no es una distancia grande, solo varía por 0.00171 de distancia, porque el *clúster* 0 se encuentra a una distancia del otro centro del 1.484872 y el *clúster* 1 a una distancia del 1.483162 del centro del *clúster* 0.

4.2.2.3 Weka

Por último, la ejecución para la obtención de los resultados de agrupamiento con la herramienta *Weka*, fue utilizando la opción *Classes to clusters evaluation* (evaluación de la clase en los *clústeres*) y se ignoró esta variable, además que el número de *clústeres* implementados al *dataset* fueron dos *clústeres*.

La figura 177 muestra el resultado obtenido con esta técnica de minería de datos, en dicha figura se muestra el número de iteraciones realizadas por el algoritmo para la obtención de los grupos de instancias, estas iteraciones realizadas fueron tres iteraciones (ver figura 177).

Figura 177
Resultados de agrupamiento obtenido en Weka.

```

kMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 891.8426053239332

Initial starting points (random):

Cluster 0: 2,20,47,4,2,1,4,3,4,3,3,5,3,4,4,4,4,4,3,4,3,3,1,1,1,1,1,2
Cluster 1: 1,21,62,1,2,1,3,3,1,1,4,5,5,3,3,5,3,4,4,5,3,2,3,3,3,2,2,4

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute      Full Data      Cluster#
              (746.0)      (471.0)      (275.0)
=====
Género         1.6314         2           1
Edad           21.4316        21.5541     21.2218
CicloEscolar   43.0067        43.9979     41.3091
DivisiónAcadémica 5.0442        4.8068      5.4509
AsignaturaCursa 2.4759         2.4268      2.56
PromedioBachillerato 1.6555        1.569       1.8036
Ítem1          2.4772         2.4268      2.5636
Ítem2          3.9343         3.9214      3.9564
Ítem3          3.0322         3.0127      3.0655
Ítem4          3.0214         2.9766      3.0982
Ítem5          3.9008         3.9448      3.8255
Ítem6          4.7922         4.8301      4.7273
Ítem7          4.2292         4.2442      4.2036
Ítem8          2.5027         2.4246      2.6364
Ítem9          2.4665         2.3949      2.5891
Ítem10         4.2265         4.2378      4.2073
Ítem11         3.7399         3.7495      3.7236
Ítem12         4.2815         4.2781      4.2873
Ítem13         4.0737         4.087       4.0509
Ítem14         4.1475         4.2229      4.0182
Ítem15         2.3954         2.3248      2.5164
Ítem16         2.3418         2.2463      2.5055
Ítem17         2.7024         2.6221      2.84
Ítem18         2.0013         1.9427      2.1018
Ítem19         1.9853         1.9108      2.1127
Ítem20         1.8391         1.8047      1.8982
Ítem21         1.7158         1.62        1.88
Ítem22         2.7145         2.6688      2.7927
    
```

Nota: Elaboración propia.

De igual forma en la figura anterior, se visualizan los datos aleatorios con los cuales el algoritmo se basó para la agrupación de las instancias, así como la distancia media al centroide de los clústeres y distancia total por cada atributo o variable que conforman el *dataset*.

Derivado de la evaluación del modelo sobre los clústeres, se obtuvieron como resultados que el número de instancias agrupadas en el *clúster 0* se conformaron por 471 instancias (63%) y el *clúster 1* conformado por 275 instancias (37%) (ver figura 178).

Figura 178

Visualización de la asignación de las instancias en los clústeres y la asignación de acuerdo con la variable ClaseEstudiante en Weka.

```
=== Model and evaluation on training set === Class attribute: ClaseEstudiante
Classes to Clusters:

Clustered Instances
0      471 ( 63%)
1      275 ( 37%)

0  1  <-- assigned to cluster
430 242 | Autorregulador
41  33 | Procrastinador

Cluster 0 <-- Autorregulador
Cluster 1 <-- Procrastinador
```

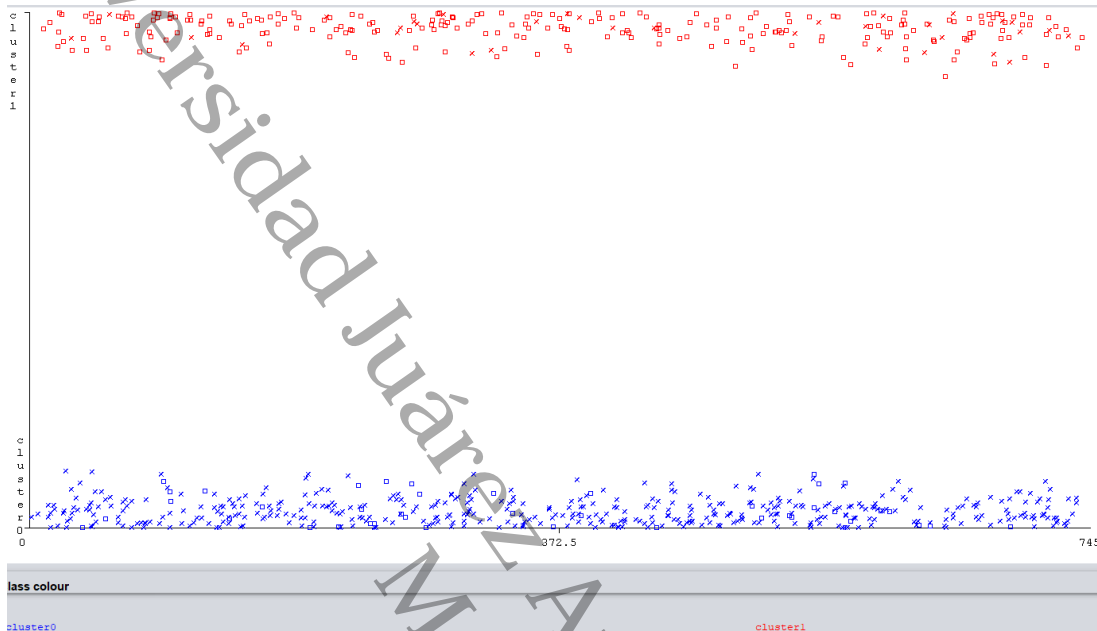
Nota: Elaboración propia.

De igual forma en la figura, se muestra la distribución de estas instancias que conforman los *clústeres* de acuerdo con la variable clase del estudiante, donde el *clúster 0* está conformado por 430 instancias de la clase autorregulador y 41 instancias de la clase procrastinador, en cambio el *clúster 1*, contiene 242 instancias de la clase autorregulador y de la clase procrastinador 33 instancias.

Por último, en la figura 179 se visualizan los dos grupos de agrupamiento, donde el *clúster 0* (grupo 1) se encuentra de color azul y el *clúster 1* (grupo uno) en color rojo, en el eje X del gráfico es el número de instancias y el eje Y los *clústeres* (ver figura 179).

Figura 179

Gráfico del resultado obtenido con la técnica de agrupamiento en Weka.



Nota: Elaboración propia.

En conclusión, con la técnica de agrupamiento realizada en tres herramientas como los son Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*, se obtuvieron resultados de agrupamiento diferentes, aunque se implementaron en el *dataset* el mismo número de *clústeres*, pero los pasos para la implementación de estos en las herramientas son diferentes.

Con las herramientas de *Microsoft Azure Machine Learning* y *Weka*, el número de instancias agrupadas en los de los *clústeres* fueron iguales, en el caso del *clúster 0*, el algoritmo en las dos herramientas clasificó 471 instancias y en el *clúster 1*, las restantes que son 275 instancias de las 746 instancias que contiene el *dataset*.

En caso contrario a lo anterior con la herramienta de Lenguaje R, el algoritmo clasificó las instancias en los dos clústeres casi de forma igualitaria, ya que el *clúster* 1 está conformado por 370 instancias y 376 instancias en el *clúster* 2.

4.2.3 Resultados de Reglas de asociación

La obtención de los resultados de la técnica de reglas de asociación, implementando el algoritmo A priori, se realizó utilizando un soporte de 0.1 y la confianza de 0.8, estos parámetros son los utilizados por defecto por el algoritmo.

De igual manera, en el caso de la herramienta *Microsoft Azure Machine Learning* y *Weka*, se indicaron en los parámetros del algoritmo que fueran 20 reglas, en cambio con la herramienta de Lenguaje R se realizó un filtrado del total de reglas, especificando que el dato de salida fuera 20 reglas; las cuales son reglas relevantes y de mayor confianza, ordenadas de forma decreciente.

A continuación, se describen los resultados obtenidos con esta técnica de minería de datos.

4.2.3.1 Lenguaje R

La ejecución del algoritmo A priori en Lenguaje R se realizó con un soporte del 0.1 y confianza de 0.8, con lo cual se obtuvo un resultado total de 13,215 reglas de asociación, de las cuales el tamaño o la longitud de cada regla variaban.

El tamaño de las reglas de asociación fue: dos reglas conformadas por solo una variable, 125 reglas con dos variables, las reglas con tres variables son 2,187 reglas, 4,412 reglas con cuatro variables, 3904 reglas con una longitud de cinco variables, 2044

reglas de seis variables, reglas conformadas por siete variables fueron 486 reglas y 12 reglas con ocho variables.

En la figura 180 se visualiza lo anteriormente descrito y se presentan los resultados obtenidos para los argumentos o parámetros que componen a este algoritmo, como lo son: *support* (soporte), *confidence* (confianza), *coverage* (cobertura), *lift* y *count* (ver figura 180).

Figura 180

Resultados del modelo de reglas de asociación obtenido en Lenguaje R.

```

set of 13215 rules

rule length distribution (lhs + rhs):sizes
  1   2   3   4   5   6   7   8
  2 168 2187 4412 3904 2044 486 12

  Min. 1st Qu.  Median    Mean 3rd Qu.  Max.
  1.000  4.000   4.000   4.527  5.000   8.000

summary of quality measures:
  support      confidence      coverage      lift      count
Min.   :0.1005   Min.   :0.8000   Min.   :0.1005   Min.   :0.8881   Min.   : 75.00
1st Qu.:0.1059   1st Qu.:0.8652   1st Qu.:0.1166   1st Qu.:1.0849   1st Qu.: 79.00
Median :0.1180   Median :0.9259   Median :0.1287   Median :1.1176   Median : 88.00
Mean   :0.1306   Mean   :0.9191   Mean   :0.1428   Mean   :1.3996   Mean   : 97.46
3rd Qu.:0.1381   3rd Qu.:0.9802   3rd Qu.:0.1528   3rd Qu.:1.6955   3rd Qu.:103.00
Max.   :0.9008   Max.   :1.0000   Max.   :1.0000   Max.   :4.6548   Max.   :672.00

mining info:
  data ntransactions support confidence
datos          746      0.1      0.8
    
```

Nota: Elaboración propia.

Al ser un gran número de reglas de asociación, se realizó un filtrado de las reglas para obtener 20 reglas de asociación, con la mayor confianza de las 13,215 reglas encontradas por algoritmo en un principio.

De acuerdo con lo anterior, en la figura 181 se presentan las 20 reglas con mayor confianza ordenadas en forma decreciente, obtenidas después de la filtración de las reglas de asociación total (ver figura 181).

Figura 181
Resultados de las reglas de asociación con Lenguaje R.

	lhs	rhs	support	confidence	coverage	lift	count
[1]	{Item1=Nunca, Item7=Siempre}	=> {Item6=Siempre}	0.1058981	1	0.1058981	1.172956	79
[2]	{Item8=Nunca, Item13=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1166220	1	0.1166220	1.110119	87
[3]	{Item7=Siempre, Item8=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1193029	1	0.1193029	1.110119	89
[4]	{Item8=Nunca, Item14=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1179625	1	0.1179625	1.110119	88
[5]	{Item8=Nunca, Item10=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1072386	1	0.1072386	1.110119	80
[6]	{Item8=Nunca, Item12=Siempre}	=> {Item6=Siempre}	0.1139410	1	0.1139410	1.172956	85
[7]	{Item8=Nunca, Item12=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1139410	1	0.1139410	1.110119	85
[8]	{Item6=Siempre, Item8=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1729223	1	0.1729223	1.110119	129
[9]	{Item10=Siempre, Item17=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1058981	1	0.1058981	1.110119	79
[10]	{Item12=Siempre, Item17=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1085791	1	0.1085791	1.110119	81
[11]	{Item6=Siempre, Item17=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1595174	1	0.1595174	1.110119	119
[12]	{DivisiónAcadémica=DACB, Item21=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1045576	1	0.1045576	1.110119	78
[13]	{Item7=Siempre, Item15=Nunca}	=> {Item6=Siempre}	0.1206434	1	0.1206434	1.172956	90
[14]	{Item5=Siempre, Item9=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1018767	1	0.1018767	1.110119	76
[15]	{Item9=Nunca, Item13=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1246649	1	0.1246649	1.110119	93
[16]	{Item7=Siempre, Item9=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1246649	1	0.1246649	1.110119	93
[17]	{Item9=Nunca, Item14=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1260054	1	0.1260054	1.110119	94
[18]	{Item9=Nunca, Item10=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1206434	1	0.1206434	1.110119	90
[19]	{Item9=Nunca, Item12=Siempre}	=> {ClaseEstudiante=Autorregulador}	0.1286863	1	0.1286863	1.110119	96
[20]	{Item6=Siempre, Item9=Nunca}	=> {ClaseEstudiante=Autorregulador}	0.1836461	1	0.1836461	1.110119	137

Nota: Elaboración propia.

A continuación, se describen los patrones de conducta visualizados en la figura anterior:

1. Cuando el ítem 1 sea nunca y el ítem 7 es siempre, entonces el ítem 6 será siempre.
2. Si el ítem 8 sea nunca y el ítem 13 siempre, entonces la clase del estudiante será autorregulador.
3. Cuando el ítem 7 sea siempre y el ítem 8 nunca, entonces la clase del estudiante será autorregulador.
4. Si el ítem 8 es nunca y el ítem 14 siempre, entonces la clase del estudiante será autorregulador.

5. Si el ítem 8 es nunca y el ítem 10 siempre, entonces la clase del estudiante será autorregulador.
6. Cuando el ítem 8 es nunca y el ítem 12 siempre, entonces el ítem 6 será siempre
7. Si el ítem 8 es nunca y el ítem 12 siempre, entonces la clase del estudiante será autorregulador.
8. Cuando el ítem 6 sea siempre y el ítem 8 nunca, entonces la clase del estudiante será autorregulador.
9. Cuando el ítem 10 sea siempre y el ítem 17 nunca, entonces la clase del estudiante será autorregulador.
10. Cuando el ítem 12 sea siempre y el ítem 17 nunca, entonces la clase del estudiante será autorregulador.
11. Si el ítem 6 es siempre y el ítem 17 nunca, entonces la clase del estudiante será autorregulador.
12. Si la División Académica es DACB y el ítem 21 nunca, entonces la clase del estudiante será autorregulador.
13. Cuando el ítem 7 sea siempre y el ítem 15 nunca, entonces el ítem 6 será siempre
14. Cuando el ítem 5 sea siempre y el ítem 9 nunca, entonces la clase del estudiante será autorregulador.

15. Si el ítem 9 sea nunca y el ítem 13 sea siempre, el estudiante será autorregulador.
16. Cuando el ítem 7 sea siempre y el ítem 9 nunca, la clase del estudiante será autorregulador.
17. Si el ítem 9 sea nunca y el ítem 14 sea siempre, el estudiante será autorregulador.
18. Si el ítem 9 sea nunca y el ítem 10 sea siempre, el estudiante será autorregulador.
19. Cuando el ítem 9 sea nunca y el ítem 12 sea siempre, entonces el estudiante será autorregulador.
20. Cuando el ítem 6 sea siempre y el ítem 9 nunca, la clase del estudiante será autorregulador.

A partir de los resultados de las reglas de asociación, se identificaron una serie de variables relevantes para la creación de las reglas de asociación y con ellos poder identificar patrones y asociaciones entre los datos del *dataset*, estas variables son las siguientes:

- Ítem 1. Cuando tengo que hacer una tarea la dejo para el último minuto.
- Ítem 5. Cuando tengo problemas para entender algo, inmediatamente trato de buscar ayuda.
- Ítem 6. Asisto a clase.
- Ítem 7. Trato de completar el trabajo asignado lo más pronto posible.

- Ítem 8. Postergo los trabajos de los cursos que no me gustan.
- Ítem 9. Postergo las lecturas de los cursos que me disgustan.
- Ítem 10. Intento mejorar mis hábitos de estudio.
- Ítem 12. Trato de motivarme para mantener mi ritmo de estudio.
- Ítem 13. Trato de terminar mis trabajos importantes con tiempo de sobra.
- Ítem 14. Me tomo el tiempo de revisar mis tareas antes de entregarlas.
- Ítem 15. Dejo para mañana lo que puedo hacer hoy.
- Ítem 17. Mientras desarrollo mis actividades del curso en línea me conecto a otras páginas que son de mi agrado.
- Ítem 21. He dejado de hacer tareas de mi curso por distraerme con el internet.

4.2.3.2 Microsoft Azure Machine Learning

Los resultados obtenidos de reglas de asociación se realizaron con los mismos parámetros de soporte y confianza, así como la obtención de 20 reglas ordenadas por la métrica o parámetro de confianza, además se realizó la transformación de los datos a valores categóricos, esto se realizó únicamente con esta herramienta.

En la figura 182 se visualizan las primeras ocho reglas de asociación de las 20 reglas obtenidas como resultado de la implementación del algoritmo A priori, donde el mínimo de soporte es 0.1 y confianza del 0.8; de igual forma se muestran los resultados de los parámetros de *support* (soporte), *confidence* (confianza) y *lift* de cada una de las reglas (ver figura 182).

Figura 182
Resultados del modelo de reglas de asociación obtenido en Microsoft Azure Machine Learning.

rows 20 columns 6

view as

id	lhs	rhs	support	confidence	lift
1	{Item8=[1.0,1.8],Item9=[1.0,1.8],Item13=[4.2,5.0]}	{ClaseEstudiante=AUTOR REGULADOR}	0.104558	1	1.110119
2	{Item7=[4.2,5.0],Item8=[1.0,1.8],Item9=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.101877	1	1.110119
3	{Item8=[1.0,1.8],Item9=[1.0,1.8],Item14=[4.2,5.0]}	{ClaseEstudiante=AUTOR REGULADOR}	0.103217	1	1.110119
4	{Item6=[4.2,5.0],Item8=[1.0,1.8],Item9=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.144772	1	1.110119
5	{Item6=[4.2,5.0],Item8=[1.0,1.8],Item16=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.115282	1	1.110119
6	{Item6=[4.2,5.0],Item8=[1.0,1.8],Item13=[4.2,5.0]}	{ClaseEstudiante=AUTOR REGULADOR}	0.115282	1	1.110119
7	{Item6=[4.2,5.0],Item8=[1.0,1.8],Item19=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.120643	1	1.110119
8	{Item7=[4.2,5.0],Item8=[1.0,1.8],Item21=[1.0,1.8]}	{ClaseEstudiante=AUTOR REGULADOR}	0.101877	1	1.110119

Nota: Elaboración propia.

Aunado a la figura anterior, en la tabla 17 se muestran las 20 reglas de asociación obtenidas en esta herramienta (ver tabla 17); de igual manera en el apéndice A, se encuentra el catálogo de las descripciones de las variables y valores de los datos, también el significado de los valores categóricos, por ejemplo, el dato 1.0,1.8 significa nunca en la escala de *Likert* y el dato 4.2,5.0 es siempre en la escala de *Likert*.

Tabla 17
Reglas de asociación obtenidas en Microsoft Azure Machine Learning.

ID	lhs	rhs	Support	Confidence	Lift
1	{Item8= [1.0,1.8), Item9= [1.0,1.8), Item13= [4.2,5.0]}	{ClaseEstudiante= AUTORREGULADOR}	0.104558	1	1.110119

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

ID	lhs	rhs	Support	Confidence	Lift
2	{Item7= [4.2,5.0], Item8= [1.0,1.8], Item9= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.101877	1	1.110119
3	{Item8= [1.0,1.8], Item9= [1.0,1.8], Item14= [4.2,5.0]}	{ClaseEstudiante= AUTORREGULADOR}	0.103217	1	1.110119
4	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item9= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.144772	1	1.110119
5	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item16= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.115282	1	1.110119
6	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item13= [4.2,5.0]}	{ClaseEstudiante= AUTORREGULADOR}	0.115282	1	1.110119
7	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item19= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.120643	1	1.110119
8	{Item7= [4.2,5.0], Item8= [1.0,1.8], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.101877	1	1.110119
9	{Item6= [4.2,5.0], Item7= [4.2,5.0], Item8= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.117962	1	1.110119
10	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item18= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.117962	1	1.110119
11	{Item8= [1.0,1.8], Item14= [4.2,5.0], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.101877	1	1.110119
12	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item14= [4.2,5.0]}	{ClaseEstudiante= AUTORREGULADOR}	0.115282	1	1.110119
13	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item10= [4.2,5.0]}	{ClaseEstudiante= AUTORREGULADOR}	0.105898	1	1.110119
14	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item12= [4.2,5.0]}	{ClaseEstudiante= AUTORREGULADOR}	0.113941	1	1.110119
15	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item20= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.135389	1	1.110119
16	{Item6= [4.2,5.0], Item8= [1.0,1.8], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.13941	1	1.110119
17	{Genero= [1.5,2.0], Item6= [4.2,5.0], Item8= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.120643	1	1.110119

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

ID	lhs	rhs	Support	Confidence	Lift
18	{Item6= [4.2,5.0], Item17= [1.0,1.8), Item19= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.136729	1	1.110119
19	{Item6= [4.2,5.0], Item17= [1.0,1.8), Item18= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.13941	1	1.110119
20	{Item6= [4.2,5.0], Item12= [4.2,5.0], Item17= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.100536	1	1.110119

Nota: Elaboración propia.

A continuación, se describen los patrones de conducta, encontrados por el algoritmo, de acuerdo con la tabla anterior:

1. Cuando el ítem 8 sea nunca, el ítem 9 nunca y el ítem 13 siempre, entonces la clase del estudiante será autorregulador.
2. Si el ítem 7 es siempre, el ítem 8 nunca y el ítem 9 es nunca, entonces la clase del estudiante será autorregulador.
3. Cuando el ítem 8 sea nunca, el ítem 9 nunca y el ítem 14 siempre, entonces la clase del estudiante será autorregulador.
4. Cuando el ítem 6 sea siempre, el ítem 8 nunca y el ítem 9 sea nunca, entonces la clase del estudiante será autorregulador.
5. Si el ítem 6 es siempre, el ítem 8 nunca y el ítem 16 sea nunca, entonces la clase del estudiante será autorregulador.
6. Cuando el ítem 6 sea siempre, el ítem 8 nunca y el ítem 13 sea siempre, entonces la clase del estudiante será autorregulador.
7. Si el ítem 6 sea siempre, el ítem 8 nunca y el ítem 19 sea nunca, entonces el estudiante será de clase autorregulador.

8. Además, si el ítem 7 siempre, el ítem 8 nunca y el ítem 21 sea nunca, entonces el estudiante es autorregulador.
9. Cuando el ítem 6 es siempre, el ítem 7 es siempre y el ítem 9 sea nunca, entonces la clase del estudiante será autorregulador.
10. De igual forma, si el ítem 6 es siempre, el ítem 8 nunca y el ítem 18 sea nunca, la clase del estudiante será autorregulador.
11. Cuando el ítem 8 sea nunca, el ítem 14 siempre y el ítem 21 nunca, el estudiante será autorregulador.
12. Cuando el ítem 6 es siempre, el ítem 8 nunca y el ítem 14 es siempre, la clase del estudiante será autorregulador.
13. Si el ítem 6 es siempre, el ítem 8 nunca y el ítem 10 es siempre, la clase del estudiante será autorregulador.
14. Si el ítem 6 es siempre, el ítem 8 nunca y el ítem 12 es siempre, entonces la clase del estudiante será autorregulador.
15. Cuando el ítem 6 es siempre, el ítem 8 nunca y el ítem 20 nunca, la clase del estudiante será autorregulador.
16. Cuando el ítem 6 es siempre, el ítem 8 nunca y el ítem 21 nunca, entonces la clase del estudiante será autorregulador.
17. Si la variable género es hombres, el ítem 6 es siempre y el ítem 8 nunca, la clase del estudiante será autorregulador.

18. Cuando el ítem 6 es siempre, el ítem 17 nunca y el ítem 19 es nunca, la clase del estudiante será autorregulador.
19. Si el ítem 6 es siempre, el ítem 17 nunca y el ítem 18 nunca, entonces la clase del estudiante será autorregulador.
20. Cuando el ítem 6 es siempre, el ítem 12 siempre y el ítem 17 sea nunca, la clase del estudiante será autorregulador.

A partir de los resultados de las reglas de asociación anteriores descritas, se identificaron una serie de variables relevantes para la creación de las reglas de asociación y con ellos se identificaron patrones y asociaciones entre los datos del *dataset*; estas variables son las siguientes:

- Ítem 6. Asisto a clase.
- Ítem 7. Trato de completar el trabajo asignado lo más pronto posible.
- Ítem 8. Postergo los trabajos de los cursos que no me gustan.
- Ítem 9. Postergo las lecturas de los cursos que me disgustan.
- Ítem 10. Intento mejorar mis hábitos de estudio.
- Ítem 12. Trato de motivarme para mantener mi ritmo de estudio.
- Ítem 13. Trato de terminar mis trabajos importantes con tiempo de sobra.
- Ítem 14. Me tomo el tiempo de revisar mis tareas antes de entregarlas.
- Ítem 16. Disfruto la mezcla de desafío con emoción de esperar hasta el último minuto para completar una tarea.

- Ítem 17. Mientras desarrollo mis actividades del curso en línea me conecto a otras páginas que son de mi agrado.
- Ítem 18. El impulso por revisar mis redes sociales mientras estoy revisando mi curso en línea es incontrolable.
- Ítem 19. Invierto más tiempo en revisar diversos sitios mientras estoy en la plataforma del curso que en desarrollar otras actividades.
- Ítem 20. Me siento nervioso y hasta desesperado si no puedo navegar
- Ítem 21. He dejado de hacer tareas de mi curso por distraerme con el internet.

4.2.3.3 Weka

Las reglas de asociación obtenidas en la ejecución del algoritmo A priori fueron 20 las que consideró más relevantes, las cuales el algoritmo las considera como las mejores reglas encontradas, éstas tienen un soporte mínimo del 0.1 y confianza del 0.8, además las reglas están ordenadas de acuerdo con el parámetro de confianza.

Aunado a lo anterior, el algoritmo A priori generó un conjunto de elementos, los cuales indican el tamaño de las reglas, a partir del número de variables que la constituyen; esto se presenta en la figura 183, donde se visualiza el tamaño de las reglas encontradas, así como los parámetros utilizados para la generación de éstas (ver figura 138).

Figura 183
Modelo de reglas de asociación obtenido en Weka.

```
Minimum support: 0.1 (75 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 89
Size of set of large itemsets L(2): 1208
Size of set of large itemsets L(3): 2862
Size of set of large itemsets L(4): 2819
Size of set of large itemsets L(5): 1549
Size of set of large itemsets L(6): 554
Size of set of large itemsets L(7): 98
Size of set of large itemsets L(8): 2
```

Nota: Elaboración propia.

De acuerdo, a la figura anterior, se aprecia el tamaño de las reglas según el número de variables que la conforman, donde el número de reglas de asociación identificadas conformadas por una variable es 89 reglas, 1,208 reglas conformadas por dos variables, 2,862 reglas conformadas por tres variables, así como el número de reglas conformadas con cuatro variables es de 2,819; 1,549 reglas con cinco variables, las reglas conformadas por seis variables es 554; 98 reglas con siete variables y dos variables conformadas por ocho variables.

La figura 184 muestra las 20 reglas de asociación obtenidas con el algoritmo A priori en la herramienta *Weka*, donde se observan los valores y parámetros de *confidence*

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

(confianza), *lift*, *leverage* y *conviction* (convicción) resultantes de cada una de las 20 reglas identificadas con mayor relevancia (ver figura 184).

Figura 184
Resultados de las reglas de asociación obtenidas con Weka.

```

1. ítem1=Casi nunca ítem16=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
2. ítem7=Siempre ítem18=Casi nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
3. ítem8=Casi nunca ítem19=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
4. ítem12=Siempre ítem17=A veces 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
5. CicloEscolar=Febrero-Agosto 2019 ítem7=Siempre ítem21=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
6. PromedioBachillerato=Menos de 9 y más de 8 ítem7=Siempre ítem14=Siempre 75 ==> ítem6=Siempre 75 <conf:(1)> lift:(1.17) lev:(0.01) [11] conv:(11.06)
7. PromedioBachillerato=Entre 9 y 10 ítem10=Siempre ítem18=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
8. ítem2=Siempre ítem5=Siempre ítem18=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
9. ítem2=Siempre ítem16=Nunca ítem18=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
10. ítem2=Siempre ítem16=Nunca ítem20=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
11. ítem6=Siempre ítem7=Siempre ítem17=Casi nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
12. ítem6=Siempre ítem8=Casi nunca ítem10=Siempre 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
13. ítem9=Nunca ítem4=Siempre ítem19=Nunca 75 ==> ítem6=Siempre 75 <conf:(1)> lift:(1.17) lev:(0.01) [11] conv:(11.06)
14. ítem6=Siempre ítem9=Casi nunca ítem15=Casi nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
15. ítem6=Siempre ítem10=Siempre ítem16=Casi nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
16. ítem6=Siempre ítem11=Casi siempre ítem16=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
17. ítem6=Siempre ítem12=Siempre ítem17=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
18. ítem6=Siempre ítem17=A veces ítem20=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
19. ítem7=Siempre ítem9=Nunca ítem20=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)
20. ítem8=Casi nunca ítem9=Casi nunca ítem20=Nunca 75 ==> ClaseEstudiante=Autorregulador 75 <conf:(1)> lift:(1.11) lev:(0.01) [7] conv:(7.44)

```

Nota: Elaboración propia.

Como se aprecia en la figura anterior, las reglas de asociación no se visualizan claramente, así como la estructura de estas, por lo cual, se realizó la tabla 18 para su mejor apreciación con las 20 reglas obtenidas de la figura 184 (ver tabla 18).

Tabla 18
Reglas de asociación obtenidas en Weka.

ID	lhr	rhs	Ocurrencia de la regla (instancias)	Confidence	Lift	Leverage	Conviction
1	Ítem1=Casi nunca Ítem16=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
2	Ítem7=Siempre Ítem18=Casi nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
3	Ítem8=Casi nunca Ítem19=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
4	Ítem12=Siempre Ítem17=A veces	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

ID	lhs	rhs	Ocurrencia de la regla (instancias)	Confidence	Lift	Leverage	Conviction
5	CicloEscolar= Febrero-Agosto 2019 Ítem7=Siempre Ítem21=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
6	PromedioBachillerato= Menos de 9 y más de 8 Ítem7=Siempre Ítem14=Siempre	Ítem6=Siempre	75 instancias	1	1.17	0.01	11.06
7	PromedioBachillerato= Entre 9 y 10 Ítem10=Siempre Ítem18=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
8	Ítem2=Siempre Ítem5=Siempre Ítem18=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
9	Ítem2=Siempre Ítem16=Nunca Ítem18=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
10	Ítem2=Siempre Ítem16=Nunca Ítem20=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
11	Ítem6=Siempre Ítem7=Siempre Ítem17=Casi nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
12	Ítem6=Siempre Ítem8=Casi nunca Ítem10=Siempre	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
13	Ítem9=Nunca Ítem14=Siempre Ítem19=Nunca	Ítem6=Siempre	75 instancias	1	1.17	0.01	11.06
14	Ítem6=Siempre Ítem9=Casi nunca Ítem15=Casi nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
15	Ítem6=Siempre Ítem10=Siempre Ítem16=Casi nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
16	Ítem6=Siempre Ítem11=Casi siempre Ítem16=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

ID	lhs	rhs	Ocurrencia de la regla (instancias)	Confidence	Lift	Leverage	Conviction
17	Ítem6=Siempre Ítem12=Siempre Ítem17=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
18	Ítem6=Siempre Ítem17=A veces Ítem20=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
19	Ítem7=Siempre Ítem9=Nunca Ítem20=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44
20	Ítem8=Casi nunca Ítem9=Casi nunca Ítem20=Nunca	ClaseEstudiante= Autorregulador	75 instancias	1	1.11	0.01	7.44

Nota: Elaboración propia.

De acuerdo, a lo anterior cada una de las reglas ocurre tanto en el antecedente como en la regla completa en 75 instancias. A continuación, se describen los patrones de conducta, encontrados por el algoritmo:

1. Cuando el ítem 1 sea casi nunca y el ítem 16 nunca, entonces la clase del estudiante será autorregulador.
2. Si el ítem 7 es siempre y el ítem 18 casi nunca, entonces la clase del estudiante será autorregulador.
3. Cuando el ítem 8 casi nunca y el ítem 19 nunca, la clase del estudiante será autorregulador.
4. Cuando el ítem 12 sea siempre y el ítem 17 sea a veces, entonces la clase del estudiante será autorregulador.

5. Si la variable ciclo escolar fue el periodo de febrero–agosto 2019, el ítem 7 siempre y el ítem 21 nunca, entonces la clase del estudiante será autorregulador.
6. Cuando la variable promedio del bachillerato fue menos de 9 y más de 8, el ítem 7 siempre y el ítem 14 siempre, entonces el ítem 6 será siempre.
7. Si la variable promedio del bachillerato fue entre 9 y 10, el ítem 10 siempre y el ítem 18 sea nunca, entonces la clase del estudiante será autorregulador.
8. Cuando el ítem 2 sea siempre, el ítem 5 siempre y el ítem 18 sea nunca, entonces la clase del estudiante será autorregulador.
9. Cuando el ítem 2 sea siempre, el ítem 16 sea nunca y el ítem 18 sea nunca, entonces la clase del estudiante será autorregulador.
10. Cuando el ítem 2 sea siempre, el ítem 16 sea nunca y el ítem 20 nunca, entonces la clase del estudiante será autorregulador.
11. Si el ítem 6 siempre, el ítem 7 sea siempre y el ítem 17 sea casi nunca, entonces la clase del estudiante será autorregulador.
12. Si el ítem 6 siempre, el ítem 8 sea casi nunca y el ítem 10 siempre, entonces la clase del estudiante será autorregulador.
13. Cuando el ítem 9 sea nunca, el ítem 14 siempre y el ítem 19 nunca, entonces el ítem 6 será siempre.
14. Si el ítem 6 siempre, el ítem 9 sea casi nunca y el ítem 15 casi nunca, entonces la clase del estudiante será autorregulador.

15. Cuando el ítem 6 es siempre, el ítem 10 es siempre y el ítem 16 sea casi nunca, la clase del estudiante será autorregulador.
16. Cuando el ítem 6 siempre, el ítem 11 es casi siempre y el ítem 16 sea nunca, la clase del estudiante será autorregulador.
17. Cuando el ítem 6 siempre, el ítem 12 siempre y el ítem 17 sea nunca, entonces la clase del estudiante será autorregulador.
18. Si el ítem 6 sea siempre, el ítem 17 a veces y el ítem 20 sea nunca, la clase del estudiante será autorregulador.
19. De igual forma, si el ítem 7 sea siempre, el ítem 9 nunca y el ítem 20 sea nunca, la clase del estudiante será autorregulador.
20. Cuando el ítem 8 sea casi nunca, el ítem 9 sea igual casi nunca y el ítem 20 sea nunca, el estudiante será autorregulador.

A partir de los resultados de las reglas de asociación anteriores descritas, se identificaron una serie de variables y datos relevantes que intervienen en las reglas de asociación; en el apéndice A, se encuentra el catálogo de las descripciones del significado de las variables y los datos.

En conclusión, la utilización de la técnica de reglas de asociación con el algoritmo A priori, en las tres herramientas como los son Lenguaje R, *Microsoft Azure Machine Learning* y *Weka*, se obtuvieron reglas diferentes, pero en algunos casos similares, solo que se contemplaban diferentes datos.

Otro de los factores identificados fue que la clase de estudiante autorregulador fue una variable constante en las reglas de asociación, es decir, todos los antecedentes de las reglas (variables o *ítems*) se asociaban con esta clase, pero en cinco en total, donde las variables se asociaban con la variable ítem 6, cuando el dato fuera siempre.

México.

Universidad Juárez Autónoma de Tabasco

Capítulo 5. Conclusiones, recomendaciones y trabajos futuros

5.1 Conclusiones

En este trabajo se identificaron patrones de conducta relacionados con las variables de procrastinación y autorregulación académica de estudiantes de licenciatura de la UJAT, empleando técnicas de minería de datos; lo cual se pudo lograr empleando las herramientas, técnicas y algoritmos utilizados para obtener estos patrones de datos.

Lo más importante para la generación de los resultados obtenidos fue la utilización de la metodología *KDD* y las técnicas de minería de datos de árboles de decisión, agrupamiento y reglas de asociación, porque con la metodología *KDD*, se siguieron una serie de pasos, para preparar el *dataset* y comenzar con la ejecución de los algoritmos a los datos y el descubrimiento de nueva información.

Así mismo, la utilización de tres herramientas para minería de datos fue de mucha ayuda, porque en cada una de ellas se descubrió nueva información.

Mediante la aplicación de los algoritmos de minería de datos, utilizados en el presente trabajo, se pudieron identificar patrones, similitudes y asociaciones entre los datos que contiene cada variable del *dataset*.

De acuerdo con lo anterior, con la técnica de árbol de decisión, se detectaron una serie de variables o *ítems* que intervienen para la clasificación de si un estudiante es procrastinador o autorregulador bajo la perspectiva académica, así como el porcentaje de instancias clasificadas correctamente, de acuerdo con la clase del estudiante antes de la implementación de minería de datos.

Por lo anterior se detectó que 16 variables, están relacionadas con la clasificación de procrastinación y autorregulación académica; pero también se identificaron seis variables que no tienen presencia al momento de clasificar las clases, estas variables son: (3) Cuando me asignan lecturas las leo la noche anterior, (4) Cuando me asignan lecturas las reviso el mismo día de clase, (11) Invierto el tiempo necesario en estudiar aun cuando el tema sea aburrido, (14) Me tomo el tiempo para revisar mis tareas antes de entregarlas, (16) Disfruto la mezcla de desafío con emoción de esperar hasta el último minuto para completar una tarea y (22) He trabajado en la plataforma con mi celular, por un lado.

De acuerdo con los resultados, la clasificación de instancias efectuadas correctamente es: 92.61% en Lenguaje R y *Weka*, y 95.97% con *Microsoft Azure Machine Learning*.

Consecutivamente, al implementar el algoritmo *K-means*, se obtuvieron dos grupos los cuales, fueron agrupados por el algoritmo según sus similitudes y la distancia existente al centro de los grupos. A partir de lo anterior se obtuvo que tanto instancias clasificadas como autorreguladores y procrastinadores comparten similitudes y relaciones.

En relación con la ejecución de esta técnica en las herramientas de *Microsoft Azure Machine Learning* y *Weka*, los grupos comparten el mismo número de instancias agrupadas en los grupos (*clústeres*) y con la herramienta de Lenguaje R sucede lo contrario, porque los grupos contienen el 49.59% y el 50.40% de instancias, es decir, cada *clúster* contiene la misma cantidad de datos.

Por último, gracias a la técnica de reglas de asociación se encontró que el *dataset* contiene 13,215 reglas asociadas con las 29 variables y 746 instancias o registros que lo conforman; además la regla de asociación puede estar conformada por una variable o hasta ocho variables.

Aunado a lo anterior se identificó que mayormente, el tamaño de una regla puede estar compuesta a partir de tres variables y hasta cinco variables, y dependiendo el valor de los parámetros considerados, en la búsqueda de reglas puede ser el número de reglas del algoritmo A priori extraído del *dataset*.

Subsecuentemente de lo anterior, se comprende también que, al utilizar diversas herramientas para minería de datos, los resultados pueden variar, ser similares o cambiar y a su vez el funcionamiento del algoritmo de una herramienta a otra herramienta puede ser diferente; este es el caso de las herramientas utilizadas en esta investigación, porque en el caso de las herramientas *Weka*, solo se configuran o cambian algunos parámetros, para poder ejecutar el algoritmo.

En el caso de la herramienta *Microsoft Azure Machine Learning*, sucede algo parecido como lo descrito anteriormente, pero con esta herramienta se crea el modelo con los recursos necesarios para poder implementar las técnicas de minería de datos, y la herramienta de Lenguaje R, ocurre lo contrario, porque con esta herramienta se tiene que programar el algoritmo a utilizar y a su vez conocer las librerías necesarias para la creación y programación del modelo del algoritmo.

En conclusión, de acuerdo con los resultados presentes, se pudieron identificar los patrones de conducta, agrupaciones y reglas de asociación presentes en el *dataset*, con relación a la procrastinación y autorregulación académica, y de igual manera al utilizar tres diferentes técnicas de minería de datos, implementadas en tres herramientas para minería de datos, los resultados que se obtuvieron fueron más amplios y en ocasiones concordaron los resultados unos con otros.

Así mismo se pudo observar que algunas instancias clasificadas con la clase autorregulador, el algoritmo la clasificaba como clase procrastinador y lo mismo sucedía con la clasificación de la clase procrastinador, pero en su mayoría las instancias y patrones de conducta identificados demuestran que la mayoría de los estudiantes son autorreguladores académicamente.

5.2 Recomendaciones

Durante el desarrollo de esta investigación, se fueron incorporando prácticas o sugerencias que fueron de mucha ayuda al momento de realizar dicha investigación; por lo consecuente en este apartado se mencionan buenas prácticas o recomendaciones que serán de ayuda a los usuarios que deseen llevar a cabo una investigación similar o que realicen trabajos futuros sobre dicho tema, a continuación, se describen dichas prácticas:

- Analizar la información que contiene el *dataset* a utilizar en la investigación, es decir, cómo se encuentra distribuida la información, el número de variables o columnas, filas y tablas, de igual forma verificar si existen datos vacíos o nulos y número de estos.

- Verificar el tipo de datos de las variables y si es necesario renombrarlas, para no tener problemas al momento de realizar la lectura en la herramienta a utilizar, ya sea herramientas de minería de datos u otras.
- Evitar los espacios en el nombre de las variables, si es posible simplificarlo o renombrarlo con nombre cortos o con nombres que se puedan identificar que datos contiene la variable.
- En el caso de la utilización de la herramienta Lenguaje R, se puede trabajar en la herramienta de *RStudio*, que es un entorno más gráfico e integrado para trabajar con el lenguaje de programación de R.
- Identificar las librerías o paquetes a utilizar e instalar, para la creación y ejecución de los modelos de las técnicas de minería de datos.
- El algoritmo A priori de la técnica de reglas de asociación, el tipo de dato de los datos del *dataset* deben ser nominal o texto, para poder ser ejecutados en la herramienta *Weka*.

5.3 Trabajos futuros

Durante el desarrollo de este trabajo de investigación han surgido algunas líneas futuras de investigación que quedan abiertas, las cuales se espera que pueda realizarse en un futuro. Estas líneas de investigación o trabajos futuros se encuentran directamente relacionados con este trabajo de tesis y han surgido durante la realización de la misma:

- Realizar minería de datos al *dataset* con otros algoritmos que contengan las técnicas de árboles de decisión, agrupamiento y reglas de asociación, de igual forma se

podrían utilizar otras herramientas y técnicas para realizar minería de datos, para visualizar si existe alguna variación en los resultados o se encuentran otros patrones de conducta de los estudiantes.

- Replicar este estudio, considerando que la población fueran los estudiantes que reciben clases presenciales en la UJAT.
- Comparar los resultados de las investigaciones, con la población de estudio en línea o a distancia con la población de estudio presencial, para identificar las diferencias de patrones, relación o similitudes que pueda haber en los resultados con respecto a la procrastinación y autorregulación académica de los estudiantes en las variaciones de educación.
- Analizar el *dataset* para identificar patrones y relaciones entre procrastinación y autorregulación académica, utilizando solamente los ítems que conforman la estructura final del instrumento validado, de las autoras García y Silva (2019).
- De igual forma se puede realizar una comparación de los resultados obtenidos de esta investigación donde se utilizaron los registros y datos del *dataset* completo, con los resultados de una investigación con las variables del cuestionario validado.
- Medir la percepción de los docentes sobre las variables de procrastinación y autorregulación académica de sus estudiantes.

Referencias

- Alegre, A. (2013). Autoeficacia y procrastinación académica en estudiantes universitarios de Lima Metropolitana. *Propósitos y representaciones*, 1(2), 57-82. DOI: 10.20511/pyr2013.v1n2.29
- Alvarado, I., Vega, Z., Cepeda, M. y Del Bosque, A. (2014). Comparación de estrategias de estudio y autorregulación en universitarios. *Revista Electrónica de Investigación Educativa*, 16(1), 137-148. Recuperado de <https://redie.uabc.mx/redie/article/view/730/897>
- Álvarez, Ó. (2010). Procrastinación general y académica en una muestra de estudiantes de secundaria de Lima metropolitana. *Persona: Revista de la Facultad de Psicología*, (13), 159-177. DOI: [10.26439/persona2010.n013.270](https://doi.org/10.26439/persona2010.n013.270)
- Amat, J. (2018). *Reglas de asociación y algoritmo Apriori con R*. *cienciadedatos.net*. Recuperado de <https://www.cienciadedatos.net/documentos/43-reglas-de-asociacion#Apriori>
- Amazon Web Services [AWS]. (s.f.). *Amazon EMR*. Recuperado de <https://aws.amazon.com/es/emr/>
- Angarita, L. (2012). Aproximación a un concepto actualizado de la procrastinación. *Revista iberoamericana de psicología: ciencia y tecnología*, 5(2), 85-94. Recuperado de <https://reviberopsicologia.iberu.edu.co/article/view/249>
- Anónimo. (2013). *Excel 2013: guía práctica para el usuario*. (1ª ed.). Buenos Aires: USERS.

- Balkis, M. & Duru, E. (2009). Prevalence of academic procrastination behavior among pre-service teachers, and its relationship with demographics and individual preferences. *Egitimde Kuram ve Uygulama*, 5(1), 18-32. Recuperado de <http://dergipark.org.tr/en/pub/eku/issue/5449/73831>
- Ballesteros, A., Sánchez-Guzmán, D. y García, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. *Lat. Am. J. Phys. Educ.*, 7(4), 662-668. Recuperado de http://www.lajpe.org/dec13/22-LAJPE_814_bis_Alejandro_Ballesteros.pdf
- Barraza, A. y Barraza, S. (2018). Evidencias de validez y confiabilidad de la escala de procrastinación académica en una población estudiantil mexicana. *Revista de Psicología y Ciencias del Comportamiento de la Unidad Académica de Ciencias Jurídicas y Sociales*, 9(1), 75-99. Recuperado de <http://www.revistapcc.uat.edu.mx/index.php/RPC/article/view/206>
- Busko, D. (1998). *Causes and consequences of perfectionism and procrastination; a structural equation model* (Tesis de Maestría Inédita). University of Guelph, Guelph Ontario, Canadá.
- Cerqueira, C., Castro, M., Tello, C. y Beltrán, P. (2016). *Relación entre procrastinación general y académica y dimensiones de personalidad en estudiantes universitarios, Chimbote – 2016*. (Tesis de grado). Universidad San Pedro, Chimbote, Perú.
- Chan, L. (2011). Procrastinación académica como predictor en el rendimiento académico en jóvenes de educación superior. *Universidad Femenina del Sagrado Corazón*,

- 7(1), 53-62. Recuperado de http://www.unife.edu.pe/publicaciones/revistas/revista_tematica_psicologia_2011/chan_bazalar.pdf
- Charris, L., Henríquez, C., Hernández, S., Jimeno, L., Guillen, O. y Moreno, S. (2018). Análisis comparativo de algoritmos de árboles de decisión en el procesamiento de datos biológicos. *Revista Investigación y Desarrollo en TIC*, 9(1), 26-34. Recuperado de <http://revistas.unisimon.edu.co/index.php/identific/article/view/3158>
- Chávez, J. y Morales, M. (2017). Procrastinación académica de estudiantes en el primer año de carrera. *Revista Electrónica sobre Cuerpos Académicos y Grupos de Investigación*, 4(8), 1-18. Recuperado de <http://www.cagi.org.mx/index.php/CAGI/article/view/154>
- Constanzo, M. (2014). Comparación de modelos de calidad, factores y métricas en el ámbito de la Ingeniería de Software. *Informes Científicos Técnicos-UNPA*, 6(1), 1-36. DOI: <https://doi.org/10.22305/ict-unpa.v6i1.89>
- Contreras, J., Molina, E. y Arteaga, P. (2010). Introducción a la programación estadística con R para profesores. (1ª ed.). Granada: Grupo de Educación Estadística, Universidad de Granada.
- Díaz-Morales, J. (2019). Procrastinación: una revisión de su medida y sus correlatos. *Revista Iberoamericana de Diagnóstico y Evaluación – e Avaliação Psicológica*, 2(15), 43-60. Recuperado de <https://www.aidep.org/sites/default/files/2019-04/RIDEP51-Art4.pdf>

- Domínguez-Lara, S. (2017a). Prevalencia de procrastinación académica en estudiantes universitarios de Lima Metropolitana y su relación con variables demográficas. *Revista de Psicología-UCSP*, 7(1), 81-95. Recuperada de <http://revistas.ucsp.edu.pe/index.php/psicologia/article/view/49>
- Domínguez-Lara, S. (2017b). Procrastinación académica, afrontamiento de la ansiedad pre-examen y rendimiento académico en estudiantes de psicología: análisis preliminar. *Cultura*, 31(), 181-193. DOI: <https://doi.org/10.24265/cultura.2017.v31.10>
- Domínguez-Lara, S. y Campos-Uscanga, Y. (2017). Influencia de la satisfacción con los estudios sobre la procrastinación académica en estudiantes de psicología: un estudio preliminar. *Liberabit*, 23(1), 123-135. DOI: <https://doi.org/10.24265/liberabit.2017.v23n1.09>
- Domínguez, S., Villegas, G. y Centeno, S. (2014). Procrastinación académica: validación de una escala en una muestra de estudiantes de una universidad privada. *Liberabit*, 20(2), 239-304. Recuperado de http://revistaliberabit.com/es/revistas/RLE_20_2_procrastinacion-academica-validacion-de-una-escala-en-una-muestra-de-estudiantes-de-una-universidad-privada.pdf
- Echeburúa, E. y de Corral, P. (2010). Adicción a las nuevas tecnologías y a las redes sociales en jóvenes: un nuevo reto. *Revista Adicciones*, 22(2), 91-95. DOI: <https://doi.org/10.20882/adicciones.196>

- Eckert, K. y Suénaga, R. (2015). Análisis de deserción-permanencia de estudiantes universitarios utilizando técnicas de clasificación en minería de datos. *Formación Universitaria*, 8(5), 3-12. DOI: 10.4067/S0718-50062015000500002
- Elvira-Valdés, M. y Pujol, L. (2012). Autorregulación y rendimiento académico en la transición secundaria-universidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 10(1), 367-378. Recuperado de <http://revistaumanizales.cinde.org.co/rlicsnj/index.php/Revista-Latinoamericana/article/view/612>
- Elvira-Valdés, M. y Pujol, L. (2014). Variables cognitivas e ingreso universitario: predictores del rendimiento académico. *Universitas Psychologica*, 13(4), 1557-1567. DOI: <https://doi.org/10.11144/Javeriana.UPSY13-4.vciu>
- Espinoza, C. (2014). *Metodología de investigación tecnológica: Pensando en sistemas*. (2ª ed.). Perú: Soluciones Gráficas.
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). The KDD Process for Extracting Useful Knowledge from Volumes of Data. *Communications of the ACM*, 39(1), 27-34. DOI: <https://doi.org/10.1145/240455.240464>
- Ferrari, J. & Díaz-Morales, J. (2007). Perceptions of self-concept and self-presentation by Procrastinators: Further Evidence. *The Spanish Journal Psychology*, 10(1), 91-96. DOI: <https://doi.org/10.1017/S113874160000634X>
- Ferrari, J., Johnson, J. & McCown, W. (1995). *Procrastination and task avoidance: theory, research, and treatment*. New York: Plenum Press.

- Ferrari, J., O'Callahan, J. & Newbegin, I. (2004). Prevalence of procrastination in the United States, United Kingdom, and Australia: Arousal and avoidance delays among adults. *North American Journal of Psychology*, 7(1), 1-6. Recuperado de <https://psycnet.apa.org/record/2005-03779-001>
- Ferrari, J. & Tice, D. (2000). Procrastination as a Self-Handicap for Men and Women: A Task-Avoidance Strategy in a Laboratory Setting. *Journal of Research in Personality*, 34(1), 73-83. DOI: <https://doi.org/10.1006/jrpe.1999.2261>
- Gamarra, G. y Farfán, M. (2019). *Análisis comparativo usando minería de datos en la predicción del rendimiento académico de adolescentes basado en emociones*. Ponencia presentada en el 17th LACCEI International Multi-Conference for Engineering, Education, and Technology. "Industry, Innovation, and Infrastructure for Sustainable Cities and Communities", Montego Bay, Jamaica.
- García, F. (2013). *Aplicación de técnicas de Minería de Datos a datos obtenidos por el Centro Andaluz de Medio Ambiente (CEAMA) (Trabajo de Máster)*. Universidad de Granada. Granada, España.
- García, M. (2012). La autorregulación académica como variable explicativa de los procesos de aprendizaje universitario. *Profesorado Revista de Currículum y Formación de Profesorado*, 16(1), 203-221. Recuperado de <http://hdl.handle.net/10481/23005>
- García, C., Agoitia, J., Acata, K., Rojas, F., Goldanski, P. y Alvarado, A. (2012). *Estudiantes de medicina con Procrastinación académica: estudio descriptivo*

- desde la teoría del análisis transaccional. Ponencia presentada en III Congreso Internacional de Educación Médica. Puerto Vallarta, Jalisco, México.
- García, C., Olvera, M., Jiménez, M. y Ramírez, M. (2014). *Procrastinación y psicoanálisis interdisciplinario: estudio descriptivo-hermenéutico llevado a cabo en estudiantes de medicina*. Ponencia presentada en el IV Congreso Internacional de Educación Médica, Puerto Vallarta, Jalisco, México.
- García, V. y Silva, M. (2020). Interacción con internet y procrastinación de estudiantes universitarios de cursos en línea. En L. Habib-Mireles (coord.). *Tecnología, Diversidad e Inclusión: repensando el modelo educativo*. (pp. 45-53). Eindhoven, NL: Adaya Press.
- García, V. y Silva, M. (2019). Procrastinación académica entre estudiantes de cursos en línea. Validación de un cuestionario. *Apertura: Revista de Innovación Educativa*, 11(2), 122-137. DOI: <http://dx.doi.org/10.32870/Ap.v11n2.1673>
- Garrote, D., Jiménez, S. y Serna, R. (2018). Gestión del tiempo y uso de las TIC en estudiantes universitarios. *Revista de Medios y Educación*, (53), 109-121. DOI: <http://dx.doi.org/10.12795/pixelbit.2018.i53.07>
- Garzón, A. y Gil, J. (2017a). El papel de la procrastinación académica como factor de la deserción universitaria. *Revista Complutense de Educación*, 28(1), 307-324. DOI: https://doi.org/10.5209/rev_RCED.2017.v28.n1.49682

- Garzón, A. y Gil, J. (2017b). Gestión del tiempo y procrastinación en la educación superior. *Universitas Psychologica*, 16(3), 1-13. DOI: <https://doi.org/10.11144/Javeriana.upsy16-3.gtpe>
- GNU. (2007a). *GNU General Public License, Versión 3*. GNU Operating System. Recuperado de <https://www.gnu.org/licenses/gpl-3.0.html>
- GNU. (2007b). *GNU Affero General Public License, versión 3*. GNU Operating System. Recuperado de <https://www.gnu.org/licenses/agpl-3.0.html>
- González, J., Calet, N., Rubio, L. y Dumitrache, C. (2018). Análisis del impacto de la procrastinación en el rendimiento académico de estudiantes universitarios usando técnicas de minería de datos. En T. Ramiro-Sánchez, M. Ramiro. Sánchez y M. Bermúdez (Ed.). *Evaluación de la Calidad de la Investigación y de la Educación Superior: Libro de Resúmenes XV FECIES*. (pp. 636). Santander, España.
- Google Cloud. (s.f.). *Cloud Datalab*. Recuperado de <https://cloud.google.com/datalab/>
- Gorbea-Portal, S. (2013). Tendencias transdisciplinarias en los estudios métricos de la información y su relación con la gestión de la información y del conocimiento. *Perspectivas em Gestão & Conhecimento*, 3(1), 13-27. Recuperado de <https://periodicos.ufpb.br/ojs2/index.php/pgc/article/view/14175>
- Gutiérrez O. y Molina, B. (2015). Identificación de técnicas de minería de datos para apoyar la toma de decisiones en la solución de problemas empresariales. *Revista Ontare*, 3(2), 33-51. DOI: <https://doi.org/10.21158/23823399.v3.n2.2015.1440>

IBM. (n.d.). *IBM SPSS Modeler*. Recuperado de <https://www.ibm.com/mx-es/products/spss-modeler>

Kantardzic, M. (2011). *Data Mining: Concepts, Models, Methods, and Algorithms*. (2nd. ed.). Hoboken, New Jersey: John Wiley & Sons, INC., Publication.

Kumar, V. (2015). *What is Intel DAAL?* Intel Data Analytics Acceleration Library. Recuperado de <https://software.intel.com/en-us/articles/what-is-intel-daal>

Lam-Figueroa, N., Contreras-Pulache, H., Mori-Quispe, E., Nizama-Valladolid, M., Gutiérrez, C., Hinostroza-Camposano, W., Torrejón, E., Hinostroza-Camposano, R., Coaquira-Condori, E. y Hinostroza-Camposano, W. D. (2011). Adicción a internet: desarrollo y validación de un instrumento en escolares adolescentes de Lima, Perú. *Revista Peruana de Medicina Experimental y Salud Publica*, 28(3), 462-469. DOI: <https://doi.org/10.17843/rpmesp.2011.283.524>

Ley Federal de Protección de datos personales en posesión de los participantes. (05 de julio de 2010). *Diario Oficial de la Federación, DOF 05-07-2010*. Recuperado del sitio de internet de la Cámara de Diputados del H. Congreso de la Unión <http://www.diputados.gob.mx/LeyesBiblio/pdf/LFPDPPP.pdf>

López, F., Macías, M. y Sandoval, E. (2019). Minería de datos: identificando causas de deserción en las instituciones públicas de educación superior de México. *Revista de Tecnología e Innovación en Educación Superior*, 1(2), 1-14. Recuperado de https://ties.unam.mx/vol1_02/art_08mineria.html

- Martínez, C. (2012). Aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión en Entel. (Tesis de maestría). Universidad de Chile, Santiago de Chile, Chile.
- Matalinares, M., Díaz, A., Rivas, L., Dioses, A., Arenas, C., Raymundo, O., Baca, D., Uceda, J., Yaringaño, J. y Fernández, E. (2017). Procrastinación y adicción a redes sociales en estudiantes universitarios de pre y post grado de Lima. *Horizontes de la Ciencia. Revista Científica*. 7(13), 63-81. Recuperado de <http://revistas.uncp.edu.pe/index.php/horizontedelaciencia/article/view/313>
- MathWorks. (s.f.-a). *Math. Graphics. Programming*. Recuperado de <https://www.mathworks.com/products/matlab.html>
- MathWorks. (s.f.-b). *Matlab for Data Science*. Recuperado de <https://www.mathworks.com/solutions/data-science.html>
- Medina, F., y Gómez, C. (2014). Funcionalidades de la minería de datos. *Revista Ingeniería y Región*, 12(2), 31-40. DOI: <https://doi.org/10.25054/22161325.728>
- Meza, J., de la Rosa, A., Rivera, J. y González, E. (2018). Evaluación de autorregulación académica en estudiantes de psicología en modalidad en línea. *Voces de la Educación*, 3(6), 126-141. Recuperado de <https://www.revista.vocesdelaeducacion.com.mx/index.php/voces/article/view/125>
- Microsoft Azure. (s.f.-a). *Conozca Azure*. Recuperado de <https://azure.microsoft.com/es-mx/overview/#what-is-azure>

Microsoft Azure. (2021a). Hoja de referencia del algoritmo de aprendizaje automático para el diseñador de aprendizaje automático de Azure. Recuperado de <https://docs.microsoft.com/en-us/azure/machine-learning/algorithm-cheat-sheet>

Microsoft Azure. (s.f.-b). *Precios de Azure Machine Learning*. Recuperado de <https://azure.microsoft.com/es-mx/pricing/details/machine-learning/>

Microsoft Azure. (2021b). What is Azure Machine Learning? Recuperado de <https://docs.microsoft.com/en-us/azure/machine-learning/overview-what-is-azure-machine-learning>

Microsoft. (2017). *Cliente de minería de datos para Excel (Complementos de minería de datos de SQL Server)*. Recuperado de <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining-client-for-excel-sql-server-data-mining-add-ins?view=sql-server-2014>

Microsoft. (2014). *Microsoft SQL Server*. Recuperado de [https://docs.microsoft.com/es-es/previous-versions/bb545450\(v=msdn.10\)](https://docs.microsoft.com/es-es/previous-versions/bb545450(v=msdn.10))

Miranda, C. y Flores, J. (2016). Autoeficacia percibida, Estilos de aprendizaje y Procrastinación académica en Estudiantes universitarios. [Edición especial] *Tlamati Sabiduría*, 7(2), 1-15. Recuperado de <http://tlamati.uagro.mx/t7e2/82.pdf>

Miranda, M. y Guzmán, J. (2017). Análisis de la deserción de estudiantes universitarios usando técnicas de minería de datos. *Formación Universitaria*, 10(3), 61-68. DOI: 10.4067/S0718-50062017000300007

- Morales, M. (2018). *Evidencias de validez de la Escala de Procrastinación Académica en estudiantes universitarios del distrito de Trujillo*. (Tesis de grado). Universidad César Vallejo, Perú.
- Morales, M. y Chávez, J. (2017). Adaptación a la vida universitaria y procrastinación académica en estudiantes de psicología. *Revista Electrónica del Desarrollo Humano para la Innovación Social*, 4(8), 1-16. Recuperado de <http://www.cdhis.org.mx/index.php/CAGI/article/view/121>
- Moreno, M., Figueroa, E. y Arrieta, D. (2015). La motivación para promover la autorregulación en la clase de inglés de negocios. En L. Hernández (Ed.). *Autorregulación académica, Proceso desde la asociación de los estudiantes*. (pp. 52-76). Durango, Dgo., México: Instituto Universitario Anglo Español.
- Mousalli-Kayat, G. (2015). Métodos y diseños de investigación cuantitativa. Mérida.
- Navarro-Mancilla, A. y Rueda-Jaimes, G. (2007). Adicción a Internet: revisión crítica de la literatura. *Revista Colombiana de Psiquiatría*, 36(4), 691-700. Recuperado de <http://psiquiatria.org.co/web/publicaciones/revista-colombiana-de-psiquiatria/volumen-36-n-4-ano-2007/>
- Núñez, A., Ferreira, H., Ayala, K., Ruiz, J. y Cuin, J. (2018). *Construcción de modelo de referencia para el desarrollo de aplicaciones WEB: propuesta metodológica*. Ponencia llevada a cabo en el Congreso Internacional de Investigación de Academia Journals Morelia 2018, Morelia, Michoacán.

Olazábal, R. y Zegarra, V. (2017). Procrastinación e intereses a los videojuegos en ingresantes a la escuela de psicología de una universidad privada. *Revista Paian*, 8(1), 4 -14. Recuperado de

<http://revistas.uss.edu.pe/index.php/PAIAN/article/view/545>

Oracle. (s.f.). *Oracle Business Intelligence*. Recuperado de

<https://www.oracle.com/mx/middleware/technologies/bi.html>

Panadero, E. y Alonso-Tapia, J. (2014). Teorías de autorregulación educativa: Una comparación y reflexión teórica. *Psicología Educativa*, 20(1), 11-22. DOI:

<https://doi.org/10.1016/j.pse.2014.05.002>

Paz, A., Aranda, R., Navarro, M., Delgado, M. y Sayas, Y. (2014). Representaciones mentales sobre la procrastinación en estudiantes de psicología de la UNMSM. *Revista Electrónica de Psicología Iztacala*. 17(3), 1148-1167. Recuperado de

<http://www.revistas.unam.mx/index.php/rep/rep/article/view/47421>

Pintrich, P. (2000). The role of goal orientation in self-regulated learning. En M. Boekaerts, P. Pintrich y M. Zeidner (Eds.). *Handbook of self-regulation* (pp. 451-502). San Diego: Academic Press.

Pressman, R. (2010). *Ingeniería de Software, Un enfoque practico*. (7ma ed.). D.F., México: McGraw-Hill.

Python. (s.f.). *Installing Python Modules*. Recuperado de

<https://docs.python.org/3/installing/index.html>

Quant, D. y Sánchez A. (2012). Procrastinación, procrastinación académica: concepto e implicaciones. *Revista Vanguardia Psicológica Clínica Teórica y Práctica*, 3(1), 45-59. Recuperado de

http://journaldatabase.info/articles/procrastinacion_procrastinacion.html

Rafael, A. y Ramírez, A. (2016). *Procrastinación y rendimiento académico en los alumnos de la Escuela Profesional de Psicología de la Universidad Peruana Unión, Filial Tarapoto, 2016*. (Tesis de grado) Universidad Peruana Unión, Perú.

Ramírez, F. y Zwerg-Villegas, A. (2012). Metodología de la investigación: más que una receta. *Revista AD-minister*, (20), 91-111. Recuperado de

<https://publicaciones.eafit.edu.co/index.php/administer/article/view/1344>

Ramos-Galarza, C., Jadán-Guerrero, J., Paredes-Nuñez, L., Bolaños-Pasquel, M., y Gómez-García, A. (2017). Procrastination, internet addiction, and academic performance in Ecuadorian college students. *Estudios Pedagógicos*, 43(3), 275-289. DOI: <https://doi.org/10.4067/S0718-07052017000300016>

Redacción España. (2020). *¿Qué es Weka y qué tiene que ver con Big Data?* B12 Tech4Business. Recuperado de https://agenciab12.mx/noticia/que-es-weka-que-tiene-que-ver-big-data?utm_source=social&utm_medium=facebook_es

Ribas, E. (2018). *¿Qué es el Data Mining o minería de datos?* IEBS. Recuperado de <https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/>

- Rodríguez, A. y Clariana, M. (2017). Procrastinación en estudiantes universitarios: su relación con la edad y el curso académico. *Revista Colombiana de Psicología*, 26(1), 45-60. DOI: <https://doi.org/10.15446/rcp.v26n1.53572>
- Rodríguez, G. (2017). Construcción y validación de las propiedades psicométricas del cuestionario de procrastinación académica para estudiantes universitarios, 2015. *Apuntes Científicos Estudiantiles de Psicología*, 1(1), 1-11. Recuperada de https://revistas.upeu.edu.pe/index.php/r_psicologia/article/view/938
- Rodríguez, Y. y Díaz, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, 3(3-4), 73-80. Recuperado de <https://rcci.uci.cu/?journal=rcci&page=article&op=view&path%5B%5D=78>
- Román, V. (2019). *Aprendizaje no supervisado en Machine Learning: Agrupamiento*. Ciencia & Datos. Recuperado de <https://medium.com/datos-y-ciencia/aprendizaje-no-supervisado-en-machine-learning-agrupaci%C3%B3n-bb8f25813edc>
- Sánchez, D. (2009). *Agentes Inteligentes; Diseño e Implementación para la Enseñanza de la Física* (Tesis Doctoral). Instituto Politécnico Nacional. México.
- Schiatti, L. (2017). *Estudio comparative de diferentes algoritmos de clustering para la estimación de grupos de evaluados que comparten debilidades conceptuales similares*. (Tesis de licenciatura). Universidad Tecnológica de Bolívar. Catargena de Indias, Colombia.
- Schouwenburg, H. C. (2004). Procrastination in academic settings: General introduction. En H. C. Schouwenburg, C. H. Lay, T. A. Pychyl, & J. R. Ferrari (Eds.). *Counseling*

- the procrastinator in academic settings* (pp. 3-17). Washington: American Psychological Association.
- Senecal, C., Koestner, R., & Vallerand, R. (1995). Self-Regulation and Academic Procrastination. *Journal of Social Psychology*, 135(5), 607-619. DOI: 10.1080 / 00224545.1995.9712234
- Simeone, O. (2018). A very brief introduction to Machine Learning with applications to communication systems. *IEEE Transactions on Cognitive Communications and Networking*, 4(4), 648-664. DOI: [10.1109/TCCN.2018.2881442](https://doi.org/10.1109/TCCN.2018.2881442)
- Solomon, L. & Rothblum, E. (1984). Academic procrastination: Frequency and cognitive-behavioral correlates. *Journal of Counseling Psychology*, 31(4), 503-509. DOI: 10.1037/0022-0167.31.4.503
- Spada, M., Hiou, K. & Nikcevic, A. (2006). Metacognitions, emotions, and Procrastination. *Journal of Cognitive Psychotherapy*, 20(3), 319-326. DOI: <https://doi.org/10.1891/jcop.20.3.319>
- Steel, P. (2007). The Nature of Procrastination: A Meta-Analytic and Theoretical Review of Quintessential Self-Regulatory Failure. *Psychological Bulletin*, 133(1), 65-94. DOI: 10.1037/0033-2909.133.1.65
- Steel, P. & Ferrari, J. (2012). Sex, education and procrastination: An epidemiological study of procrastinators' characteristics from a global sample. *European Journal of Personality*, 27(1), 51-58. DOI: 10.1002/per.1851

Unir Revista. (2019). *Lenguaje R, ¿qué es y por qué es tan usado en Big Data?* uniR: La Universidad en Internet. Recuperado de <https://www.unir.net/ingenieria/revista/lenguaje-r-big-data/>

Valdiviezo, A. (2017). *Procrastinación en estudiantes de derecho y administración de una universidad privada de Trujillo*. (Tesis de grado). Universidad Privada Antenor Orrego, Trujillo, Perú.

Valle, A., Núñez, J., Cabanach, R., González-Pienda, J., Rodríguez, S., Rosário, P., Cerezo, R. & Muñoz-Cadavid, M. (2008). Self-regulated profiles and academic achievement. *Revista Psicothema*, 20(4), 724-731. Recuperado de <http://www.psycothema.com/psycothema.asp?ID=3547>

Vallejo, D. (2016). *Clustering de documentos con restricciones de tamaño*. (Tesis de Máster). Universidad Politécnica de Valencia. Valencia, España.

Vallejos, S. (2015). *Procrastinación académica y ansiedad frente a las evaluaciones en estudiantes universitarios*. (Tesis de grado). Pontificia Universidad Católica del Perú, Perú.

Vargas, L., Farfán, J., Rodríguez, M., Aramayo, F., Flores, H., Bustamante, S. y López, V. (2016). Comparación de las principales herramientas de Data Mining y Análisis de Sábanas Telefónicas. En L. Malbernat, J. Finochietto y G. Bacigalup (comps.). *II Jornadas Argentinas de Tecnología, Innovación y Creatividad – II JATIC 2016*. Mar del Plata, Buenos Aires.

Vázquez, I. (2016). *Procrastinación y estrés académico en estudiantes de una institución educativa estatal*. (Tesis de grado). Universidad Señor de Sipán, Perú.

Vázquez, R. (2016). Procrastinación en alumnos de preparatoria. *ALTAmira, revista académica*, 4(13), 18-25. DOI: 10.15418/ALTAMIRA1302

Wilson, B. & Bguyen, T. (2012). Belonging to Tomorrow: An Overview of Procrastination. *International Journal of Psychological Studies*, 4 (1), 211-217. DOI: [10.5539/ijps.v4n1p211](https://doi.org/10.5539/ijps.v4n1p211)

Witten, I., Frank, E. & Hall, M. (2011). *Data Mining: practical machine learning tools and techniques*. (3rd ed.). USA: El Sevier.

Yarlequé, L., Javier, L., Monroe, J., Nuñez, E., Navarro, L., Padilla, M., Matalinares, M., Navarro, L. R. y Campos, J. (2016). Procrastinación, estrés y bienestar psicológico en estudiantes de educación superior de Lima y Junín. *Horizonte de la ciencia* 6(10), 173-184. Recuperado de <http://revistas.uncp.edu.pe/index.php/horizontedelaciencia/article/view/367>

Glosario

A

AGPL: Affero General Public License

AI: Inteligencia Artificial

API: Interfaz de Programación de Aplicaciones

B

BI: Business Intelligence

C

CLI: Interfaz de Línea de Comando

CPU: Unidades de Procesamiento Central

D

DM: Data Mining

E

EAIL: Escala de Adicción a Internet de Lima

EDAOM: Estilos de Aprendizaje y Orientación Motivacional

EPA: Escala de Procrastinación Académica

F

FURPS: Functionality, Usability, Reliability, Performance y Supportability

G

GPL: General Public License

GPU: Unidades de procesamiento de gráficos

GUI: Interfaz Gráfica de Usuario

I

IBM: International Business Machines Corporation

IDE: Entorno de Desarrollo Integrado

IEEA: Inventario de Estrategias de Estudio y Autorregulación

K

KDD: Knowledge Discovery in Databases

M

ML: Machine Learning

MLOps: Machine Learning Operations

O

OLAP: OnLine Analytical Processing

P

PASS: Procrastination Assessment Scale-Students

PCR: Planeación, Control y Reflexión

R

RAM: Memoria de Acceso Aleatorio

RBAC: Role Based Access Control

RMDA: Interfaz de Red de Alto Rendimiento Opcionales

S

SIES: Sistema de Información de la Educación Superior

SUAyED: Sistema de Universidad Abierta y Educación a Distancia

T

TAN: Tree Augmented Network

TMBQ: Time Management Behavior Questionnaire

U

UCN: Universidad Católica del Norte de Chile

UJAT: Universidad Juárez Autónoma de Tabasco

UPNT: Uso Problemático de Nuevas Tecnologías

USD: Dólar estadounidense

v

vCPU: Núcleo de Procesamiento Virtual

México.

Autónoma de Tabasco.

Apéndice A. Catálogo y descripción de las variables y datos del *dataset*

La tabla 19 muestra la descripción de las variables que conforman el *dataset* de la investigación, además de la descripción o representación de los datos de este.

Tabla 19
Catálogo de las variables del *dataset*.

#1 Variable Género		
Nombre	Género	
Definición	Representa el género de los estudiantes	
Rango	1 - 2	
Valor	Etiquetas	Valores categóricos
1	Hombres	1.0,1.5
2	Mujeres	1.5,2.0
#2 Variable Edad		
Nombre	Edad	
Definición	Representa la edad de los estudiantes	
Rango	17 - 46 y 189	
Valor	Etiquetas	Valores categóricos
17	17 años	17.0,25.20
18	18 años	17.0,25.20
19	19 años	17.0,25.20
20	20 años	17.0,25.20
21	21 años	17.0,25.20
22	22 años	17.0,25.20
23	23 años	17.0,25.20
24	24 años	17.0,25.20
25	25 años	17.0,25.20
26	26 años	25.20,33.40
27	27 años	25.20,33.40
28	28 años	25.20,33.40
29	29 años	25.20,33.40
30	30 años	25.20,33.40
31	31 años	25.20,33.40
32	32 años	25.20,33.40
37	37 años	33.40,41.60
38	38 años	33.40,41.60
39	39 años	33.40,41.60
46	46 años	41.60,49.80
189	189 años	181.0,189.0
#3 Variable CicloEscolar		
Nombre	Ciclo escolar	
Definición	Representa el ciclo escolar que los estudiantes están cursando	

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Rango	1 - 68	Etiquetas	Valores categóricos
1		01/02/2019	1.0,1.99
2		2	1.99,2.97
3		3	2.97,3.96
4		4	3.96,4.94
5		5	4.94,5.93
6		6	5.93,6.91
7		7	6.91,7.90
8		8	7.90,8.88
9		9	8.88,9.87
10		10	9.87,10.90
11		11	10.90,11.80
12		132	11.80,12.80
13		2010	12.80,13.80
14		2014	13.80,14.80
15		2015	14.80,15.80
16		2016	15.80,16.80
17		2017	16.80,17.80
18		2018	17.80,18.70
19		2019	18.70,19.70
20		10mo. semestre	19.70,20.70
21		12vo. semestre	20.70,21.70
22		2014-2019	21.70,22.70
23		2014-2021	22.70,23.70
24		2015-2019	23.70,24.60
25		2015-2020	24.60,25.60
26		2016-	25.60,26.60
27		2016-2019	26.60,27.60
28		2016-2020	27.60,28.60
29		2016-2021	28.60,29.60
30		2017-2019	29.60,30.60
31		2017-2021	30.60,31.50
32		2017-2022	31.50,32.50
33		2017-B	32.50,33.50
34		2018-2019	33.50,34.50
35		2018-2022	34.50,35.50
36		2018-2024	35.50,36.50
37		2018-B	36.50,37.50
38		2019-01	37.50,38.40
39		2019-2020	38.40,39.40
40		2019-A	39.40,40.40
41		2019-B	40.40,41.40
42		2019-largo	41.40,42.40
43		2do. semestre	42.40,43.40
44		3ro. semestre	43.40,44.40

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

45	4to. semestre	44.40,45.30
46	5to. semestre	45.30,46.30
47	6to. semestre	46.30,47.30
48	7mo. semestre	47.30,48.30
49	8vo. semestre	48.30,49.30
50	9no. semestre	49.30,50.30
51	Agosto 2018-Junio 2019	50.30,51.20
52	Agosto-Diciembre 2019	51.20,52.20
53	Ciclo Febrero-Agosto 2019	52.20,53.20
54	Ciclo largo 2018-2019	53.20,54.20
55	Ciclo largo 2019	54.20,55.20
56	Educación superior	55.20,56.20
57	Enero-Agosto 2019	56.20,57.20
58	Enero-Julio 2019	57.20,58.10
59	Enero-Junio 2018	58.10,59.10
60	Enero-Junio 2019	59.10,60.10
61	Febrero-Agosto 2017	60.10,61.10
62	Febrero-Agosto 2019	61.10,62.10
63	Febrero-Julio 2019	62.10,63.10
64	Febrero-Junio 2019	63.10,64.10
65	Largo	64.10,65.0
66	Licenciatura	65.0,66.0
67	Nivel superior	66.0,67.0
68	Preparatoria	67.0,68.0

#4 Variable División Académica

Nombre	División Académica	
Definición	Representa la división académica al que pertenecen los estudiantes	
Rango	1 - 12	
Valor	Etiquetas	Valores categóricos
1	DAMC	1.0,1.92
2	DACBIOL	1.92,2.83
3	DACS	2.83,3.75
4	DACB	3.75,4.67
5	DAIA	4.67,5.58
6	DAEA	5.58,6.50
7	DACEA	6.50,7.42
8	DACA	7.42,8.33
9	DAMJM	8.33,9.25
10	DACSYH	9.25,10.20
11	DACYTI	10.20,11.10
12	DAMR	11.10.,12.0

#5 Variable AsignaturaCursa

Nombre	Asignatura que cursa el estudiante
Definición	Representa las asignaturas que los estudiantes estén cursando en línea y están agrupadas en cinco grupos

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Derechos Humanos, Sociedad y	1.0,1.8
2	Medio Ambiente	1.8,2.6
3	Filosofía y Ética profesional	2.6,3.4
4	Comunicación Oral y Escrita	3.4,4.2
5	Habilidades del Pensamiento Tecnologías de la Información y la Comunicación	4.2,5.0
#6 Variable Promedio Bachillerato		
Nombre	Promedio obtenido en el bachillerato	
Definición	Representa el promedio obtenido en el bachillerato de los estudiantes al finalizar sus estudios en este nivel educativo	
Rango	1 - 4	
Valor	Etiquetas	Valores categóricos
1	Entre 9 y 10	1.0,1.5
2	Menos de 9 y más de 8	2.0,2.5
3	Menos de 8 y más de 7	2.5,3.0
4	Menos de 7	1.5,2.0
#7 Variable Ítem1		
Nombre	Cuando tengo que hacer una tarea, la dejo para el último minuto.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0
#8 Variable Ítem2		
Nombre	Me preparo con anticipación para los exámenes.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0
#9 Variable Ítem3		
Nombre	Cuando me asignan lecturas las leo la noche anterior.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0
#10 Variable Ítem4		
Nombre	Cuando me asignan lecturas las reviso el mismo día de clase.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0
#11 Variable Ítem5		
Nombre	Cuando tengo problemas para entender algo, inmediatamente trato de buscar ayuda.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0
#12 Variable Ítem6		
Nombre	Asisto a clase.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0
#13 Variable Ítem7		
Nombre	Trato de completar el trabajo asignado lo más pronto posible.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#14 Variable Ítem8

Nombre	Postergo los trabajos de los cursos que no me gustan.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#15 Variable Ítem9

Nombre	Postergo las lecturas de los cursos que me disgustan.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#16 Variable Ítem10

Nombre	Intento mejorar mis hábitos de estudio.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#17 Variable Ítem11

Nombre	Invierto el tiempo necesario en estudiar aun cuando el tema sea aburrido.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#18 Variable Ítem12

Nombre	Trato de motivarme para mantener mi ritmo de estudio.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#19 Variable Ítem13

Nombre	Trato de terminar mis trabajos importantes con tiempo de sobra.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#20 Variable Ítem14

Nombre	Me tomo el tiempo de revisar mis tareas antes de entregarlas.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#21 Variable Ítem15

Nombre	Dejo para mañana lo que puedo hacer hoy.	
Definición	Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#22 Variable Ítem16

Nombre Disfruto la mezcla de desafío con emoción de esperar hasta el último minuto para completar una tarea.

Definición Ítem perteneciente al cuestionario de EPA de Busko, que miden el nivel de procrastinación.

Rango 1 - 5

Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#23 Variable Ítem17

Nombre Mientras desarrollo mis actividades del curso en línea me conecto a otras páginas que son de mi agrado (a través de la computadora o mi celular).

Definición Ítem agregado para conocer el uso de los estudiantes hacen de internet durante la realización de las actividades del curso en línea, denominado dentro del grupo de VGM.

Rango 1 - 5

Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#24 Variable Ítem18

Nombre El impulso por revisar mis redes sociales mientras estoy revisando el curso en línea es incontrolable.

Definición Ítem agregado para conocer el uso de los estudiantes hacen de internet durante la realización de las actividades del curso en línea, denominado dentro del grupo de VGM.

Rango 1 - 5

Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#25 Variable Ítem19

Nombre	Invierto más tiempo en revisar diversos sitios mientras estoy en la plataforma del curso que en desarrollar las actividades.	
Definición	Ítem agregado para conocer el uso de los estudiantes hacen de internet durante la realización de las actividades del curso en línea, denominado dentro del grupo de VGM.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#26 Variable Ítem20

Nombre	Me siento nervioso y hasta desesperado si no puedo navegar libremente en el Internet mientras hago mis actividades.	
Definición	Ítem agregado para conocer el uso de los estudiantes hacen de internet durante la realización de las actividades del curso en línea, denominado dentro del grupo de VGM.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#27 Variable Ítem21

Nombre	He dejado de hacer tareas de mi curso por distraerme con el Internet.	
Definición	Ítem agregado para conocer el uso de los estudiantes hacen de internet durante la realización de las actividades del curso en línea, denominado dentro del grupo de VGM.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8
2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#28 Variable Ítem22

Nombre	Trabajo en la plataforma con mi celular, por un lado.	
Definición	Ítem agregado para conocer el uso de los estudiantes hacen de internet durante la realización de las actividades del curso en línea, denominado dentro del grupo de VGM.	
Rango	1 - 5	
Valor	Etiquetas	Valores categóricos
1	Nunca	1.0,1.8

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

2	Casi nunca	1.8,2.6
3	A veces	2.6,3.4
4	Casi siempre	3.4,4.2
5	Siempre	4.2,5.0

#29 Variable ClaseEstudiante		
Nombre	¿El estudiante es procrastinador o autorregulador?	
Definición		
Rango	1 - 2	
Valor	Etiquetas	Valores categóricos
1	Procrastinador	PROCRASTINADOR
2	Autorregulador	AUTORREGULADOR

Apéndice B. Reglas de asociación obtenidas en la prueba uno en la herramienta

Microsoft Azure Machine Learning

ID	lhs	rhs	Support	Confidence	Lift
1	{Item6= [4.2,5.0], Item16= [1.0,1.8), Item21= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.206434	1	1.110119
2	{Item6= [4.2,5.0], Item14= [4.2,5.0], Item19= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.203753	1	1.110119
3	{Item6= [4.2,5.0], Item18= [1.0,1.8), Item19= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.273458	1	1.110119
4	{Item6= [4.2,5.0], Item12= [4.2,5.0], Item19= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.206434	1	1.110119
5	{Item6= [4.2,5.0], Item19= [1.0,1.8), Item20= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.286863	1	1.110119
6	{Item6= [4.2,5.0], Item19= [1.0,1.8), Item21= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.298928	1	1.110119
7	{Genero= [1.5,2.0], Item6= [4.2,5.0], Item19= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.238606	1	1.110119
8	{Item14= [4.2,5.0], Item20= [1.0,1.8), Item21= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.213137	1	1.110119
9	{Genero= [1.5,2.0], Item14= [4.2,5.0], Item21= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.206434	1	1.110119
10	{Item6= [4.2,5.0], Item14= [4.2,5.0], Item21= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.264075	1	1.110119
11	{Item6= [4.2,5.0], Item12= [4.2,5.0], Item18= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.205094	1	1.110119
12	{Item6= [4.2,5.0], Item18= [1.0,1.8), Item20= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.301609	1	1.110119
13	{Item6= [4.2,5.0], Item18= [1.0,1.8), Item21= [1.0,1.8)}	{ClaseEstudiante= AUTORREGULADOR}	0.301609	1	1.110119

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

ID	lhs	rhs	Support	Confidence	Lift
14	{Item6= [4.2,5.0], Item11= [3.4,4.2], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.214477	1	1.110119
15	{PromedioBachillerato= [1.0,1.5], Item6= [4.2,5.0], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.256032	1	1.110119
16	{Item6= [4.2,5.0], Item20= [1.0,1.8], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.365952	0.99635	1.106068
17	{Genero= [1.5,2.0], Item6= [4.2,5.0], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.345845	0.996139	1.105833
18	{Genero= [1.5,2.0], Item6= [4.2,5.0], Item20= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.316354	0.995781	1.105435
19	{Item18= [1.0,1.8], Item20= [1.0,1.8], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.293566	0.995455	1.105073
20	{Genero= [1.5,2.0], Item20= [1.0,1.8], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.285523	0.995327	1.104932

Apéndice C. Reglas de asociación obtenidas en la prueba dos en la herramienta

Microsoft Azure Machine Learning

ID	lhs	rhs	Support	Confidence	Lift
1	{Item6= [4.2,5.0], Item18= [1.0,1.8], Item20= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.301609	1	1.110119
2	{Item6= [4.2,5.0], Item18= [1.0,1.8], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.301609	1	1.110119
3	{Item6= [4.2,5.0], Item20= [1.0,1.8], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.365952	0.99635	1.106068
4	{Genero= [1.5,2.0], Item6= [4.2,5.0], Item21= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.345845	0.996139	1.105833
5	{Genero= [1.5,2.0], Item6= [4.2,5.0], Item20= [1.0,1.8]}	{ClaseEstudiante= AUTORREGULADOR}	0.316354	0.995781	1.105435
6	{PromedioBachillerato= [1.0,1.5], Item6= [4.2,5.0], ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2]}	0.376676	0.975694	1.028062
7	{Item1= [2.6,3.4), Item6= [4.2,5.0], ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2]}	0.321716	0.97561	1.027973
8	{Item3= [2.6,3.4), Item6= [4.2,5.0], ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2]}	0.329759	0.964706	1.016484
9	{Item6= [4.2,5.0], Item7= [4.2,5.0], ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2]}	0.357909	0.963899	1.015634
10	{Item6= [4.2,5.0], Item10= [4.2,5.0], ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2]}	0.347185	0.962825	1.014502
11	{Item6= [4.2,5.0], Item12= [3.4,4.2), ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2]}	0.310992	0.962656	1.014324

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

ID	lhs	rhs	Support	Confidence	Lift
12	{Item2= [3.4,4.2), Item6= [4.2,5.0], ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2)}	0.357909	0.960432	1.01198
13	{Item6= [4.2,5.0], Item14= [4.2,5.0], ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2)}	0.356568	0.960289	1.01183
14	{PromedioBachillerato= [2.0,2.5), Item6= [4.2,5.0], ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2)}	0.312332	0.954918	1.006171
15	{Item6= [4.2,5.0], Item13= [3.4,4.2), ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2)}	0.329759	0.953488	1.004664
16	{Item6= [4.2,5.0], Item19= [1.0,1.8), ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2)}	0.324397	0.952756	1.003893
17	{Item6= [4.2,5.0], Item11= [3.4,4.2), ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2)}	0.347185	0.952206	1.003313
18	{Genero= [1.5,2.0], Item6= [4.2,5.0], Item20= [1.0,1.8)}	{Edad= [17.0, 25.2)}	0.301609	0.949367	1.000322
19	{Item6= [4.2,5.0], Item20= [1.0,1.8), ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2)}	0.424933	0.949102	1.000042
20	{Item6= [4.2,5.0], Item21= [1.0,1.8), ClaseEstudiante= AUTORREGULADOR}	{Edad= [17.0, 25.2)}	0.461126	0.947658	0.998521

Anexo A. Ley Federal de Protección de Datos Personales en Posesión de los

Particulares: Capítulo I Disposición Generales

Artículo 1.- La presente Ley es de orden público y de observancia general en toda la República y tiene por objeto la protección de los datos personales en posesión de los particulares, con la finalidad de regular su tratamiento legítimo, controlado e informado, a efecto de garantizar la privacidad y el derecho a la autodeterminación informativa de las personas.

Artículo 2.- Son sujetos regulados por esta Ley, los particulares sean personas físicas o morales de carácter privado que lleven a cabo el tratamiento de datos personales, con excepción de:

- I. Las sociedades de información crediticia en los supuestos de la Ley para Regular las Sociedades de Información Crediticia y demás disposiciones aplicables, y
- II. Las personas que lleven a cabo la recolección y almacenamiento de datos personales, que sea para uso exclusivamente personal, y sin fines de divulgación o utilización comercial.

Artículo 3.- Para los efectos de esta Ley, se entenderá por:

- I. Aviso de Privacidad: Documento físico, electrónico o en cualquier otro formato generado por el responsable que es puesto a disposición del titular, previo al tratamiento de sus datos personales, de conformidad con el artículo 15 de la presente Ley.

- II. Bases de datos: El conjunto ordenado de datos personales referentes a una persona identificada o identificable.
- III. Bloqueo: La identificación y conservación de datos personales una vez cumplida la finalidad para la cual fueron recabados, con el único propósito de determinar posibles responsabilidades en relación con su tratamiento, hasta el plazo de prescripción legal o contractual de éstas. Durante dicho periodo, los datos personales no podrán ser objeto de tratamiento y transcurrido éste, se procederá a su cancelación en la base de datos que corresponde.
- IV. Consentimiento: Manifestación de la voluntad del titular de los datos mediante la cual se efectúa el tratamiento de los mismos.
- V. Datos personales: Cualquier información concerniente a una persona física identificada o identificable.
- VI. Datos personales sensibles: Aquellos datos personales que afecten a la esfera más íntima de su titular, o cuya utilización indebida pueda dar origen a discriminación o conlleve un riesgo grave para éste. En particular, se consideran sensibles aquellos que puedan revelar aspectos como origen racial o étnico, estado de salud presente y futuro, información genética, creencias religiosas, filosóficas y morales, afiliación sindical, opiniones políticas, preferencia sexual.
- VII. Días: Días hábiles.

- VIII. **Disociación:** El procedimiento mediante el cual los datos personales no pueden asociarse al titular ni permitir, por su estructura, contenido o grado de desagregación, la identificación del mismo.
- IX. **Encargado:** La persona física o jurídica que sola o conjuntamente con otras trate datos personales por cuenta del responsable.
- X. **Fuente de acceso público:** Aquellas bases de datos cuya consulta puede ser realizada por cualquier persona, sin más requisito que, en su caso, el pago de una contraprestación, de conformidad con lo señalado por el Reglamento de esta Ley.
- XI. **Instituto:** Instituto Federal de Acceso a la Información y Protección de Datos, a que hace referencia la Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental.
- XII. **Ley:** Ley Federal de Protección de Datos Personales en Posesión de los Particulares.
- XIII. **Reglamento:** El Reglamento de la Ley Federal de Protección de Datos Personales en Posesión de los Particulares.
- XIV. **Responsable:** Persona física o moral de carácter privado que decide sobre el tratamiento de datos personales.
- XV. **Secretaría:** Secretaría de Economía.
- XVI. **Tercero:** La persona física o moral, nacional o extranjera, distinta del titular o del responsable de los datos.

XVII. Titular: La persona física a quien corresponden los datos personales.

XVIII. Tratamiento: La obtención, uso, divulgación o almacenamiento de datos personales, por cualquier medio. El uso abarca cualquier acción de acceso, manejo, aprovechamiento, transferencia o disposición de datos personales.

XIX. Transferencia: Toda comunicación de datos realizada a persona distinta del responsable o encargado del tratamiento.

Artículo 4.- Los principios y derechos previstos en esta Ley, tendrán como límite en cuanto a su observancia y ejercicio, la protección de la seguridad nacional, el orden, la seguridad y la salud públicos, así como los derechos de terceros.

Artículo 5.- A falta de disposición expresa en esta Ley, se aplicarán de manera supletoria las disposiciones del Código Federal de Procedimientos Civiles y de la Ley Federal de Procedimiento Administrativo.

Para la substanciación de los procedimientos de protección de derechos, de verificación e imposición de sanciones se observarán las disposiciones contenidas en la Ley Federal de Procedimiento Administrativo.

Anexo B. Cuestionario aplicado a los estudiantes para la recolección de datos

Cuestionario de autorregulación académica

Instrucciones: El propósito del siguiente cuestionario es conocer las características de su autorregulación académica como estudiante. Se preguntarán datos tanto personales y académicos, su participación es voluntaria y anónima. Usted decide si desea contestar, pero no le tomará más de 10 minutos.

***Obligatorio**

Género. *

Mujer Hombre

Edad. *

Ciclo escolar. *

División Académica. *

Educación y Artes (DAEA)

Ciencias Económico Administrativas (DACEA)

Ciencias y Tecnologías de la Información (DACyTI)

Ciencias Básicas (DACB)

Ingeniería y Arquitectura (DAIA)

Ciencias de la Salud (DACS)

Multidisciplinaria de Jalpa de Méndez (DAMJM)

Multidisciplinaria de Comalcalco (DAMC)

Multidisciplinaria de Ríos (DAMR)

Ciencias Biológicas (DACBIOL)

Ciencias Sociales y Humanidades (DACSyH)

Ciencias Agropecuarias (DACA)

Asignatura que cursa. *

Promedio obtenido en el bachillerato. *

1. Cuando tengo que hacer una tarea, la dejo para el último minuto. *

Nunca	Casi nunca	A veces	Casi siempre	Siempre
-------	------------	---------	--------------	---------

2. Me preparo con anticipación para los exámenes. *

Nunca	Casi nunca	A veces	Casi siempre	Siempre
-------	------------	---------	--------------	---------

3. Cuando me asignan lecturas las leo la noche anterior. *

Nunca	Casi nunca	A veces	Casi siempre	Siempre
-------	------------	---------	--------------	---------

4. Cuando me asignan lecturas las reviso el mismo día de clase. *

Nunca	Casi nunca	A veces	Casi siempre	Siempre
-------	------------	---------	--------------	---------

5. Cuando tengo problemas para entender algo, inmediatamente trato de buscar ayuda. *

Nunca	Casi nunca	A veces	Casi siempre	Siempre
-------	------------	---------	--------------	---------

6. Asisto a clase. *

Nunca Casi nunca A veces Casi siempre Siempre

7. Trato de completar el trabajo asignado lo más pronto posible. *

Nunca Casi nunca A veces Casi siempre Siempre

8. Postergo los trabajos de los cursos que no me gustan. *

Nunca Casi nunca A veces Casi siempre Siempre

9. Postergo las lecturas de los cursos que me disgustan.

Nunca Casi nunca A veces Casi siempre Siempre

10. Intento mejorar mis hábitos de estudio. *

Nunca Casi nunca A veces Casi siempre Siempre

11. Invierto el tiempo necesario en estudiar aun cuando el tema sea aburrido. *

Nunca Casi nunca A veces Casi siempre Siempre

12. Trato de motivarme para mantener mi ritmo de estudio. *

Nunca Casi nunca A veces Casi siempre Siempre

13. Trato de terminar mis trabajos importantes con tiempo de sobra. *

Nunca Casi nunca A veces Casi siempre Siempre

14. Me tomo el tiempo para revisar mis tareas antes de entregarlas. *

Nunca Casi nunca A veces Casi siempre Siempre

15. Dejo para mañana lo que puedo hacer hoy. *

Nunca Casi nunca A veces Casi siempre Siempre

16. Disfruto la mezcla de desafío con emoción de esperar hasta el último minuto para completar una tarea. *

Nunca Casi nunca A veces Casi siempre Siempre

17. Mientras desarrollo mis actividades del curso en línea, me conecto a otras páginas que son de mi agrado (a través de la computadora o mi celular). *

Nunca Casi nunca A veces Casi siempre Siempre

18. El impulso por revisar mis redes sociales mientras estoy revisando mi curso en línea es incontrolable. *

Nunca Casi nunca A veces Casi siempre Siempre

19. Invierto más tiempo en revisar diversos sitios mientras estoy en la plataforma del curso que en desarrollar otras actividades. *

Nunca Casi nunca A veces Casi siempre Siempre

20. Me siento nervioso y hasta desesperado si no puedo navegar libremente en internet mientras hago mis actividades. *

Nunca Casi nunca A veces Casi siempre Siempre

21. He dejado de hacer tareas de mi curso por distraerme con el internet. *

Nunca Casi nunca A veces Casi siempre Siempre

22. Trabajo en la plataforma con mi celular, por un lado. *

Nunca Casi nunca A veces Casi siempre Siempre

Anexo C. Código para la creación del modelo de árbol de decisión en Lenguaje R

##LECTURA DEL DATASET##

```
procras_atorre1 = read.csv("../Minería-Tesis/procras_atorre1.csv") #Leer dataset
```

```
View(procras_atorre1) #Visualización del dataset
```

```
summary(procras_atorre1) #Resumen de las variables del dataset
```

```
datos <- procras_atorre1 #Renombre del dataset
```

##CARGA DE LIBRERIAS DEL PAQUETE DE ARBOL DE CLASIFICACIÓN C5.0##

```
install.packages("C50")
```

```
library(C50)
```

##CAMBIO DEL TIPO DE DATO DE CARACTER A FACTOR DE LA COLUMNA

CLASE##

```
clase <- vector(length = dim(datos)[1])
```

```
clase [datos$ClaseEstudiante=="Autorregulador"] <-"AUTORREGULADOR"
```

```
clase [datos$ClaseEstudiante=="Procrastinador"] <-"PROCRASTINADOR"
```

```
datos$ClaseEstudiante <- factor(clase)
```

##CREACIÓN DE LOS SETS DE ENTRENAMIENTO Y PRUEBA##

```
set.seed(125) #Establecer semilla
```

```
tamano.total <- nrow(datos) #Devuelve el número de filas o columnas presentes en el dataset
```

```
tamano.entreno <- round(tamano.total*0.8) #Obtención de un vector o subconjunto del 80% de los datos
```



```
datos.indices <- sample(1:tamano.total , size=tamano.entreno) #Obtención de datos de  
manera aleatoria
```

```
datos.entreno <- datos[datos.indices,] #Set de entrenamiento
```

```
datos.test <- datos[-datos.indices,] #Set de prueba
```

```
##CREACIÓN DEL MODELO DE ARBOL DE DECISIÓN##
```

```
modelo <- C5.0(ClaseEstudiante ~ .,data = datos) #Arbol de decisión
```

```
summary(modelo) # Información sobre el modelo
```

```
plot(modelo) # Gráfico
```

```
#Para detallar un nodo en particular
```

```
plot(modelo, subtree=3) # Ejemplo, muestra un nodo en particular en este caso el  
nodo 3
```

```
modelo
```

```
##PREDICCIÓN##
```

```
prediccion <- predict(modelo,newdata=datos) #Validación de la capacidad de  
predicción del arbol con el set de prueba
```

```
res=data.frame (predicciones=prediccion,reales=datos$ClaseEstudiante) #Validación  
de resultados predictivos con reales
```

```
##MATRIZ DE CONFUSIÓN##
```

```
tabla <- table (prediccion, datos$ClaseEstudiante)
```

```
tabla #Visualización de la matriz de confusión
```

```
##CALCULO DEL PROCENTAJE DE ACIERTOS##
```

Identificación de patrones de conducta, asociadas con procrastinación y autorregulación académica en estudiantes de educación superior, aplicando técnicas de minería de datos

$100 * \text{sum}(\text{diag}(\text{tabla})) / \text{sum}(\text{tabla})$ # % **correctamente clasificados**

Universidad Juárez Autónoma de Tabasco
México.

Anexo D. Código para la creación del modelo de agrupamiento en Lenguaje R

##LECCURA DEL DATASET DE PRUEBA##

```
procras_atorre1 = read.csv("../Minería-Tesis/procras_atorre1.csv") #Leer dataset
```

```
View(procras_atorre1) #Visualización del dataset
```

```
summary(procras_atorre1) #Resumen de las variables del dataset
```

```
datos <- procras_atorre1 #Renombre del dataset
```

##CARGA DE LIBRERIAS PARA AGRUPAMIENTO##

```
library(tidyverse)
```

```
library(cluster)
```

```
library(factoextra)
```

```
library(NbClust)
```

```
library(tidyr)
```

```
library(ggplot2)
```

##NORMALIZACIÓN DE LOS DATOS##

```
datos.scale <- as.data.frame(scale(datos[1:28])) #Normalizar datos para que ninguna variable tenga más peso que otra
```

```
head(datos.scale) #Visualizar las primeras filas de datos
```

##CALCULO DE LA MATRIZ DE DISTANCIAS##

```
m.distancia <- get_dist(datos.scale,method = "euclidean") #El metodo euclidean es el más utilizado
```

```
fviz_dist(m.distancia, gradient = list(low = "blue", mid = "white", high = "red")) #Grafico de
```

la matriz de distancia

##ESTIMACIÓN DEL NÚMERO DE CLÚSTER PARA EL MODELO##

```
fviz_nbclust(datos.scale, kmeans, method = "wss")
```

```
fviz_nbclust(datos.scale, kmeans, method = "silhouette")
```

```
fviz_nbclust(datos.scale, kmeans, method = "gap_stat")
```

##CREACIÓN DEL CLUSTERS##

```
set.seed(456) #Establecer semilla
```

```
agru <- kmeans(datos.scale, centers = 2) #Realización del modelo de agrupamiento o
```

cluster

```
agru #Información del modelo
```

```
names(agru) #Nombre de los objetos que contiene el modelo
```

```
str(agru) #Para obtener más información
```

#Tambien se puede visualizar la información del modelo de la siguiente manera

```
agru$cluster #Asignación de observaciones a los cluteres
```

```
agru$totss #Inercia total
```

```
agru$betweenss #Inercia inter grupos
```

```
agru$withinss #Inercia intra grupos
```

```
agru$tot.withinss #Inercia intra grupos (total)
```

##PLOTAR LOS CLUSTER DEL MODELO##

```
fviz_cluster(agru, data = datos.scale)
```

```
fviz_cluster(agru, data = datos.scale, ellipse.type = "norm")
```

##INSPECCIONAR LOS RESULTADOS##

```
plot(datos$DivisiónAcadémica, datos$Ítem16, col=agru$cluster ,xlab = "grupo1", ylab =  
"grupo2")
```

```
aggregate(datos[,1:28] ,by = list(agru$cluster),mean)
```

```
text(datos$DivisiónAcadémica, datos$Ítem16, labels = datos$ClaseEstudiante, pos = 1,  
col = agru$cluster,cex = .7)
```

```
ggplot(datos, aes(x = datos$DivisiónAcadémica, y = datos$Ítem16)) + geom_point()
```

##ASIGNACIÓN DE LOS CLUSTERES AL DATASET##

```
tabla <- data.frame (datos) %>%
```

```
  mutate(Cluster = agru$cluster)
```

```
#tabla #Visualización de la tabla
```

##MATRIZ DE CONFUSIÓN DE LOS AGRUPAMIENTOS##

```
cluster=agru$cluster
```

```
table(tabla$ClaseEstudiante,cluster)
```

Anexo E. Código para la creación del modelo de reglas de asociación en Lenguaje

R

##LECTURA DEL DATASET DE PRUEBA##

```
procras_autorre = read.csv("../Minería-Tesis/procras_autorre.csv") #Leer dataset
```

```
View(procras_autorre) #Visualización del dataset
```

```
summary(procras_autorre) #Resumen de las variables del dataset
```

```
datos <- procras_autorre #Renombre del dataset
```

##CARGA DE LIBRERIAS DEL PAQUETE ARULES##

```
library(arules)
```

##REGLAS DE ASOCIACIÓN##

```
reglas <- apriori(data = datos,
```

```
  parameter = list(support = 0.1,
```

```
    confidence = 0.8,
```

```
    target = "rules")) #Especificación que se creen reglas
```

```
summary(reglas) #Visualización de la información de modelo
```

```
inspect(sort(x = reglas, decreasing = TRUE, by = "confidence")) #Visualización de las
```

todas las reglas de asociación obtenidas del modelo ordenas por orden descendente de confianza

```
filtrado_reglas <- sort(reglas, decreasing = TRUE, by = "confidence")[1:20] #Filtrado de
```

las 20 reglas de asociación con mayor confianza

```
inspect(filtrado_reglas) #Visualización de las reglas
```