



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



ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE LA ASIGNATURA DE PROGRAMACIÓN MEDIANTE TÉCNICAS DE MINERÍA DE DATOS

Trabajo recepcional bajo la modalidad de Tesis

Que para obtener el grado de

Maestro en Tecnologías para el Aprendizaje y el Conocimiento

Presenta:

Jesús Alberto Rodríguez Sanarao

Directora:

Dra. Erika Yunuen Morales Mateos

Jurado Revisor:

Dr. Arturo Corona Ferreira

Dr. Juan de Dios González Torres

Dr. Eddy Arquímedes García Alcocer

Cuerpo Académico:

Tecnologías para el análisis de datos y desarrollo de aplicaciones

TÍTULO DE LA TESIS:

TÍTULO EN ESPAÑOL

ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE LA ASIGNATURA DE PROGRAMACIÓN MEDIANTE TÉCNICAS DE MINERÍA DE DATOS

TÍTULO EN INGLÉS

ANALYSIS OF THE ACADEMIC PERFORMANCE OF THE PROGRAMMING SUBJECT THROUGH DATA MINING TECHNIQUES

PRESENTA: JESÚS ALBERTO RODRÍGUEZ SANARAO

Durante el trabajo de investigación de esta tesis se desarrollaron diversas actividades que permitieron llegar a su conclusión manera satisfactoria, entre las que destacan:

- Una estancia de Investigación en Consejo de Ciencia y Tecnología del Estado de Tabasco donde se desarrolló un diagnóstico post COVID-19 en el estado de Tabasco del 01 de junio 2020 al 31 de mayo de 2021
- Publicación de artículo titulado: “Tareas y Herramientas de minería de Datos en el ámbito educativo” en la revista “Investigación Aplicada un Enfoque en la Tecnología” con ISSN: 3594-035X, 14 de diciembre de 2021

"2022, Año de Ricardo Flores Magón"

Cunduacán, Tabasco a 04 de febrero de 2022
Oficio No. 126/DACYTI/CP/2022

Asunto: Dirección de Tesis

Dra. Erika Yunuen Morales Mateos
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 "**Análisis del rendimiento académico de la asignatura de programación mediante Minería de Datos**", a realizar por el **C. Jesús Alberto Rodríguez Sanarao**, para obtener el grado de Maestro en Tecnologías para el Aprendizaje y el Conocimiento.

No omito manifestarle que dispone de 30 días naturales a partir de esta fecha de acuerdo al artículo 71 sección III del Reglamento de Posgrado para realizar la aceptación u observaciones pertinentes.

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

Atentamente


MTE. Oscar Alberto González González
Director



C. p. Dr. Eddy Arquímedes García Alcocer, Encargado del Despacho de la Coordinación de Posgrado
Alumno, Jesús Alberto Rodríguez Sanarao
Archivo
MTE/OAGG/EAGA

+

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

www.ujat.mx



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



Cunduacán, Tabasco., a 03 de febrero de 2022

Asunto: Liberación de dirección de tesis.

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

Por medio de la presente me permito comunicarle que después de haber concluido la dirección de la Tesis: "ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE LA ASIGNATURA DE PROGRAMACIÓN MEDIANTE TÉCNICAS DE MINERÍA DE DATOS", elaborada por el *C. Jesús Alberto Rodríguez Sanrao*, de la Maestría en Tecnologías para el Aprendizaje y el Conocimiento, consideramos que puede continuar con los trámites para la obtención del grado.

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

Atentamente

Dra. Erika Yunuen Morales Mateos

c.c.p. Dr. Eddy Arquímedes García Alcoer. Encargado del despacho de la Coordinación de Posgrado.
Directores de Tesis.
Estudiante.

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



Cunduacán, Tabasco, a 10 de febrero de 2022.

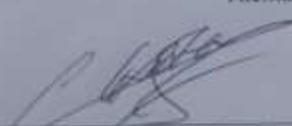
Asunto: Respuesta de Jurado.

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

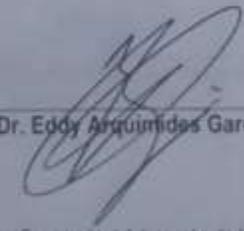
En atención a los oficios girados por usted, en los que se nos designa como parte del jurado para efectuar la revisión de la tesis titulada "ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE LA ASIGNATURA DE PROGRAMACIÓN MEDIANTE TÉCNICAS DE MINERÍA DE DATOS", realizada por el C. Jesús Alberto Rodríguez Samrao, estudiante de la Maestría en Tecnologías para el Aprendizaje y el Conocimiento, nos permitimos informarle que, en virtud de que ha atendido las observaciones realizadas, otorgamos nuestra aprobación para que continúe los trámites para la obtención del grado.

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

Atentamente integrantes del jurado


Dr. Arturo Corona Ferreira


Dr. Juan de Dios Gonzalez Torres


Dr. Eddy Arquimedes Garcia Alcocer

c.c.p. Dr. Eddy Arquimedes Garcia Alcocer, Encargado del despacho de la Coordinación de Posgrado
Directores de tesis
Estudiante

"2022, Año de Ricardo Flores Magón"

Cunduacán, Tabasco a 08 de febrero de 2022
Oficio No. 133/DACYTI/CP/2022

Asunto: Autorización de impresión de Tesis

C. Jesús Alberto Rodríguez Sanarao
Matricula: 191H12007

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 **"Análisis del rendimiento académico de la asignatura de programación mediante Minería de Datos"**, para presentar examen y obtener el Grado de Maestra en Administración en Tecnologías para el Aprendizaje y el Conocimiento.

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

Atentamente


MTE. Oscar Alberto González González
Director



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

MTE. DAGGEAGA

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

Carta de Autorización

A quien corresponda:

El que suscribe **Jesús Alberto Rodríguez Sanarao** 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, "**ANÁLISIS DEL RENDIMIENTO ACADÉMICO DE LA ASIGNATURA DE PROGRAMACIÓN MEDIANTE 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 03 días del mes de febrero del año 2022.



Jesús Alberto Rodríguez Sanarao

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

Agradecimientos

En primer lugar, deseo expresar mi agradecimiento a mi directora de esta tesis, por la dedicación y apoyo que me ha brindado en la creación de este trabajo, por el respeto a mis sugerencias e ideas y por la dirección y el rigor que ha facilitado a las mismas. Gracias por la confianza ofrecida.

Gracias a mi familia y a mis amigos, que siempre me han prestado un gran apoyo moral y humano, necesarios en todo momento. Pero, sobre todo, gracias a una gran amiga la Lic. Anel Guzmán que me ha acompañado desde hace un largo tiempo, por su apoyo, comprensión y solidaridad con un servidor. Sin su apoyo este trabajo nunca se habría escrito y, por eso, este trabajo es también el suyo. A todos, muchas gracias.

Dedicatorias

Dedico este trabajo a mis amigos más cercanos los cuales no son muchos, pero quienes con el pasar del tiempo no solo ha demostrado ser incondicionales, sino también los considero familia. Así mismo terminar este proyecto no ha sido un trabajo fácil, pero gracias al apoyo de mi asesora quien estuvo conmigo como guía, siempre pendiente y conocedora para aclarar mis dudas me permitió llegar a la culminación de este trabajo para la obtención de grado.

México.

Universidad Juárez Autónoma de Tabasco.^{vii}

Resumen

En este proyecto se realiza un análisis de la relación entre el compromiso estudiantil con las pruebas realizadas a lo largo del semestre y que técnicas de la minería de datos serían útiles para unos mejores resultados de dicho análisis. Se empleó la metodología para el Descubrimiento de Conocimientos en Bases de Datos (Knowledge Discovery in Databases - KDD), la cual es un proceso que no solo incluye la obtención de modelos o patrones, sino también la interpretación de estos. La población fue una muestra de 25 estudiantes de la materia de programación de una universidad del Sureste mexicano. Los resultados fueron un modelo de correlaciones entre las variables resultantes de la prueba UWES creado con el lenguaje de programación R, de igual forma un modelo de tres clusters los cuales se realizaron con la ayuda del algoritmo K-means con el software WEKA y por último con este mismo software unas reglas de asociación con el algoritmo apriori.

PALABRAS CLAVE: Minería de datos, Compromiso estudiantil, Correlaciones, Cluster, Reglas de asociación.

Introducción

En el desarrollo de este proyecto se adentrará en compromiso estudiantil y las técnicas de la minería de datos que ayudan a conocer mejor el rendimiento académico de los estudiantes en cuanto a la materia de programación en una universidad del suroeste mexicano. Dicha investigación se estructura en 5 capítulos.

Primero se dan a conocer los antecedentes y la problemática a solucionar, así mismo los objetivos a cumplir, la justificación y la metodología a utilizar para la realización de este proyecto.

Después se muestra el marco teórico el cual consta de marco referencial, conceptual, tecnológico y legal, los cuales proporcionaran una amplia definición de los conceptos para ayudar a una mejor comprensión de la temática planteada.

Luego se puede observar al desarrollo de la metodología utilizada, la cual es una de las partes más importantes de este proyecto ya que inicia desde la abstracción del escenario, como la limpieza y transformación de los datos hasta un estadístico descriptivo para conocer mejor los datos recolectados.

Enseguida se pueden observar las técnicas de minería de datos aplicadas, así como los modelos resultantes de estas y el análisis de dichos modelos para una mejor comprensión de los resultados.

Y, por último, se presentan las conclusiones y trabajos futuros en el cual se le da respuesta a la pregunta de investigación, y pauta para nuevos proyectos relacionados con la temática de este.

Índice general

Índice general.....	x
Índice de figuras.....	xii
Índice de tablas.....	xiii
Capítulo I. Generalidades.....	1
1.1 Antecedentes.....	1
1.2 Planteamiento del Problema.....	2
1.2.1 Definición del problema.....	2
1.2.2 Delimitación de la investigación.....	3
1.2.3 Preguntas de investigación.....	3
1.3 Objetivos.....	3
1.3.1 Objetivo General.....	3
1.3.2 Objetivos específicos.....	4
1.4 Justificación.....	4
1.5 Metodología utilizada.....	4
1.5.1 Población de estudio.....	4
1.5.2 Instrumentos.....	5
Capítulo II. Marco teórico.....	6
2.1 Marco referencial.....	6
2.2 Marco conceptual.....	9
2.2.1 Minería de datos.....	9
2.2.2 Compromiso estudiantil.....	15
2.3 Marco Tecnológico.....	16
2.3.1 Weka.....	16
2.3.2 R.....	17
2.3.3 El ambiente R.....	18
	x

2.4 Marco Legal	19
2.4.1 Licencia pública general GNU	19
Capítulo III. Desarrollo metodológico.....	21
3.1 Abstracción del escenario.....	21
3.2 Instrumento UWES- S.....	21
3.3 Recolección de los datos	23
3.4 Limpieza y transformación.....	24
3.5 Proceso de descubrimiento de conocimiento en bases de datos (KDD)	25
3.6 Estadístico descriptivo de las variables que integran el estudio.....	28
Capítulo IV. Análisis de minería de datos	31
4.1 Técnica de minería de datos: Correlación	31
4.1.1 Paquete: Psych 2.1.9.....	31
4.1.2 Resultados	32
4.2 Técnica de minería de datos: Clusterización.....	32
4.2.1 Algoritmo K means	33
4.2.2 Resultados	35
4.2.3 Evaluación e interpretación de los resultados	35
4.3 Técnica de minería de datos: Reglas de asociación	36
4.3.1 Algoritmo Apriorí	36
4.3.2 Resultados	37
Capítulo V. Conclusiones y trabajos futuros.....	40
Referencias.....	42
Glosario	45
Anexos.....	46
Anexo A. Test aplicado a los estudiantes.....	47

Índice de figuras

Figura 3. 1 Datos recolectados para el análisis.	24
Figura 3. 2 Datos con los valores de vigor, dedicación, absorción y compromiso.	24
Figura 3. 3 Datos finales para el análisis.....	25
Figura 3. 4 Proceso de KDD.....	26
Figura 3. 5 Fases del proceso de descubrimiento en base de datos, KDD.....	28
Figura 3. 6 Comparativa entre el Pretest y la Calificación final.	29
Figura 3. 7 Comparativa entre la Calificación final y el promedio general.....	30
Figura 4. 1 Mapa de correlaciones entre variables.....	32
Figura 4. 2 Clusters resultantes mediante el algoritmo K-means.....	35
Figura 4. 3 Algoritmo de búsqueda de conjuntos de items (Apriori).....	37

|

Índice de tablas

Tabla 3. 1 Características del grupo de estudio.....	21
Tabla 3. 2 Ítems del test UWES-S.....	22
Tabla 3. 3 Escala de medidas de los atributos considerados.....	23
Tabla 3. 4 Estadísticas descriptivas del conjunto de datos.....	29

Capítulo I. Generalidades

1.1 Antecedentes

Una de las dimensiones más importantes en el proceso de enseñanza aprendizaje lo constituye el rendimiento académico del estudiante, Artunduaga Murillo (2008) menciona que el rendimiento académico es un indicador de eficacia y calidad educativa, de tal forma que las manifestaciones de fracaso como el bajo rendimiento académico, la repitencia y la deserción, expresan deficiencias en un sistema educativo universitario. Por lo cual las causas del fracaso estudiantil deben buscarse más allá del estudiante mismo debido a que no es éste el único responsable de su fracaso.

Por lo tanto, debe ser necesario plantear un enfoque preventivo, lo que según Fullana (1996), significa tomar en consideración que la intervención educativa debe llevarse a cabo antes de que se haya alcanzado la situación de fracaso escolar. Esto quiere decir, que disminuir las tasas de deserción académica y de repitencia, así como aumentar el nivel de aprovechamiento en los estudios, es posible si se hace diagnóstico e intervención educativa, desde un enfoque de la prevención.

En la actualidad existen un gran índice de deserción y repitencia en los estudiantes que cursan la materia de programación a nivel universitario, y las instituciones no saben a que se debe. Cada uno de los estudiantes cursantes de esta materia en su transcurso van dejando una gran cantidad de datos que la universidad almacena, en sus sistemas de bases de datos.

Con el surgimiento de nuevas tecnologías, surgieron también mejores técnicas para el análisis de grandes cantidades de datos que antes no tenían ningún propósito nada más que ser almacenados, estas nuevas tecnologías buscan en grandes volúmenes de datos información importante que puede ser de mucha utilidad para las instituciones y así poder buscar una forma de ir reduciendo este problema en sus estudiantes.

Por consiguiente, para poder realizar el diagnóstico fue necesario el análisis de una gran cantidad de datos relacionados con los alumnos, los cuales estaban almacenados en las Bases de Datos institucionales, es aquí donde adquirió una gran importancia la minería de datos (Data Mining-DM), ya que es parte fundamental del Descubrimiento de Conocimientos en Bases de Datos (Knowledge Discovery in Databases - KDD), esto para el descubrimiento de los Patrones o Modelos de Comportamiento de los Alumnos para poder evitar la situación de deserción o fracaso escolar debido a un bajo rendimiento académico.

1.2 Planteamiento del Problema

1.2.1 Definición del problema

En una universidad del suroeste mexicano los estudiantes en el primer año de su carrera deben cursar la materia de Programación, dicha materia es impartida en aulas con equipo de cómputo con conexión a internet con lo cual ellos pueden trabajar con las tecnologías más actualizadas; sin embargo se entrevistó a varios docentes de la materia todos mencionan que al final de sus cursos cuentan como un gran índice de reprobación y es de ahí de dónde surge la inquietud por indagar las causas que llevan a los alumnos a un bajo rendimiento escolar en esta materia. En los últimos tiempos se tienen datos aproximados sobre el índice de reprobación de esta materia el cual es del 30% y un índice de deserción del 9%. Debido a esto los estudiantes se atrasan en su trayectoria académica lo que conlleva a una baja eficiencia terminal en su carrera.

Aquí es donde cabe destacar que los estudiantes generaron los datos en el transcurso de su materia, dichos datos fueron recolectados y almacenados en una base de datos, y si esta se analiza se obtienen las analíticas de aprendizaje, y estas son de mucha utilidad en la búsqueda de información para mejorar el proceso de enseñanza-aprendizaje de los estudiantes.

1.2.2 Delimitación de la investigación

1.2.2.1 Alcances

- Se aplicó la minería de datos, en los datos recolectados de estudiantes de la materia de programación de nivel superior para descubrir algunos indicadores y elementos relacionados con su rendimiento académico en esta materia.
- Se desarrolló un informe con los resultados obtenidos de los modelos para la ayuda a la toma de decisiones.

1.2.2.2 Limitaciones

- Cantidad de datos recolectados del grupo de estudio que cursan la materia de Programación a nivel superior.
- El tamaño de la población y la duración de esta, cumplan con los elementos suficientes para lograr el objetivo de la investigación.

1.2.3 Preguntas de investigación

¿Qué técnicas de minería de datos permitirían conocer los patrones de comportamiento de los estudiantes en la materia de Programación?

1.3 Objetivos

1.3.1 Objetivo General

Desarrollar un modelo de minería de datos para conocer la situación académica de estudiantes universitarios en la materia de Programación.

1.3.2 Objetivos específicos

- Identificar las variables de los factores que influyen en el rendimiento académico de los estudiantes.
- Analizar las variables de los factores que influyen en el rendimiento académico de los estudiantes.
- Analizar las políticas de evaluación que servirán de base para el modelo.
- Desarrollar el modelo de los factores que influyen en el rendimiento académico de los estudiantes.
- Presentar los resultados en formato de tesis.

1.4 Justificación

El desarrollo de este proyecto estuvo enfocado en el análisis de los perfiles de los estudiantes y los factores que influyen en su desarrollo académico, específicamente en la materia de programación. Los estudiantes desde su ingreso a la universidad están generando grandes cantidades de datos, desde sus datos personales hasta de las materias que cursan cada semestre, por lo que en este proyecto se utilizaron técnicas de minería de datos que permitieron a través de un modelo conocer los patrones de comportamiento de los estudiantes. Esto en beneficio de los estudiantes, ya que esta información les servirá a las autoridades correspondientes en la toma de decisiones en la implementación de estrategias enseñanza-aprendizaje.

1.5 Metodología utilizada

1.5.1 Población de estudio.

El desarrollo de este estudio se llevará a cabo en una universidad del suroeste mexicano en la materia de Programación, a un grupo de 25 alumnos que cursan el primer año de su carrera.

1.5.2 Instrumentos.

La recolección de los datos se llevará a cabo por medio de la utilización de un cuestionario el cual consta de 58 preguntas de opción múltiple dividido en 3 secciones (Ver anexo A), el cual ya se encuentra validado y con un nivel de confiabilidad aceptable, para que la información recabada refleje lo más fidedignamente la realidad y así poder obtener una investigación de calidad.

México.

Universidad Juárez Autónoma de Tabasco.⁵

Capítulo II. Marco teórico

2.1 Marco referencial

Derivado del análisis de investigaciones previas similares al estudio que se realizó cabe destacar que cada una de ellas nos aportara conocimiento previamente construido por la información teórica ya existente en su momento.

El mejoramiento de la calidad académica a nivel universitario no necesariamente debe estar enfocado en el mejoramiento del sistema de enseñanza aprendizaje, sino que debe atender otras variables ligadas a la calidad académica, como es el caso que nos mencionan los autores:

Acosta et. al. (2018), en su artículo titulado “Determinación de perfiles de rendimiento académico en la UNNE con Minería de Datos Educacional”, en el cual proponen evaluar a los estudiantes mediante técnicas de minería de datos con el objetivo de obtener los perfiles de estos y así poder determinar patrones que conduzcan al éxito o fracaso de los estudiantes en el primer año de su carrera universitaria en las materias de Álgebra Y Matemáticas I.

Y así poder detectar de forma temprana a los alumnos que tendrán dificultades en sus estudios, en este análisis del desempeño académico de los estudiantes no sólo se evaluó teniendo en cuenta los resultados de las instancias de evaluaciones previstas por la asignatura, sino que también se incluyeron otros factores culturales, sociales y/o económicos que afectan el rendimiento del alumno.

Para la recolección de esta última información se apoyaron de una aplicación web que permitió contar con una Encuesta On-Line compuesta por preguntas referentes a situación familiar e historial de estudios secundarios, entre otras cuestiones. Dichos datos fueron almacenados en un Data Warehouse por la necesidad de proporcionar una fuente única de datos limpia y consistente para propósitos de apoyo a la toma de decisiones y la necesidad de hacerlo sin afectar a los sistemas operacionales.

De igual forma en el artículo Análisis de la Deserción de Estudiantes Universitarios usando Técnicas de Minería de Datos (Miranda y Guzmán, 2017), utilizan técnicas de minería de datos como redes bayesianas, árboles de decisión y redes neuronales, para determinar cuáles son y que importancia tienen las variables que llevan a un estudiante a abandonar sus estudios. El tamaño de su muestra es de 9,195 estudiantes, la información de ellos se encontraba almacenada en la base de datos institucional y por último para la aplicación de las técnicas de la minería de datos utilizaron el software libre Weka.

Como resultados ellos pudieron determinar el clasificador mediante el algoritmo BayesNet en el software Weka que el abandono encontrado por este clasificador es de un 33,9% y la retención de un 66,1%, estos porcentajes corresponden al resultado de la clasificación para la variable clase (abandono y no abandono). Otra de sus clasificaciones mediante el algoritmo de árbol de decisión determino entre sus resultados obtenidos que: para la variable clase (abandono y retención), un 21,7% se clasifica como abandonado y un 78,3% como retenido. En tanto que aquellos estudiantes que hoy tienen beneficios estudiantiles (créditos, becas), en el promedio o menos que este valor tiene un 89,3% de probabilidad de permanecer en su carrera universitaria y un 10,7% de abandonar.

Por otro lado, quienes tienen mayores o menores beneficios a la media (extremos) desertan con una probabilidad de un 28,1% de probabilidad de abandonar sus estudios frente a un 71,9% que no lo haría. Este grupo de estudiantes corresponde al 63,1% de los estudiantes analizados y demuestra que quienes poseen un beneficio económico tienen una mayor posibilidad de permanecer y terminar sus carreras. También desarrollaron una red neuronal mediante el algoritmo Perceptrón Multicapa con un error admisible de 0,0001, que de acuerdo a la curva ROC, se puede decir que la clasificación fue realizada correctamente en un 83% de los casos y con una alta exactitud para los casos positivos clasificados con un 73% de precisión y un 88% para el ratio de casos negativos clasificados de forma exacta.

El estudio Minería de datos en los Sistemas de gestión de Aprendizaje en la Educación Universitaria (Hidalgo, 2018), describe la minería de datos educativa aplicada a los sistemas de

gestión del aprendizaje (LMS) con el objetivo de identificar patrones de comportamiento, esto servirá para la elección de recursos y actividades más adecuadas. El tamaño de su población es de 125 estudiantes de un curso de B-learning en la plataforma MOODLE, también utilizaron las herramientas informáticas Excel 2016, Statistic Program for Social Sciences (SPSS) y Rapidmining 7.5.

También crearon 3 grupos mediante la técnica de clustering y con el empleo del algoritmo de k medias obtuvieron el vector de desempeño promedio, obteniendo una distancia considerable entre estos grupos, concluyendo que existe una correlación entre el nivel de actividad y su rendimiento académico, como también recopila los perfiles de comportamiento y los compara con su nivel de actividad, con el objetivo de crear grupos con características similares garantizando que los recursos y actividades se adapten a las necesidades de los estudiantes.

Barragan y Gonzales, (2017), en su artículo “Hallazgos sobre la educación en Bogotá con base en la Encuesta Multipropósito 2014”, utilizaron los datos de la Encuesta Multipropósito del año 2014 financiada por la Secretaría Distrital de Planeación de la ciudad de Bogotá y desarrollada por el Departamento Nacional de Estadísticas. Dicha encuesta fue aplicada a 61, 725 personas. A estos datos recolectados se le aplicó la técnica de minería de datos descriptiva llamada árboles de decisión utilizando el algoritmo de Clasificación y Regresión (C&R) por medio del software estadístico SPSS versión 23, y como resultado obtuvieron que las dos principales razones para no estudiar son la falta de dinero y la necesidad de trabajar.

Pérez-Gutiérrez (2019), en su estudio Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico logro establecer una comparación entre las técnicas de minería de datos utilizadas para la identificación de la deserción estudiantil, utilizando un registro académico de un periodo de 7 años. Las técnicas comparadas fueron los Árboles de decisión, regresión logística y Naive Bayes, también se utilizó la herramienta Watson Analytics de IBM para comparar su usabilidad y precisión para un usuario no experto. Al final pudo demostrar que con el uso de algoritmos simples es suficiente para alcanzar niveles ideales de precisión y así ayudar a la disminución de la deserción estudiantil.

El artículo Minería de datos educativos para la predicción personalizada del rendimiento académico (Campo-Ávila, et al., 2017) informa cómo predecir el rendimiento académico alcanzado por los estudiantes en la materia Teoría de Autómatas y Lenguajes Formales, a partir de la realización de controles intermedios. Para ello se han aplicado y comparado distintos tipos de algoritmos de aprendizaje (vecinos más cercanos, árboles de decisión, multclasificadores). Todo el proceso de control y evaluación de los estudiantes durante el curso se ha llevado a cabo a través de la herramienta web denominada SIETTE, desarrollada en su departamento.

2.2 Marco conceptual

2.2.1 Minería de datos

Jiménez y Alvares (2010), definen la minería de datos o también conocida como Descubrimiento de Conocimiento en Bases de datos (sus siglas en inglés son “KDD – Knowledge Discovery in Databases”). Recientemente, se ha incrementado el interés en utilizar la minería de datos en el estudio educacional, centrándose en el desarrollo de métodos de descubrimiento que utilicen los datos de plataformas educacionales y en el uso de esos métodos para comprender mejor a los estudiantes y el entorno en el que aprenden. Los métodos empleados en la minería de datos en la educación suelen diferir de los métodos más generalistas, explotando explícitamente los múltiples niveles de jerarquía presentes en los datos. Métodos psicométricos suelen ser integrados con métodos de aprendizaje máquina y textos de minería de datos para lograr los objetivos como el campo que nos permite descubrir información nueva y potencialmente útil de grandes cantidades de datos. Así como también la clasifican según su enfoque.

2.2.1.1 Enfoques de la Minería de datos

Hay una gran variedad de métodos empleados habitualmente en el ámbito de la educación en la minería de datos. Estos métodos están comprendidos en las siguientes categorías: predicción, agrupamiento, minería de relaciones, inferencia a través de modelos, y destilación de datos para la interpretación por parte de un ser humano. Las tres primeras categorías son universales para distintos tipos de minería de datos (aunque en algunos casos con distintos nombres). Las

categorías cuarta y quinta consiguen una particular importancia dentro de la minería de datos educacionales. A continuación, se presenta la clasificación de acuerdo con los enfoques propuestos por Jiménez y Alvares (2010).

1.- Predicción

En predicción, el objetivo es desarrollar un modelo que pueda inferir una variable a partir de alguna combinación de otras variables incluidas en los datos. La predicción requiere etiquetas para la variable de salida para un conjunto de datos limitado, donde una etiqueta suponga una información fiable sobre el valor de la variable de salida en casos específicos. De todas maneras, en algunos casos es importante considerar el grado en el que estas etiquetas puedan ser aproximadas o inciertas. La predicción tiene dos usos clave comprendidos en la minería de datos educacionales. En algunos casos, métodos de predicción pueden ser usados para estudiar qué características de un modelo son importantes para una predicción, dando información sobre la construcción subyacente. Este es un enfoque común en programas de investigación que tratan de predecir resultados educacionales sin predecir anteriormente factores intermedios. En un segundo tipo de uso, los métodos de predicción son utilizados para predecir cuál será el valor de salida en contextos donde no es deseable obtener una etiqueta para esa construcción.

En general, existen tres tipos de predicción: clasificación, regresión y estimación de densidad. En clasificación, el valor predicho es una variable categórica o binaria. Algunos métodos populares de clasificación incluyen árboles de decisión, regresión logística (modelo de regresión para variables dependientes o de respuestas binomialmente distribuidas) y máquinas de soporte vector. En regresión, el valor predicho es una variable continua. Algunos métodos populares de regresión en la minería de datos educacionales incluyen la regresión lineal, redes neuronales y regresión sobre máquina de soporte vector. En estimación de densidad, la variable predicha es una función de densidad de probabilidad. Estimadores de densidad pueden estar basados en una variedad de funciones de kernel, incluyendo funciones gaussianas. Para cada tipo de predicción, las variables de entrada pueden ser categóricas o continuas.

2.- Agrupamiento

En agrupamiento, el objetivo es encontrar puntos de datos que se agrupen de manera natural, repartiendo el conjunto original de datos en un conjunto de 'clusters'. El agrupamiento es particularmente útil en casos donde las categorías más comunes de los datos no son conocidas. Si un conjunto de clusters es óptimo, en cada categoría, cada punto será más similar a los puntos pertenecientes a su cluster que a puntos pertenecientes a otros grupos. Los clusters pueden ser creados con distinta granularidad. Los algoritmos de agrupamiento pueden comenzar sin hipótesis previas sobre los grupos de datos (como el algoritmo k-means con inicio aleatorio), o empezar desde una hipótesis específica, posiblemente generada en estudios previos con un conjunto de datos distinto. Un algoritmo de agrupamiento puede postular que cada punto debe pertenecer únicamente a un cluster (como en el algoritmo k-means), o puede decidir que algunos de los puntos pertenezcan a varios o ningún cluster (como en los modelos de mezcla de gaussianas). La calidad de un conjunto de grupos o clusters suele ser evaluada tomando como referencia la medida en la cual el conjunto de clusters se ajusta a los datos, relativo a cuánto se espera que se ajusten únicamente por casualidad dado el número de clusters, usando métricas estadísticas tales como el criterio de información bayesiano.

3.- Minería de Relaciones

En la minería de relaciones, el objetivo es descubrir relaciones entre variables en un conjunto de datos con un gran número de variables. Una forma de realizar esto es localizando las variables más fuertemente relacionadas con una única variable de interés, o también mediante el descubrimiento de las relaciones más fuertes entre dos variables. De manera general, existen cuatro tipos de minería de relaciones: minería de reglas de asociación, minería de correlación, minería de patrones de secuencias y minería de datos causales.

4.- Descubrimiento mediante Modelos

En el descubrimiento mediante modelos, se desarrolla un modelo mediante predicción, agrupamiento o, en algunos casos, ingeniería del conocimiento (usando métodos de razonamiento

humano en vez de métodos automatizados). Este modelo es entonces utilizado como un componente en otro análisis, como predicción o minería de datos. En el caso de predicción, las predicciones hechas por el modelo creado son usadas como variables de entrada en la predicción de una nueva variable. En el caso de la minería de relaciones, se estudian las relaciones entre las predicciones del modelo creado y variables adicionales. Esto permite al investigador estudiar la relación entre una construcción compleja oculta y construcciones observables. A menudo, el descubrimiento mediante modelos enfatiza la validación generalizada de un modelo de predicción a través de varios contextos

2.2.1.2 Tipos de Modelos de Minería de Datos

Hernández et. al. (2004), Define que la minería de datos tiene como objetivo analizar los datos para extraer conocimiento. Este conocimiento puede ser en forma de relaciones, patrones o reglas inferidos de los datos y (previamente) desconocidos, o bien en forma de una descripción más concisa (es decir, un resumen de éstos). Estas relaciones o resúmenes constituyen el modelo de los datos analizados. Existen muchas formas diferentes de representar los modelos y cada una de ellas determina el tipo de técnica que puede usarse para inferirlos.

En la práctica, los modelos pueden ser de dos tipos: predictivos y descriptivos.

Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés, que denominamos variables objetivo o dependientes, usando otras variables o campos de la base de datos, a las que nos referiremos como variables independientes o predictivas. Los modelos descriptivos, en cambio, identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos.

2.2.1.3 Tareas de la minería de datos

Dentro de la minería de datos hemos de distinguir tipos de tareas, cada una de las cuales puede considerarse como un tipo de problema a ser resuelto por un algoritmo de minería de datos. Esto

significa que cada tarea tiene sus propios requisitos, y que el tipo de información obtenida con una tarea puede diferir mucho de la obtenida con otra (Hernández et. al., 2004).

Entre las tareas predictivas encontramos la clasificación y la regresión, mientras que el agrupamiento (clustering), las reglas de asociación, las reglas de asociación secuenciales y las correlaciones son tareas descriptivas

1.- Clasificación

La clasificación es quizá la tarea más utilizada. En ella, cada instancia (o registro de la base de datos) pertenece a una clase, la cual se indica mediante el valor de un atributo que llamamos la clase de la instancia. Este atributo puede tomar diferentes valores discretos, cada uno de los cuales corresponde a una clase. El resto de los atributos de la instancia (los relevantes a la clase) se utilizan para predecir la clase. El objetivo es predecir la clase de nuevas instancias de las que se desconoce la clase. Más concretamente, el objetivo del algoritmo es maximizar la razón de precisión de la clasificación de las nuevas instancias, la cual se calcula como el cociente entre las predicciones correctas y el número total de predicciones (correctas e incorrectas).

2.-Regresión

La regresión es también una tarea predictiva que consiste en aprender una función real que asigna a cada instancia un valor real. Esta es la principal diferencia respecto a la clasificación; el valor a predecir es numérico. El objetivo en este caso es minimizar el error (generalmente el error cuadrático medio) entre el valor predicho y el valor real.

3.- Agrupamiento

El agrupamiento (clustering) es la tarea descriptiva por excelencia y consiste en obtener grupos "naturales" a partir de los datos. Hablamos de grupos y no de clases, porque, a diferencia de la clasificación, en lugar de analizar datos etiquetados con una clase, los analiza para generar esta etiqueta. Los datos son agrupados basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizando la similitud entre los distintos grupos. Es decir, se forman

grupos tales que los objetos de un mismo grupo son muy similares entre sí y, al mismo tiempo, son muy diferentes a los objetos de otro grupo.

Al agrupamiento también se le suele llamar segmentación, ya que parte o segmenta los datos en grupos que pueden ser o no disjuntos. El agrupamiento está muy relacionado con la sumarización, que algunos autores consideran una tarea en sí misma, en la que cada grupo formado se considera como un resumen de los elementos que lo forman para así describir de una manera concisa los datos.

4.- Correlaciones

Las correlaciones son una tarea descriptiva que se usa para examinar el grado de similitud de los valores de dos variables numéricas. Una fórmula estándar para medir la correlación lineal es el coeficiente de correlación r , el cual es un valor real comprendido entre -1 y 1 . Si r es 1 (respectivamente, -1) las variables están perfectamente correlacionadas (perfectamente correlacionadas negativamente), mientras que si es 0 no hay correlación. Esto quiere decir que cuando r es positivo, las variables tienen un comportamiento similar (ambas crecen o decrecen al mismo tiempo) y cuando r es negativo si una variable crece la otra decrece. El análisis de correlaciones, sobre todo las negativas, puede ser muy útil para establecer reglas de ítems correlacionados.

5.- Reglas de asociación

Las reglas de asociación son también una tarea descriptiva, muy similar a las correlaciones, que tiene como objetivo identificar relaciones no explícitas entre atributos categóricos. Pueden ser de muchas formas, aunque la formulación más común es del estilo "si el atributo X toma el valor d entonces el atributo Y toma el valor b ". Las reglas de asociación no implican una relación causa-efecto, es decir, puede no existir una causa para que los datos estén asociados. Este tipo de tarea se utiliza frecuentemente en el análisis de la cesta de la compra, para identificar productos que son frecuentemente comprados juntos, información esta que puede usarse para ajustar los

inventarios, para la organización física del almacén o en campañas publicitarias. Las reglas se evalúan usando dos parámetros: precisión y soporte (cobertura).

Un caso especial de reglas de asociación, que recibe el nombre de reglas de asociación secuenciales, se usa para determinar patrones secuenciales en los datos. Estos patrones se basan en secuencias temporales de acciones y difieren de las reglas de asociación en que las relaciones entre los datos se basan en el tiempo.

2.2.1.4 Técnicas de minería de datos

Hernández et. al. (2004), mencionan que la minería de datos es un campo interdisciplinar por lo cual existen diferentes paradigmas detrás de las técnicas utilizadas: técnicas de inferencia estadística, árboles de decisión, redes neuronales, inducción de reglas, aprendizaje basado en instancias, algoritmos genéticos, aprendizaje bayesiano, programación lógica inductiva y varios tipos de métodos basados en núcleos, entre otros. Cada uno de estos paradigmas incluye diferentes algoritmos y variaciones de estos, así como otro tipo de restricciones que hacen que la efectividad del algoritmo dependa del dominio de aplicación, no existiendo lo que podríamos llamar el método universal aplicable a todo tipo de aplicación.

2.2.2 Compromiso estudiantil

El compromiso estudiantil ha sido ampliamente estudiado en las recientes décadas, esto debido a su influencia sobre el éxito académico de los estudiantes, influyendo sobre sus calificaciones o rendimiento. Por tanto, las investigaciones sobre este compromiso han adquirido relevancia al contribuir con información desde el estudiantado, en momentos en que un número creciente de sistemas educativos a lo largo del mundo emprenden reformas para mejorar sus servicios (Chacón et al. 2018).

El compromiso estudiantil permite escudriñar el éxito académico en la universidad desde la perspectiva de los propios estudiantes y desde los efectos positivos que tiene sobre ellos. Este compromiso vuelve más probable la decisión de continuar con sus estudios, por medio de la integración académica y social que el mismo compromiso fomenta (Díaz, 2008).

Este compromiso se refiere al tiempo y energía que los estudiantes dedican a realizar actividades cuyo propósito central está puesto en su éxito educativo, y las implicancias que esto tiene en término de las acciones que las universidades pueden, consecuentemente, llevar a cabo para promoverlo. En ese sentido, en este trabajo asumimos una perspectiva comportamental consistente con las investigaciones más relevantes sobre el tema (KAHU, 2013).

De igual manera, la orientación académica del compromiso estudiantil se basa en las acciones del estudiante que se alinean esencialmente con el aprendizaje, derivando en logro académico. En esta orientación, todos los esfuerzos se relacionan con los aprendizajes que los estudiantes desarrollan a partir de las actividades que realizan en el aula. Precisamente, en este alcance, un estudiante está dispuesto a invertir tiempo y esfuerzo para conseguir las metas académicas, de la mano con el nivel de calificaciones alcanzadas. Hacia dicha meta resulta crucial la intervención de los profesores, factor determinante en el aprendizaje efectivo, quienes deben organizar sus prácticas pedagógicas para favorecer dicho compromiso (Aspeé et al. 2018).

2.3 Marco Tecnológico

2.3.1 Weka

Weka (Waikato Environment for Knowledge Analysis, en español «entorno para análisis del conocimiento de la Universidad de Waikato), es un software de aprendizaje automático de código abierto al que se puede acceder a través de una interfaz gráfica de usuario, aplicaciones de terminal estándar o una API de Java. Se usa ampliamente para la enseñanza, la investigación y las aplicaciones industriales, contiene una gran cantidad de herramientas integradas para tareas de aprendizaje automático estándar y, además, brinda acceso transparente a cajas de herramientas conocidas como scikit-learn, R y DeepLearning4j (Waikato, 2020).

Weka soporta varias tareas estándar de minería de datos, especialmente, preprocesamiento de datos, clustering, clasificación, regresión, visualización, y selección.

Las características de Weka son:

- Está disponible libremente bajo la licencia pública general de GNU.
- Es muy portable porque está completamente implementado en Java y puede correr en casi cualquier plataforma.
- Contiene una extensa colección de técnicas para preprocesamiento de datos y modelado.
- Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario.

2.3.2 R

R es un lenguaje y entorno para computación estadística y gráficos. Es un proyecto GNU que es similar al lenguaje S y al entorno que fue desarrollado en los Laboratorios Bell (anteriormente AT&T, ahora Lucent Technologies) por John Chambers y sus colegas. R puede considerarse como una implementación diferente de S. Hay algunas diferencias importantes, pero gran parte del código escrito para S se ejecuta sin modificaciones bajo R (R- Project, 2020).

R proporciona una amplia variedad de técnicas estadísticas (modelos lineales y no lineales, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, agrupamiento, ...) y gráficas, y es altamente extensible. El lenguaje S es a menudo el vehículo elegido para la investigación en metodología estadística, y R proporciona una ruta de código abierto para participar en esa actividad.

Una de las fortalezas de R es la facilidad con la que se pueden producir trazados bien diseñados con calidad de publicación, incluidos los símbolos matemáticos y las fórmulas donde sea necesario. Se ha tenido mucho cuidado con los valores predeterminados para las elecciones de diseño menores en gráficos, pero el usuario conserva el control total.

R está disponible como software libre bajo los términos de la Free Software Foundation 's Licencia Pública General de GNU en forma de código fuente. Compila y se ejecuta en una amplia variedad de plataformas UNIX y sistemas similares (incluidos FreeBSD y Linux), Windows y MacOS.

2.3.3 El ambiente R

R es un conjunto integrado de instalaciones de software para la manipulación de datos, el cálculo y la visualización gráfica (R- Project, 2020. Incluye

- Una instalación efectiva de manejo y almacenamiento de datos,
- Un conjunto de operadores para cálculos en matrices, en particular matrices,
- Una colección grande, coherente e integrada de herramientas intermedias para el análisis de datos,
- Facilidades gráficas para el análisis y visualización de datos en pantalla o en papel, y
- Un lenguaje de programación bien desarrollado, simple y efectivo que incluye condicionales, bucles, funciones recursivas definidas por el usuario e instalaciones de entrada y salida.

El término "medio ambiente" tiene la intención de caracterizarlo como un sistema totalmente planificado y coherente, en lugar de una acumulación incremental de herramientas muy específicas e inflexibles, como suele ser el caso con otro software de análisis de datos.

R, como S, está diseñado en torno a un verdadero lenguaje informático, y permite a los usuarios agregar funcionalidades adicionales definiendo nuevas funciones. Gran parte del sistema está escrito en el dialecto R de S, lo que facilita a los usuarios seguir las elecciones algorítmicas realizadas. Para tareas computacionalmente intensivas, los códigos C, C ++ y Fortran se pueden vincular y llamar en tiempo de ejecución. Los usuarios avanzados pueden escribir código C para manipular objetos R directamente.

Muchos usuarios piensan en R como un sistema de estadísticas. Preferimos considerarlo como un entorno dentro del cual se implementan técnicas estadísticas. R se puede extender (fácilmente) a través de paquetes. Hay alrededor de ocho paquetes suministrados con la distribución R y muchos más están disponibles a través de la familia CRAN de sitios de Internet que cubren una amplia gama de estadísticas modernas.

R tiene su propio formato de documentación similar a LaTeX, que se utiliza para proporcionar documentación completa, tanto en línea en varios formatos como en papel.

2.4 Marco Legal

2.4.1 Licencia pública general GNU

Según la Free Software Foundation, Inc (2007), La Licencia Pública General GNU es una licencia gratuita con copyleft para software y otros tipos de obras.

Las licencias para la mayoría del software y otros trabajos prácticos están diseñadas para quitarle la libertad de compartir y cambiar los trabajos. Por el contrario, la Licencia pública general de GNU pretende 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

Cuando hablamos de software libre, nos referimos a la libertad, no al precio. Nuestras Licencias Públicas Generales están diseñadas para garantizar que usted 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 pueda cambiar el software o usar partes de él en nuevos programas gratuitos, y que sepa que puede hacer estas cosas.

Para proteger sus derechos, debemos evitar que otros le nieguen estos derechos o le pidan que los renuncie. Por lo tanto, tiene ciertas responsabilidades si distribuye copias del software o si lo modifica: responsabilidades de respetar la libertad de los demás.

Por ejemplo, si distribuye copias de dicho programa, ya sea gratis o pagando una tarifa, debe transmitir a los destinatarios las mismas libertades que recibió. Debe asegurarse de que ellos también reciban o puedan obtener el código fuente. Y debe mostrarles estos términos para que conozcan sus derechos.

Para la protección de los desarrolladores y autores, la GPL explica claramente que no hay garantía para este software gratuito. Tanto por el bien de los usuarios como de los autores, la GPL exige que las versiones modificadas se marquen como cambiadas, para que sus problemas no se atribuyan erróneamente a los autores de versiones anteriores.

Por último, cada programa está constantemente amenazado por las patentes de software. Los estados no deben permitir que las patentes restrinjan el desarrollo y uso de software en computadoras de propósito general, pero en aquellos que lo hacen, deseamos evitar el peligro

especial de que las patentes aplicadas a un programa libre puedan convertirlo efectivamente en propietario. Para evitar esto, la GPL asegura que las patentes no se pueden usar para hacer que el programa no sea libre.

Universidad Juárez Autónoma de Tabasco.
México.

Capítulo III. Desarrollo metodológico

3.1 Abstracción del escenario

En una universidad del suroeste están interesados en conocer cómo se relaciona el compromiso de sus estudiantes con su calificación al final del curso, por tal motivo se utiliza un grupo de estudio de 25 estudiantes de los cuales el 80% son masculinos y el 20% femeninos como se puede observar en la tabla 2.1, a dichos estudiantes el primer día de su curso se les aplicó el test UWES-S el cual consta de 17 ítems 6 de vigor, 5 de dedicación y 6 de absorción, con los cuales se puede conocer el valor de compromiso de estos, con lo cual al final del curso se podrá corroborar que tan relacionado está el compromiso con las notas finales de los estudiantes, las cuales se pueden observar cómo quedaron distribuidas al final del curso en la tabla 3.1.

Tabla 3. 1 Características del grupo de estudio.

<i>Variables</i>	<i>Valores</i>	<i>N</i>	<i>%</i>
Sexo	M (Masculino)	20	80
	F (Femenino)	5	20
Calificación Final	5	4	16
	6	4	16
	7	5	20
	8	8	32
	9	4	16

3.2 Instrumento UWES- S

El compromiso estudiantil se midió utilizando la versión del UWES para estudiantes UWES-S (Schaufeli y Bakker, 2003), el cual es un test de autoevaluación de 17 ítems los cuales se pueden ver en la Tabla 3.2 y que miden Vigor, Dedicación y Absorción:

- Vigor (seis ítems): Se refieren a los altos niveles de energía y resiliencia, la voluntad de dedicar esfuerzos, no fatigarse con facilidad, y la persistencia frente a las dificultades.
- Dedicación (cinco ítems): Se refieren al sentido o significado del trabajo, a sentirse entusiasmado y orgulloso por su labor, y sentirse inspirado y retado por el trabajo.
- Absorción (seis ítems): Se refieren a estar felizmente inmerso en sus labores y presentar dificultad para dejarlo, de tal forma que el tiempo pasa rápidamente y uno se olvida de todo a su alrededor.

Todos los ítems se puntuaron en una escala de calificación de frecuencia de siete puntos que iba de 0 (nunca) a 6 (todos los días). Estudios nacionales e internacionales revelan coeficientes alfa de Cronbach para las tres subescalas que oscilan entre .68 y .81. Los coeficientes de confiabilidad oscilan entre cero y uno, cero significa nula confiabilidad y uno representa el nivel máximo de confiabilidad, si se obtiene como resultado un valor de 0.25 de coeficiente, esto indica baja confiabilidad; si el valor es 0.50, la confiabilidad es media o regular; si supera el 0.75 es aceptable; finalmente, si el resultado es mayor a 0.90 se considera elevada (Hernández et al., 2010).

Tabla 3. 2 Ítems del test UWES-S

<i>N</i>	<i>Ítem</i>	<i>Variable</i>
1	Mis tareas como estudiante me hacen sentir lleno de energía (VI1)	P1
2	Creo que mi carrera tiene significado (DE1)	P2
3	El tiempo “pasa volando” cuando realizo mis tareas como estudiante (AB1)	P3
4	Me siento fuerte y vigoroso cuando estoy estudiando o voy a las clases (VI2)	P4
5	Estoy entusiasmado con mi carrera (DE2)	P5
6	Olvido todo lo que pasa alrededor de mi cuando estoy abstraído con mis estudios (AB2)	P6
7	Mis estudios me inspiran cosas nuevas (DE3)	P7
8	Cuando me levanto por la mañana me apetece ir a clase o estudiar (VI3)	P8

9	Soy feliz cuando estoy haciendo tareas relacionadas con mis estudios (AB3)	P9
10	Estoy orgulloso de hacer esta carrera (DE4)	P10
11	Estoy inmerso en mis estudios (AB4)	P11
12	Puedo seguir estudiando durante largos periodos de tiempo (VI4)	P12
13	Mi carrera es retadora para mi (DE5)	P13
14	Me “dejo llevar” cuando realizo mis tareas como estudiante (AB5)	P14
15	Soy muy “resistente” para afrontar mis tareas como estudiante (VI5)	P15
16	Es difícil para mí separarme de mis estudios (AB6)	P16
17	En mis tareas como estudiante no paro incluso si no me encuentro bien (VI6)	P17

Para el establecimiento de normas estadísticas para el UWES-S, se decidió utilizar cinco categorías: “Muy bajo”, “Bajo”, “Promedio”, “Alto” y “Muy alto”. La Tabla 3.3 despliega los puntajes para el UWES-17 (Schaufeli y Bakker, 2003).

Tabla 3. 3 Escala de medidas de los atributos considerados

Categoría	Vigor	Dedicación	Absorción	Puntaje total
Muy bajo	≤2.00	≤1.60	≤1.60	≤1.93
Bajo	2.01-3.20	1.61-3.00	1.61-2.75	1.94-3.06
Promedio	3.21-4.80	3.01-4.90	2.76-4.40	3.07-4.66
Alto	4.81-5.60	4.91-5.79	4.41-5.35	4.67-5.53
Muy alto	≥5.61	≥5.80	≥5.36	≥5.54

3.3 Recolección de los datos

Para poder lograr el objetivo se utilizó un tipo de muestra no probabilística del tipo dirigida y por conveniencia (Hernández et al., 2010) pues se tomaron en cuenta los estudiantes de dos distintas carreras y que cursaban la materia de programación 1, esto solo para conocer cómo está relacionado el compromiso estudiantil con las notas finales del curso, se recopilaron los

siguientes datos de los 25 estudiantes como son el sexo, los resultados de la prueba UWES-S y la calificación final del curso a como se pueden observar en la Figura 3.1.

1	Centro	Calificación Test	Calificación Clase	Evaluación	Calificación Final	SEXO	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	Presencia	
1	LSC	SEIS	DEZ	OCHO	NOVE	H	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	OCHO
2	LSC	OCHO	DEZ	SETE	OCHO	H	DOS	OCHO	TRES	UNO	OCHO	SEIS	TRES	OCHO											
3	LSC	SETE	DEZ	OCHO	SEIS	H	UNO	TRES	OCHO	TRES	SEIS	SEIS	CUATRO	OCHO	SEIS	CUATRO	SEIS	TRES	DOS	SEIS	TRES	OCHO	OCHO	OCHO	OCHO
4	LSC	NOVE	NOVE	SETE	OCHO	H	TRES	CUATRO	DOS	TRES	OCHO	OCHO	CUATRO	OCHO	OCHO	SEIS	SEIS	OCHO	OCHO	SEIS	OCHO	OCHO	SEIS	OCHO	NOVE
5	LSC	NOVE	DEZ	OCHO	OCHO	M	TRES	OCHO	CUATRO	CUATRO	TRES	SEIS	OCHO	OCHO	SEIS	CUATRO	SEIS	OCHO	CUATRO	SEIS	OCHO	CUATRO	SEIS	OCHO	OCHO
6	LSC	SETE	NOVE	SEIS	OCHO	H	DOS	OCHO	OCHO	OCHO	CUATRO	CUATRO	OCHO	CUATRO	CUATRO	OCHO	CUATRO	CUATRO	CUATRO	CUATRO	TRES	DOS	CUATRO	OCHO	SETE
7	LSC	SEIS	DEZ	SETE	SETE	H	TRES	OCHO	TRES	DOS	CUATRO	OCHO	DOS	CUATRO	SEIS	TRES	OCHO	CUATRO	DOS	DOS	OCHO	UNO	OCHO	OCHO	SETE
8	LSC	NOVE	NOVE	SEIS	SETE	H	TRES	TRES	CUATRO	SEIS	CUATRO	CUATRO	OCHO	TRES	SEIS	TRES	SEIS	TRES	SEIS	TRES	OCHO	DOS	TRES	UNO	OCHO
9	LSC	OCHO	DEZ	OCHO	NOVE	H	DOS	CUATRO	TRES	DOS	CUATRO	SEIS	DOS	CUATRO	SEIS	CUATRO	SEIS	TRES	DOS	CUATRO	CUATRO	OCHO	SEIS	OCHO	OCHO
10	LSC	OCHO	NOVE	OCHO	OCHO	H	DOS	CUATRO	DOS	TRES	OCHO	OCHO	CUATRO	TRES	OCHO	CUATRO	OCHO	TRES	TRES	OCHO	DOS	OCHO	DOS	UNO	OCHO
11	LSC	SETE	DEZ	SETE	SETE	H	SEIS	SEIS	SEIS	OCHO	SEIS	SEIS	SEIS	SEIS	SEIS	OCHO	SEIS	OCHO	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	NOVE
12	LSC	SEIS	OCHO	SEIS	SETE	H	OCHO	SEIS	SEIS	DOS	OCHO	SEIS	TRES	SEIS	OCHO										
13	LSC	OCHO	DEZ	OCHO	OCHO	H	DOS	SEIS	CUATRO	OCHO	DOS	SEIS	TRES	OCHO											
14	LSC	OCHO	DEZ	OCHO	OCHO	H	DOS	SEIS	UNO	OCHO	SEIS	OCHO	OCHO												
15	LSC	SETE	OCHO	SETE	OCHO	H	TRES	SEIS	SEIS	TRES	CUATRO	OCHO	CUATRO	CUATRO	TRES	TRES	CUATRO	CUATRO	TRES	TRES	CUATRO	TRES	TRES	OCHO	OCHO
16	LSC	NOVE	DEZ	OCHO	SETE	H	TRES	OCHO	TRES	CUATRO	OCHO	SEIS	CUATRO	CUATRO	SEIS	CUATRO	SEIS	OCHO	OCHO						
17	LSC	NOVE	DEZ	OCHO	NOVE	H	UNO	CUATRO	TRES	OCHO	TRES	OCHO	DOS	TRES	OCHO	CUATRO	OCHO	TRES	TRES	UNO	OCHO	UNO	DOS	OCHO	OCHO
18	LSC	SEIS	OCHO	OCHO	OCHO	H	OCHO	SEIS	OCHO	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	OCHO	SEIS	CUATRO	TRES	SEIS	SEIS	SEIS	SEIS	OCHO	NOVE
19	LSC	SEIS	OCHO	SETE	OCHO	H	SEIS	SEIS	TRES	UNO	SEIS	OCHO													
20	LSC	SEIS	DEZ	OCHO	OCHO	H	TRES	SEIS	TRES	TRES	OCHO	SEIS	SEIS	TRES	TRES	TRES	OCHO	SEIS	SEIS	OCHO	TRES	TRES	TRES	SETE	SETE
21	LSC	OCHO	DEZ	SEIS	SEIS	H	OCHO	OCHO	OCHO	OCHO	SEIS	CUATRO	SEIS	OCHO	SEIS	CUATRO	SEIS	OCHO	SEIS	OCHO	SEIS	OCHO	SEIS	SEIS	OCHO
22	LSC	SETE	DEZ	OCHO	OCHO	H	TRES	SEIS	TRES	TRES	OCHO	SEIS	SEIS	TRES	TRES	TRES	OCHO	SEIS	SEIS	OCHO	TRES	TRES	TRES	SETE	SETE
23	LSC	OCHO	DEZ	SEIS	SEIS	H	OCHO	OCHO	OCHO	OCHO	SEIS	CUATRO	SEIS	OCHO	SEIS	CUATRO	SEIS	OCHO	SEIS	OCHO	SEIS	OCHO	SEIS	SEIS	OCHO
24	MA	OCHO	NOVE	DOS	OCHO	H	TRES	SEIS	CUATRO	SEIS	TRES	TRES	TRES	TRES	TRES	CUATRO	TRES	TRES	CUATRO	TRES	TRES	CUATRO	TRES	SETE	SETE
25	MA	OCHO	DEZ	OCHO	OCHO	H	UNO	OCHO	OCHO	OCHO	SEIS	TRES	TRES	TRES	TRES	OCHO	TRES	TRES	OCHO	TRES	DOS	TRES	UNO	NOVE	NOVE
26	LSC	SEIS	OCHO	CUATRO	OCHO	H	TRES	SEIS	OCHO	TRES	TRES	SEIS	CUATRO	TRES	SEIS	SEIS	SEIS	SEIS	CUATRO	TRES	OCHO	CUATRO	OCHO	SEIS	OCHO

Figura 3. 1 Datos recolectados para el análisis.

Fuente: Elaboración propia.

3.4 Limpieza y transformación

Durante esta etapa se procedió a realizar el promedio de los 17 ítems del test, los cuales 6 son de vigor, 5 de dedicación y 6 de absorción, para posteriormente con estos 3 valores de obtuviera el valor de compromiso estudiantil, quedando la tabla de los datos a como se observa en la Figura 3.2.

Evaluación	Calificación	SEXO	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	Vigor	Dedicación	Absorción	Compromiso	Promedio	
OCHO	NOVE	H	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	SETE	
SETE	OCHO	H	DOS	OCHO	TRES	UNO	UNO	SEIS	TRES	TRES	TRES	TRES	TRES	OCHO	OCHO	CUATRO	TRES	OCHO							
OCHO	SEIS	H	UNO	TRES	OCHO	TRES	DOS	SEIS	CUATRO	OCHO	SEIS	CUATRO	SEIS	TRES	SEIS	TRES	SEIS	TRES	OCHO	OCHO	TRES	TRES	SEIS	CUATRO	OCHO
SETE	OCHO	H	TRES	CUATRO	CUATRO	DOS	TRES	OCHO	OCHO	CUATRO	OCHO	SEIS	OCHO	OCHO	TRES	SEIS	OCHO	CUATRO	TRES	CUATRO	OCHO	OCHO	OCHO	OCHO	OCHO
SETE	OCHO	H	OCHO	OCHO	OCHO	SEIS	SEIS	OCHO	OCHO	SEIS	SEIS	OCHO	SEIS	OCHO	SEIS	OCHO	SEIS	SEIS	OCHO	SEIS	SEIS	OCHO	OCHO	OCHO	NOVE
OCHO	OCHO	H	TRES	OCHO	CUATRO	CUATRO	TRES	SEIS	OCHO	OCHO	SEIS	CUATRO	SEIS	OCHO	CUATRO	SEIS	OCHO	SEIS	OCHO	CUATRO	OCHO	OCHO	OCHO	OCHO	OCHO
SEIS	OCHO	H	DOS	OCHO	OCHO	OCHO	CUATRO	OCHO	OCHO	CUATRO	CUATRO	CUATRO	CUATRO	TRES	SEIS	OCHO	SEIS	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	CUATRO	SETE
SETE	SETE	H	TRES	OCHO	TRES	DOS	CUATRO	OCHO	DOS	CUATRO	SEIS	TRES	OCHO	CUATRO	SEIS	SEIS	OCHO	SEIS	OCHO	TRES	CUATRO	CUATRO	CUATRO	CUATRO	SETE
SEIS	SETE	H	TRES	TRES	CUATRO	OCHO	CUATRO	CUATRO	OCHO	TRES	SEIS	SEIS	SEIS	OCHO	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	OCHO
OCHO	NOVE	H	DOS	CUATRO	TRES	DOS	CUATRO	SEIS	DOS	CUATRO	SEIS	TRES	SEIS	CUATRO	CUATRO	SEIS	SEIS	CUATRO	TRES	OCHO	CUATRO	CUATRO	CUATRO	OCHO	OCHO
OCHO	OCHO	H	SEIS	CUATRO	DOS	TRES	OCHO	OCHO	CUATRO	TRES	OCHO	OCHO	TRES	TRES	OCHO	SEIS	CUATRO	TRES	OCHO	SEIS	CUATRO	CUATRO	CUATRO	CUATRO	SETE
SETE	SETE	H	SEIS	SEIS	SEIS	OCHO	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	NOVE								
SEIS	SETE	H	OCHO	SEIS	SEIS	DOS	DOS	SEIS	SEIS	TRES	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	OCHO
CUATRO	SEIS	H	OCHO	SEIS	CUATRO	CUATRO	DOS	SEIS	TRES	TRES	CUATRO	CUATRO	CUATRO	TRES	TRES	CUATRO	CUATRO	SETE							
OCHO	OCHO	H	SEIS	SEIS	TRES	UNO	OCHO	SEIS	OCHO	SEIS	OCHO	SEIS	OCHO	SEIS	OCHO	SEIS	OCHO	SEIS	OCHO	SEIS	OCHO	OCHO	OCHO	OCHO	OCHO
SETE	OCHO	H	TRES	OCHO	TRES	CUATRO	SEIS	SEIS	CUATRO	CUATRO	TRES	TRES	CUATRO	CUATRO	TRES	TRES	OCHO	OCHO	TRES	UNO	CUATRO	CUATRO	CUATRO	CUATRO	OCHO
OCHO	SETE	H	TRES	OCHO	TRES	CUATRO	SEIS	SEIS	CUATRO	CUATRO	SEIS	CUATRO	SEIS	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO	OCHO
OCHO	NOVE	H	UNO	CUATRO	TRES	OCHO	TRES	OCHO	DOS	TRES	OCHO	OCHO	TRES	TRES	UNO	OCHO	UNO	OCHO	UNO	OCHO	SEIS	TRES	TRES	OCHO	OCHO
OCHO	OCHO	H	OCHO	SEIS	OCHO	CUATRO	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	NOVE								
SETE	OCHO	H	SEIS	SEIS	TRES	UNO	SEIS	SEIS	SEIS	SEIS	SEIS	OCHO	SEIS	OCHO	SEIS	SETE									
OCHO	OCHO	H	TRES	SEIS	TRES	TRES	OCHO	SEIS	TRES	OCHO	TRES	TRES	OCHO	TRES	TRES	CUATRO	SEIS	CUATRO	SETE						
DOS	OCHO	H	UNO	OCHO	OCHO	OCHO	SEIS	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	TRES	OCHO							
CUATRO	OCHO	H	TRES	SEIS	OCHO	TRES	TRES	SEIS	CUATRO	TRES	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	SEIS	OCHO

Figura 3. 2 Datos con los valores de vigor, dedicación, absorción y compromiso.

Fuente: Elaboración propia.

Posteriormente, después de anexar los valores de vigor, dedicación, absorción y compromiso, se optó por eliminar los valores recolectados en la prueba, ya que estos no serían necesarios para proseguir en el análisis, de igual forma se procedió a la discretización de los datos quedando como resultante la Figura 3.3 a como se puede observar.

1	Test	Classcraft	Eval	Final	Sex	Vigor	Dedicacion	Absorcion	Comp	Prom
2	SEIS	DIEZ	OCHO	NUEVE	H	TRES	TRES	TRES	TRES	SIETE
3	OCHO	DIEZ	SIETE	OCHO	M	CUATRO	TRES	CINCO	CUATRO	SIETE
4	SIETE	DIEZ	OCHO	SEIS	H	TRES	TRES	SEIS	CUATRO	OCHO
5	NUEVE	NUEVE	SIETE	OCHO	H	TRES	CUATRO	CINCO	CUATRO	OCHO
6	NUEVE	NUEVE	SIETE	OCHO	H	SEIS	CINCO	CINCO	CINCO	NUEVE
7	NUEVE	DIEZ	OCHO	OCHO	M	CINCO	CUATRO	CINCO	CINCO	OCHO
8	SIETE	NUEVE	SEIS	OCHO	H	CUATRO	CUATRO	CUATRO	CUATRO	SIETE
9	SEIS	DIEZ	SIETE	SIETE	H	CUATRO	TRES	CUATRO	CUATRO	SIETE
10	NUEVE	NUEVE	SEIS	SIETE	H	TRES	TRES	CUATRO	TRES	OCHO
11	OCHO	DIEZ	OCHO	NUEVE	H	CUATRO	TRES	CINCO	CUATRO	OCHO
12	CINCO	NUEVE	CINCO	CINCO	H	TRES	CUATRO	CUATRO	CUATRO	SIETE
13	SIETE	DIEZ	SIETE	SIETE	M	SEIS	SEIS	SEIS	SEIS	NUEVE
14	SEIS	OCHO	SEIS	SIETE	H	CUATRO	CINCO	SEIS	CINCO	OCHO
15	SEIS	DIEZ	CINCO	SEIS	H	TRES	TRES	CUATRO	CUATRO	SIETE
16	OCHO	DIEZ	CINCO	OCHO	H	TRES	CINCO	CINCO	CUATRO	OCHO
17	SIETE	OCHO	SIETE	OCHO	H	CUATRO	CUATRO	CUATRO	CUATRO	OCHO
18	NUEVE	DIEZ	CINCO	SIETE	H	CUATRO	CUATRO	CINCO	CINCO	OCHO
19	NUEVE	DIEZ	OCHO	NUEVE	H	TRES	TRES	TRES	TRES	OCHO
20	SEIS	OCHO	CINCO	OCHO	H	CINCO	CINCO	CINCO	CINCO	NUEVE
21	SEIS	OCHO	SIETE	OCHO	H	CINCO	SEIS	CINCO	SEIS	SIETE
22	SIETE	DIEZ	CINCO	CINCO	H	CUATRO	CINCO	CUATRO	CUATRO	SIETE
23	OCHO	DIEZ	SEIS	SEIS	H	CINCO	CINCO	SEIS	CINCO	SEIS
24	CINCO	NUEVE	TRES	CINCO	M	CUATRO	CUATRO	TRES	CUATRO	SIETE
25	CINCO	DIEZ	CINCO	SEIS	M	UNO	CUATRO	CUATRO	TRES	NUEVE
26	SEIS	OCHO	CUATRO	CINCO	H	CUATRO	CUATRO	CINCO	CUATRO	SEIS

Figura 3. 3 Datos finales para el análisis.

Fuente: Elaboración propia.

3.5 Proceso de descubrimiento de conocimiento en bases de datos (KDD)

Se define el KDD como "el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos". En esta definición se resumen cuáles deben ser las propiedades deseables del conocimiento extraído (Hernández et. al., 2004):

- Válido: hace referencia a que los patrones deben seguir siendo precisos para datos nuevos (con un cierto grado de certidumbre), y no sólo para aquellos que han sido usados en su obtención.
- Novedoso: que aporte algo desconocido tanto para el sistema y preferiblemente para el usuario.

- Potencialmente útil: la información debe conducir a acciones que reporten algún tipo de beneficio para el usuario.
- comprensible: la extracción de patrones no comprensibles dificulta o imposibilita su interpretación, revisión, validación y uso en la toma de decisiones. De hecho, una información incomprensible no proporciona conocimiento (al menos desde el punto de vista de su utilidad).

Como se deduce de la anterior definición, el KDD es un proceso complejo que incluye no sólo la obtención de los modelos o patrones (el objetivo de la minería de datos), sino también la evaluación y posible interpretación de estos, tal y como se refleja en la Figura 3.4.



Figura 3. 4 Proceso de KDD

Fuente: Hernández et. al. (2004).

Así, los sistemas de KDD permiten la selección, limpieza, transformación y proyección de los datos; analizar los datos para extraer patrones y modelos adecuados; evaluar e interpretar los patrones para convertirlos en conocimiento; consolidar el conocimiento resolviendo posibles conflictos con conocimiento previamente extraído; y hacer el conocimiento disponible para su uso. Esta definición del proceso clarifica la relación entre

el KDD y la minería de datos: el KDD es el proceso global de descubrir conocimiento útil desde las bases de datos mientras que la minería de datos se refiere a la aplicación de los métodos de aprendizaje y estadísticos para la obtención de patrones y modelos. Al ser la fase de generación de modelos, comúnmente se asimila KDD con minería de datos.

El KDD es un proceso iterativo e interactivo. Es iterativo ya que la salida de alguna de las fases puede hacer volver a pasos anteriores y porque a menudo son necesarias varias iteraciones para extraer conocimiento de alta calidad. Es interactivo porque el usuario, o más generalmente un experto en el dominio del problema, debe ayudar en la preparación de los datos, validación del conocimiento extraído.

El proceso de KDD se organiza en torno a cinco fases como se observa en la Figura 3.5. En la fase de integración y recopilación de datos se determinan las fuentes de información que pueden ser útiles y donde conseguirlas. A continuación, se transforman todos los datos a un formato común, frecuentemente mediante un almacén de datos que consiga unificar de manera operativa toda la información recogida, detectando y resolviendo las inconsistencias. Este almacén de datos facilita enormemente la "navegación" y visualización previa de sus datos, para discernir qué aspectos puede interesar que sean estudiados.

Dado que los datos provienen de diferentes fuentes, pueden contener valores erróneos o faltantes. Estas situaciones se tratan en la fase de selección, limpieza y transformación, en la que se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Además, se proyectan los datos para considerar únicamente aquellas variables o atributos que van a ser relevantes, con el objetivo de hacer más fácil la tarea propia de minería y para que los resultados de esta sean más útiles. La selección incluye tanto una criba o fusión horizontal (filas / registros) como vertical (columnas / atributos). Las dos primeras fases se suelen englobar bajo el nombre de "preparación de datos". En la fase de minería de datos, se decide cual es la tarea para realizar (clasificar, agrupar, etc.) y se elige el método que se va a utilizar. En la fase de evaluación e interpretación se evalúan los patrones y se analizan por los expertos, y si es necesario se vuelve a las fases anteriores para una nueva iteración. Esto incluye resolver posibles

conflictos con el conocimiento que se disponía anteriormente. Finalmente, en la fase de difusión se hace uso del nuevo conocimiento y se hace partícipe de el a todos los posibles usuarios.

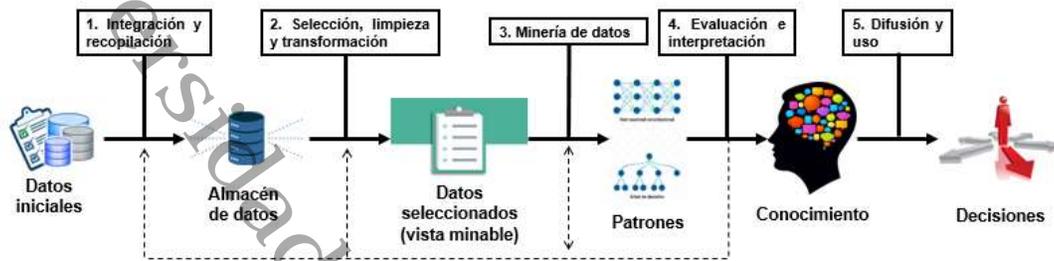


Figura 3. 5 Fases del proceso de descubrimiento en base de datos, KDD
Fuente: Hernández et. al. (2004).

3.6 Estadístico descriptivo de las variables que integran el estudio

En la tabla 3.4 se observan los resultados obtenidos de la población estudiada, como el valor de la variable del test aplicado al inicio del curso con el id "Test" con valores máximos de 9 y mínimos de 5, de igual manera la variable con el id "Classcraft" para la cual se utilizaron los valores de una prueba realizada mediante la plataforma Classcraft para la cual los valores máximos fueron de 10 y mínimos de 8, así mismo las variables "Eval" con un máximo de 8 y mínimo de 3, y "Final" con máximo de 9 y mínimo de 5, corresponden a la calificación del examen ordinario de la materia de programación y del promedio final obtenido en dicha materia respectivamente. En cuanto a las variables obtenidas con ayuda de la prueba UWES-S las cuales son vigor, dedicación, absorción y compromiso cabe destacar que los valores de estas oscilaban entre 0 y 6, y de las cuales destacan sus promedios de 3.9, 4.1, 4.6 y 4.2 respectivamente. Por último, se utilizó una variable llamada "Final", la cual corresponde al promedio general de los estudiantes en cuanto al avance de su carrera, con valores máximos de 9, mínimos de 6 y promedio de 7.6.

Tabla 3. 4 Estadísticas descriptivas del conjunto de datos

Variable	Descripción	Mínimo	Máximo	Promedio
Test	Test aplicado al inicio del curso	9	5	7.1
Classcraft	Calificación obtenida en una prueba gamificada en la plataforma Classcraft	10	8	9.4
Eval	Calificación obtenida en el examen ordinario del curso	8	3	6.2
Final	Calificación obtenida al final del curso	9	5	7.1
Vigor	Valor del vigor obtenido con el UWES-S	6	1	3.9
Dedicación	Valor de la dedicación obtenido con el UWES-S	6	3	4.1
Absorción	Valor de la absorción obtenido con el UWES-S	6	3	4.6
Comp	Valor de compromiso estudiantil	6	3	4.2
Promedio	Promedio general en el kardex	9	6	7.6

Con los datos obtenidos anteriormente también se realizó a una gráfica comparativa respecto a los datos obtenidos en el Pretest y en la Calificación Final a como se puede observar en la Figura 3.6, cabe destacar el comportamiento de los estudiantes ya que en su mayor parte estos obtuvieron una calificación igual o mejor al final que la obtenida en el pretest.

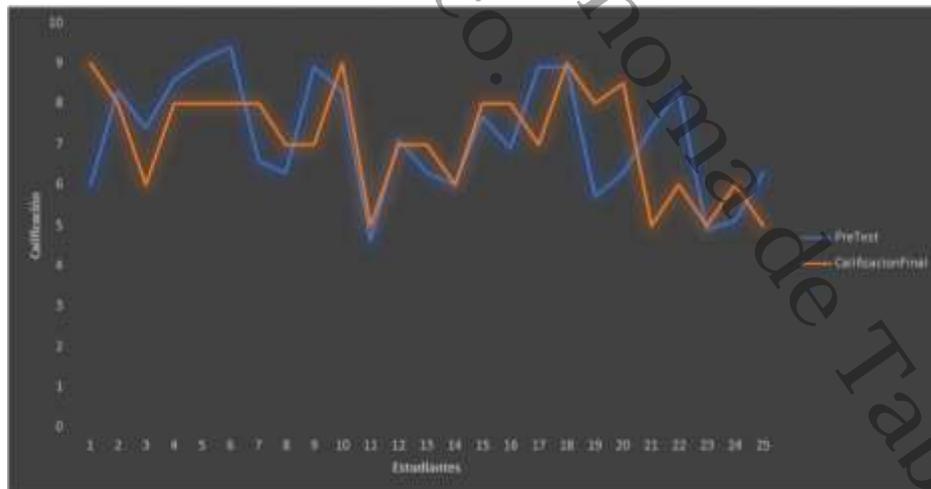


Figura 3. 6 Comparativa entre el Pretest y la Calificación final.

Fuente: Elaboración propia

Así mismo, se realizó una comparativa entre la calificación final y el promedio general de los estudiantes donde la mayoría obtuvo la calificación final debajo del promedio general a como se puede observar en la Figura 3.7.

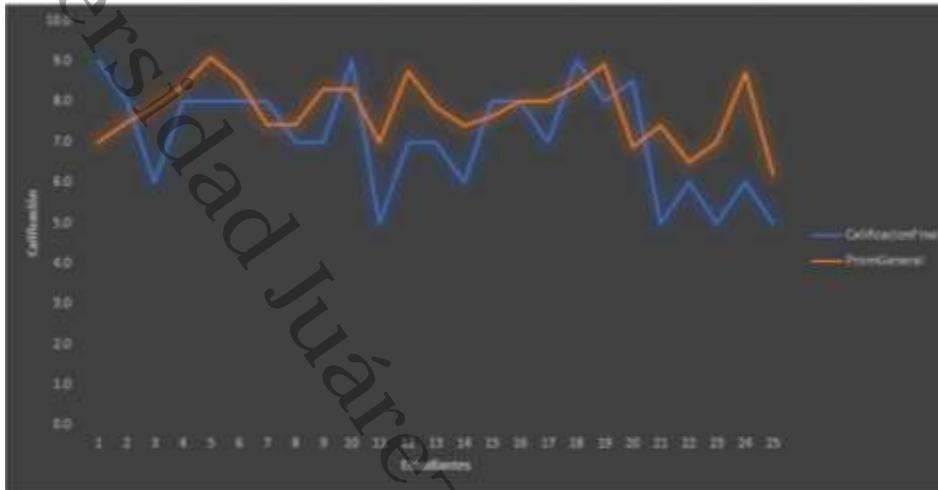


Figura 3. 7 Comparativa entre la Calificación final y el promedio general.

Fuente: Elaboración propia

Capítulo IV. Análisis de minería de datos

4.1 Técnica de minería de datos: Correlación

Históricamente, se ha utilizado el coeficiente de correlación para medir la dependencia entre dos series de datos numéricos. La correlación es una medida entre 1 y -1 bastante sencilla que permite establecer el grado de similitud entre las dos series de valores ya que entre más cercano a 1 es mayor grado de correlación y más alejado de este es menor grado de correlación según Hernández et. al. (2004).

4.1.1 Paquete: Psych 2.1.9

Una caja de herramientas de propósito general para la personalidad, la teoría psicométrica y la psicología experimental a la cual se adaptan muy bien las variables vigor, dedicación, absorción y compromiso obtenidas en esta investigación mediante el test UWES-S. Las funciones son principalmente para el análisis multivariado y la construcción de escalas mediante el análisis factorial, el análisis de componentes principales, el análisis de conglomerados y el análisis de confiabilidad, aunque otras proporcionan estadísticas descriptivas básicas. La Teoría de Respuesta al Ítem se realiza utilizando análisis factorial de correlaciones tetracóricas y policóricas. Las funciones para analizar datos en múltiples niveles incluyen estadísticas dentro y entre grupos, incluidas correlaciones y análisis factorial. Se incluyen funciones para simular y probar elementos particulares y estructuras de prueba. Varias funciones sirven como interfaz útil para el modelado de ecuaciones estructurales. Se crean visualizaciones gráficas de diagramas de ruta, análisis factorial y modelos de ecuaciones estructurales utilizando gráficos básicos. Algunas de las funciones están escritas para respaldar un libro sobre teoría psicométrica, así como publicaciones sobre investigación de la personalidad esto según Revelle (2021).

4.1.2 Resultados

Con la ayuda del paquete Psych 2.1.9 se obtuvo el resultado de las correlaciones entre las variables vigor, dedicación, absorción y compromiso para lo cual se obtuvo como resultado el mapa de correlaciones entre variables como se observa en la figura 4.1., en la cual entre más intenso sea el color del cuadro y más cercano al número uno sean los valores mayor será la correlación entre estas variables la cuales serán analizadas de mayor a menor iniciando con vigor y compromiso con un valor de 0.83, dedicación y compromiso con un valor de 0.78, absorción y compromiso con valor de 0.71, dedicación y vigor con valor de 0.57, absorción y vigor con valor de 0.52 y dedicación y absorción con valor de 0.5. Con dichos valores se puede observar que la correlación entre estas variables está por encima de la media.

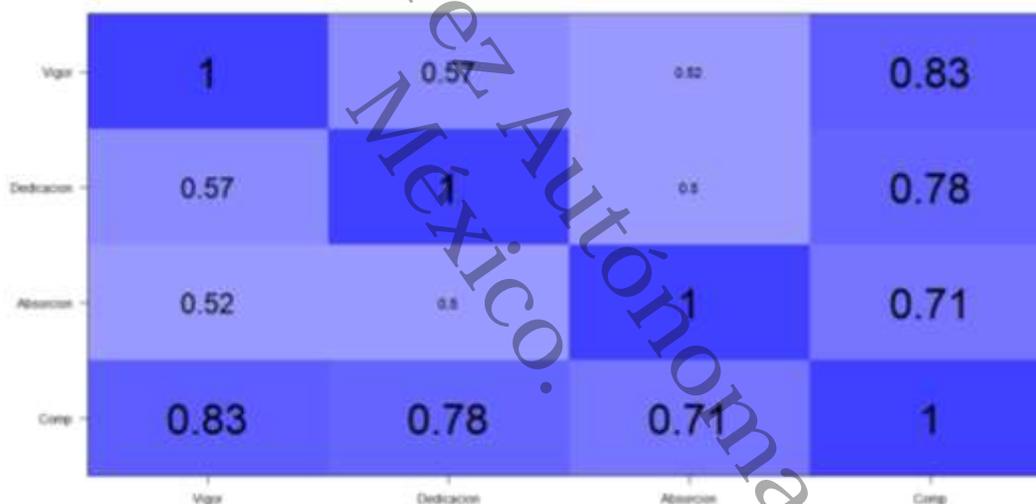


Figura 4. 1 Mapa de corraciones entre variables

Fuente: Elaboración Propia

4.2 Técnica de minería de datos: Clusterización

Para Hernández et.al. (2004), dentro de la minería de datos hemos de distinguir tipos de tareas, cada una de las cuales puede considerarse como un tipo de problema a ser resuelto por un

algoritmo de minería de datos. Esto significa que cada tarea tiene sus propios requisitos, y que el tipo de información obtenida con una tarea puede diferir mucho de la obtenida con otra.

Las distintas tareas pueden ser predictivas o descriptivas. Entre las tareas predictivas encontramos la clasificación y la regresión, mientras que el agrupamiento (clustering), las reglas de asociación, las reglas de asociación secuenciales y las correlaciones son tareas descriptivas.

En este caso se utilizó la tarea de minería de datos de agrupamiento (clustering) la cual es una tarea descriptiva por excelencia y consiste en obtener grupos "naturales" a partir de los datos. Hablamos de grupos y no de clases, porque, a diferencia de la clasificación, en lugar de analizar datos etiquetados con una clase, los analiza para generar esta etiqueta. Los datos son agrupados basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizando la similitud entre los distintos grupos. Es decir, se forman grupos tales que los objetos de un mismo grupo son muy similares entre sí y, al mismo tiempo, son muy diferentes a los objetos de otro grupo. Al agrupamiento también se le suele llamar segmentación, ya que parte o segmenta los datos en grupos que pueden ser o no disjuntos. El agrupamiento está muy relacionado con la sumari|zacion, que algunos autores consideran una tarea en sí misma, en la que cada grupo formado se considera como un resumen de los elementos que lo forman para así describir de una manera concisa los datos (Hernández et.al., 2004).

4.2.1 Algoritmo K means

El algoritmo K medias (del inglés Kmeans) es un algoritmo de clasificación no supervisada (clusterización) que agrupa objetos en k grupos basándose en sus características esto según Hernández et.al. (2004). Para lo cual el agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo o cluster. Se suele usar la distancia cuadrática.

El algoritmo consta de tres pasos:

- Inicialización: una vez escogido el número de grupos, k, se establecen k centroides en el espacio de los datos, por ejemplo, escogiéndolos aleatoriamente.

- Asignación objetos a los centroides: cada objeto de los datos es asignado a su centroide más cercano.
- Actualización centroides: se actualiza la posición del centroide de cada grupo tomando como nuevo centroide la posición del promedio de los objetos pertenecientes a dicho grupo.

Se repiten los pasos 2 y 3 hasta que los centroides no se mueven, o se mueven por debajo de una distancia umbral en cada paso.

El algoritmo k-means resuelve un problema de optimización, siendo la función a optimizar (minimizar) la suma de las distancias cuadráticas de cada objeto al centroide de su cluster.

Los objetos se representan con vectores reales de d dimensiones (x_1, x_2, \dots, x_n) y el algoritmo k-means construye k grupos donde se minimiza la suma de distancias de los objetos, dentro de cada grupo $S = \{S_1, S_2, \dots, S_k\}$ a su centroide. El problema se puede formular de la siguiente forma:

$$\min_{\mathbf{S}} E(\boldsymbol{\mu}_i) = \min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2 \quad (1)$$

donde \mathbf{S} es el conjunto de datos cuyos elementos son los objetos \mathbf{x}_j representados por vectores, donde cada uno de sus elementos representa una característica o atributo. Tendremos k grupos o clusters con su correspondiente centroide $\boldsymbol{\mu}_i$.

En cada actualización de los centroides, desde el punto de vista matemático, imponemos la condición necesaria de extremo a la función $E(\boldsymbol{\mu}_i)$ que, para la función cuadrática (1) es

$$\frac{\partial E}{\partial \boldsymbol{\mu}_i} = 0 \implies \boldsymbol{\mu}_i^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{\mathbf{x}_j \in S_i^{(t)}} \mathbf{x}_j$$

y se toma el promedio de los elementos de cada grupo como nuevo centroide.

Las principales ventajas del método k-means son que es un método sencillo y rápido. Pero es necesario decidir el valor de k y el resultado final depende de la inicialización de los centroides.

En principio no converge al mínimo global sino a un mínimo local.

4.2.2 Resultados

Una vez aplicada la tarea de la minería de datos de clusterización y en algoritmo k-means por medio del software WEKA 3.8.4 se obtuvo el siguiente resultado de la Agrupación la cual consta de tres clusters a como se puede apreciar en la Figura 4.1.

Attribute	Cluster#			
	Full Data (25.0)	0 (12.0)	1 (9.0)	2 (4.0)
SEXO	M	M	M	M
Vigor	3.88	3.25	4.7778	3.75
Dedicacion	4.112	3.3667	5.0444	4.25
Absorcion	4.616	4.225	5.4111	4
Compromiso	4.252	3.6833	5.1222	4
Calificación Final	7.16	7.5833	7.5556	5

Figura 4. 2 Clusters resultantes mediante el algoritmo K-means
Fuente: Elaboración propia.

4.2.3 Evaluación e interpretación de los resultados

El cluster cero agrupa 12 estudiantes, con una calificación final promedio de 7.16 y los cuales en su mayoría son hombres, sin embargo, de los tres grupos es el que tiene el mayor número de estudiantes y los valores más bajos de vigor, dedicación y compromiso, de 3.25, 3.36 y 3.68 respectivamente. Mientras que la absorción la tiene en un valor medio de 4.22 y la calificación final en el valor más alto de los tres grupos con valor de 7.58.

El cluster uno incluye 9 estudiantes en su mayoría hombres, con los valores más altos de vigor, dedicación, absorción y compromiso, con los valores 4.77, 5.04, 5.41 y 5.12 respectivamente. Siendo estos los valores más altos de los tres cluster, no obstante, en cuanto a calificación final son el grupo con el valor medio de 7.55.

Por último, el cluster dos, cuenta con 4 estudiantes en su mayoría hombres con el valor más bajo de calificación final el cual no es aprobatorio, más sin embargo en comparación con los otros dos

clustes estos cuentan con los valores medios de vigor, dedicación y compromiso, 3.75, 4.25 y 4 respectivamente. En este cluster existen los dos valores más bajos de los tres grupos los cuales la absorción y la calificación final con valores de 4 y 5 respectivamente.

4.3 Técnica de minería de datos: Reglas de asociación

El aprendizaje de reglas de asociación se basa en su confianza y cobertura. Los algoritmos de aprendizaje trabajan en la búsqueda de reglas que cumplan unos requisitos mínimos en estas medidas, el aprendizaje de reglas de asociación se divide normalmente en dos fases: la extracción de los conjuntos de items que cumplan con la cobertura requerida desde los datos, y la generación de las reglas a partir de estos conjuntos. Las reglas de asociación son una técnica de clasificación de objetos, basándose de un conjunto de R reglas del tipo “Si... Entonces...”, la primera parte de la regla es el antecedente y la segunda el consecuente de la primera. Para conocer la calidad de la regla se usa el soporte y la cobertura. La cobertura de una regla es el número de instancias que la regla predice correctamente y la confianza mide el porcentaje de veces que la regla se cumple cuando se puede aplicar (Hernández et.al., 2004).

4.3.1 Algoritmo Apriori

Un algoritmo de aprendizaje de reglas de asociación muy simple y popular es el algoritmo Apriori. El funcionamiento de este algoritmo se basa en la búsqueda de los conjuntos de items con determinada cobertura. Para ello, en primer lugar, se construyen simplemente los conjuntos formados por sólo un ítem que superan la cobertura mínima. Este conjunto de conjuntos se utiliza para construir el conjunto de conjuntos de dos items, y así sucesivamente hasta que se llegue a un tamaño en el cual no existan conjuntos de ítems con la cobertura requerida. La selección de candidatos de este algoritmo consiste en formar, dado un conjunto de items de tamaño i , los posibles candidatos de tamaño $i+1$. Si se dispone de un grupo Z formado por conjuntos de items de tamaño i , para la selección de candidatos, se toma cada pareja $\{x,y\}$ que formen un conjunto de items de tamaño $i+1$, donde x y y son conjuntos de items de Z , en la Figura 4.2 se puede observar el algoritmo a priori (Hernández et. al., 2004).

```

ALGORITMO Apriori(D: datos, MinC: cobertura minima)
i=0
Rellena_Item(Ci) //Incluye en C0 todos los items de tamaño
1
MIENTRAS Ci ≠ ∅
  PARA CADA x=elemento de Ci
    SI Cobertura(x) ≥ MinC ENTONCES Li = Li ∪ x
  FINPARA
  Ci+1=Selecciona_Candidatos (Li)
  i=i+1;
FIN MIENTRAS
RETORNA C
FIN ALGORITMO

```

Figura 4. 3 Algoritmo de búsqueda de conjuntos de items (Apriori).
Fuente: Hernández et. al., (2004), Pág. 241.

4.3.2 Resultados

A continuación, se describen y explican las reglas consideradas como más relevantes obtenidas de usando el algoritmo a priori y usando el lenguaje R para cada una de ellas específica un soporte del 0.15 y una confianza del 0.9 en el conjunto de datos:

Regla 1:

SI Eval=8 **ENTONCES** Classcraft=10

La regla 1 indica que si la calificación en la evaluación es igual a ocho entonces la calificación en el Classcraft será de diez.

Regla 2:

SI Dedicación=3 **Y** Compromiso=4 **ENTONCES** Classcraft=10

La regla 2 indica que si la dedicación es igual a tres y el compromiso a cuatro entonces la calificación en el Classcraft será de diez.

Regla 3:

SI Vigor=4 **Y** Prom=7 **5** **ENTONCES** Comp=4

La regla 3 indica que si el vigor es tres y el promedio es 7 entonces el valor del compromiso del estudiante será de cuatro.

Regla 4:

SI Absorcion=4 **Y** Prom=7 **ENTONCES** Comp=4

La regla 4 indica que, si los valores de la absorción son de cuatro y el promedio de siete, entonces el valor del compromiso del estudiante será de cuatro.

Regla 5:

SI Test=8 ENTONCES Classcraft=10

La regla 5 indica que si la calificación del test es de ocho entonces la calificación en el classcraft será de diez.

Regla 6:

SI Final=6 ENTONCES Classcraft=10

La regla 6 indica que si la calificación final es de seis entonces la calificación en el classcraft será de diez.

Regla 7:

SI Final=5 ENTONCES Comp=4

La regla 7 indica que si la calificación final es de cinco entonces el valor del compromiso es de cuatro.

Regla 8:

SI Eval=8 Y Dedicación=3 ENTONCES Classcraft=10

La regla 8 indica que si la evaluación es de ocho y la dedicación de tres entonces la calificación en el classcraft será de diez.

Regla 9:

SI Eval=8 Y Prom=8 ENTONCES Classcraft=10

La regla 9 indica que se la evaluación es de ocho y el promedio de ocho entonces la calificación en el classcraft será de diez.

Regla 10:

SI Dedicacion=3 Y Prom=7 ENTONCES Classcraft=10

La regla 10 indica que si la dedicación es de tres y el promedio de siete entonces la calificación en el classcraft será de diez.

Una vez descritas todas las reglas se puede decir que en su mayoría las variables relacionadas al compromiso estudiantil se encuentran altamente relacionadas con las variables que miden las calificaciones, entre todas estas reglas antes mencionadas una de las más destacadas es aquella que relaciona a los estudiantes que obtuvieron calificación final con un valor de cinco, por consiguiente, su valor en compromiso es de cuatro. Así mismo, otra de las reglas destacadas es la que relaciona a los estudiantes con cuatro como valor de absorción y siete en promedio, pues tendrán cuatro en compromiso. Por último, la regla donde cuatro es el valor de vigor y promedio con valor de siete, por consiguiente, el cuatro será el valor de compromiso.

México.

Universidad Juárez Autónoma de Tabasco.

Capítulo V. Conclusiones y trabajos futuros

En la presente tesis se ha desarrollado una investigación sobre el compromiso estudiantil en la materia de programación empleando el instrumento UWES-S, y con la ayuda de las técnicas de la minería de datos dar respuesta a la pregunta: ¿Qué técnicas de minería de datos permitirían conocer los patrones comportamiento de los estudiantes en la materia de Programación?. Para la cual la respuesta sería las técnicas de correlaciones, clustes y reglas de asociación. Para dar respuesta a esta pregunta se emplearon técnicas descriptivas de minería de datos con las cuales se obtuvieron modelos que permitieron dar respuesta a la interrogante. Uno de los modelos de representación visual obtenido mediante la técnica de correlaciones y las variables encargadas de medir el compromiso de los estudiantes, en el cual se observó una correlación positiva y que estas variables están altamente correlacionadas ya que se encuentran muy por encima de la media a lo cual son valores altamente aceptables para nuestro estudio, esto tomando en cuenta los valores de la formula estándar para medir la correlación r según Hernández et. al. (2004), el cual es un valor real comprendido entre -1 y 1 . Si r es 1 las variables están perfectamente correlacionadas, mientras que si es 0 no hay correlación.

Así mismo, el modelo de clusters el cual dividió toda la población de estudio en tres grupos con la mayor similitud entre ellos y deferencias con los otros grupos, lo cual es la característica principal de esta técnica, y con esto se logro enfocarse en las similitudes y diferencias mas relevantes de nuestro grupo de estudio y así poder conocer mejor su comportamiento y algunas causas de este.

Por último, para conocer un poco más el comportamiento de los estudiantes nos ayudó el modelo de reglas de asociación el cual consto de diez reglas con un soporte del 0.15 y una confianza del 0.9 en el conjunto de datos para los cuales la cobertura de una regla es el número de instancias que la regla predice correctamente y la confianza mide el porcentaje de veces que la regla se cumple cuando se puede aplicar esto según Hernández et. al. (2004). Dichas reglas ayudaron a

conocer más la relación entre las variables de estudio debido a su naturaleza de “SI... ENTONCES” y así conocer como una variable se relacionaba con otra.

Estos modelos obtenidos de los datos recolectados de los estudiantes ayudaron a conocer algunos de los factores que afectan el rendimiento académico de los estudiantes en programación ya que la calificación final obtenida en la materia se encuentra por debajo del promedio general de los estudiantes, los cuales en el inicio la materia tienen un compromiso muy bueno, mas sin embargo el comportamiento de las calificaciones de los estudiantes en el test inicial y el examen final mantienen un comportamiento muy parecido en los resultados obtenidos.

Entre las principales líneas de investigación que pudieran continuarse de este proyecto se encuentran la aplicación de otras técnicas de minería de datos descriptivas como pueden ser los árboles de decisión, redes neuronales, inducción de reglas, los cuales ayudarían a conocer el comportamiento de los datos y así conocer mejor a los estudiantes, también se podría aplicar en otras materias que se encuentren relacionadas con la programación y así ver como estas influyen en el rendimiento académico de los estudiantes. Todo esto para que posteriormente se pudieran aplicar las técnicas predictivas para estimar valores futuros o desconocidos con las técnicas de clasificación y regresión, para disminuir los bajos indicadores identificados con las técnicas descriptivas.

Referencias

- Aspeé, J. E., González, J. A., & Cavieres-Fernández, E. A. (2018). El compromiso estudiantil en educación superior como agencia compleja. *Formación universitaria*, 11(4), 95-108.
- Chacón, J. A., Fernández, E. C., & Campos, J. G. (2018). El compromiso estudiantil ciudadano y su vínculo con el éxito educativo en la educación superior. *Opción: Revista de Ciencias Humanas y Sociales*, (86), 393-422.
- Díaz, C., Modelo Conceptual para la Deserción Estudiantil Universitaria Chilena, doi: 10.4067/S0718-07052008000200004, *Estudios Pedagógicos*, 34(2), 65-86 (2008)
- KAHU, E. 2013. "Framing student engagement in higher education". *Studies in Higher Education*, 38: 758-773.
- Acosta, J; La Red, D. y Primorac, C. (2018). Determinación de perfiles de rendimiento académico en la UNNE con Minería de Datos Educacional. *XX Workshop de Investigadores en Ciencias de la Computación*, 1,078-1,082.
- Artunduaga Murillo, M. (2008). VARIABLES QUE INFLUYEN EN EL RENDIMIENTO ACADÉMICO EN LA UNIVERSIDAD. *Departamento MIDE (Métodos de Investigación y Diagnóstico en Educación)*.
- Barragán, S.P., González, L. (2017). Hallazgos sobre la educación en Bogotá con base en la encuesta Multipropósito 2014. *Sophia*, 13 (2), 13-29.
- Campo-Ávila, J.; Ramos-Jiménez, G.; Morales-Bueno, R. y Baena-García, M. (2017). Minería de datos educativos para la predicción personalizada del rendimiento académico. *Conferencia Internacional de Procesamiento de la Información*.

Coelho, F. y Zita, A. (2020). En: *Significados.com*. Recuperado de

<https://www.significados.com/script/>

Fullana, N. J. (1998). LA BÚSQUEDA DE FACTORES PROTECTORES DEL FRACASO ESCOLAR EN NIÑOS EN SITUACION DE RIESGO MEDIANTE UN ESTUDIO DE CASO. *Revista de Investigación Educativa*, 47-70

Free Software Foundation, Inc (2007). LICENCIA PÚBLICA GENERAL GNU. Recuperado el día 30 de enero del 2022 del sitio: <https://www.gnu.org/licenses/gpl-3.0.html>

Hernández Orallo, J.; Ramírez Quintana, M. J. y Ferri Ramírez, C. (2004). Introducción a la minería de datos. Pearson Educación. S.A., Madrid.

Hernández-Sampieri, R., Fernández, C., & Batista, P. (2010). Metodología de la Investigación. 5ta. Edición. México: Editorial Mc Graw Hill. Hernández, C.

Hidalgo Cajó, B. G. (2018). Minería de datos en los Sistemas de gestión de Aprendizaje en la Educación Universitaria. *Campus Virtuales*, 7(2), 115-128.

Jiménez Galindo, A. y Álvarez Garcia, H. (2010). Minería de Datos en la Educación. Inteligencia en Redes de Comunicación.

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.

Pérez-Gutiérrez, B. (2019). Comparación de técnicas de minería de datos para identificar indicios de deserción estudiantil, a partir del desempeño académico. *Revista UIS Ingenierías*, 19(1), 93-204.

Revelle W (2021). *psych: Procedimientos para la Investigación Psicológica, Psicométrica y de la Personalidad* . Universidad del Noroeste, Evanston, Illinois. Paquete R versión 2.1.9, <https://CRAN.R-project.org/package=psych> .

R-Project (2020). ¿Qué es la R?. La fundación. Recuperado de <https://www.r-project.org/about.html>

Schaufeli, W.; Bakker, A. UWES – Utrecht Work Engagement Scale Preliminary Manual. *Revista Psicogente*. Vol. 9. Núm. 1. (2003)

Wakaito (2020). WEKA. Recuperado de <https://www.cs.waikato.ac.nz/ml/weka/#>

Glosario

A

API: Application Programming Interface.

D

DM: Data Mining.

G

GPL: General Public License

I

IBM: International Business Machines

K

KDD: Knowledge Discovery in Databases

S

SPSS: Statistic Program for Social Sciences

W

WEKA: Waikato Environment for Knowledge Analysis

Universidad Juárez Autónoma de Tabasco.
México.

A N E X O S

Anexo A. Test aplicado a los estudiantes

Test

Test diagnóstico para los alumnos próximos a cursar la materia de Programación 1

*Obligatorio

1. Correo electrónico *
-

Datos Generales

2. 1. Sexo *

Marca solo un óvalo.

- Hombre
 Mujer

3. 2. Estado civil *

Marca solo un óvalo.

- Soltero (a)
 Casado (a)
 Separado (a) / divorciado (a)
 Unión libre

4. 3. ¿Cuál es su situación laboral actual? * *Marca solo un óvalo.*

- Trabaja
- No trabaja

5. 4. ¿Cómo financia sus estudios? * *Marca solo un óvalo.*

- Beca de la universidad
- Trabajo propio
- Ayuda de padres u otras personas Otros: _____
- _____

Encuesta de Bienestar en el Contexto Académico

Las siguientes preguntas se refieren a los sentimientos de las personas en el trabajo. Por favor, lea cuidadosamente cada pregunta y decida si se ha sentido de esta forma. Si nunca se ha sentido así conteste '0' (cero), y en caso contrario indique cuántas veces se ha sentido así teniendo en cuenta el número que aparece en la siguiente escala de respuesta (de 1 a 6).

Valores de 1 a 6:

Nunca	Casi nunca	Algunas veces	Regularmente	Bastante veces	Casi siempre	Siempre
0	1	2	3	4	5	6
Ninguna vez	Pocas veces al año	Una vez al mes o menos	Pocas veces al mes	Una vez por semana	Pocas veces por semana	Todos los días

6. 1. En mis tareas como estudiante no paro incluso si no me encuentro bien: *

Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

7. 2. Mis estudios me inspiran cosas nuevas: *

Marca solo un óvalo.

	0	1	2	3	4	5	6
Nunca	<input type="radio"/>	Siempre					

8. 3. Me “dejo llevar” cuando realizo mis tareas como estudiante: *

Marca solo un óvalo.

	0	1	2	3	4	5	6
Nunca	<input type="radio"/>	Siempre					

9. 4. Es difícil para mí separarme de mis estudios: *

Marca solo un óvalo.

	0	1	2	3	4	5	6
Nunca	<input type="radio"/>	Siempre					

10. 5. Puedo seguir estudiando durante largos períodos de tiempo: *

Marca solo un óvalo.

	0	1	2	3	4	5	6
Nunca	<input type="radio"/>	Siempre					

11. 6. Estoy entusiasmado con mi carrera: *

Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

12. 7. Cuando me levanto por la mañana me apetece ir a clase o estudiar: *

Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

13. 8. Soy muy “resistente” para afrontar mis tareas como estudiante:

*

Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

14. 9. Estoy orgulloso de hacer esta carrera: *

Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

15. 10. Estoy inmerso en mis estudios: *

Marca solo un óvalo.

		0	1	2	3	4	5	6
Nunca	<input type="radio"/>	Siempre						

16. 11. Creo que mi carrera tiene significado: *

Marca solo un óvalo.

		0	1	2	3	4	5	6
Nunca	<input type="radio"/>	Siempre						

17. 12. Me siento fuerte y vigoroso cuando estoy estudiando o voy a las clases: *

Marca solo un óvalo.

		0	1	2	3	4	5	6
Nunca	<input type="radio"/>	Siempre						

18. 13. Mis tareas como estudiante me hacen sentir lleno de energía: *

Marca solo un óvalo.

		0	1	2	3	4	5	6
Nunca	<input type="radio"/>	Siempre						

19. 14. Olvido todo lo que pasa alrededor de mí cuando estoy abstraído con mis estudios: * Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

20. 15. Mi carrera es retadora para mí: *

Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

21. 16. Soy feliz cuando estoy haciendo tareas relacionadas con mis estudios: *

Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

22. 17. El tiempo “pasa volando” cuando realizo mis tareas como estudiante: *

Marca solo un óvalo.

0 1 2 3 4 5 6

Nunca Siempre

23. 1. Lenguaje artificial que puede ser usado para controlar el comportamiento de una maquina * *Marca solo un óvalo.*

- Algoritmo
- Lenguaje de programación
- Diagrama de flujo
- Pseudocódigo

24. 2. Conjunto de pasos ordenados y finitos que permiten resolver un problema *

Marca solo un óvalo.

- Algoritmo
- Lenguaje de programación
- Diagrama de flujo
- Pseudocódigo

25. 3. Es la representación gráfica de un algoritmo *

Marca solo un óvalo.

- Algoritmo
- Lenguaje de programación
- Diagrama de flujo
- Pseudocódigo

26. 4. Es la combinación del lenguaje natural (español, inglés o cualquier otro idioma), símbolos y términos utilizados dentro de la programación *

Marca solo un óvalo.

- Algoritmo
-
-
-

Lenguaje de programación

Diagrama de flujo

Pseudocódigo

27. 5. Característica de un algoritmo que se refiere a que debe tener un punto particular de inicio y fin * *Marca solo un óvalo.*

- Finito
- Preciso
- Definido
- General

28. 6. Característica de un algoritmo que se refiere a que debe tener un orden entre los pasos. * *Marca solo un óvalo.*

- Finito
- Preciso
- Definido
- General

29. 7. Característica de un algoritmo que se refiere a que no debe ser ambiguo *

Marca solo un óvalo.

- Finito
- Preciso
- Definido
- General

30. 8. Característica de un algoritmo que se refiere a que debe tolerar cambios que se puedan presentar en la definición del problema * *Marca solo un óvalo.*

- Finito
- Preciso
- Definido
- General

31. 9. Con ellos podemos dar nombre a variables, constantes, tipos de dato, nombres de funciones o procedimientos, etcétera. * *Marca solo un óvalo.*

- Identificadores
- Tipos enteros
- Tipos reales o de coma flotante
- Datos lógicos o booleanos

32. 10. Son aquellos números que no tienen fracciones o decimales, pueden ser negativos o positivos * *Marca solo un óvalo.*

- Identificadores
- Tipos enteros
- Tipos reales o de coma flotante
- Datos lógicos o booleanos

33. 11. Contienen una coma (un punto) decimal, tal como 3.1416, pueden ser positivos y negativos formando el subconjunto de los números reales *

Marca solo un óvalo.

- Identificadores
-
-
-

Tipos enteros

Tipos reales o de coma flotante

Datos lógicos o booleanos

34. 12. Sólo pueden tomar uno de dos valores: verdadero (true) o falso (false) *

Marca solo un óvalo.

- Identificadores
- Tipos enteros
- Tipos reales o de coma flotante Datos lógicos o booleanos
-

35. 13. Dato cuyo valor puede cambiar durante el desarrollo del algoritmo o ejecución del programa * *Marca solo un óvalo.*

- Identificadores
- Tipos enteros
- Constante
- Variable

36. 14. La siguiente afirmación "una variable que es de un tipo sólo puede tomar valores que correspondan a ese tipo" es:

Selecciona todas las opciones que correspondan.

- Verdadera
- Falsa

37. 15. Dato que permanece sin cambio durante el desarrollo del algoritmo o durante la ejecución del programa, es decir valores fijos que no pueden ser alterados por el usuario *

Marca solo un óvalo.

- Identificadores
- Tipos enteros
- Constante Variable
-

38. 16. Símbolos que permiten relacionar dos datos en una expresión y evaluar el resultado * *Marca solo un óvalo.*

- Identificador
- Operadores Constante
- Variable
-

39. 17. Operadores que pueden ser utilizados con tipos enteros o reales y sirven para realizar operaciones aritméticas básicas * *Marca solo un óvalo.*

- Aritméticos.
- Relacionales.
- Lógicos.
- Asignación.

40. 18. Permite evaluar una expresión y asignar el resultado en una variable *

Marca solo un óvalo.

- Aritméticos.
- Relacionales.
-
-

Lógicos.

Asignación.

41. 19. ¿A qué grupo de operadores pertenece el símbolo: < ? *

Marca solo un óvalo.

- Relacionales.
- Lógicos.
- Asignación.

42. 20. Operadores que trabajan con valores booleanos *

Marca solo un óvalo.

- Aritméticos.
- Relacionales.
- Lógicos.
- Asignación.

43. 21. La siguiente afirmación "todas las expresiones entre paréntesis se evalúan primero" es: *

Selecciona todas las opciones que correspondan.

- Verdadera
- Falsa

44. 22. La siguiente afirmación "las expresiones con paréntesis anidados se evalúan de dentro hacia afuera" es: *

Selecciona todas las opciones que correspondan.

Verdadera

Falsa

45. 23. Su significado se encuentra definido dentro del lenguaje de programación pues tiene un uso específico * *Marca solo un óvalo.*

- Identificador
- Constante
- Algoritmo
- Palabra reservada

46. 24. Se identifican los datos que deberán introducirse para resolver un determinado problema. Son los datos que el programa le pide al usuario *

Marca solo un óvalo.

- Entrada
- Proceso
- Salida

47. 25. Se realizan todos los pasos para resolver el problema utilizando los datos de entrada. Es decir, cálculos y controles de flujo de información. * *Marca solo un óvalo.*

- Entrada
- Proceso Salida
-

48. 26. Muestra al usuario el resultado del problema. En caso de que el procedimiento sea una función, la salida sería el valor que regresa dicha función. *

Marca solo un óvalo.

- Entrada
-
-

Proceso

Salida

49. 27. Seleccione la respuesta de correcta de la siguiente expresión:

$$8 + 5 * 9 - 6 *$$

Marca solo un óvalo.

- 47
- 111
- 120

50. 28. Seleccione la respuesta de correcta de la siguiente expresión:

$$(5 > 3) *$$

Marca solo un óvalo.

- Verdadero
- Falso

51. 29. Seleccione la respuesta de correcta de la siguiente expresión:

$$(5 > 3 \ \&\& \ 8 < 4) *$$

Marca solo un óvalo.

- Verdadero
- Falso

52. 30. Seleccione la respuesta de correcta de la siguiente expresión:

$$(2 > 3 \ || \ 6 < 9) *$$

Marca solo un óvalo.

Verdadero

Falso

53. 31. Seleccione la respuesta de correcta de la siguiente expresión: !

$(4 > 2)$ *

Marca solo un óvalo.

Verdadero Falso

México.

Autónoma de Tabasco.⁶¹

32. Símbolos gráficos utilizado para indicar Inicio y final del diagrama deflujo. *

Marca solo un óvalo.

Universidad Juárez Autónoma de Tabasco.
México.

Universidad Juárez Autónoma de Tabasco.



Opción 1

Opción 2



Opción 3

Opción 4

33. Símbolos gráficos utilizado para indicar un proceso en el diagrama de flujo. *

Marca solo un óvalo.

Universidad Juárez Autónoma de Tabasco.
México.

Universidad Juárez Autónoma de Tabasco.



Opción 1

Opción 2



Opción 3

Opción 4

34. Símbolos gráficos utilizado para indicar una decisión en el diagrama de flujo. *

Marca solo un óvalo.

Universidad Juárez Autónoma de Tabasco.
México.

Universidad Juárez Autónoma de Tabasco.



Opción 1

Opción 2



Opción 3

Opción 4

35. Selecciona la opción con el orden correcto del siguiente algoritmo *

- a) Inicio
- b) Imprime "Teclea el valor de a, b, c, e, d"
- c) suma $a + b + c + d$
- d) Acepta a, b, c, d
- e) producto $a * b * c * d$
- f) $media = (a + b + c + d) / 4$
- g) Imprime producto, suma, media
- h) Declara variables: a, b, c, d, producto, suma, media
- i) Fin

Marca solo un óvalo.

- a,h,b,d,c,e,f,g,i
- a,b,c,d,e,f,g,h,i
- a,b,d,c,e,f,g,h,i
- a,h,b,c,e,f,g,d,i

Información Complementaria

58. ¿Cual es tu promedio académico general? *

Se recomienda consultar este dato en su reporte de calificaciones académico, ya que este dato debe ser preciso.

1.

¡GRACIAS!



Google no creó ni aprobó este contenido.

Google

Formularios

Universidad Juárez Autónoma de Tabasco.
México.