



ANÁLISIS DE VARIABLES DEFINIDAS EN UN MODELO DE ESTUDIANTE DE UN ENTORNO DE APRENDIZAJE ADAPTATIVO SOBRE LA PLATAFORMA MOODLE, POR MEDIO DE LA APLICACIÓN DE MINERÍA DE DATOS.

LEIDY VIVIANA OSORIO JIMÉNEZ

**UNIVERSIDAD PONTIFICIA BOLIVARIANA
ESCUELA DE INGENIERÍA
FACULTAD DE INGENIERÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN
MAESTRÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN
MEDELLÍN - COLOMBIA
2019**



ANÁLISIS DE VARIABLES DEFINIDAS EN UN MODELO DE ESTUDIANTE DE UN ENTORNO DE APRENDIZAJE ADAPTATIVO SOBRE LA PLATAFORMA MOODLE, POR MEDIO DE LA APLICACIÓN DE MINERÍA DE DATOS.

LEIDY VIVIANA OSORIO JIMÉNEZ

Trabajo de grado para optar al título de magíster en tecnologías de la información y la comunicación

Dirigido por

JEIMY VÉLEZ RAMOS

PhD. EN TECNOLOGÍAS DE LA INFORMACIÓN

Asesor

SERGIO AUGUSTO CARDONA

PhD. EN INGENIERÍAS

UNIVERSIDAD PONTIFICIA BOLIVARIANA

ESCUELA DE INGENIERÍA

FACULTAD DE INGENIERÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

MAESTRÍA EN TECNOLOGÍAS DE INFORMACIÓN Y COMUNICACIÓN

MEDELLÍN

2019

DECLARACIÓN ORIGINALIDAD

“Declaro que esta tesis (o trabajo de grado) no ha sido presentada para optar a un título, ya sea en igual forma o con variaciones, en esta o cualquier otra universidad”. Art. 82 Régimen Discente de Formación Avanzada, Universidad Pontificia Bolivariana.



FIRMA AUTOR _____

Leidy Viviana Osorio Jiménez

C.C: 1094907844 de Armenia

Medellín y Julio 2019

DEDICATORIA

A Dios, por ser mi guía en el camino.

A mi familia, por su apoyo incondicional.

AGRADECIMIENTOS

Quiero expresar mi agradecimiento a Dios, mi familia y a las personas e instituciones que me han apoyado durante mi proceso de formación en mi Maestría.

Agradezco de corazón a mi directora, PhD. Jeimy Beatriz Vélez Ramos, por su apoyo, orientación, paciencia y guía para la elaboración de este trabajo de maestría.

Al Doctor Sergio Augusto Cardona Torres, por su acompañamiento durante todo este tiempo en todo el proceso de formación y generación de resultados de este trabajo, además de su comprensión y paciencia.

Al Departamento Administrativo de Ciencia Tecnología e Innovación Colciencias por la convocatoria 569 de 2012, mediante su apoyo financiero del proyecto de investigación “Entorno de aprendizaje adaptativo basado en un modelo de estudiante multidimensional” y parte de mis estudios de maestría.

CONTENIDO

1. INTRODUCCIÓN.....	7
2. PLANTEAMIENTO DEL PROBLEMA.....	9
2.1. PROBLEMA.....	9
3. JUSTIFICACIÓN	10
4. OBJETIVOS	11
4.1. OBJETIVO GENERAL	11
4.2. OBJETIVOS ESPECÍFICOS.....	11
5. MARCO REFERENCIAL.....	12
5.1. MARCO CONTEXTUAL	12
5.2. MARCO CONCEPTUAL	12
5.2.1. COMPETENCIAS.....	12
5.2.2. SISTEMAS ADAPTATIVOS EDUCATIVOS	12
5.2.3. MODELO DE ESTUDIANTE	13
5.2.4. SISTEMAS DE ADMINISTRACIÓN DE APRENDIZAJE	13
5.2.5. MINERÍA DE DATOS.....	14
5.3. HERRAMIENTA PARA LA MINERÍA DE DATOS	16
5.4. MARCO LEGAL	16
5.5. ESTADO DEL ARTE.....	16
6. METODOLOGÍA.....	21
6.1. ELABORACIÓN DEL MARCO REFERENCIAL.....	21
6.2. ANÁLISIS DE DATOS	22
6.2.1. ETAPAS DE LA METODOLOGÍA CRISP.....	22
6.2.1.1. ENTENDIMIENTO DEL NEGOCIO.....	22
6.2.1.2. VARIABLES CONSIDERADAS PARA EL ANÁLISIS	23
6.2.1.4. ENTENDIMIENTO DE LOS DATOS	27
6.2.1.5. MODELAMIENTO	31
6.2.1.6. EVALUACIÓN.....	31
6.2.1.7. DESPLIEGUE.....	32

7.	PRESENTACIÓN Y ANÁLISIS DE RESULTADOS	33
7.1.	GENERACIÓN DE LA BASE DE DATOS PARA EL ANÁLISIS	33
7.2.	PREPARACIÓN DE LOS DATOS	33
7.2.1.	PASO 1: LIMPIEZA DE DATOS	34
7.2.2.	PASO 2: ELIMINACIÓN DE VARIABLES IRRELEVANTES	34
7.2.3.	PASO 3: VARIABLES SELECCIONADAS	37
7.3.	EJECUCIÓN DE TÉCNICAS DE MINERÍA DE DATOS	38
7.3.1.	EXPERIMENTO 1: ANÁLISIS DE PERFILES DE LOS ESTUDIANTES EN CURSOS BAJO LA METODOLOGÍA DE PROYECTOS FORMATIVOS	39
8.	RESULTADO DE LAS TÉCNICAS	49
8.1.3.	EVALUACIÓN Y ANÁLISIS	51
	VARIABLES FINALES	54
9.	CONCLUSIONES	57
10.	TRABAJOS FUTUROS	58
11.	REFERENCIAS	59
12.	ANEXOS A	63
13.	ANEXOS B	64

LISTA DE FIGURAS

Ilustración 1: Descubrimiento de Conocimiento BD (KDD), definido por (Fayyad, Piatetsky-Shapiro, and Smyth 1996)	14
Ilustración 2: Modelo de curso adaptativo (Duque, Ovalle, and Jiménez 2007).....	17
Ilustración 3: Características del modelo (González, Duque, and Ovalle 2008)	17
Ilustración 4: Modelo del Estudiante (Duque, Ovalle, and Jiménez 2007)	18
Ilustración 5: Metodología de investigación	21
Ilustración 6: Metodología CRISP – DM (Chapman, Clinton, and Keber, 2000).....	22
Ilustración 7: Variables consideradas para el análisis.(Cardona 2017)	24
Ilustración 8: Características de la categoría aspectos didácticos.	25
Ilustración 9: Base de datos	33
Ilustración 10: análisis estadístico – Género.....	34
Ilustración 11: Proporción de las variables [Elaboración Propia]	38
Ilustración 12: perfilamiento de los estudiantes WEKA	40
Ilustración 13: Importancia de las variables	¡Error! Marcador no definido.
Ilustración 14: Datos por clúster.....	41
Ilustración 15: porcentaje de datos por Clúster.....	41
Ilustración 16: Árbol de decisión factores de influencia sobre la nota definitiva	45
Ilustración 17: Resultado Prueba Mann-Whitney para el género y el rendimiento académico	51
Ilustración 18: Resultado Prueba Mann-Whitney entre autoevaluación y rendimiento académico ..	52
Ilustración 19: Resultado Categoría Evaluación Ward.....	53
Ilustración 20: Resultado Categoría Participación	53
Ilustración 21: Resultado Categoría Diseño Moodle.....	54
Ilustración 22: Resultado Categoría Coherencia con las actividades.....	54

LISTA DE TABLAS

Tabla 1: Matriz de componentes rotados	27
Tabla 2: Descripción de variables.....	30
Tabla 3: Evaluador de Atributos [Elaboración Propia].....	37
Tabla 4: Distancias entre los centros de clúster.....	40
Tabla 5: Evaluación de Resultados – Clústeres.....	42
Tabla 6: Evaluación de Resultados - Experimento 1.....	44
Tabla 7: Evaluación de resultados PCA.....	46
Tabla 8: Resultado Final.....	50
Tabla 9: Comparativo Resultados Variables finales	55

RESUMEN

La creciente demanda de formación a través de plataformas de educación virtual podría ser el camino hacia procesos de aprendizaje cada vez más significativos, aunque, para ello sea necesario que el docente cuente con información relevante para guiar el proceso de aprendizaje de forma personalizada. Los sistemas adaptativos educativos utilizan modelos de estudiantes para identificar, a partir de las características de los estudiantes, perfiles de cada uno de ellos que puedan ser utilizados para ofrecerles contenidos, navegación, etc., personalizados. Entre los retos de estos sistemas se encuentra el poder identificar las variables que mayor incidencia tienen en los procesos de aprendizaje para ser consideradas en los procesos de personalización.

Por otra parte, existen nuevos enfoques educativos, tales como el socioformativo, que promueven la formación integral a partir del desarrollo de competencia mediante metodologías enfocadas a resolver problemas del contexto promoviendo el mejoramiento continuo y el compromiso ético (Tobón, 2017). Este tipo de formación se viene soportando en plataformas de educación virtual libre; tales como Moodle. Sin embargo, estas plataformas no cuentan con componentes que permitan implementar cursos personalizados según diferentes aspectos considerados relevantes en dicho proceso de aprendizaje.

A partir de lo anterior, en este trabajo se presenta un análisis de las variables definidas en un modelo de estudiante de un componente adaptativo para la plataforma Moodle, con la aplicación de una técnica de minería sobre los datos de los alumnos. De esta manera se identifica patrones y perfilamientos en usuarios, que da un aporte al diseño de los modelos de estudiante al considerar las variables con mayor incidencia sobre los procesos de aprendizaje.

PALABRAS CLAVE:

Modelo del estudiante; minería de datos, sistemas adaptativos educativos ; Moodle; perfilamiento, socioformación

ABSTRACT

The growing demand for training through virtual education platforms could be the way towards increasingly significant learning processes, although for this the necessary sea that the teacher has relevant information to guide the learning process in a personalized way. The adaptive and educational systems for student models to identify, based on the characteristics of the students, profiles of each one that can be used to receive content, navigation, etc., personalized.

Among the challenges of these systems is the power identified the variables that have the greatest incidence in the learning processes to be considered in the personalization processes. On the other hand, there are new educational approaches, such as the Socioformat, that promote the integral formation of the development of competence through methodologies focused on solving problems in the context, promoting continuous improvement and ethical commitment (Tobón, 2017). This type of training is supported by free virtual education platforms; stories like Moodle. However, these platforms do not have the components that are transformed into this learning process.

Based on the above, this paper presents an analysis of the variables defined in a student model of an adaptive component for the Moodle platform, with the application of a mining technique on student data. In this way, patterns and profiling are identified in users, which support the design of student models when considering the variables with the greatest impact on learning processes.

KEY WORDS:

Student's model; data mining, Adaptive Courses; Moodle; profiling; socioformation.

1. Introducción

Con la evolución de las Tecnologías de la información y las comunicaciones (TIC), surgen herramientas que dan soporte a los procesos de formación y contribuyen a mejorar la educación. Estas herramientas introducen mejoras como la adaptatividad, que busca la personalización de contenidos, actividades, navegación, entre otras, para ayudar a los estudiantes a optimizar y a hacer más efectivo su proceso de aprendizaje. Sin embargo, algunos sistemas adaptativos solo tienen en cuenta aspectos como el nivel de conocimiento (Ferreira, Salcedo, Kotz, and Barrientos, 2012). Además en el trabajo de (González, Duque, and Ovalle, 2008) consideran aspectos como el desempeño, la participación, entre otros, los cuales podrían contribuir a mejorar los sistemas adaptativos y los resultados obtenidos a partir de su uso.

Por otra parte, las plataformas de aprendizaje generan una gran cantidad de información acerca de los estudiantes y la experiencia al utilizar este tipo de entornos. Algunos de los datos obtenidos son las interacciones que realiza el estudiante en las plataformas durante su participación en: charlas, consultas, glosarios, tareas, encuestas, lección, cuestionarios, fórum, taller, diario, etc. En el caso de Moodle, una de las plataformas de gestión de cursos virtuales más utilizada, aunque es posible obtener estadísticas del uso de la plataforma, no implementa funcionalidades adicionales que apoyen el proceso de aprendizaje de los participantes. No siempre los datos generados se convierten en información relevante para el docente y el estudiante, la cual podría servir de soporte en la toma de decisiones en función de rendimiento de los estudiantes (Casales, Rojas, and Paulí, 2008). Así como, en la identificación de patrones y perfilamientos de usuarios, que orienten el diseño de los modelos de estudiante en sistemas adaptativos, con la finalidad de identificar variables con mayor incidencia sobre los procesos de aprendizaje y de esta forma desarrollar servicios de personalización más efectivos.

Algunos trabajos han explotado el potencial de los datos a través de técnicas de minería que permiten obtener información para mejorar los procesos de formación. En el trabajo realizado por Jácome (2017) donde utilizó la herramienta Weka para analizar diferentes modelos didácticos y de esta manera establecer niveles de conocimiento y satisfacción de los estudiantes, entre otros aspectos. En el trabajo realizado por Martínez, Karanik, Giovannini, Báez, y Torre (2016), se utilizan técnicas de minería de clustering y reglas de asociación sobre datos generados en las asignaturas de algoritmos y estructuras de datos. En dicha investigación se consideraron variables sociales, económicas, actitudinales y culturales, las cuales fueron comparadas con las notas finales de los estudiantes, generando las variables de mayor incidencia en este estudio. Los logs generados en las plataformas virtuales también han sido analizados para obtener información relevante en procesos de formación en línea.

Hernández, Tobón and Guzmán (2015), analizaron las interacciones realizadas por estudiantes en Moodle y pudieron establecer que existe relación entre lograr un mejor desempeño académico y participar en actividades como: análisis de saberes previos, análisis de casos y coevaluación. Por otra parte, uno de los elementos esenciales en el desarrollo de tecnologías en la educación es el pedagógico. En este sentido, se vienen desarrollando nuevos enfoques y modelos educativos orientados hacia la sociedad del conocimiento, con el fin de orientar las reformas educativas y transformar los procesos de gestión del talento humano. Uno de ellos es la socioformación, un enfoque que busca el desarrollo del talento y la formación integral mediante metodologías enfocadas en resolver problemas del contexto mediante proyectos transversales, promoviendo el mejoramiento continuo y el compromiso ético (Tobón 2017).

La formación desde la socioformación se viene soportando en plataformas de educación virtual, como Moodle, como lo expresa Tobón (2017) estos sistemas han integrado el aprendizaje adaptativo, el cual se concibe como el ajuste y la personalización del proceso de formación considerando tanto los niveles de conocimiento, como la participación, la colaboración y la resolución de problemas a partir de productos concretos por parte de los estudiantes, asignando actividades complementarias o diferenciadas en función de estos aspectos. Sin embargo, la plataforma tecnológica Moodle carece de funcionalidades de análisis de la información generada, para uso en procesos de personalización.

Este análisis podría contribuir a la identificación de variables de mayor utilidad y beneficio para la formación por competencias, en entornos de aprendizaje virtual.

Con el ánimo de contribuir mejoras en las plataformas de educación virtual, este trabajo presenta los resultados de la aplicación de una técnica de minería de datos en el análisis de las variables definidas en un modelo de estudiante, así como, de otras variables incorporadas en el estudio para evaluar la experiencia de los estudiantes en el proceso de formación. De esta manera fue posible identificar patrones y perfilamientos en usuarios, que orientan el diseño de los modelos de estudiante en sistemas adaptativos educativos y contribuyen al desarrollo de procesos de formación más pertinentes y efectivos. Este trabajo se ha organizado de la siguiente manera: En el apartado número dos se presenta el planteamiento del problema, en el apartado número tres se describen la justificación del trabajo, en el apartado número cuatro se presentan los objetivos de investigación, en el apartado número cinco se describe el marco referencial, en el apartado número seis se presenta la metodología, en el apartado número siete se presenta el análisis de resultados , seguido en el apartado número ocho las conclusiones y en el número nueve se describe los trabajos futuros, finalmente, se presentan las referencias utilizadas para soportar el trabajo.

2. Planteamiento del problema

2.1. Problema

Con la evolución de las Tecnologías de la información y comunicación (TIC), surgen herramientas que dan soporte a los procesos de formación y contribuyen a mejorar la educación. Estas tecnologías, plantean mejoras como la adaptatividad, que buscan la personalización de contenidos, actividades, navegación, entre otras, para ayudar a los estudiantes a optimizar y a hacer más efectivo su proceso de aprendizaje. Sin embargo, en algunos sistemas que actualmente utilizan la adaptatividad como ELE-Tutor (Ferreira et al. 2012), solo se tienen en cuenta aspectos como el nivel de conocimiento y se desconoce otros factores importantes que podrían llegar a influir en el proceso educativo de una persona, tales como, desempeño, participación, entre otros (González and Duque 2008).

Por otra parte, las plataformas de aprendizaje generan una gran cantidad de información acerca de los estudiantes y la experiencia al utilizar este tipo de entornos, algunos de los datos son las interacciones con las actividades (charlas, consultas, glosarios, tareas, encuestas, lección, cuestionarios, fórum, taller, diario, etc.). En el caso de Moodle, una de las plataformas de gestión de cursos virtuales más utilizada, aunque es posible obtener estadísticas del uso de la plataforma, no implementa funcionalidades adicionales que apoyen la toma de decisiones sobre el proceso de aprendizaje de los participantes. No siempre los datos generados se convierten en información relevante para el docente y el estudiante, la cual podría servir de soporte en la toma de decisiones en función de rendimiento de los estudiantes (Casales, Rojas, and Paulí 2008). Así como, a la identificación de patrones y perfilamientos de usuarios, que orienten el diseño de los modelos de estudiante en sistemas adaptativos, con la finalidad de identificar variables con mayor incidencia sobre los procesos de aprendizaje y de esta forma desarrollar servicios de personalización más efectivos.

Además de la parte tecnológica, uno de los elementos esenciales es el desarrollo de tecnologías en la Educación es el pedagógico, en este sentido, los enfoques pedagógicos orientados a la formación por competencias han estado orientando reformas en los sistemas educativos en varios países. Uno de los enfoques educativos emergentes que promueven la formación de competencias en el contexto de la sociedad del conocimiento es el Socioformativo, este enfoque busca la formación integral a partir del desarrollo de competencias mediante metodologías enfocadas a resolver problemas del contexto promoviendo el mejoramiento continuo y el compromiso ético (Tobón, 2017). Este tipo de formación se viene soportando en plataformas de educación virtual libre; tales como Moodle. Sin embargo, este soporte tecnológico carece de funcionalidades de análisis de la información generada, para uso en procesos de personalización. Este análisis podría contribuir a la identificación de variables de mayor utilidad y beneficio para procesos de personalización de la formación por competencias, en entornos de aprendizaje virtual.

Esto nos lleva a plantear las siguientes preguntas de investigación:

- ¿Qué técnicas de minería de datos podrían ser utilizadas para identificar patrones de comportamientos y perfilamientos de usuarios a partir de la información generada por un sistema de información de aprendizaje adaptativo?
- ¿Qué variables de un modelo de estudiante tienen mayor incidencia en el proceso de aprendizaje?

3. Justificación

La llegada de las tecnologías de la información y comunicación (TIC) en la educación virtual y presencial ha generado gran impacto y crecimiento, existe una gran tendencia a utilizar sistemas de aprendizaje tales como los sistema para la gestión del aprendizaje (LMS) que soportan la formación; cada día estas tecnologías se vuelven una necesidad en la sociedad donde los cambios, el incremento de los conocimientos y la demanda de una educación de alta calidad son una exigencia permanente, es allí donde las TIC se convierten en una herramienta para fortalecer y mejorar el desarrollo de la educación.

Aunque los LMS son muy utilizados para soportar sistemas E-learning, mucha de la información generada por ellos, durante el uso de los estudiantes en los cursos, no es analizada para producir desarrollos tecnológicos y pedagógicos orientados a mejorar procesos de enseñanza – aprendizaje. Áreas como el modelado de usuario y los sistemas adaptativos pueden utilizar la información para realizar el análisis de los datos y de esta forma contribuir a procesos de personalización más satisfactorios y efectivos para los estudiantes. Desde el punto de vista pedagógico, el análisis de esta información podría ser utilizada para mejorar el diseño y planeación de los cursos, la retroalimentación dada a los estudiantes y los procesos de evaluación por competencias, entre otros.

Dado que los sistemas actuales de enseñanza virtual proporcionan información estática, definida inicialmente en los cursos (C. R. Morales, Soto, and Martínez 2005) y, grandes volúmenes de información sobre las interacciones de los estudiantes en los cursos, es difícil interpretar estos datos ya que se presentan generalmente en formatos Logs. Por lo anterior, es necesario un tratamiento especial para su comprensión, por ello, este proyecto pretende aplicar técnicas de minería de datos descriptivas, las cuales permitan descubrir patrones e información oculta, con el fin de dar un apoyo a los docentes en la toma de decisiones.

A partir de lo anterior, este proyecto plantea, un análisis de las variables definidas en un modelo de estudiante de un componente adaptativo, mediante la aplicación de las técnicas de minería de datos Clustering y Análisis de secuencias, sobre los datos de los alumnos que proporciona dicha plataforma. De esta manera se espera identificar patrones y perfilamientos en usuarios, que orienten el diseño de los modelos de estudiante, al considerar, las variables con mayor incidencia sobre los procesos de aprendizaje.

4. Objetivos

4.1. Objetivo general

Contribuir a la validación de un modelo de estudiante de un sistema adaptativo mediante la aplicación de una técnica de minería de datos que permita analizar las variables definidas e identificar las características de mayor incidencia en el rendimiento académico de los estudiantes.

4.2. Objetivos específicos

- Extraer la información a partir de las variables definidas en el modelo del estudiante de acuerdo a las características y necesidades de un sistema adaptativo educativo.
- Preparar los datos generados en un curso virtual por competencias para realizar el análisis de la información del modelo del estudiante.
- Ejecutar técnicas de minería de datos sobre las variables definidas en el modelo del estudiante.
- Validar las técnicas de minería de datos implementadas a partir de las variables definidas en el modelo del estudiante de un curso virtual en Moodle.
- Proponer variables para el diseño de un modelo de estudiante a partir del análisis de los resultados obtenidos de la aplicación de la técnica de minería de datos.

5. Marco referencial

5.1. Marco contextual

Este proyecto de tesis surge en el marco del proyecto “Entorno de Aprendizaje Adaptativo basado en un Modelo de Estudiante Multidimensional”, el cual fue financiado por Colciencias, la Universidad Pontificia Bolivariana (UPB) y el centro universitario CIFE (Ciencia e Innovación para la Formación y el Emprendimiento).

Este trabajo se realizó con los datos de cursos presenciales, diseñados bajo el enfoque Socioformativo para la formación por competencias y, realizó un análisis de las variables definidas en un modelo de estudiante multidimensional de un componente adaptativo, mediante la aplicación de las técnicas de minería agrupación y análisis de secuencias.

5.2. Marco conceptual

El marco conceptual es el resultado de una investigación documental sobre los elementos conceptuales que dan soporte a este trabajo. Se revisaron los ejes principales de la formación por competencias, los sistemas adaptativos y la minería de datos. A continuación se describen los aspectos relacionados con la evaluación de competencias, y los fundamentos teóricos de los sistemas adaptativos, el modelo de estudiante, y los aspectos relacionados con los sistemas de gestión de aprendizaje y la minería de datos.

5.2.1. Competencias

Desde la década de 1970 se resalta la importancia de reconocer el desempeño de los trabajadores en sus actividades y se comienza a emplear el concepto de competencias para evaluar la productividad en situaciones laborales específicas (McClelland 1973). También comienza a analizarse cómo el enfoque de competencias afecta el currículo y el aprendizaje de los estudiantes (Gibbons 1980). En la década de 1980, las competencias aparecen en el área de la gestión del talento humano, y en la década de 1990 se comienzan a aplicar en la educación como referente para mejorar la calidad de la formación, por influencia de la lingüística y la psicolingüística, la gestión del talento, la globalización y los sistemas de aseguramiento de la calidad (McClelland 1973) (Tobón 2017). En los trabajos de (Bunk 1994), (Biemans et al. 2005), (Tobón 2017), (Díaz-barriga 2006), se presenta el desarrollo histórico de las competencias en el contexto educativo.

Las definiciones de competencia integran diversas dimensiones de las personas, relacionan aspectos éticos y enfatizan en las actitudes, habilidades y conocimientos. Sin embargo, es posible identificar elementos comunes en ellas como menciona (Cardona, Vélez, and Tobón 2013) “se basan en el desempeño, están orientadas a tareas, actividades y/o problemas del contexto, se deben abordar con criterios y evidencias, y consideran los diversos saberes de la actuación humana en donde se integran escenarios de actuación y múltiples disciplinas”.

Uno de los enfoques para el desarrollo de competencias es el enfoque Socioformativo, en el cual las competencias son actuaciones integrales para resolver problemas del contexto con idoneidad, mejoramiento continuo y compromiso ético (Tobón 2017). La metodología de formación y evaluación de competencias de este enfoque está basada en proyectos formativos. Un proyecto formativo es un conjunto articulado de actividades que se van desplegando en el tiempo para resolver un problema contextualizado y contribuir a formar una o varias competencias, con base en criterios y evidencias (Tobón 2013).

5.2.2. Sistemas adaptativos educativos

La adaptación permite una experiencia individual de aprendizaje a los estudiantes, a partir de criterios predefinidos de personalización, con el propósito de mejorar el proceso de formación. Puesto que el comportamiento del sistema se adapta a una persona, este tipo de adaptación es también denominado personalización (Paramythis and Loidl-Reisinger 2004), (Boticario, Santos, and Van Rosmalen 2006).

La idea fundamental de los sistemas adaptativos es identificar las características y necesidades, para mejorar la experiencia del usuario. Lo anterior, tal como se plantea en (Vélez, 2009): “determinar qué características se tendrán en cuenta para el modelo, como se representarán dichas características, como se actualiza el modelo de usuario y que adaptación se aplicará de acuerdo al modelo del usuario y al dominio en el que se está trabajando”. (Brusilovsky 1994) define la arquitectura de los sistemas adaptativos y plantean los siguientes componentes:

- El Modelo del Usuario guarda la información del usuario que un SHA mantiene en un registro permanente.
- El Modelo del Dominio describe cómo se enlaza y estructura la información del dominio, en términos de fragmentos, páginas, conceptos, etc., los cuales pueden ser abstractos o no.
- El Modelo de la Adaptación está compuesto por reglas que definen cómo se combinan el modelo del dominio y el modelo del usuario para proveer de adaptación al sistema.
- El Motor Adaptativo efectúa la tarea de adaptación en la cual genera recursos educativos personalizados, actividades y formación según las necesidades de los estudiantes.

Esta investigación hace énfasis en el modelo de usuario, el cual se nombra modelo de estudiante puesto que es la denominación específica en el campo de la educación.

5.2.3. Modelo de estudiante

Un modelo de estudiante es una representación abstracta de un estudiante en el sistema, puede ser utilizado para: promover la reflexión en los estudiantes en su proceso de aprendizaje, mejorar la precisión del modelo, ayudar a los estudiantes a planear y monitorear su aprendizaje (Bull and McKay 2004). Así mismo, puede usarse con diversos propósitos: adaptar recursos de aprendizaje, recomendar contenidos, sugerir problemas y actividades individualizadas, ofrecer recomendaciones cuando identifique errores, generar estudios de caso, etc., de acuerdo con su nivel de competencia. En (Millán Valldeperas 2000), se destacan propósitos como: determinar si el alumno está preparado para continuar con el siguiente tema del currículum, y para elegir este tema, generar explicaciones que el alumno pueda entender (al nivel de detalle adecuado a sus conocimientos actuales), ofrecer ayudas sin que el estudiante lo solicite y, seleccionar la estrategia tutorial más apropiada dado el nivel de conocimiento actual.

La selección de las características de los estudiantes en un proceso de personalización es la actividad inicial para la construcción del modelo del estudiante. Estas características pueden ser estáticas y dinámicas. Estáticas como los datos personales, los cuales son capturados antes del proceso de aprendizaje y generalmente no se modifican durante las sesiones de aprendizaje. Las características dinámicas pueden modificarse durante la sesión de aprendizaje y están asociadas a la interacción de los estudiantes con el sistema. Esas características distintivas están asociadas a aspectos como los estilos de aprendizaje (Graf 2009)(Schiaffino, Garcia, and Amandi 2008). Así mismo, es evidente el desarrollo de herramientas informáticas orientadas a identificar características individuales de los estudiantes, para personalizar la experiencia de aprendizaje(Gouli, Papanikolaou, and Grigoriadou 2002).

5.2.4. Sistemas de administración de aprendizaje

Los LMS (Learning Management System) son los ambientes que más se utilizan en procesos de formación virtual. Estos sistemas permiten la gestión de contenidos para el aprendizaje, poseen mecanismos de comunicación entre profesor y estudiante, y dan soporte a la evaluación de los estudiantes. En la mayoría de los casos, los usuarios de un LMS pertenecen a grupos heterogéneos con diferencias individuales, en ocasiones, contrarias, de acuerdo a sus características y necesidades individuales. Es por ello que cada vez más la investigación en e-learning se centra más en el rol desempeñado por el estudiante dentro de su proceso de aprendizaje(Limongelli, Sciarrone,

and Vaste 2011), en el cual, la personalización de las experiencias de aprendizaje está relacionada directamente con la eficiencia y efectividad de su propio proceso de aprendizaje.

Diversas investigaciones reportan problemas con relación a la personalización en los LMS (Sciences and Ili 2012)(Brusilovsky 2004) (Graf and Kinshuk 2009), en donde se destaca que generalmente no se ofrecen servicios personalizados y a los estudiantes se les da acceso al mismo conjunto de recursos educativos, ignorando diferencias individuales de los alumnos tales como: nivel de competencias, intereses, necesidades, características y motivaciones.

El LMS Moodle es a nivel mundial una de las plataformas que más se utiliza dentro de las instituciones académicas para la gestión de cursos virtuales(Cardona, Vélez, and Tobón 2013), esta plataforma cuenta con la mejor funcionalidad global y para aspectos de adaptación (Graf 2009).

5.2.5. Minería de datos

Teniendo en cuenta que día tras día se generan datos que se incrementan exponencialmente en las empresas mediante los programas computacionales, surge la necesidad de obtener información relevante que genere “conocimiento” y sea un apoyo en la toma de decisiones de cada entidad; Hall, Witten y Frank definen la minería de datos como “el proceso automático y semiautomático de descubrir patrones en los datos” (Witten, Frank, and Hall 2011).

El descubrimiento de conocimiento en base de datos (KDD) es definido por (Fayyad, Piatetsky-Shapiro, and Smyth 1996) como “el proceso no trivial de identificar patrones en datos, los cuales sean válidos, novedosos, potencialmente útiles y comprensibles”.

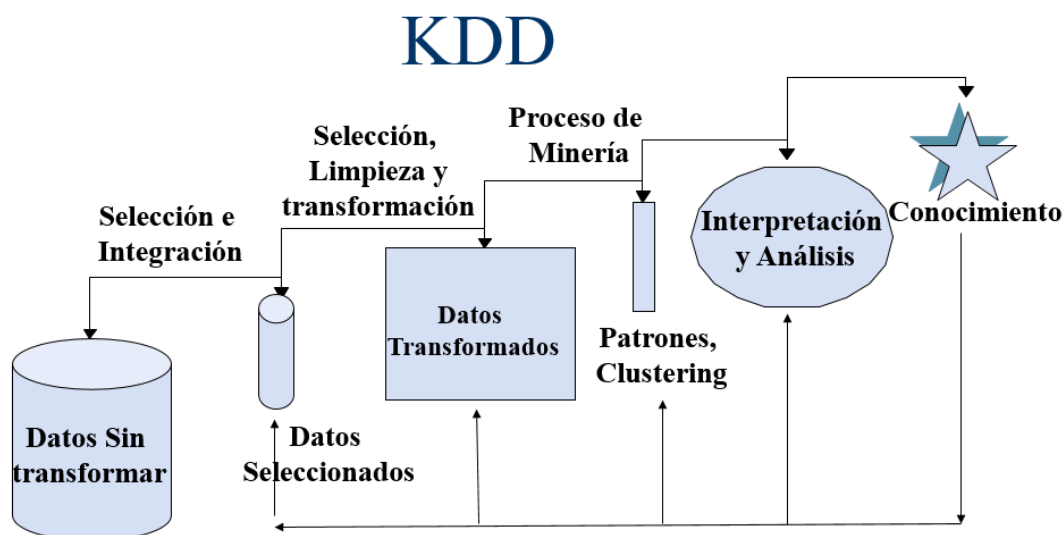


Ilustración 1: Descubrimiento de Conocimiento BD (KDD), definido por (Fayyad, Piatetsky-Shapiro, and Smyth 1996).

En los sistemas de aprendizaje virtual se generan grandes volúmenes de datos de los estudiantes, por esta razón se aplican técnicas de minería de datos las cuales permiten descubrir patrones e información oculta con el fin de obtener el perfilamiento de usuarios e información relevante y soportar la toma de decisiones. Además, estas técnicas se aplican para explorar, analizar y visualizar datos de los sistemas virtuales sobre actividades de los usuarios en la web y descubrir con mayor precisión como los estudiantes interactúan y aprenden (Huapaya et al. 2012).

A continuación, se describen las técnicas de minería de datos más utilizadas en los sistemas E-learning:

- **Agrupamiento o Clustering:** La técnica de agrupamiento es utilizada para formar clústeres o grupos de usuarios siendo similares con respecto a las características definidas, basándose en el comportamiento de navegación de un usuario. Esta técnica es utilizada para encontrar información relevante en la web, personalizar y adaptar sus contenidos basados en observaciones del sistema y por las propias valoraciones acumuladas dadas por los estudiantes (C. R. Morales, Soto, and Martínez 2005).
 - **Clustering probabilístico EM:** Este algoritmo particiona los N objetos en K particiones (K siendo un valor arbitrario) en donde un objeto ira al clúster con la media más cercana. El algoritmo asigna K centros aleatoriamente, luego asigna los objetos al centro más cercano. El centro se recalcula como la media de los puntos que tiene asignado, una vez actualizado se vuelven a reasignar los objetos al más cercano y así hasta tener convergencia (Bouckaert, Frank, and Hall, 2010).

Se utiliza el Clustering probabilístico EM el cual asigna una distribución de probabilidad a cada instancia que indica la probabilidad de que pertenezca a cada uno de los grupos. EM puede decidir cuántos clústeres crear mediante validación cruzada.

La validación cruzada realizada para determinar el número de agrupaciones:

- El número de grupos se establece en 1
 - El conjunto de entrenamiento se divide aleatoriamente en 10 pliegues.
 - EM se realiza 10 veces utilizando los 10 pliegues de la forma CV habitual.
 - El loglikelihood se promedia en todos los 10 resultados.
 - Si loglikelihood ha aumentado, la cantidad de clústeres se incrementa en 1 y el programa continúa en el paso 2 El número de pliegues se fija en 10, siempre que el número de instancias en el conjunto de entrenamiento no sea menor 10. Si este es el caso, el número de pliegues se establece igual al número de instancias.
- **Regresión:** Permite predecir el valor número de alguna variable mediante la aplicación de modelos de regresión logística (Jácome León 2017).
 - **Descubrimiento de reglas de asociación:** Las reglas de asociación permiten descubrir relaciones entre un conjunto de datos, esta técnica se ha utilizado principalmente en los e-commerce, pero en los sistemas e-learning permite descubrir relaciones o asociaciones entre diferentes páginas visitadas (C. R. Morales, Soto, and Martínez 2005).
 - **Soporte:** Este es el indicador del número de casos que cubre la regla. (Molina and García 2008)
 - **Confianza:** Indica el número de casos que predice la regla correctamente, esta se expresa como: el cociente entre el número de casos en que se cumple la regla y el número de casos en que se aplica, ya que se cumplen las premisas (Molina and García 2008).
 - **Formulas:** (Molina and García 2008)

$$\text{soporte}(A \Rightarrow B) = P(A \cap B)$$

$$\text{confianza}(A \Rightarrow B) = P(B | A) = \frac{P(A \cap B)}{P(A)}$$

- **Análisis de secuencias:** teniendo en cuenta que una secuencia es una lista de eventos ordenados (Luzcando 2017), esta técnica descubre secuencia en los conjuntos de datos mediante las técnicas de redes neuronales.

5.3. Herramienta para la minería de datos

Existen diferentes herramientas gratuitas y de pago, que permite realizar los algoritmos de la minería de datos, entre los cuales se tiene SPSS Modeler (McCormick et al. 2013) (), Enterprise Miner (McCormick et al. 2013), y algunas gratuitas como WEKA (Bouckaert, Frank, and Hall 2010) y Rapidminer (Hofmann and Klinkenberg 2013).

La herramienta que se ha seleccionado para este trabajo ha sido WEKA debido a que es de código libre.

WEKA es una herramienta libre que permite el análisis de datos, está implementada en el lenguaje de programación JAVA, fue desarrollada por un grupo de investigadores de la universidad de Waikato en nueva Zelanda. (Jácome León 2017).

5.4. Marco legal

5.5. Estado del arte

A continuación, se presentan trabajos relacionados con la definición de un modelo de estudiante, evaluación del aprendizaje, estimación de niveles de aprendizaje y técnicas de minería de datos utilizadas en la web, los cuales sirven como punto de referencia para el desarrollo de este trabajo de investigación.

Los sistemas adaptativos permite realizar adaptación de aplicaciones basados en IMS en sistemas e-learning como lo presenta(López-Moratalla 2005), este trabajo implementa una adaptación en la capa de presentación utilizando un sistema de modelado de usuario mediante estereotipos que permiten la generación de cursos adaptados pedagógicamente al estudiante. Otro aporte relevante es el de (Duque, Ovalle, and Jiménez 2007) el cual presenta un modelo de adaptación de cursos virtuales orientados por metas como se muestra en la Ilustración 2, es importante mencionarlo ya que personaliza contenido según las características de cada estudiante especialmente considerando el aspecto psicopedagógico. La adaptación de los cursos se realiza mediante la aplicación de técnicas de inteligencia artificial, al final, presenta la plataforma SICAD “Sistema Inteligente de sistemas adaptativos educativos” construida a partir de herramientas libres.

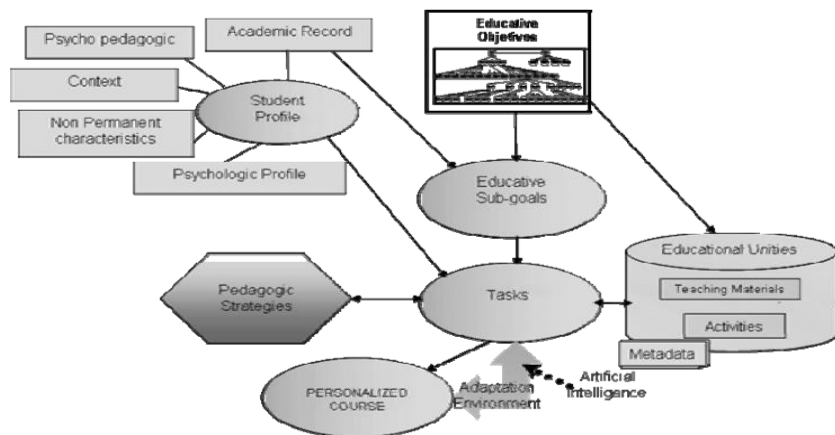


Ilustración 2: Modelo de curso adaptativo (Duque, Ovalle, and Jiménez 2007)

Otro sistema adaptativo es el recomendador **CIECoF**, se basa en los resultados de una evaluación por competencias (Salcines et al. 2008). El sistema aplica minería de datos distribuida con el objetivo de mejorar los cursos virtuales e-learning, permitiendo que los docentes compartan los resultados del análisis de los datos de los cursos. Este trabajo fue utilizado como referente para determinar las técnicas de minería a implementar en el análisis de datos en la educación virtual y para establecer la forma como se determinan patrones mediante técnicas de asociación que permite realizar recomendaciones más efectivas del contenido de los cursos.

Los sistemas adaptativos en la educación virtual definen características particulares de los alumnos denominado modelo de estudiante, algunas de las variables definidas son el nivel de comprensión de un tema, estilos de aprendizaje, gustos del usuario a la hora de aprender, características psicológicas como el estado de ánimo, las metas, aspectos del entorno socioeconómico y demográfico, entre otros. En la Ilustración 3 se presentan algunas características relevantes (González, Duque, and Ovalle 2008), así mismo, es una base sobre la forma en que éstas características se pueden inicializar, actualizar y también mantener en un curso



Ilustración 3: Características del modelo (González, Duque, and Ovalle 2008)

En el trabajo presentado por (Kritikou et al. 2008), se explica la manera en que un usuario puede navegar en un sistema de gestión de aprendizaje en la web, proponiendo una arquitectura de un sistema e-learning y describe los principales componentes de un estudiante; Gonzáles (2008) también selecciona características principales del estudiante y dependiendo de su variación adapta contenido dinámicamente a un curso, esta investigación es una referencia en la definición de características ya que la actualización de las variables definidas en la ilustración 4, se realiza con técnicas de inteligencia artificial, las cuales están analizando constantemente la interacción usuario-sistema mezclado con otras variables para hallar el valor apropiado que permita la modificación automática del modelo.

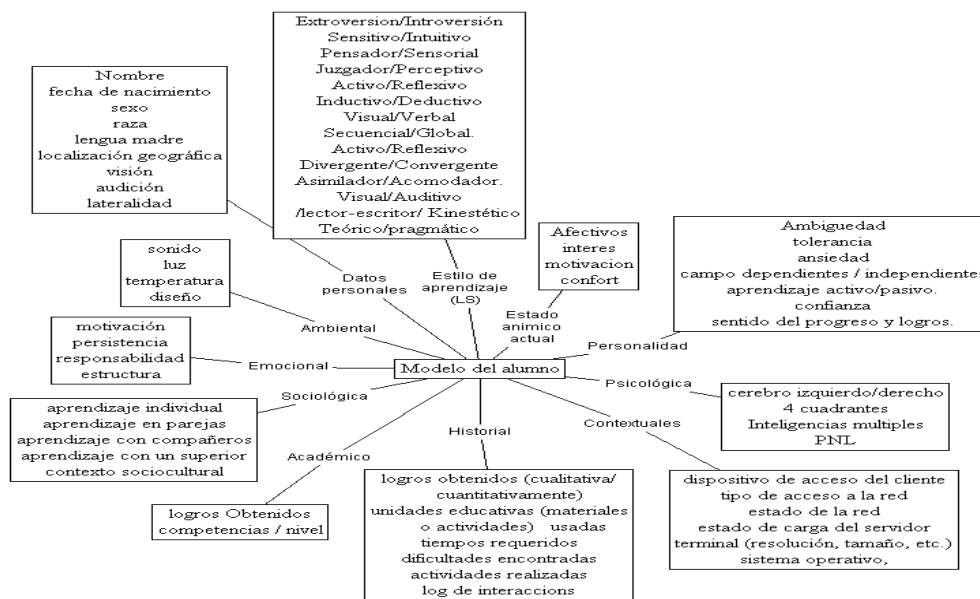


Ilustración 4: Modelo del Estudiante (Duque, Ovalle, and Jiménez 2007)

Otro aporte importante para para este trabajo es la definición de un modelo de estudiante en un entorno virtual de aprendizaje adaptativo realizado por (Veléz 2009), donde se realiza la validación de un modelo integral del estudiante, esto se refiere a la integración de los diferentes aspectos que se pueden considerar para el desarrollo del modelo de estudiante. Es una referencia para este proyecto ya que agrupa las características de los estudiantes en dimensiones, entre las cuales menciona la individual: sexo, edad, país, estado civil, conocimientos previos, etc., contexto (tecnológico y accesibilidad) y de interacción (personal y de comunidad). En las de interacción personal tenemos: tiempo de visita en el curso, tipo de material, participación, número de sesiones. Entre las de interacción de comunidad se encuentran: interacción con otros usuarios, fecha inicio interacción, usuarios con quienes interactúa, tipo de interacción, duración de interacciones, etc. También (Méndez 2009) propone de modelo del usuario y, la forma como se realiza la adaptación de contenido en un curso.

En el trabajo realizado por Betancourt, Moreno y Ovalle (2009) presentaron un modelo para la recomendación y recuperación de objetos de aprendizaje de un Sistema Tutorial Inteligente, con el fin de aumentar la interacción de docentes y estudiantes en los cursos. Las recomendaciones para los docentes se basan en las temáticas que realizan en los cursos y para los estudiantes en sus características principales (modelo de usuario). Este trabajo presenta cursos virtuales acordes a las necesidades de los alumnos, además, explica cómo genera el modelo del estudiante mediante preguntas categorizadas para definir mejor las recomendaciones generadas (Betancur, Moreno, and Ovalle C. 2009).

Existen sistemas tutoriales inteligentes como **ELE-TUTOR** que es un sistema inteligente para el español como lengua extranjera, este sistema obtiene los datos de los estudiantes que interactúan con el sistema, ofreciendo una herramienta como apoyo a los docentes para mejorar la estrategia de enseñanza de forma personalizada (Ferreira et al. 2012), es importante conocer este tipo de herramientas y como son implementadas en los sistemas de educación personalizada. (Cataldi and Lage 2010) define un modelo del estudiante en el sistema ZULMA, considerando características psicológicas, velocidad de aprendizaje, desarrollo de la tarea, habilidad de aprendizaje, nivel de habilidades, métodos de enseñanza y mapa de conocimientos, se tuvo en cuenta la forma como se construye y actualiza el modelo de estudiante.

Petropoulou, Kasimatis, Dimopoulos y Retalis proponen un sistema adaptativo sobre la plataforma Moodle, donde se presenta una herramienta de evaluación "Learning Analytics Enriched Rubric (LAE-R) (L)". Esta investigación es una guía ya que realiza el análisis de información y permite la evaluación del rendimiento de los estudiantes, enfocándose en la evaluación por rúbricas, implementando una técnica de evaluación muy notada y utilizada en la educación. Esta técnica es una mezcla de criterios y niveles de calificación de unos indicadores en un entorno de e-learning basada en LMS (Petropoulou et al. 2014).

Wu He (2013), presenta los resultados donde aplicó técnicas de minería de datos y minería de textos. Es importante resaltar el análisis que realizaron en dos conjuntos de datos diferentes y luego llevaron a cabo un análisis de correlación en profundidad de los cursos de formación, incluyendo preguntas en línea y mensajes de chat, respectivamente. Esto permitió determinar cómo se aplica la minería de datos y de texto en una gran cantidad de datos ayudando a comprender y revelar patrones en los comportamientos de aprendizaje de los estudiantes.

Para implementar análisis de datos en este proyecto es importante conocer el estado actual de la aplicación de la minería de datos a los sistemas E-learning, entre las cuales se tiene alguna técnica más sobresaliente como: La técnica de Clustering o agrupamiento la cual es utilizada para formar clústeres o grupos de usuarios basándose en el comportamiento de navegación de un estudiante. Esta técnica es utilizada para encontrar informaciones relevantes en la web, personalizar y adaptar sus contenidos basados en observaciones del sistema y por las propias valoraciones acumuladas dadas por los alumnos. En la investigación de Elena Gaudioso y Luis Talavera (Talavera and Gaudioso 2004) se analizan los datos obtenidos en un curso, utilizan el modelo probabilístico de Naive Bayes con el objetivo de descubrir patrones que reflejan comportamientos de los usuarios en el sistema.

Bogarín, Romero y Cerezo realizan un agrupamiento previo en el análisis de datos donde se generan los grupos con características similares, después de esto aplican la minería de procesos para encontrar grupos específicos en el comportamiento de los alumnos, es importante resaltar este tipo de análisis ya que es una alternativa con la cual se genera información más relevante, ya que se realiza un análisis más detallado de las variables (Bogarín Vega, Romero Morales, and Cerezo Menéndez 2015).

Por otra parte, se tienen las reglas de asociación que las cuales permiten descubrir relaciones entre un conjunto de datos, esta técnica se ha utilizado principalmente en los e-commerce, pero en los sistemas e-learning permite descubrir relaciones o asociaciones entre diferentes páginas visitadas. Osmar Zaïne (Zaïane 2001) propone la recomendación de actividades de aprendizaje en línea o atajos en un curso web, basado en los historiales de acceso y mejorar el proceso de aprendizaje en línea.

Por último, en el análisis de esta investigación tenemos el Análisis de secuencias la cual es una técnica que descubre secuencia en los conjuntos de datos, permite determinar secuencias de las páginas visitadas en una sesión de un usuario. Un trabajo que se debe resaltar es el de Wei Wang (Su et al., 2006) el cual lo realiza sobre entornos SCORM ayudando a los docentes a comprender los motivos de un buen y mal aprendizaje de los estudiantes, primero se utiliza técnicas de minería de secuencias de patrones, posteriormente agrupan a los estudiantes mediante clustering y finalmente generan un árbol de actividades personalizadas (Morales, C. R., Soto, S. V., and Martínez, 2005).

La aplicación de las técnicas de minería de datos permite descubrir patrones e información oculta para dar un apoyo en la toma de decisiones. En la investigación "*Minería De Datos Educativa En Ambientes Virtuales De Aprendizaje* (Huapaya et al. 2012)" se muestra cómo se debe realizar el proceso para el análisis de dichos datos como la limpieza de los datos, identificación de usuarios, identificación de sesión, transformación de los datos, integración y reducción de los datos cuando se aplique la técnica de minería para identificar patrones.

En otras investigaciones donde se aplica la minería de datos en entornos educativos a procesos b-learning como Monsalve, Aponte y Hoyos los cuales describen la forma en que se aplican los conceptos de la minería de datos educativos a información extraída de varias fuentes de datos de una institución educativa en procesos de b-learning (Monsalve, Aponte, and Hoyos 2013). Es importante para este proyecto ya que se explica el proceso de ingeniería de software en el desarrollo de una herramienta informática, en la cual, se aplica el modelo de análisis construido por los autores y se utilizan técnicas de minería de datos para el análisis de la información. Con base en el análisis cuantitativo y cualitativo de la información obtenida, se generaron recomendaciones para mejorar el uso de la plataforma virtual con que cuenta la institución. Además, se realiza un proceso para filtrar los datos y aplicar el proceso de análisis mediante la ejecución de diferentes algoritmos de minería para la identificar patrones y determinar un perfilamiento de los estudiantes.

Es importante resaltar estas investigaciones porque brindan bases para el análisis de datos en sistemas educativos, los cuales permitan clasificar estudiantes y así poder mejorar los estilos y formas de aprendizaje, también conocer herramientas que se estén implementado en Moodle teniendo en cuenta las características relevantes del estudiante. Las técnicas que se utilizan son regresión, clustering, patrón de la minería, reglas de asociación (Shaun, Baker, and Eds, 2008). Además de conocer los casos prácticos de la implementación de la minería web, la investigación “Minería De Uso Web: Presentación De Caso Práctico Para El Mejoramiento De La Educación, Basada En El Uso De Plataformas Virtuales” de (Curso and Alfaro 2010), se hace sobre una asignatura de programación y primer nivel de ingeniera de sistemas de la universidad de Córdoba, es importante resaltar ya que se implementa una metodología en la cual se realizan la recuperación de información mediante la minería de datos.

Otra manera de mejorar la experiencia en la educación en los sistemas E-learnign es el aprendizaje en la red, donde se utilizan entornos de aprendizaje ayudando a los estudiantes a desarrollar sus competencias mediante la colaboración, realización de actividades y compartiendo información (Sloep and Berlanga Flores 2011).

En el trabajo realizado por (Martínez et al. 2016), el cual propone utilizar técnicas de minera de datos de clustering y reglas de asociación sobre datos generados en las asignaturas de Algoritmos y estructuras de dato. En dicha investigación se consideraron variables sociales, económicas, actitudinales y culturales, las cuales fueron comparadas con las notas finales de los estudiantes, generando las variables de mayor incidencia en este estudio.

En la Tesis Doctoral de Luzcando en su propuesta de un modelo de comportamiento colectivo de estudiantes para un sistema inteligente de tutoría dirigido al entrenamiento personal, se presenta un modelo de comportamiento colectivo, el cual se propone para realizar predicciones en las acciones de los alumnos, este modelo parte de los logs generados plataforma virtual de Biotecnología de la universidad politécnica de Madrid; el objetivo de esta investigación es proveer una mejor retroalimentación en el proceso de aprendizaje a los estudiantes, partiendo del modelo que se genera a partir de los logs del estudiante del cual se realizan las predicciones (Luzcando 2017).

En el trabajo realizado por Jácome (2009) se presenta el estudio de los diferentes modelos didácticos propuestos por diferentes autores, generando un análisis de los datos mediante la herramienta Weka con aplicación de minería de datos. El análisis permitió identificar las variables de mayor relación de los estudiantes dentro de la faculta de ciencias administrativas y económicas de la universidad. Con esto pudo establecer el nivel real del conocimiento y dominio, niveles de satisfacción y el aprendizaje de cada alumno (Jácome León 2017).

6. Metodología

Se establecieron las etapas que se muestran en la ilustración 5, donde se presenta la metodología de investigación para el desarrollo de este proyecto.

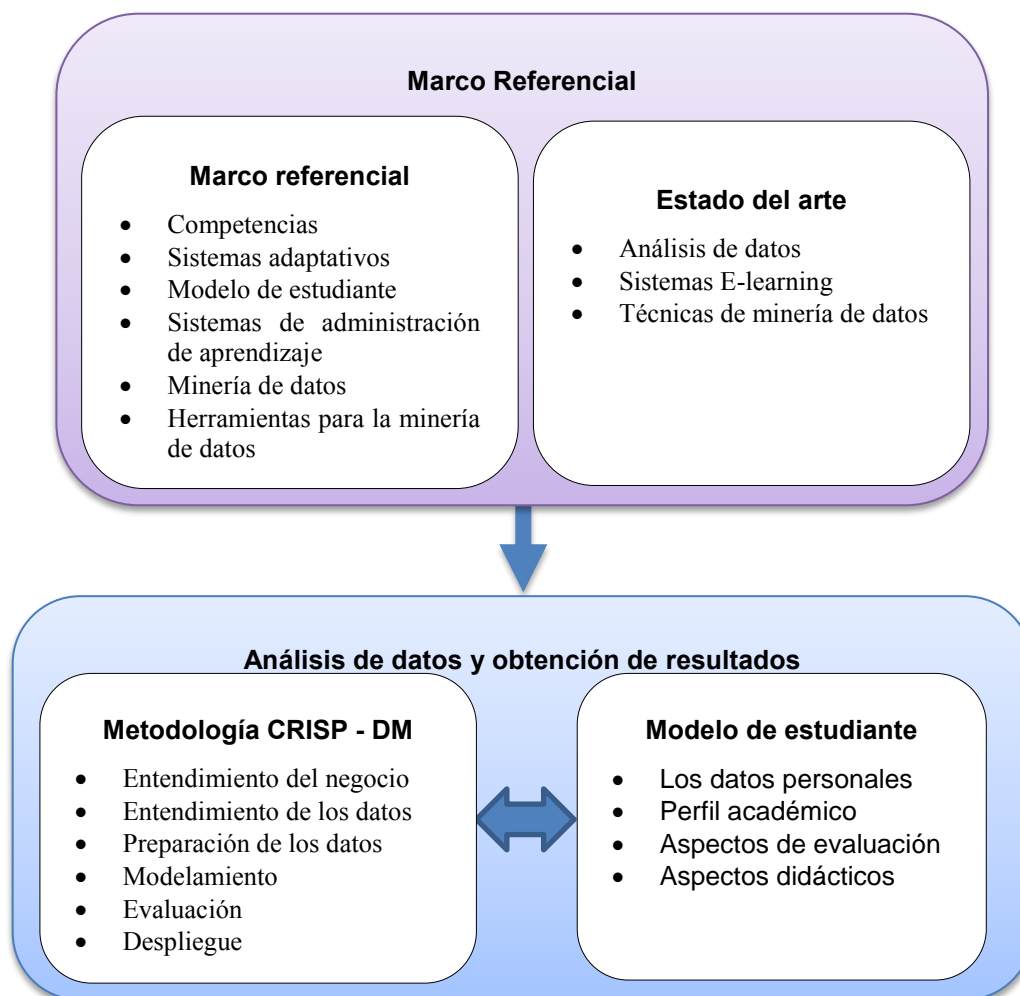


Ilustración 5: Metodología de investigación

6.1. Elaboración del marco referencial

En esta fase se presentan todos los elementos teóricos que guiarán este trabajo. Se realizó una investigación de tipo documental mediante una búsqueda sistemática de los trabajos desarrollados en las áreas objeto de estudio.

Las actividades realizadas en esta etapa fueron:

- Estructuración del marco referencial donde se realizó la investigación de trabajos relacionados con diferentes temas en Competencias, sistemas adaptativos educativos, modelo de estudiante, sistemas de administración de aprendizaje, minería de datos, herramientas para minería de datos.

- Elaboración del estado del arte en análisis de datos, sistemas E-learning, técnicas de minería de datos implementadas en sistemas adaptativos educativos.

6.2. Análisis de datos

Existen diferentes modelos para la implementación de minería de datos como SEMMA, P3TQ (Catalyst) y CRISP-DM convirtiéndose esta última en la guía más utilizada en el desarrollo de proyectos de minería de datos desde del 2007 por su nivel de detalle en la descripción de las actividades (Moine, Haedo, and Gordillo, 2011).

Para el desarrollo de este trabajo se utilizó la metodología Cross Industry Standard Process for Data Mining “CRISP – DM”, debido a que profundiza en los detalles de las tareas y actividades que se van a ejecutar en cada etapa del proceso de minería de datos, además fomenta la interoperabilidad de las diferentes herramientas durante todas las fases en el proceso de desarrollo, del mismo modo, es neutral con respecto al software de implementación ver Ilustración No 1. Esta metodología incluye todo el proceso desde el entendimiento del negocio hasta la validación de los resultados mediante la combinación con técnicas de análisis estadísticos. Inicialmente se presentan las características de la muestra de estudiantes que hicieron parte del estudio (Moine, Haedo, and Gordillo, 2011).

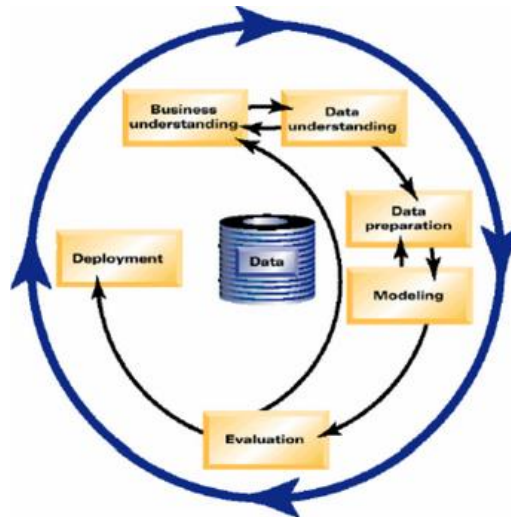


Ilustración 6: Metodología CRISP – DM (Chapman, Clinton, and Keber, 2000)

6.2.1. Etapas de la metodología CRISP

A continuación, se detalla el proceso seguido en cada una de las fases propuestas:

6.2.1.1. Entendimiento del negocio

En esta fase se identificaron las variables consideradas en el modelo de usuario del componente adaptativo a utilizar. El componente adaptativo fue desarrollado en el marco del proyecto de investigación titulado “Entorno de aprendizaje virtual adaptativo soportado por un modelo de estudiante multidimensional”, cuyo objetivo principal estuvo orientado a contribuir al desarrollo de los sistemas adaptativos educativos. El modelo de estudiante incorporó las variables nivel de competencia y estilo de aprendizaje para personalizar las actividades y los contenidos de los estudiantes.

El conjunto de datos surge de la investigación realizada por Cardona (2017) en la cual se analizó la posible incidencia de los cursos bajo la metodología por proyectos formativos, en un ambiente de aprendizaje virtual adaptativo, sobre el rendimiento académico de los estudiantes. Los objetivos de esa investigación se centraron en conocer la opinión de los estudiantes con respecto a la

implementación de una metodología de evaluación de competencias, además, analizar los resultados de un proceso de evaluación de competencias y la posible incidencia que puede tener la implementación de una metodología de evaluación en el rendimiento académico de los estudiantes soportado en un ambiente de aprendizaje virtual adaptativo (Cardona, 2017).

En este estudio se tuvieron en cuenta variables dependientes e independientes. Las variables dependientes se especifican como características en los sujetos que se verán modificadas por efecto de la variable independiente (Barbosa, 2010). Las variables independientes se asociaron a los aspectos de evaluación, perfil académico y datos personales, entre las que se destacan: edad, generó, tipo de colegio de egreso, estrato socioeconómico y nota definitiva. Se explica más en detalle en el ítem 2.1.1. Variables consideradas para el análisis.

Las variables independientes son aquellas que al ser manipuladas producen un efecto sobre las variables dependientes. Las variables dependientes fueron: (1) rendimiento académico, que corresponde al logro del nivel de competencia, el cual se obtuvo a partir de los resultados de un proceso de evaluación, y (2) opinión del estudiante, con respecto a la metodología de evaluación de competencias que se denominará aspectos didácticos.

Del mismo modo, los aspectos didácticos considerados fueron obtenidos del instrumento de la encuesta orientado a conocer la opinión de los estudiantes con relación a la implementación de la metodología de proyectos formativos soportada en un ambiente virtual de aprendizaje. El cuestionario se compone de preguntas de corte cuantitativo.

El conjunto de datos de este estudio correspondió a estudiantes de la institución de Administración y Mercadotecnia (EAM), en el programa de Ingeniería de Software, en cursos del área de programación de computadores bajo la metodología de proyectos formativos. El número de individuos del estudio empírico fue de 262, los estudiantes pertenecían al segundo semestre de ingeniería de software, la intervención tuvo una duración de 12 semanas académicas, durante el segundo periodo del 2016.

La selección de los sujetos se realizó mediante un muestreo aleatorio simple, por medio de la función aleatoria programada en Excel. Cada sujeto de la población tuvo igual probabilidad de ser seleccionado de forma independiente de los otros sujetos de la población. Se tuvo en cuenta como criterio de inclusión para la selección, que el estudiante no fuera repitente del curso para un total de 249 registros. Para el cálculo del tamaño de la muestra se tuvo en cuenta que el nivel de confianza fue del 95%. El error máximo aceptado del 3%, la proporción p del 0.5 y q del 0.5.

La muestra aleatoria probabilística de 133 estudiantes de un total de 262 registros. En dicha muestra, la distribución de género fue de 48,1% mujeres y el 51,9% hombres. El estrato socioeconómico de los estudiantes mostró que el 10.7% son de estrato muy bajo, el 30.6% de estrato bajo, el 32.2% corresponde al estrato medio, un 16% al estrato medio alto y el 10.5% al estrato alto. Con relación a la edad se encontró que los estudiantes de la muestra están entre los 18 y los 22 años. El 47.6% están entre los 18 y 19 años, e 29.3% están entre los 20 y 21 años, y el 23% entre los 22 y 23 años. La edad promedio de los estudiantes es de 20.4 años. La moda muestra que la edad que más se repite es 20 años. La desviación del 1.66, muestra poca variabilidad en la edad de los estudiantes.

6.2.1.2. Variables consideradas para el análisis

La segunda fase de la metodología CRISP - DM, tiene como propósito el entendimiento del negocio. Para ello, se analizaron algunas de las características y variables propuestas por Cardona (2017), así mismo, se consideraron variables adicionales que posiblemente pueden tener incidencia sobre el rendimiento académico del estudiante, ver ilustración No 7. En el [anexo 1](#) se presentan las variables su descripción y el dominio de cada atributo.

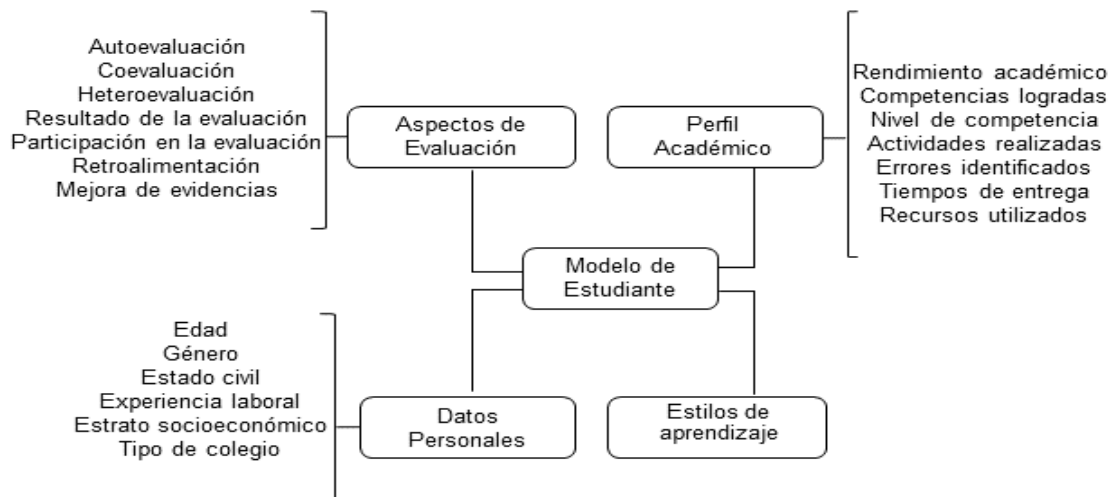


Ilustración 7: Variables consideradas para el análisis.(Cardona 2017)

Las categorías aspectos de evaluación, perfil académico y datos personales fueron extraídas del componente adaptativo desarrollado en el marco del proyecto mencionado previamente.

De acuerdo con lo que se presenta en la Ilustración No 7, del componente adaptativo se analizó:

- Los datos personales hacen referencia a toda la información personal de un estudiante como edad, género, estado civil, experiencia laboral, entre otras, que pueden tener influencia en el proceso de educación. Estas variables se inicializan al comienzo del curso el curso en la plataforma Moodle.
- El estilo de aprendizaje se enfoca principalmente en la enseñanza del estudiante. Son preferencias y tendencias altamente individualizadas de una persona que influyen en su aprendizaje (Cardona 2017).
- El perfil académico hace referencia a información académica de un estudiante como evidencias, actividades, tiempos de entrega, errores, recursos utilizados en el proceso de formación. Estas variables se actualizan a medida que los estudiantes presentan las actividades en su proceso de formación mediante la plataforma Moodle.
- Aspectos de evaluación agrupa las características relevantes en la valoración de los logros por competencias. En esta categoría se consideró el rendimiento académico como variable principal de un estudiante siguiendo la metodología mencionada anteriormente.

Adicionalmente, Cardona (2017) consideró la categoría de aspectos didácticos. Los datos para esta categoría fueron tomados mediante un instrumento que incluye las características relacionadas con la metodología de proyectos formativos, tal como se presenta en la ilustración No 8.

El instrumento consta de las siguientes 6 secciones:

- Participación: Cumplimiento de los elementos de la metodología de proyectos formativos.
- Coherencia de las actividades: Comprensión y participación en las actividades del proceso de evaluación.
- Evaluación: Contribución de las actividades del proceso de evaluación al rendimiento académico.
- Utilidad de los recursos: Utilidad de los recursos y las actividades de aprendizaje, ofrecidas en la plataforma Moodle.
- Diseño curso Moodle: Opinión sobre el diseño del curso en la Plataforma Moodle



Ilustración 8: Características de la categoría aspectos didácticos.

6.2.1.3. Análisis de factores

En el trabajo de Cardona (2017) la fiabilidad del instrumento se verificó con base en las apreciaciones de 31 profesores universitarios. Cada uno de los profesores valoró la pertinencia de los ítems del instrumento, en una escala de 1 a 5. Para verificar la consistencia interna de los ítems del instrumento, se realizó un análisis de fiabilidad basado en el coeficiente alfa de Cronbach. Los ítems del instrumento tuvieron discriminación positiva para el coeficiente alfa, y, por lo tanto, un alto grado de consistencia interna. El coeficiente resultante fue 0.887, por lo que la fiabilidad puede considerarse aceptable, teniendo como referente que una encuesta de este tipo puede estar desde un 0.70 (Morales, Urosa, and Blanco, 2003), citado en (Olmos 2008). El alfa de Cronbach en caso de eliminar algún elemento sigue teniendo discriminación positiva. Así mismo, la varianza de la escala es homogénea en caso de eliminarse algún elemento (Cardona 2017).

Se realizó la técnica para la validez de las variables mediante análisis factorial, el cual se utiliza para identificar propiedades homogéneas de las variables independientes que se están analizando. El objetivo fue establecer si los ítems del instrumento se agrupan en características comunes a ellos. Para el análisis factorial se verificaron los supuestos estadísticos y se analizaron los factores extraídos producto de la rotación de componentes. Los supuestos estadísticos para el análisis factorial fueron:

- El coeficiente de correlaciones entre las variables independientes es alto.
- La prueba de esfericidad Bartlett mostró que el nivel de significación de p-valor = 0.000, permite afirmar que se encontraron relaciones estadísticamente significativas entre ítems del instrumento.
- El valor KMO de 0.702, permite afirmar que es posible realizar el análisis factorial.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.702
Bartlett's Test of Sphericity	Approx. Chi-Square	3065.132
	df	1128
	Sig.	.000

- La diagonal principal de la matriz de correlaciones tiene coeficientes cercanos a 1.0, a excepción de las variables (p14=0.351 y p26=0.385) y los coeficientes fuera de esa diagonal son bajos. Los resultados obtenidos permiten afirmar que los supuestos estadísticos se cumplen para la extracción de los factores.

Var	Componente				
	1	2	3	4	5
p16: La coevaluación de evidencias contribuye al rendimiento académico	.839				
p13: Utilizo la autoevaluación para la mejora continua de mis evidencias	.823				
p14: Realizo coevaluación y autoevaluación de las evidencias, con base en los criterios de evaluación definidos en la rúbrica.	.762				
p15: La autoevaluación de evidencias contribuye al rendimiento académico.	.621				.381
P1: Las competencias a desarrollar en el proyecto formativo son claras a partir de las evidencias definidas.		.832			
p3: El proyecto formativo se enfoca en la solución de un problema del contexto y su nivel de reto está acorde con la formación de los estudiantes		.798			
P2. La metodología de proyectos formativos contribuye al desarrollo de las competencias del curso		.765			
p10: Los criterios de las rúbricas facilitan la evaluación de las evidencias del proyecto.		.687			
p31.: Me siento a gusto con la metodología de proyectos formativos.	.358	.418			
p25 El diseño gráfico del curso en Moodle permite una navegación adecuada.			.830		
p28: La interfaz del curso facilita la comprensión de los elementos del proyecto formativo.	.326		.814		
p27: Me siento a gusto con el diseño del curso en la plataforma Moodle.			.784		
p26.: El diseño gráfico del curso en Moodle es estético			.706		
p29.: El diseño de los recursos y las actividades es estéticamente agradable.			.686		
p23: Las actividades de aprendizaje ofrecidas en Moodle están acorde con mis necesidades de aprendizaje.				.858	
p22: Los recursos de aprendizaje ofrecidos en Moodle están acorde con mis necesidades de aprendizaje				.835	
p21: Los recursos de aprendizaje ofrecidos en Moodle son coherentes con las actividades de aprendizaje del proyecto formativo				.813	
p24: Las actividades de aprendizaje ofrecidas en Moodle son coherentes mi nivel de competencia				.647	.325
p20: Los recursos de aprendizaje ofrecidos en Moodle son útiles para la solución del proyecto formativo.			.412	.500	
p32: La metodología de evaluación incentiva a mejorar mi rendimiento académico		.355			.770
p34: La metodología de proyectos aumenta mi motivación para trabajar de forma colaborativa.					.695
p35: La autoevaluación me incentiva a mejorar mi rendimiento académico.					.658

p30: La metodología de proyectos formativos incentiva a mejorar el rendimiento académico.	.363				.472
---	------	--	--	--	------

Tabla 1: Matriz de componentes rotados

El método de extracción de los componentes se realizó mediante el análisis de componentes principales. El método de extracción generó 9 factores. Se utilizaron los primeros cinco factores que explican el 70.2% de la variabilidad total. Los factores que se agruparon fueron: metodología de proyectos formativos, proceso de evaluación, diseño del curso en Moodle, adaptación de recursos y actividades de aprendizaje y proceso de aprendizaje, el trabajo de Cardona (2017) permite observar que las variables agrupadas fueron coherentes con su investigación.

- El factor 1 está compuesto por variables asociadas al proceso de evaluación (p16, p13, p14 y p15).
- El factor 2 asoció los ítems (p1, p3, p2, p10, p31), estos están relacionados con la metodología de proyectos formativos.
- El factor 3 compuesto por los ítems (p25, p28, p27, p26, p29), los cuales están relacionados con el diseño de la plataforma Moodle.
- El factor 4 compuesto por los ítems (p23, p22, p21, p24, p20), los cuales están relacionados con la utilidad de los recursos y las actividades de aprendizaje de Moodle.
- El factor 5 compuesto por los ítems (p32, p34, p35, p30), los cuales se asocian con la opinión del estudiante sobre el proceso de aprendizaje y el rendimiento académico.

6.2.1.4. Entendimiento de los datos

En la recopilación de los datos de los estudiantes del curso bajo la metodología de proyecto formativo se describe los campos en la siguiente tabla:

Nombre	Descripción	Tipo
Rendimiento académico	Nota definitiva de los estudiantes en el curso	Numérico
Datos Personales		
Edad	Edad	Numérico
Género	Masculino Femenino	Nominal
Estrato	0 - Bajo 1 - Medio bajo 2 - Medio 3 - Medio Alto 4 - Alto 5 - Alto-alto	Nominal
Estado civil	Soltero Casado/Unión libre Separado / Divorciado Viudo	Nominal
Trabaja	Si No	Nominal
Personas Cargo	Número de personas a cargo	Numérico
Repitente	Si No	Nominal

Promedio Universidad	Esta variable permite tener en cuenta el promedio que llevan los estudiantes en carreras anteriores o en el transcurso de su carrera.	Numérico
egresado colegio	Publico Privado	Nominal
promedio Colegio	Promedio del alumno en el colegio	Numérico
ICFES	Examen	Numérico
Lugar residencia	Armenia Calarcá Circasia La tebaida Montenegro Pijao	Nominal
Nivel educativo madre	Primaria Secundaria Universidad Especialización Maestría	Nominal
Nivel educativo padre	Primaria Secundaria Universidad Especialización Maestría	Nominal
Numero hermanos	Numero de hermanos	Numérico
Posición hermanos	Primero Segundo Tercero Cuarto Quinto Posterior	Nominal
Tipo vivienda	Arrendada Familiar Propia	Nominal
Cumplimiento de los elementos de proyectos formativos		
Contribución Metodología PF	Selección de 1 – 5 La metodología de proyectos formativos contribuye al desarrollo de las competencias a formar	Numérico
Complejidad del PF	Selección de 1 – 5 El proyecto formativo se enfoca en la solución de un problema del contexto y su nivel de reto está acorde con mi nivel de formación	Numérico
Participación en el PF	Selección de 1 – 5 Participo en la definición del problema del contexto del proyecto formativo	Numérico
Realiza las actividades del PF	Selección de 1 – 5 Realizo las actividades correspondientes a cada una de las fases del proyecto formativo	Numérico
Trabajo colaborativo	Selección de 1 – 5 La metodología de proyectos formativos promueve el trabajo colaborativo y la autogestión	Numérico
Comprensión y participación en las actividades del proceso de evaluación		

Orientan Elaboración de las evidencias	Selección de 1 – 5 Orientan la elaboración de las evidencias del proyecto formativo	Numérico
Elaboración de las evidencias	Selección de 1 – 5 Realiza las evidencias propuestas basado en los criterios definidos en el/los mapa(s) de aprendizaje,	Numérico
Realizo autoevaluación	Selección de 1 – 5 Realizo autoevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s) de aprendizaje	Numérico
Realizo coevaluación	Selección de 1 – 5 Realizo coevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s)	Numérico
Retroalimentación del profesor	Selección de 1 – 5 La retroalimentación del profesor promueve la reflexión sobre las competencias desarrolladas, los logros y los aspectos por mejorar	Numérico
Retroalimentación de los compañeros	Selección de 1 – 5 La retroalimentación de los compañeros muestra objetividad en la valoración de las evidencias y fomenta una cultura democrática en la evaluación	Numérico
Contribución de las actividades del proceso de evaluación al rendimiento académico		
contribución de la autoevaluación al rendimiento académico	Selección de 1 – 5 La autoevaluación de evidencias contribuye al rendimiento académico	Numérico
contribución de la coevaluación al rendimiento académico	Selección de 1 – 5 La coevaluación de evidencias contribuye al rendimiento académico	Numérico
contribución de la Heteroevaluación al rendimiento académico	Selección de 1 – 5 La heteroevaluación de evidencias contribuye al rendimiento académico.	Numérico
contribución de metodología PF al rendimiento académico	Selección de 1 – 5 La metodología de proyectos formativos incentiva a mejorar el rendimiento académico	Numérico
contribución de la evaluación al rendimiento académico	Selección de 1 – 5 La metodología de evaluación contribuye a mejorar mi rendimiento académico	Numérico
Utilidad de los recursos y las actividades de aprendizaje en la plataforma Moodle		
Utilidad de los recursos de aprendizaje	Selección de 1 – 5 Los recursos de aprendizaje ofrecidos en Moodle Utilidad de los recursos de aprendizaje: Los recursos de aprendizaje ofrecidos en Moodle son útiles para la solución del proyecto formativo.	Numérico
coherencia de los recursos para el aprendizaje	Selección de 1 – 5 Los recursos de aprendizaje ofrecidos en Moodle son coherentes con las actividades de aprendizaje del proyecto formativo	Numérico

Coherencia de los recursos con las necesidades	Selección de 1 – 5 Los recursos de aprendizaje ofrecidos en Moodle están acordes con mis necesidades de aprendizaje	Numérico
coherencia de las Actividades con las necesidades	Selección de 1 – 5 Las actividades de aprendizaje ofrecidas en Moodle están acordes con mis necesidades de aprendizaje	Numérico
coherencia con el nivel de competencia	Selección de 1 – 5 Las actividades de aprendizaje ofrecidas en Moodle son coherentes mi nivel de competencia	Numérico
Opinión sobre el diseño del curso en la Plataforma Moodle		
Navegación del curso	Selección de 1 – 5 El diseño gráfico del curso en Moodle permite una navegación adecuada.	Numérico
Diseño gráfico	Selección de 1 – 5 El diseño gráfico del curso en Moodle es estético	Numérico
Experiencia del curso en la plataforma	Selección de 1 – 5 Me siento a gusto con el diseño del curso en la plataforma Moodle	Numérico
Interfaz del curso	Selección de 1 – 5 La interfaz del curso facilita la comprensión de los elementos del proyecto formativo	Numérico
Diseño es estética agradable	Selección de 1 – 5 El diseño de los recursos y las actividades es estéticamente agradable	Numérico
Mejora Evidencia		
Mejora Evidencia	Selección de 1 – 5 Realiza las sugerencias propuestas por el docente	Numérico
Retroalimentación	Retroalimentación de las evidencias	Numérico
Periodo	Periodo académico	Numérico
rendimiento académico	Nota definitiva de los estudiantes en el curso	Numérico
Aspectos de la Evaluación		
Autoevaluación	Nota de la evaluación realizada por cada estudiante	Numérico
Coevaluación	Evaluación definida por el docente	Numérico
Heteroevaluación	Calificación entre los estudiantes	Numérico
Recomendaciones	Comentarios sobre el curso.	Texto

Tabla 2: Descripción de variables

En la fase de entendimiento de datos de la metodología CRISP-MD, se realizó toda la preparación de la muestra de estudio. El objetivo de esta fase fue obtener los datos con la calidad necesaria para realizar el proceso de minería de datos. Esta información se extrajo de la plataforma Moodle y del instrumento aplicado a los estudiantes. Adicionalmente, se obtuvo información relacionada con el rendimiento académico, representado por la nota definitiva del estudiante en el curso.

En la fase de entendimiento de datos, se realizó la preparación de los datos siguiendo cuatro pasos:

- Paso 1. Limpieza de los datos: se utilizó la herramienta DQ-Analyzer la cual generó el perfilamiento de datos, permitiendo evidenciar problemas de calidad como: datos nulos, datos duplicados, errores en formatos, valores extremos. También para los valores nulos se utilizó el valor de la media para rellenar estos campos (Pushkarev, Neumann, Varol, and Talburt, 2010)

- Paso 2. Eliminación de variables: Mediante la herramienta WEKA, se aplicaron los evaluadores de atributos. Esto permitió identificar las variables más relevantes con relación al rendimiento académico de los estudiantes (Bouckaert, Frank, and Hall, 2010).
- Paso 3. Resultado evaluador de atributos: se consideraron los atributos que presentaron más de tres (3) repeticiones en los diferentes métodos implementados, teniendo en cuenta su correlación con el rendimiento académico de los estudiantes.
- Paso 4. Variables seleccionadas: Se definieron las variables para la implementación de las técnicas de minería de datos.

6.2.1.5. Modelamiento

En la fase de modelamiento de la metodología CRISP-MD, se continuó con el proceso de descubrimiento de la información de valor por medio de técnicas de minería de datos. Para ello se realizó el modelamiento con la herramienta WEKA especializada en la implementación de algoritmos de minería de datos, donde se utilizó el algoritmo K-Means de la técnica de agrupación siendo una técnica muy citada en los trabajos sobre agrupamiento, además se parte de k grupos conocidos, para este proyecto se definieron dos grupos (aprobados y reprobados) (Duda, Hart, and Stork, 2001).

Este algoritmo toma cada registro y lo sitúa en el centroide más cercano, vuelve a calcular el centroide de los grupos y distribuye los objetos nuevamente, este proceso se ejecuta las veces que sea necesario hasta que ya no existan cambios en los grupos. Como resultado se formaron dos grupos de estudiantes con características similares, se aclara que en la herramienta se configuraron dos clústeres $K=2$, esto fue basado en los estudiantes aprobados y reprobados de los cursos.

Además, se ejecutó el algoritmo A-Priori de la técnica de reglas de asociación de minería que permitió descubrir relaciones entre las variables de los estudiantes y el rendimiento académico; existen diferentes algoritmos como: Tertius, Apriori y PredictiveApriori, aunque estos pueden realizar la misma tarea, para este estudio se seleccionó Apriori, el criterio de selección para seleccionar el algoritmos fue filtrar aquellos que buscara grupos de elementos comunes en las transacciones, tuvieran una fácil representación e interpretación (Duda, Hart, and Stork, 2001).

A partir de la ejecución de las técnicas de minería de datos, se consideró el rendimiento académico como variable clase para determinar el perfilamiento de los diferentes tipos de estudiantes que realizaron el curso bajo la metodología de proyectos formativos. El rendimiento académico puede ser conceptualizado como la valoración cuantitativa y cualitativa del logro de competencia alcanzado en el proceso de enseñanza-aprendizaje, dentro de un contexto de desempeño específico (Cardona, 2017).

Al rendimiento académico se asocian diversos factores que pueden tener influencia en él, generalmente se asocian factores demográficos, factores socioeconómicos, aspectos de personalidad del estudiante, formación previa del estudiante, metodologías de enseñanza, sistema de evaluación, motivación escolar, entre otros. Son diversas las investigaciones que analizan los factores que inciden en el rendimiento académico (Acevedo, 2011), (Arribas, 2012;), (Garbanzo, 2013), (Suárez, Fernández and Muñiz, 2014;) y (Villegas Barahona, 2015;).

6.2.1.6. Evaluación

En la fase de evaluación de la metodología CRISP-MD se analizaron las variables de acuerdo con los resultados de las técnicas de minería de datos utilizadas y su relación con el rendimiento académico de los estudiantes, en cursos bajo la metodología de proyectos formativos. Para esto se realizó una validación mediante la aplicación de técnicas estadística a los datos, las cuales permitieron observar las variables con mayor relevancia tanto en la minería de datos, como estadísticamente, en el proceso de aprendizaje de un estudiante (Martínez, Karanik, Giovannini, Báez and Torre, 2016).

Se evaluó el modelo de estudiante y los algoritmos implementados con el objetivo de validar el proceso y los resultados generados, se aplicaron técnicas estadísticas al conjunto de datos iniciales con la intención de encontrar relación con los resultados generados por el análisis factorial, las técnicas de minería de agrupación y reglas de asociación.

6.2.1.7. Despliegue

El objetivo fue explicar los resultados que se generaron en las fases anteriores; se identificaron las variables que influyeron en el rendimiento académico de los estudiantes, a partir del análisis de resultados obtenidos. Se identificaron aquellas que presentaron mayor incidencia en el proceso de minería de datos.

7. Presentación y análisis de resultados

En esta sección se presentan los resultados generados en la implementación de las técnicas de minería de datos, con la intención de proponer variables a considerar en el ambiente adaptativo educativo de un curso bajo la metodología de proyectos formativos.

7.1. Generación de la base de datos para el análisis

La recolección de los datos para el análisis de las variables se realizó mediante la extracción de la información de los cursos elaborados bajo la metodología de proyectos formativos En un ambiente virtual de aprendizaje soportado en la plataforma Moodle, además se extrajeron los datos de una encuesta elaborada en la tesis doctoral de (Cardona 2017) con el objetivo de conocer la opinión de los estudiantes sobre su experiencia con la metodología, esta información se consideró como atributos complementarios “Aspectos Didácticos” para entender e identificar oportunidades de mejora en el sistema adaptativo. El detalle de la encuesta se presenta en el [anexo 2](#).

En esta investigación se consideró conveniente excluir del análisis la categoría “El estilo de aprendizaje” del modelo de estudiante propuesto en (Cardona 2017), ya que los datos obtenidos eran insuficientes para realizar el análisis.

Esta información fue digitalizada para dar inicio al proceso de preparación de los datos. Para ello, se creó una Base de datos en Excel como se muestra en la Ilustración 7, en la cual se almacenó y organizó toda la información de cada estudiante. El detalle de los datos se presenta en el [anexo 4](#).

Edad	Genero	Estrato	ESOLTERO	CASADO	UNION	LIBRE	Trabajo	Personas_Car	Repite	NO	promedio_universid	egresado_cole	promedio_Cole	ICFES	Lugar_residencia	Nivel_educativo	Nivel_educativo_madre	Nivel_educativo_padre	Numero_hermanos	Posicion_hermanos	Tipo_vivienda
17	MUJER	3	SOLTERO	SI	0	NO	0	NO	4.2	PUBLICO	42	299	CALARCA	UNIVERSIDAD	UNIVERSIDAD	2	SEGUNDO	APRENDIZAJE			
22	MUJER	1	SOLTERO	NO	0	NO	4.4	PUBLICO	4.0	202	ARMENIA	SECUNDARIA	SECUNDARIA	1	PRIMERO	APRENDIZAJE					
19	MUJER	3	SOLTERO	NO	0	NO	4.3	PUBLICO	4.5	327	CALARCA	SECUNDARIA	SECUNDARIA	4	QUINTO	PROPIA					
19	MUJER	3	SOLTERO	NO	0	NO	4	PUBLICO	4.5	327	CALARCA	SECUNDARIA	SECUNDARIA	1	SEGUNDO	PROPIA					
19	HOMBRE	3	SOLTERO	SI	0	NO	3.8	PUBLICO	4.2	295	ARMENIA	PRIMARIA	PRIMARIA	3	CUARTO	PROPIA					
19	MUJER	4	SOLTERO	NO	0	NO	4.5	PRIVADO	4.4	380	ARMENIA	SECUNDARIA	SECUNDARIA	2	PRIMERO	PROPIA					
19	HOMBRE	3	SOLTERO	NO	0	NO	4.4	PUBLICO	4.5	360	ARMENIA	SECUNDARIA	SECUNDARIA	2	SEGUNDO	PROPIA					
19	HOMBRE	3	SOLTERO	NO	3	NO	3.9	PUBLICO	4.5	360	RIVICONSANTO	PRIMARIA	SECUNDARIA	1	PRIMERO	PROPIA					
23	HOMBRE	3	SOLTERO	NO	9	NO		PRIVADO	417	SOFRENTO	PRIMARIA	PRIMARIA	3	CUARTO	APRENDIZAJE						
21	HOMBRE	3	SOLTERO	NO	0	NO		PUBLICO			ARMENIA	UNIVERSIDAD	ESPECIALIZACION	1	PRIMERO	PROPIA					
19	MUJER	2	SOLTERO	NO	0	NO		PUBLICO			ARMENIA	SECUNDARIA	SECUNDARIA	1	SEGUNDO	PROPIA					
19	MUJER	3	SOLTERO	SI	0	NO	4.0	PRIVADO	4.2		ARMENIA	SECUNDARIA	SECUNDARIA	1	SEGUNDO	APRENDIZAJE					
19	MUJER	3	SOLTERO	NO	0	NO	4.3	PUBLICO	4.8	320	ARMENIA	SECUNDARIA	SECUNDARIA	1	SEGUNDO	APRENDIZAJE					
19	MUJER	1	SOLTERO	NO	0	NO	4.2	PUBLICO	255		PUJAO	SECUNDARIA	SECUNDARIA	1	SEGUNDO	APRENDIZAJE					
21	MUJER	3	SOLTERO	NO	0	NO		PRIVADO			NIR	SECUNDARIA	SECUNDARIA	1	PRIMERO	APRENDIZAJE					
20	HOMBRE	3	SOLTERO	NO	0	NO	3.5	PUBLICO	4.0	280	PUERTOSPECIO	PRIMARIA	PRIMARIA	0		PROPIA					
19	MUJER	2	SOLTERO	NO	0	NO	3.8	PUBLICO	4.0		CALARCA	UNIVERSIDAD	SECUNDARIA	2	TERCERO	PROPIA					
20	MUJER	3	SOLTERO	NO	1	NO	3.0	PRIVADO			LOS ANDRES	SECUNDARIA	UNIVERSIDAD	3	SEGUNDO	PROPIA					
22	MUJER	1	SOLTERO	SI	0	NO	3.3	PUBLICO			ARMENIA	PRIMARIA	PRIMARIA	4	SEGUNDO	PROPIA					
17	MUJER	1	SOLTERO	SI	0	NO	2.5	PUBLICO	428		LAS COLINAS	SECUNDARIA	SECUNDARIA	1	SEGUNDO	PROPIA					
21	MUJER		SOLTERO	SI	0	NO	3.3	PUBLICO	4.0	48	ALAMOS	PRIMARIA	PRIMARIA	1	SEGUNDO						
20	MUJER		SOLTERO	NO	0	NO	3.7	PUBLICO	4	45	MONTEBLANCO	SECUNDARIA	PRIMARIA	1	PRIMERO	PROPIA					
20	MUJER	3	SOLTERO	SI	0	NO	3.5	PUBLICO	4.5	58	MONTEBLANCO	UNIVERSIDAD	UNIVERSIDAD	4	CUARTO	PROPIA					
23	HOMBRE	5	CASADO	UNION	LIBRE	SI	1	NO	3.5	PRIVADO			ARMENIA	SECUNDARIA	UNIVERSIDAD	2	SEGUNDO	APRENDIZAJE			
43	MUJER	3	SOLTERO	SI	1	NO	3.0	PUBLICO	257		ARMENIA	PRIMARIA	PRIMARIA	6	PRIMERO	APRENDIZAJE					
23	HOMBRE	4	CASADO	UNION	LIBRE	SI	3	NO	4	PUBLICO			ARMENIA	PRIMARIA	SECUNDARIA	3	PRIMERO	APRENDIZAJE			
17	HOMBRE	2	SOLTERO	NO	0	SI	3.5	PUBLICO	3.0	257	LA PATRIA	SECUNDARIA	SECUNDARIA	1	PRIMERO	PROPIA					
20	MUJER	2	SOLTERO	NO	0	NO		PUBLICO			NIR	PRIMARIA	SECUNDARIA	1	SEGUNDO	PROPIA					
21	MUJER	2	CASADO	UNION	LIBRE	SI	1	NO			CALARCA	SECUNDARIA	SECUNDARIA	3	CUARTO	PROPIA					
20	MUJER	2	SOLTERO	NO	0	NO		PRIVADO	40		NIR	SECUNDARIA	SECUNDARIA	1	PRIMERO	APRENDIZAJE					
21	MUJER	2	SOLTERO	NO	0	NO		PUBLICO	45		ARMENIA	SECUNDARIA	SECUNDARIA	2	TERCERO	PROPIA					
23	MUJER	5	CASADO	UNION	LIBRE	SI	0	NO			ARMENIA	SECUNDARIA	SECUNDARIA	1	PRIMERO	APRENDIZAJE					
19	MUJER	3	SOLTERO	SI	3	NO	3.0	PUBLICO	4.0		ARMENIA	SECUNDARIA	SECUNDARIA	3	SEGUNDO	APRENDIZAJE					
23	HOMBRE	3	SOLTERO	SI	0	NO		PUBLICO	4	204	ARMENIA	E	ESPECIALIZACION	1	SEGUNDO	PROPIA					

Ilustración 9: Base de datos

7.2. Preparación de los datos

En el punto anterior se extrajeron los datos más relevantes del modelo de estudiante y los datos de la encuesta que son atributos complementarios “Aspectos Didáctico”, considerando lo que se expone en el trabajo de (Kohavi and John 1997), donde se definen tres tipos de relevancia para los atributos: Fuertemente relevantes, Débilmente relevantes e Irrelevante, además Langley (Langley 1994) propone reducir el número de variables antes de aplicar algún algoritmo de clasificación, con el objetivo de generar mayor precisión y menor ruido en los resultados de las técnicas de minería.

El conjunto total de datos fue de 262 del cual se obtuvo una muestra aleatoria probabilística con la información de 133 estudiantes, una confianza del 90% y un margen de error del 5%. Además, se consideró como criterio para la muestra la calidad de los datos, que existieran el 80% de los registros

de cada estudiante. La distribución de género fue de 48,1% mujeres y el 51,9% hombres y la edad promedio de los alumnos fue de 20,5 años. El detalle del conjunto total de datos se presenta en el anexo 3: “Datos iniciales estudiantes proyectos formativos”.

Por lo tanto, se analizaron los datos generados de las características mencionadas anteriormente, teniendo en cuenta que podían incidir en el rendimiento académico, y de esta manera poder establecer variables de mayor relevancia en el proceso de aprendizaje en un ambiente adaptativo de un estudiante (Martínez et al. 2016), como se explica a continuación.

7.2.1. Paso 1: Limpieza de datos

Con el fin de conocer la calidad de los datos se realizó un diagnóstico de estos utilizando el software DQ Analizar; esta herramienta permitió realizar el perfilamiento de datos y ayudó a identificar problemas de calidad como: datos nulos, datos duplicados, errores en formatos, valores extremos.

Se realiza las descripciones estadísticas de cada uno de los datos

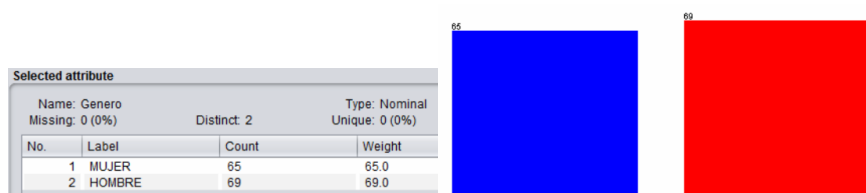


Ilustración 10: análisis estadístico – Género

Esta variable contiene valores posibles Mujer y Hombre, para un total de 65 Mujeres y 69 hombres.

Posterior al análisis de calidad de los datos, se realizó el análisis de correlación generado por WEKA, también se presentó el ranking con las variables más relevantes correlacionadas con la clase rendimiento académico “Definitiva”.

El detalle de la muestra se presenta en el [anexo 5: Preparación y selección de atributos.](#)

7.2.2. Paso 2: Eliminación de variables irrelevantes

En la selección de atributos tiene como objetivo realizar un ranking de características que permite disminuir la dimensionalidad permitiendo crear modelos que definen el grupo observado. Se resalta que no existe una técnica que permita seleccionar siempre los mejores resultados, generalmente se requiere la intervención de expertos.

Kohavi y John (Kohavi and John 1997) presentan tres tipos de relevancia para los atributos: Fuertemente relevantes, Débilmente relevantes e Irrelevante, además Langley (Langley 1994) propone reducir el número de variables antes de aplicar algún algoritmo de clasificación, con el objetivo de generar mayor precisión y menor ruido en los resultados de las técnicas de minería.

Se realizó la evaluación de atributos para identificar cuáles eran más relevantes, los métodos que se utilizaron para la selección de estas variables parten del conjunto total de datos, los cuales determinaron la correlación entre atributos y la clase dependiente rendimiento académico “Definitiva”.

A continuación, se presentan los evaluadores utilizados por WEKA:

- **CfsSubsetEval (CFS):** Este Evaluador genera un ranking de atributos de acuerdo a la correlación con la clase dependiente “Definitiva”, realiza una búsqueda de las variables que presentan una alta correlación con la clase, y una baja Inter correlación con otras variables; una desventaja de este método es que en algunas ocasiones se pueden eliminar características que no están relacionados con la clase pero si con otros atributos si (Gutiérrez García 2016).

- **CorrelationAttributeEval:** Este Evaluador realiza un análisis de relación entre el atributo y la clase dependiente “Definitiva”, los atributos que son categóricos los analiza como indicadores mediante un promedio ponderado (Bouckaert, Frank, and Hall 2010).
- **GainRatioAttributeEval:** Evalúa el valor de un atributo midiendo la relación de beneficio con respecto a la clase dependiente “Definitiva” (Bouckaert, Frank, and Hall 2010).
- **InfoGainAttributeEval:** Realiza la evaluación de los atributos con el valor de la ganancia con respecto a la clase “Definitiva” (Bouckaert, Frank, and Hall 2010).
- **OneRAttributeEval:** Evalúa el valor de los atributos utilizando el método de clasificación ONER (Bouckaert, Frank, and Hall 2010).
- **ReliefFAttributeEval:** Evalúa las variables mediante un muestreo repetido de una instancia, considerando el valor de un atributo para la instancia más cercana de la misma clase “Definitiva” (Bouckaert, Frank, and Hall 2010).
- **SymmetricalUncertAttributeEval:** Evalúa el valor de un atributo mediante la medición de la incertidumbre simétrica con respecto a la clase “Definitiva” (Bouckaert, Frank, and Hall 2010).

Evaluador de atributos	Algoritmo de clasificación	Método de Búsqueda	Clase	Variables Seleccionadas
CfsSubsetEval (CFS)		BestFirst	Definitiva	Edad Género Trabaja Personas Cargo ICFES Contribución Metodología PF Trabajo colaborativo Realizo coevaluación Diseño grafico Heteroevaluación
CorrelationAttributeEval		Ranker	Definitiva	<ul style="list-style-type: none"> • 48 heteroevaluación • 49 coevaluación • 11 ICFES • 47 autoevaluación • 2 género • 7 repitente • 17 tipo vivienda • 19 complejidad del PF • 34 coherencia de los recursos para el aprendizaje

				<ul style="list-style-type: none"> • 36 coherencia de las Actividades con las necesidades • 23 orientan Elaboración de las evidencias • 43 mejora Evidencia • 20 participación en el PF • 24 realizo autoevaluación • 16 posición hermanos
GainRatioAttributeEval		Ranker	Definitiva	<ul style="list-style-type: none"> • 48 heteroevaluación • 47 autoevaluación • 49 coevaluación • 7 repitente • 12 lugar residencia • 10 promedio Colegio • 8 promedio universidad • 4 estado civil • 16 posición hermanos • 36 coherencia de las Actividades con las necesidades • 19 complejidad del PF • 34 coherencia de los recursos para el aprendizaje • 38 navegación del curso • 39 diseño grafico
InfoGainAttributeEval		Ranker	Definitiva	<ul style="list-style-type: none"> • 48 heteroevaluación • 47 autoevaluación • 49 coevaluación • 10 promedio Colegio • 8 promedio universidad • 36 coherencia de las Actividades con las necesidades • 16 posición hermanos • 34 coherencia de los recursos para el aprendizaje • 38 navegación del curso • 28 contribución de la autoevaluación al rendimiento académico • 3 estrato • 24 realizo autoevaluación • 19 complejidad del PF • 45 periodo
OneRAttributeEval	OneR	Ranker	Definitiva	<ul style="list-style-type: none"> • 48 heteroevaluación • 47 autoevaluación • 49 coevaluación • 20 participación en el PF • 39 diseño grafico • 17 tipo vivienda • 23 orientan Elaboración de las evidencias • 18 contribución Metodología PF • 16 posición hermanos

				<ul style="list-style-type: none"> • 15 número hermanos • 30 contribución de la heteroevaluación al rendimiento académico • 5 trabaja • 2 género • 7 repitente
ReliefFAtributeEval		Ranker	Definitiva	<ul style="list-style-type: none"> • 48 heteroevaluación • 47 autoevaluación • 49 coevaluación • 36 coherencia de las Actividades con las necesidades • 25 realizo coevaluación • 20 participación en el PF • 21 realiza las actividades del PF • 19 complejidad del PF • 41 interfaz del curso • 38 navegación del curso • 34 coherencia de los recursos para el aprendizaje • 8 promedio universidad • 39 diseño grafico • 26 retroalimentación del profesor
SymmetricalUncertAttributeEval		Ranker	Definitiva	<ul style="list-style-type: none"> • 48 heteroevaluación • 47 autoevaluación • 49 coevaluación • 10 promedio Colegio • 8 promedio universidad • 16 posición hermanos • 36 coherencia de las Actividades con las necesidades • 34 coherencia de los recursos para el aprendizaje • 38 navegación del curso • 28 contribución de la autoevaluación al rendimiento académico • 19 complejidad del PF • 24 realizo autoevaluación • 3 estrato • 20 participación en el PF

Tabla 3: Evaluador de Atributos [Elaboración Propia]

Para este proyecto se utilizó el **CfsSubsetEval (CFS)** el cual genera un ranking de atributos de acuerdo a la correlación con la clase dependiente "Definitiva".

7.2.3. Paso 3: Variables seleccionadas

Finalizada la fase de procesamiento donde se realizó la limpieza de datos, normalización, calidad de datos, selección de atributos se presentan las variables seleccionadas: 19 variables incluida la clase:

Variables			
1	Definitiva (clase)	11	Nivel educativo padre
2	Heteroevaluación	12	Nivel educativo madre
3	género	13	Estado civil
4	coevaluación	14	Estrato
5	Posición hermanos	15	Egresado colegio
6	Edad	16	Personas Cargo
7	Autoevaluación	17	ICFES
8	Tipo vivienda	18	Trabaja
9	Mejora Evidencia	19	Autoevaluación
10	Lugar residencia		

Tabla 4: Variables seleccionadas

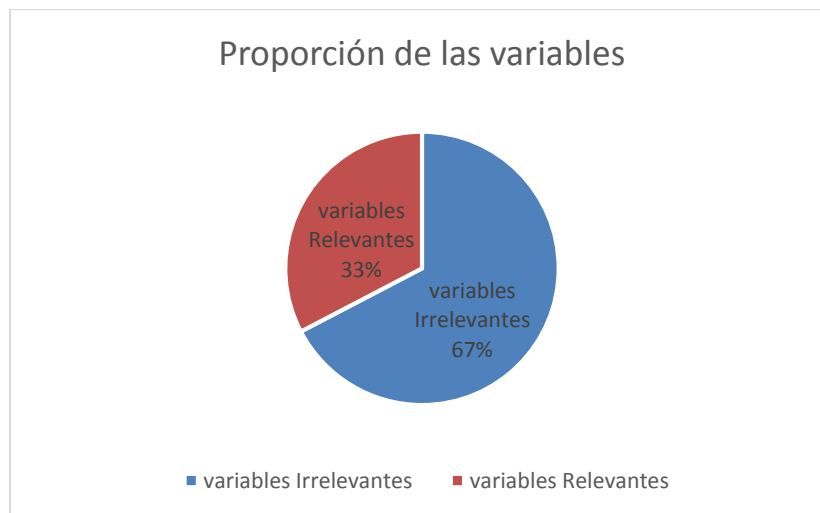


Ilustración 11: Proporción de las variables [Elaboración Propia]

El proceso en detalle que se realizó en la preparación y selección de atributos de acuerdo a su relevancia para el modelo de minería de datos se presenta en el [anexo 5](#).

7.3. Ejecución de técnicas de minería de datos

El objetivo de esta fase es aplicar las técnicas de minería en el conjunto de datos obtenido en la fase de preprocesamiento, para detectar información oculta y patrones. Con cada uno de los patrones se pretende encontrar las variables más relevantes de cada uno y finalmente, realizar una votación

entre las técnicas utilizadas teniendo en cuenta el número de veces que el factor fue encontrado como el más relevante. En el **anexo 7** se puede ver los experimentos en detalle.

Los algoritmos que se utilizan para determinar los factores que más influyen en la nota definitiva de los cursos bajo la metodología de proyectos formativos son:

- **Experimento 1:** Perfilamiento de los estudiantes. Clústeres
- **Experimento 2:** Selección de factores con Árboles de Decisión
- **Experimento 3:** Selección de factores con PCA
- **Experimento 4:** Selección de factores con regresión logística
- **Experimento 5:** Reglas de asociación

Teniendo en cuenta las variables descritas anteriormente, luego de aplicar los algoritmos de la minería de datos se obtuvieron los siguientes:

7.3.1. **EXPERIMENTO 1: Análisis de perfiles de los estudiantes en cursos bajo la metodología de proyectos formativos**

El objetivo es establecer el perfil de los estudiantes que realizaron el curso bajo la metodología de proyecto formativo, permitiendo formar grupos de acuerdo a las características.

Método: K-means y EM

Primero para aplicar el algoritmo Simple K-means se debe determinar el número de grupos mediante el algoritmo de **EM**.

Clustered Instances: Log likelihood: -23.78086

0	45 (34%)
1	35 (26%)
2	20 (15%)
3	34 (25%)

Después de obtener los grupos, se aplica el algoritmo de Simple K-means con dos, tres, cuatro y cinco grupos donde se selecciona el que tiene menor cohesión como se presenta a continuación:

- **Con 2 cluster:** Within cluster sum of squared errors: 527.7053244110108
- **Con 3 cluster:** Within cluster sum of squared errors: 497.4589292221793
- **Con 4 cluster:** Within cluster sum of squared errors: 460.9013850414691
- **Con 5 cluster:** Within cluster sum of squared errors: 435.90998566120703

Teniendo en cuenta lo anterior, se concluye que la agrupación óptima es con 5 clústeres para recorrer el algoritmo, debido a que la cohesión es menor en este caso. En la ilustración 12 se presentan los resultados generados por Weka de cada uno de los clústeres.

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (134.0)	Cluster#				
		0 (28.0)	1 (22.0)	2 (37.0)	3 (26.0)	4 (21.0)
Edad	20.4254	19.5	21.1818	21.7568	19.3462	19.8571
Genero	HOMBRE	HOMBRE	HOMBRE	MUJER	MUJER	HOMBRE
Estrato	2.7388	2.7857	2.8636	2.6757	2.9615	2.381
Estado_civil	SOLTERO	SOLTERO	SOLTERO	SOLTERO	SOLTERO	SOLTERO
Trabaja	NO	NO	NO	SI	NO	NO
Personas_Cargo	0.7388	0.6429	0.8636	0.7027	0.5385	1.0476
egresado_colegio	PUBLICO	PRIVADO	PUBLICO	PUBLICO	PUBLICO	PUBLICO
ICFES	299.5	333.8214	308.1818	308.9459	275.8077	257.3333
Lugar_residencia	ARMENIA	ARMENIA	ARMENIA	ARMENIA	ARMENIA	ARMENIA
Nivel_edcativo_madre	SECUNDARIA	PRIMARIA	SECUNDARIA	SECUNDARIA	SECUNDARIA	PRIMARIA
Nivel_educativo_padre	SECUNDARIA	SECUNDARIA	SECUNDARIA	SECUNDARIA	UNIVERSIDAD	PRIMARIA
Posicion_hermanos	SEGUNDO	SEGUNDO	PRIMERO	SEGUNDO	PRIMERO	PRIMERO
Tipo_vivienda	PROPIA	ARRENDADA	PROPIA	PROPIA	PROPIA	PROPIA
MejoraEvidencia	2.3134	2.1071	2.4545	1.973	3.2308	1.9048
Definitiva	'(2.76-3.32]'	'(2.76-3.32]'	'(3.88-4.44]'	'(-inf-2.76]'	'(2.76-3.32]'	'(3.32-3.88]'
Autoevaluacion	3.8694	3.6036	4.1909	4	3.6192	3.9667
Heteroevaluacion	3.3813	3.3607	4.0045	2.9784	3.4308	3.4048
coevaluacion	4.2284	4.2679	4.4455	4.1892	4.0538	4.2333

Ilustración 12: perfilamiento de los estudiantes WEKA

En la tabla 3 se presenta la distancia entre los centros de cada clúster, calculado por Weka mediante el algoritmo de distancia euclidiana.

Cluster	1	2	3	4	5
1		110.373	166.470	243.755	54.086
2	110.373		56.108	133.414	56.371
3	166.470	56.108		77.316	112.458
4	243.755	133.414	77.316		189.757
5	54.086	56.371	112.458	189.757	

Tabla 5: Distancias entre los centros de clúster

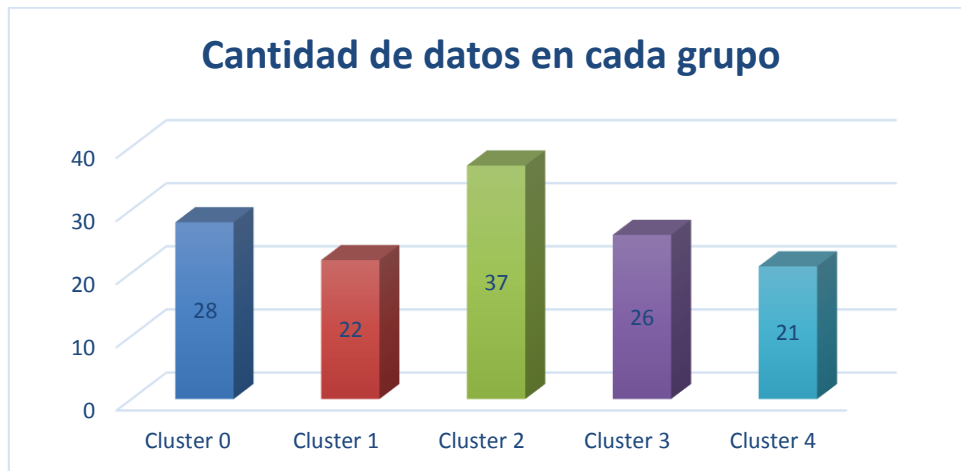


Ilustración 13: Datos por clúster

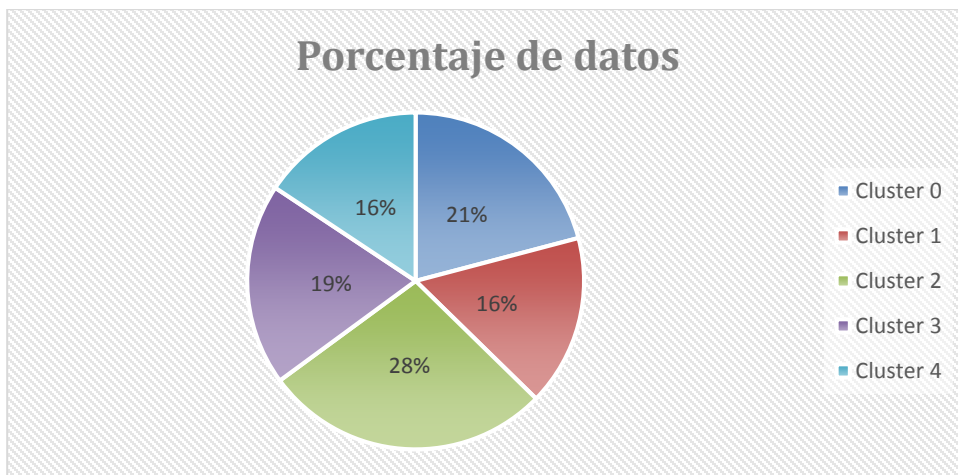


Ilustración 14: porcentaje de datos por Clúster

Clúster	Descripción
0	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (2.7 – 3.32], siendo el 21 % de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes hombres una edad promedio de 19.5 años, con estrado 2 – 3, hombres solteros que no trabajan y pueden tener entre 0 y 1 persona a cargo. Son egresado de un colegio privado con un promedio de ICFES 333.8, residente en armenia con un nivel educativo de primaria para la madre y secundaria para el padre, es el segundo en la posición de los hermanos y tiene una vivienda arriendo.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que en el ítem mejora la evidencia el promedio fue de 2.1, su autoevaluación de 3.6, su heteroevaluación en 3.3. y la coevaluación en 4.2.</p>

1	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (3.8 – 4.4], siendo el 16% de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes hombres una edad promedio de 21.1 años, con estrado 2 – 3, hombres solteros que no trabajan y pueden tener entre 0 y 1 persona a cargo. Son egresados de un colegio público con un promedio de ICFES 308.9, residente en armenia con un nivel educativo de secundaria para la madre y el padre, es el primero en la posición de los hermanos y tiene una vivienda propia.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que aprobaron el curso, además, en el ítem mejora la evidencia el promedio fue de 2.4, su autoevaluación de 4.1, su heteroevaluación en 4. y la coevaluación en 4.4.</p>
2	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (0 – 2.76], siendo el 28% de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes Mujeres una edad promedio de 21.7 años, con estrado 2 – 3, Mujeres solteras que si trabajan y pueden tener entre 0 y 1 persona a cargo. Son egresados de un colegio público con un promedio de ICFES 308.9, residente en armenia con un nivel educativo de secundaria para la madre y el padre, es el segundo en la posición de los hermanos y tiene una vivienda propia.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que no aprobaron el curso, además, en el ítem mejora la evidencia el promedio fue de 1.9, su autoevaluación de 4, su heteroevaluación en 2.9. y la coevaluación en 4.1.</p>
3	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (2.76– 3.332], siendo el 19% de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes Mujeres una edad promedio de 19 años, con estrado 2 – 3, Mujeres solteras que no trabajan y pueden tener entre 0 y 1 persona a cargo. Son egresados de un colegio público con un promedio de ICFES 275.8, residente en armenia con un nivel educativo de secundaria para la madre y universitaria para el padre, es el primero en la posición de los hermanos y tiene una vivienda propia.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que no aprobaron el curso, además, en el ítem mejora la evidencia el promedio fue de 3.2, su autoevaluación de 3.6, su heteroevaluación en 3.4. y la coevaluación en 4.</p>
4	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (3.32 – 3.88], siendo el 16% de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes hombres una edad promedio de 19.8 años, con estrado 2 – 3, hombres solteros que no trabajan y pueden tener 1 persona a cargo. Son egresados de un colegio público con un promedio de ICFES 257.3, residente en armenia con un nivel educativo de primaria para la madre y el padre, es el primero en la posición de los hermanos y tiene una vivienda propia.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que no aprobaron el curso, además, en el ítem mejora la evidencia el promedio fue de 1.9, su autoevaluación de 3.9, su heteroevaluación en 3.4 y la coevaluación en 4.2.</p>

Tabla 6: Evaluación de Resultados – Clústeres

Según los resultados presentados en la tabla anterior, se puede interpretar que los estudiantes con un rendimiento académico más bajo son aquellos que:

- Estudiantes mujeres
- Estudiantes que trabajan
- Estudiantes que obtuvieron en su heteroevaluación una nota menor a 3.
- Estudiantes que obtuvieron en su coevaluación una nota menor 4.2.

Ahora, los estudiantes con un rendimiento académico mayor son aquellos que:

- Estudiantes hombres
- Estudiantes que no trabajan
- Estudiantes que son primeros en la posición de sus hermanos
- Estudiantes que obtuvieron en su autoevaluación una nota de 4.1
- Estudiantes que obtuvieron en su heteroevaluación una nota 4.
- Estudiantes que obtuvieron en su coevaluación mayor que 4.2.

7.3.2. EXPERIMENTO 2: SELECCIÓN DE FACTORES CON ÁRBOLES DE DECISIÓN

En el presente experimento se analizan las variables que más influyen en la nota definitiva mediante arboles de decisión. El modelado se realiza con el algoritmo J48 y su validación cruzada con de grupos y balanceo de los datos para mejorar la calidad de aprendizaje.

Modelado: J48, con nodos binarios y un mínimo número de instancias de objetos de 400.

Para aplicar este algoritmo se creó una variable categórica con “aprobado” para los estudiantes que su nota definitiva fue mayor o igual a 3 (≥ 3) y “reprobado” para los que obtuvieron una nota menor a 3 (< 3).

DETAILED ACCURACY BY CLASS				
TP Rate	FP Rate	Precision	Recall	Class
1.000	0.036	0.991	1.000	aprobado
0.964	0.000	1.000	0.964	reprobado
0.993	0.028	0.993	0.993	Weighted Avg
F-Measure	MCC	ROC Area	PRC Area	Class
0.995	0.977	0.982	0.991	aprobado
0.982	0.977	0.982	0.972	reprobado
0.992	0.977	0.982	0.987	Weighted Avg
CONFUSIÓN MATRIX				
a	b	a = Aprobado		
106	0	b = Reprobado		
1	27			
Correctly Classified Instances 133 99.2537 %				

Incorrectly Classified Instances 1 0.7463 %

Tabla 7: Evaluación de Resultados - Experimento 1

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	133	99.2537 %
Incorrectly Classified Instances	1	0.7463 %
Kappa statistic	0.9771	
Mean absolute error	0.0075	
Root mean squared error	0.0864	
Relative absolute error	2.2377 %	
Root relative squared error	21.2354 %	
Total Number of Instances	134	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.036	0.991	1.000	0.995	0.977	0.982	0.991	Aprobado
	0.964	0.000	1.000	0.964	0.982	0.977	0.982	0.972	Reprobado
Weighted Avg.	0.993	0.028	0.993	0.993	0.992	0.977	0.982	0.987	

=== Confusion Matrix ===

```
a  b  <-- classified as
106  0 | a = Aprobado
  1 27 | b = Reprobado
```

J48 unpruned tree

```
-----
Heteroevaluación <= 2.7: Reprobado (26.0)
Heteroevaluación > 2.7
| Personas Cargo <= 1.0: Aprobado (91.0)
| Personas Cargo > 1.0
|| Egresado colegio = PUBLICO: Aprobado (12.0)
|| Egresado colegio!= PUBLICO
||| Género = MUJER: Aprobado (2.0)
||| Género!= MUJER: Reprobado (3.0/1.0)
Number of Leaves:    5
Size of the tree:    9
```

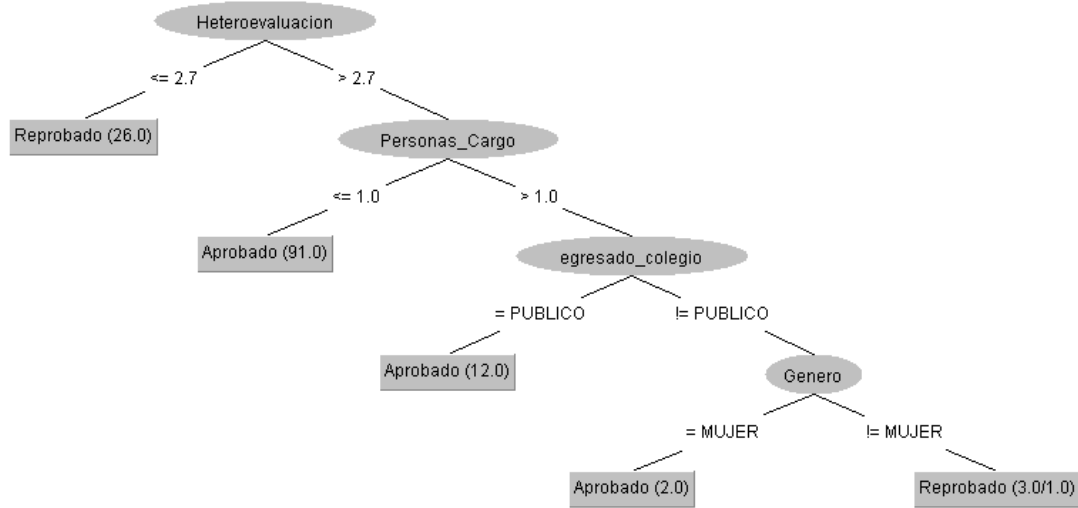


Ilustración 15: Árbol de decisión factores de influencia sobre la nota definitiva

Análisis de resultados

Como se presenta en la tabla 3, el 99.25% de los datos están clasificados de forma correcta. El área ROC para cada grupo es mayor a 0.7 lo que indica que tiene un alto grado de confiabilidad. Además, sobre la diagonal principal de la matriz de confusión se tiene la mayor parte de los datos, esto muestra que el resultado es cercano a la realidad.

De acuerdo con estos resultados, se puede indicar la nota definitiva es influenciada en gran medida por las variables presentadas en la siguiente tabla:

Variables
Heteroevaluación
Personas a cargo
Egresado colegio
Género

7.3.3. EXPERIMENTO 3: Análisis de componentes principales PCA

El objetivo de este experimento es observar las variables que más influyen en el rendimiento académico mediante PCA.

Modelado: Principal Component (PCA)

Ranked attributes:

```

0.9172 1 0.375Edad+0.347Estado_civil=CASADO/UNIONLIBRE-0.327Lugar_residencia=ARMENIA-0.299Estado_civil=SOLTERO+0.277Posicion_hermanos=PRIMERO...
0.8451 2 -0.371Tipo_vivienda=PROPIA-0.347Estado_civil=SOLTERO+0.341Tipo_vivienda=ARRENDADA-0.299Nivel_educativo_padre=UNIVERSIDAD+0.294Estado_civil
0.7831 3 -0.41Tipo_vivienda=ARRENDADA+0.366Tipo_vivienda=PROPIA-0.334Nivel_educativo_padre=SECUNDARIA+0.274Nivel_educativo_padre=PRIMARIA-0.24Estado
0.7291 4 0.341Nivel_educativo_madre=PRIMARIA-0.316Lugar_residencia=ARMENIA+0.292Posicion_hermanos=SEGUNDO-0.28Nivel_educativo_padre=UNIVERSIDAD-0.26
0.6778 5 0.407Nivel_educativo_madre=SECUNDARIA-0.333Nivel_educativo_madre=UNIVERSIDAD-0.299Genero=HOMBRE+0.256Tipo_vivienda=FAMILIAR+0.249Nivel_educ
0.6312 6 0.385Nivel_educativo_madre=SECUNDARIA-0.315Nivel_educativo_padre=SECUNDARIA+0.313Nivel_educativo_padre=PRIMARIA-0.283Nivel_educativo_madre=ES
0.5872 7 -0.409Posicion_hermanos=SEGUNDO+0.38 coevaluacion-0.307Nivel_educativo_madre=ESPECIALIZACION-0.3Nivel_educativo_padre=ESPECIALIZACION+0.269
0.547 8 0.41 Posicion_hermanos=QUINTO-0.353Estado_civil=SEPARADO/DIVORCIADO+0.349Autoevaluacion+0.311Tipo_vivienda=FAMILIAR-0.258Lugar_residencia
0.5082 9 0.394Nivel_educativo_padre=MAESTRIA-0.383Estrato-0.282Lugar_residencia=CALARCA-0.262Nivel_educativo_padre=UNIVERSIDAD-0.249Heteroevaluacion
0.4704 10 -0.45Posicion_hermanos=CUARTO+0.316Nivel_educativo_madre=MAESTRIA-0.293egresado_colegio=PRIVADO-0.277Nivel_educativo_madre=UNIVERSIDAD+0.28
0.4339 11 -0.325Nivel_educativo_padre=UNIVERSIDAD+0.277Tipo_vivienda=PROPIA-0.273Tipo_vivienda=ARRENDADA+0.239Genero=HOMBRE+0.235Nivel_educativo_padr
0.4 12 -0.427Lugar_residencia=MONTENEGRO-0.339Lugar_residencia=CIRCASIA+0.297Lugar_residencia=ARMENIA-0.294Trabaja=NO-0.259Nivel_educativo_madre=
0.367 13 -0.374Heteroevaluacion-0.362Estado_civil=SEPARADO/DIVORCIADO-0.331Lugar_residencia=LA TEBAIDA-0.255Personas_Cargo+0.251Nivel_educativo_mad
0.3349 14 -0.527Lugar_residencia=CIRCASIA+0.285Lugar_residencia=MONTENEGRO+0.274Nivel_educativo_madre=UNIVERSIDAD+0.257Lugar_residencia=PIJAO+0.253T
0.303 15 -0.414Estado_civil=VIUDO+0.362Posicion_hermanos=OTRO-0.285Trabaja=NO-0.279Heteroevaluacion+0.273Nivel_educativo_padre=PRIMARIA...
0.2739 16 0.522Posicion_hermanos=OTRO+0.429Estado_civil=VIUDO+0.266Nivel_educativo_madre=MAESTRIA+0.253egresado_colegio=PRIVADO-0.249Heteroevaluacion
0.2485 17 0.373Nivel_educativo_padre=ESPECIALIZACION+0.335MejoraEvidencia+0.333Lugar_residencia=CALARCA-0.282Lugar_residencia=LA TEBAIDA+0.271Nivel
0.2242 18 0.533Posicion_hermanos=TERCERO-0.304Estado_civil=SEPARADO/DIVORCIADO-0.278Nivel_educativo_padre=MAESTRIA+0.24 egresado_colegio=PRIVADO-0.2
0.2007 19 -0.57Lugar_residencia=PIJAO+0.367Lugar_residencia=MONTENEGRO+0.243Personas_Cargo-0.24Estado_civil=VIUDO+0.229Posicion_hermanos=SEGUNDO...
0.1788 20 -0.475MejoraEvidencia-0.403Lugar_residencia=LA TEBAIDA+0.277Nivel_educativo_madre=MAESTRIA+0.239Posicion_hermanos=CUARTO-0.2Trabaja=NO...
0.1575 21 0.421Estado_civil=VIUDO-0.377Posicion_hermanos=OTRO-0.373Lugar_residencia=PIJAO-0.251Nivel_educativo_madre=PRIMARIA+0.227Lugar_residencia=
0.1374 22 0.332Nivel_educativo_madre=MAESTRIA+0.301egresado_colegio=PRIVADO+0.283Tipo_vivienda=FAMILIAR+0.276Lugar_residencia=PIJAO+0.27 Estrato...

```

Variables	
1	Edad
2	Estado civil
3	Lugar residencia
4	Posición hermanos
5	Trabaja
6	Nivel educativo madre
7	Nivel educativo padre
8	Heteroevaluación
9	egresado colegio
10	Estrato

Tabla 8: Evaluación de resultados PCA

Seleccionando 10 variables que generó el método PCA en la tabla 10 con un porcentaje de 91.91%.

7.3.4. Experimento 4: SELECCIÓN DE FACTORES CON REGRESIÓN LOGÍSTICA

El objetivo de este experimento es encontrar las variables que más influyen en el rendimiento académico mediante regresión logística.

Se realizó una categorización por medio de rangos mediante Dcretize en WEKA de la variable para el modelado.

Modelado: Simple Logistic, con un balanceo de datos y una validación cruzada de 10 grupos.

DETAILED ACCURACY BY CLASS

TP Rate	FP Rate	Precision	Recall	Class
0.833	0.73	0.714	0.833	(-inf : 2.76]
0.756	0.118	0.738	0.756	(2.76 : 3.32]
0.46	0.102	0.522	0.462	(3.32 : 3.88]
0.73	0.054	0.739	0.739	(3.38 : 4.44]
0.85	0.009	0.944	0.850	(4.44 : 5]
0.72	0.080	0.723	0.724	Weighted Avg
F-Measure	MCC	ROC Area	PRC Area	Class
0.769	0.717	0.924	0.841	(-inf : 2.76]
0.747	0.634	0.904	0.785	(2.76 : 3.32]
0.490	0.377	0.809	0.472	(3.32 : 3.88]
0.739	0.685	0.969	0.885	(3.38 : 4.44]
0.895	0.879	0.997	0.986	(4.44 : 5]
0.722	0.644	0.914	0.781	Weighted Avg

Variables relevantes:

- Personas a cargo
- Heteroevaluación
- Posición hermanos
- Mejora evidencia
- Autoevaluación
- Estrato
- Nivel educativo padre
- Nivel educativo madre
- Lugar de residencia
- Edad
- Egresado colegio
- género

7.3.3. EXPERIMENTO 5: REGLAS DE ASOCIACIÓN A PRIORI

La implementación del algoritmo Apriori en este estudio, involucra las medidas de confianza y soporte de cada regla. Las 10 reglas generadas fueron revisadas con el objetivo de descartar los siguientes problemas:

- a) **Reglas Redundantes:** Son reglas que poseen dos (2) o más características similares o que tienden a la deducción lógica simple.
- b) **Reglas Innecesarias:** Son reglas que no cumplen el objetivo del trabajo de grado ya que contienen consecuentes diferentes.
- c) **Reglas con una confianza baja:** Son aquellas reglas con medida probabilística de confianza menor del 0.65 (65%).

Se resalta que, si alguna regla no cumple con todos los criterios anteriormente mencionados, la regla es descartada.

Método: A priori, 10 reglas con un factor de confiabilidad de 0.5

- Personas Cargo='(-inf-0.9]' Heteroevaluación='(2.76-3.04]' 25 ==> Definitiva='(2.76-3.32]' 25 conf.:(1)
Si los estudiantes tienen entre 0 y 1 persona a cargo, y tiene Heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.3], la confiabilidad de esta regla en la prueba es del 100%.
- Estado civil=SOLTERO Personas Cargo='(-inf-0.9]' Heteroevaluación='(2.76-3.04]' 25 ==> Definitiva='(2.76-3.32]' 25 conf.:(1)
Si los estudiantes tienen un estado civil soltero, persona a cargo entre 0 y 1, y tiene Heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.2], la confiabilidad de esta regla en la prueba es del 100%.
- Estado civil=SOLTERO egresado colegio=PUBLICO Heteroevaluación='(2.76-3.04]' 23 ==> Definitiva='(2.76-3.32]' 23 conf.:(1)
Si los estudiantes tienen un estado civil soltero, son egresados de un colegio público, y tiene Heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.2], la confiabilidad de esta regla en la prueba es del 100%.
- Trabaja=NO Heteroevaluación='(2.76-3.04]' 22 ==> Definitiva='(2.76-3.32]' 22 [conf:\(1\)](#)
Si los estudiantes trabajan y tiene Heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.2], la confiabilidad de esta regla en la prueba es del 100%.
- Personas Cargo='(-inf-0.9]' Lugar residencia=ARMENIA Heteroevaluación='(2.76-3.04]' 22 ==> Definitiva='(2.76-3.32]' 22 [conf:\(1\)](#)
Si los estudiantes tienen entre 0 y 1 persona a cargo, su lugar de residencia es en Armenia y tiene Heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.3], la confiabilidad de esta regla en la prueba es del 100%.
- Estado civil=SOLTERO Personas Cargo='(-inf-0.9]' Lugar residencia=ARMENIA Heteroevaluación='(2.76-3.04]' 22 ==> Definitiva='(2.76-3.32]' 22 [conf:\(1\)](#)
Si los estudiantes tienen un estado civil soltero, persona a cargo entre 0 y 1, su lugar de residencia es Armenia, y tienen una heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.2], la confiabilidad de esta regla en la prueba es del 100%.
- Nivel educativo madre=SECUNDARIA heteroevaluación='(2.76-3.04]' 21 ==> Definitiva='(2.76-3.32]' 21 [conf:\(1\)](#)
Si los estudiantes tienen un nivel educativo de la madre Secundaria y tienen una heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.2], la confiabilidad de esta regla en la prueba es del 100%.
- Estado civil=SOLTERO Trabaja=NO heteroevaluación='(2.76-3.04]' 21 ==> Definitiva='(2.76-3.32]' 21 [conf:\(1\)](#)

Si los estudiantes tienen un estado civil soltero, No trabajan, y tienen una heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.2], la confiabilidad de esta regla en la prueba es del 100%.

- Trabaja=NO Lugar residencia=ARMENIA heteroevaluación='(2.76-3.04]' 21 ==> Definitiva='(2.76-3.32]' 21 [conf:\(1\)](#)

Si los estudiantes no trabajan, su lugar de residencia es Armenia y tienen una heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.2], la confiabilidad de esta regla en la prueba es del 100%.

- Estado civil=SOLTERO Trabaja=NO Lugar residencia=ARMENIA heteroevaluación='(2.76-3.04]' 21 ==> Definitiva='(2.76-3.32]' 21 [conf:\(1\)](#)

Si los estudiantes son solteros, No trabajan, su lugar de residencia es Armenia y tienen una heteroevaluación (2.7 – 3.04]; entonces tendrán una nota definitiva entre (2.7 – 3.2], la confiabilidad de esta regla en la prueba es del 100%.

El objetivo de este experimento es observar las variables que más influyen en el rendimiento académico mediante el algoritmo A priori.

Variables	
1	Trabaja
2	Personas a cargo
3	Estado civil
4	Lugar de residencia
5	Nivel educativo madre
6	Tipo vivienda
7	Egresado colegio

8. Resultado de las técnicas

Después de aplicar los diferentes algoritmos de minería de datos, se realiza una evaluación con los resultados anteriores si es posible llegar a una conclusión referente a que las variables pueden influir en el rendimiento académico.

En la tabla 9, se presentan todos los resultados que se generaron de cada uno de los algoritmos utilizados en la minería de datos, esto con la finalidad de encontrar patrones que permitan concluir que factores tienen más influencia en el rendimiento académico de cursos bajo la metodología de proyectos formativos.

VARIABLE	EXPERIMENTOS DE MINERÍA DE DATOS					
	CfsSubset Eval	Arboles de decisión	PCA	Regresión logística	Reglas de asociación	Frecuencia
Definitiva (clase)						
Heteroevaluación	x	x	x	x		4
Género		x		x		2
coevaluación	x					1
Posición hermanos	x		x	x		3
Edad			x	x		2
Autoevaluación	x			x		2
Tipo vivienda	x				x	2
Mejora Evidencia	x			x		2
Lugar residencia			x	x	x	3
Nivel educativo padre			x	X		2
Nivel educativo madre			x	X	x	3
Estado civil			x		x	2
Estrato			x	x		2
Egresado colegio		x	x	x	x	4
Personas Cargo		x		x	x	3
ICFES						0
Trabaja	x		x		x	3
Sexo	x					1

Tabla 9: Resultado Final.

A continuación, se presentan en orden mayor frecuencia los 7 factores más importantes según los experimentos realizados, los que más influyen en el rendimiento académico:

- **Factor 1: Heteroevaluación**
- **Factor 2: Egresado Colegio**
- **Factor 3 Posición hermano**
- **Factor 4: lugar de residencia**
- **Factor 5: nivel educativo madre**
- **Factor 6: trabaja**
- **Factor 7: personas a cargo**

En el anexo 7 se puede ver en detalle el diagrama de dispersión de los clústeres, además de todo el proceso que se realizó en la implementación de las técnicas de minería de datos.

8.3.3. Evaluación y Análisis

Inicialmente se realizaron las pruebas normalidad y homocedasticidad como se presentan en el anexo 8 donde se realizan las pruebas de normalidad de residuales y se muestra porque no se utilizó la Anova.

Para realizar la validación y proponer las variables de mayor incidencia en el proceso de aprendizaje de un estudiante en un curso por proyectos formativos, se complementó el análisis de los resultados de las técnicas de minería de datos, con estadística no paramétrica. El propósito de esta estadística es identificar la posible incidencia de las variables que interviene en el estudio. En este estudio se aplicaron pruebas no paramétricas de Mann-Whitney para muestras independientes para todas las variables categóricas y no categóricas. Además, esta técnica se aplicó teniendo en cuenta que no requiere el cumplimiento de los supuestos estadísticos de normalidad e igualdad de varianzas. A continuación, se presentan las hipótesis a estudiar en este trabajo:

La primera hipótesis que estudiar fue: H_1 : “Existe diferencia estadísticamente significativa de los resultados del rendimiento académico en función del género”.

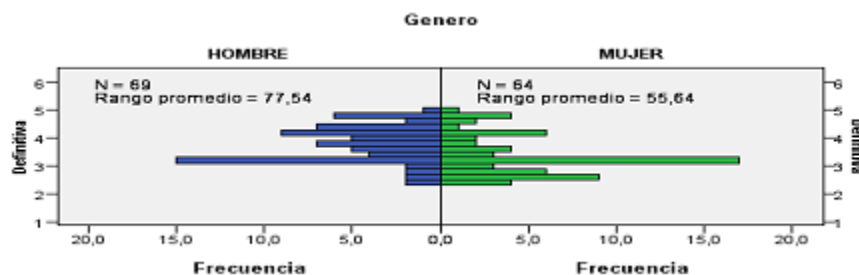


Ilustración 16: Resultado Prueba Mann-Whitney para el género y el rendimiento académico

Inicialmente se analizó la posible incidencia de los resultados del rendimiento académico, con relación a la variable género. Los resultados de la prueba no paramétrica de Mann-Whitney generaron un p-valor = 0.002, a partir de lo cual se puede afirmar que la variable género incide en el rendimiento de los estudiantes. La prueba de rango promedio muestra que los hombres tienen un mejor rendimiento académico con relación a la mujer, en la Ilustración No 12 se presenta el resultado de la prueba no paramétrica.

La segunda hipótesis analizó la relación de la autoevaluación con el rendimiento académico: H_2 : “Existe diferencia estadísticamente significativa de los resultados del rendimiento académico en función de la autoevaluación”.

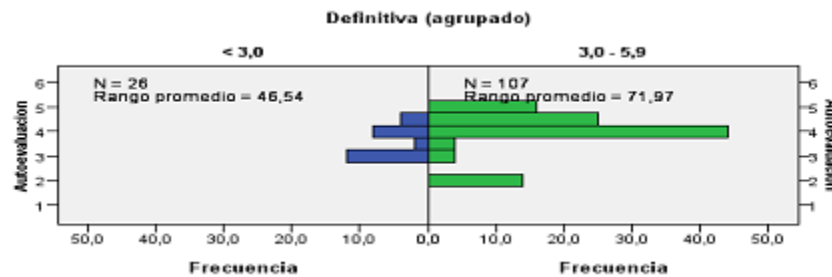


Ilustración 17: Resultado Prueba Mann-Whitney entre autoevaluación y rendimiento académico

A continuación, se analizó la posible incidencia de la autoevaluación con el rendimiento académico. El resultado de la prueba no paramétrica generó un p-valor = 0.0023, a partir de lo cual se puede afirmar que la variable autoevaluación también tiene incidencia sobre el rendimiento académico de los estudiantes. La prueba de rango promedio muestra que los estudiantes que tienen una mejor percepción de la autoevaluación también tienen un mejor desempeño académico. En Ilustración No 12, se presenta los resultados de la prueba de rango promedio entre la autoevaluación y el rendimiento académico.

Adicionalmente se realizó un análisis factorial a las características consideradas en aspectos didácticos ver detalle en la Ilustración No 8, el cual agrupó en categorías las variables adicionales de la opinión de los estudiantes. El resultado de este estudio incluye algunas de las variables que se tuvieron en cuenta para el análisis de minería de datos. Para el análisis factorial, se verificaron los supuestos estadísticos, se realizó la extracción de las categorías mediante la rotación de componentes y finalmente se analizó cada uno de los componentes.

Los supuestos estadísticos indican una alta correlación entre las variables, la prueba de esfericidad Bartlett indica un p-valor = 0.000, por tanto, la matriz de correlaciones no es una matriz identidad. La prueba KMO presenta un valor de 0.807, por tanto, es posible realizar el análisis factorial. Finalmente, el índice de adecuación muestra de la matriz anti-imagen presenta para todos los casos, coeficientes superiores a 0.72, por tanto, la muestra es adecuada para el análisis factorial. La estructura de los componentes del análisis factorial muestra que, a través de los cuatro primeros componentes, se puede explicar el 66.4% de la variabilidad total.

Después de identificar los componentes, se realizó un análisis estadístico entre el rendimiento académico y cada componente deducido del análisis factorial, es decir: evaluación, participación, diseño del curso en Moodle y coherencia con las actividades. Estas características fueron presentadas en la Ilustración No 8.

La tercera hipótesis analizó la relación de la categoría evaluación con el rendimiento académico: H_3 : *“Existe diferencia estadísticamente significativa de los resultados del rendimiento académico en función de la evaluación”*.

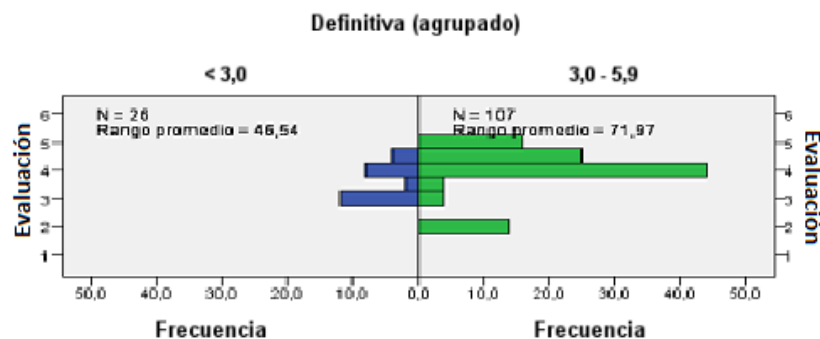


Ilustración 18: Resultado Categoría Evaluación Ward.

Dada la característica de los factores, para todos los casos se aplicó la prueba no paramétrica de Mann-Whitney para muestras independientes. Inicialmente se analizó la categoría evaluación ver Ilustración No 8, el resultado de la prueba muestra un p-valor = 0.494, a partir de lo cual se puede decir que el factor evaluación no tiene incidencia sobre el rendimiento académico del estudiante. En la Ilustración No 13, se presenta los resultados de la prueba de rango promedio entre la evaluación y el rendimiento académico.

La cuarta hipótesis analizó la relación de la categoría participación con el rendimiento académico: H_4 : "Existe diferencia estadísticamente significativa de los resultados del rendimiento académico en función de la participación".

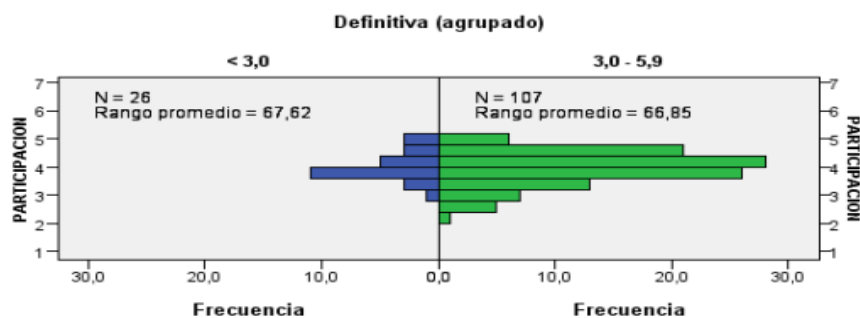


Ilustración 19: Resultado Categoría Participación

La siguiente variable fue la participación del estudiante en las actividades de formación y cuyo p-valor= 0.132, esto permite afirmar que esta variable no tiene incidencia en el rendimiento académico de los estudiantes. En la Ilustración No 14, la prueba de rango media determina que no existe relación significativa entre la participación y el rendimiento académico.

La quinta hipótesis analizó H_5 : "Existe diferencia estadísticamente significativa de los resultados del rendimiento académico en función del Diseño del curso en Moodle".

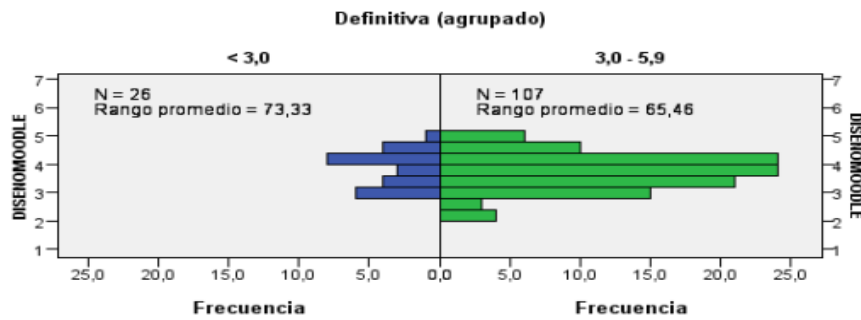


Ilustración 20: Resultado Categoría Diseño Moodle

Con relación al diseño del curso, la prueba no paramétrica generó un p-valor = 0,927, con lo que se indica que no existe una incidencia relevante entre la opinión de los estudiantes con respecto al diseño del curso y su rendimiento académico. En la Ilustración No 15, se presenta la prueba de rango promedio, en la cual se identifica un indicio en el cual los estudiantes con menor rendimiento académico son los que tienden a valorar mejor el diseño de la plataforma Moodle.

La sexta hipótesis analizó H_6 : "Existe diferencia estadísticamente significativa de los resultados del rendimiento académico en función de la coherencia de las actividades de aprendizaje".

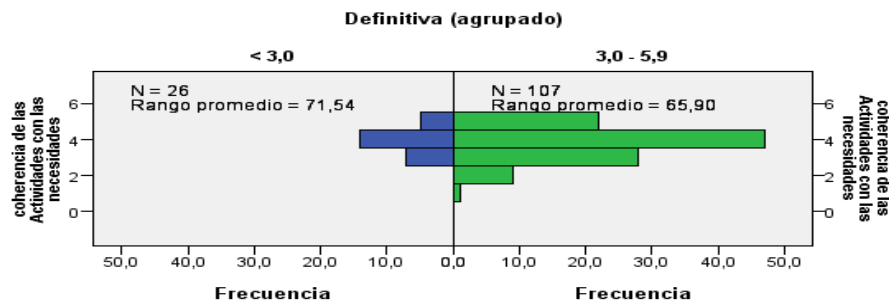


Ilustración 21: Resultado Categoría Coherencia con las actividades

La prueba no paramétrica generó para el factor utilidad de la plataforma Moodle un p-valor = 0.348 y para la categoría coherencia de las actividades de aprendizaje, un p-valor= 0.474, a partir de lo cual se puede afirmar que ninguna de ellas tiene incidencia en el rendimiento académico de los estudiantes. En la Ilustración No 16, se presenta el resultado de la prueba de rango promedio para la categoría coherencia de las actividades de aprendizaje.

Con base en los anteriores resultados, fue posible identificar las variables que inciden de forma significativa en el rendimiento académico de un estudiante, cuando se aplica la metodología de proyectos formativos. La implementación del análisis estadístico complementario se puede observar en la hipótesis H_1 y H_2 las variables género y autoevaluación presentan incidencia sobre el rendimiento académico.

VARIABLES FINALES

Luego de los análisis en la implementación de las técnicas de minería de datos y el análisis estadístico complementario, se encontró que las siguientes variables tienen mayor incidencia en el rendimiento académico de los estudiantes y que por lo tanto se proponen para ser consideradas en el diseño de modelos de estudiantes para sistemas adaptativos:

VARIABLE	EXPERIMENTOS DE MINERÍA DE DATOS
----------	----------------------------------

	Frecuencia Técnicas de minería de datos	Análisis estadístico complementario
Definitiva (clase)		
Heteroevaluación	4	
Género	1	X
coevaluación	1	
Posición hermanos	3	
Edad	2	
Autoevaluación	2	
Tipo vivienda	2	
Mejora Evidencia	2	
Lugar residencia	3	
Nivel educativo padre	2	
Nivel educativo madre	3	
Estado civil	2	
Estrato	2	
Egresado colegio	4	
Personas Cargo	3	
ICFES	0	
Trabaja	3	
Sexo	1	
Autoevaluación		X

Tabla 10: Comparativo Resultados Variables finales

Como se puede observar en la tabla No. 8, las variables que resultan de ambos estudios son género y autoevaluación, por lo cual, considerando los trabajos previos de Cardona (2017), Hernández, Tobón, Ortega and Ramírez (2018), podrían ser las de mayor relevancia para el diseño de modelos de usuario en sistemas adaptativos. Sin embargo, se sugiere también considerar el resto de las

variables contenidas en dicha tabla con el propósito de someterlas a nuevos estudios que nos permitan tener más elementos de juicio.

9. Conclusiones

En este trabajo se aplicaron técnicas de minería de datos, las cuales proporcionaron herramientas para determinar las características de mayor incidencia en el rendimiento académico de los estudiantes.

La aplicación de las técnicas de minería de datos siguió la metodología CRISP-MD, la cual permitió realizar el análisis de un conjunto de datos de los cursos, bajo la metodología proyectos formativos en un ambiente de aprendizaje virtual adaptativo. De esta manera se obtuvieron las variables, que, según el estudio realizado, fueron las de mayor incidencia en el rendimiento académico de los estudiantes en formación y por lo tanto deberían ser consideradas en el diseño de modelos de usuarios en sistemas adaptativos.

De la misma manera, los resultados de este trabajo se orientaron en primer lugar a establecer la posible relación entre el proceso de aprendizaje de un curso y el rendimiento académico de un estudiante. Los resultados generados muestran una correlación positiva entre el rendimiento académico, la heteroevaluación obtenida durante el proceso de formación y el tipo de colegio del cual es egresado el estudiante. Con lo anterior se podría indicar la importancia guiar el proceso en la formación de un curso representado por su heteroevaluación para obtener mejores resultados al finalizar el curso.

En segundo lugar, se mostró que la variable posición entre hermanos, lugar de residencia, nivel educativo de la madre, trabaja y personas a cargo puede tener incidencia con el rendimiento académico, para este caso de estudio bajo la metodología de proyectos formativos. Del mismo modo, se identificó que los estudiantes que tenían una mejor heteroevaluación, además que sean de un colegio público y su posición entre hermanos es primero para este estudio presentan mejor rendimiento académico.

Por último, dado que el objetivo de este trabajo era establecer las características de mayor incidencia en el rendimiento académico de los estudiantes, consideradas en la implementación de las técnicas de minería y la aplicación estadística al conjunto de datos iniciales. Los resultados muestran que existe diferencia estadísticamente significativa en la variable autoevaluación y el rendimiento académico, además es coherente con los resultados de las técnicas de minerías de datos.

Como se presenta en el trabajo de Cardona (2017), el cual mostró la autoevaluación tuvo un papel fundamental para los estudiantes y su proceso de aprendizaje y afirma que los estudiantes generaron capacidades de autorregulación. Del mismo modo, Hernández, Tobón, Ortega and Ramírez (2018) presentan es su trabajo que, para tener un mejor rendimiento académico en el proceso de formación, es necesario que el estudiante realice evaluaciones de sus entregas a partir de criterios definidos durante el proceso definido autoevaluación.

En aras de tener mayores elementos de juicio, se sugiere realizar nuevos estudios que aporten a la veracidad de los resultados y evaluar la consistencia en otros escenarios. Puesto que, la autoevaluación es un proceso elemental para el desarrollo del pensamiento y la construcción de la sociedad de conocimiento. Además, existen muchos sistemas de formación en línea que son basados en la autoevaluación por lo cual se debe continuar con la investigación en el tema (Hernández, Tobón and Guzmán, 2015).

Aunque en estudios previos no se han encontrado resultados de minería de datos relacionados con la heteroevaluación, posición entre hermanos, la autoevaluación y el rendimiento académico; en este trabajo se identifican posibles incidencias entre dichas variables. Es necesario realizar otros estudios y así poder establecer condiciones donde se puedan tener elementos en común y poder hacer dichas comparaciones.

10. Trabajos futuros

- Como trabajo futuro, se propone analizar los datos de otros cursos basados en la metodología de proyectos formativos, con objetivo de aumentar la cantidad de datos y observar la correlación existente entre las variables
- Se espera un mejoramiento incremental de los perfiles con nuevas técnicas de minería de datos y poder adicionar nuevas variables al modelo.

11. Referencias

- Betancur, Daniel, Julián Moreno, and Demetrio A. Ovalle C. 2009. "Modelo Para La Recomendación Y Recuperación De Objetos De Aprendizaje En Entornos Virtuales De Enseñanza/Aprendizaje." *Avances en Sistemas e Informática* 6(1): 45–56. <http://revistas.unal.edu.co/index.php/avances/article/view/14443%5Cnhttp://redalyc.uaemex.mx/pdf/1331/133112608005.pdf>.
- Biemans, Harm et al. 2005. "Competence-Based VET in the Netherlands : Background and Pitfalls." *Vocational Education and Training* 56(4): 523–38.
- Bogarín Vega, Alejandro, Cristóbal Romero Morales, and Rebeca Cerezo Menéndez. 2015. "Aplicando Minería de Datos Para Descubrir Rutas de Aprendizaje Frecuentes En Moodle Applying Data Mining to Discover Common Learning Routes in Moodle 73." *Revista de Educación Mediática y TIC* 5(1): 73–92.
- Boticario, J, O Santos, and P Van Rosmalen. 2006. "Issues in Developing Adaptive Learning Management Systems for Higher Education Institutions." In *ADALE Workshop. Adaptive Hypermedia 2006 Conference*, Dublin, 55–71.
- Bouckaert, Remco, Eibe Frank, and Mark Hall. 2010. "WEKA Manual for Version 3-7-13." *WEKA Manual for Version 3-7-13*: 327.
- Brusilovsky, Peter. 1994. "The Construction and Application of Student Models in Intelligent Tutoring Systems." *The construction and application of student models in intelligent tutoring systems* 32(1): 70–89.
- . 2004. "A Distributed Architecture for Adaptive E-Learning." *Proceedings of the 13th international world wide web conference (www 2004 Proceedings of the 13th international World Wide Web conference)*: 104–13. <http://www.sis.pitt.edu/~peterb/papers/www2004final.html>.
- Bull, Susan, and Mark McKay. 2004. "An Open Learner Model for Children and Teachers: Inspecting Knowledge Level of Individuals and Peers" eds. James C Lester, Rosa Maria Vicari, and Fábio Paraguacu. *Intelligent Tutoring Systems (ITS'2004)*: 646–55.
- Bunk, G. 1994. "La Transmisión de Las Competencias En La Formación y Perfeccionamiento Profesionales de La RFA." *Revista Europea* 1: 8–14.
- Cardona, Sergio. 2017. "Metodología Para La Evaluación de Competencias Soportada En Un Ambiente de Aprendizaje Virtual Adaptativo." Universidad Pontificia Bolivariana.
- Cardona, Sergio, Jeimy Vélez, and Sergio Tobón. 2013. "Metodología de Proyectos Formativos Aplicada a Un Curso de Lógica Matemática." *8th Latin American Conference on Learning Objects and Technologies LACLO* 4(1): 1–12.
- Casales, Reynaldo, José Rojas, and Grismilda Paulí. 2008. "Algunas Experiencias Didácticas En El Entorno de La Plataforma Moodle." *Revista de Informática Educativa y Medios Audiovisuales* 5(10): 1–10.
- Cataldi, Zulma, and Fernando J Lage. 2010. "Modelado Del Estudiante En Sistemas Tutores Inteligentes." *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología* 5: 29–38.
- Corso, Cynthia Lorena, and Sofía Lorena Alfaro. 2010. "Minería de Uso Web: Presentación de Caso Práctico Para El Mejoramiento de La Educación, Basada En El Uso de Plataformas Virtuales." *XII Workshop de Investigadores en Ciencias de la Computación*: 177–82. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0> (November 24, 2014).
- Díaz-barriga, Ángel. 2006. "El Enfoque de Competencias En La Educación. ¿Una Alternativa o Un Disfraz de Cambio?" *Perfiles educativos* XXVIII(111): 7–36.

- Duque, Néstor D, Demetrio A Ovalle, and Jovani A Jiménez. 2007. "Modelo Adaptativo Para Cursos Virtuales Basado En Técnicas de Planificación Inteligente Adaptive Model for Virtual Courses Based on Intelligent Planning Techniques." 4(1): 39–46.
- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. 1996. "From Data Mining to Knowledge Discovery in Databases." *AI Magazine* 17(3): 37.
- Ferreira, Anita, Pedro Salcedo, Gabriela Kotz, and Fernanda Barrientos. 2012. "La Arquitectura de ELE-TUTOR: Un Sistema Tutorial Inteligente Para El Español Como Lengua Extranjera." *Revista Signos* 45(79): 102–31. <http://cm.redalyc.org/articulo.oa?id=157023091001%5Cnhttp://www.redalyc.org/pdf/1570/157023091001.pdf%5Cnhttp://www.redalyc.org/9081/articulo.oa?id=157023091001>.
- Gibbons, Robert. 1980. "On Competence: A Critical Analysis of Competence-Based Reforms in Higher Education." *The Journal of Higher Education* 51(6): 695–97.
- González, H. M, N. D Duque, and D. A Ovalle. 2008. "Modelo Del Estudiante Para Sistemas Adaptativos de Educación Virtual." *Revista Avances en Sistemas e Informática* 5(1): 199–206. http://www.researchgate.net/publication/28230402_Modelo_del_estudiante_para_sistemas_adaptativos_de_educacin_virtual/file/60b7d5233239bd868e.pdf (November 24, 2014).
- González, H M, and Duque. 2008. "Modelo Del Estudiante Para Sistemas Adaptativos de Educación Virtual Student Model for Adaptive Systems of Virtual Education." *Avances en Sistemas e Informática*. <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:No+Title#0> (November 24, 2014).
- Gouli, Evangelia, Kyprisia Papanikolaou, and Maria Grigoriadou. 2002. "Personalizing Assessment in Adaptive Educational Hypermedia Systems." *LECTURE NOTES IN COMPUTER SCIENCE* 2347: 153–63. <http://www.springerlink.com/content/6bf0t8uat75n8kvf/>.
- Graf, Sabine. 2009. "Adaptivity in Learning Management Systems Focussing on Learning Styles." *Proceedings of the 2009 IEEE/WIC/ACM International ...* 3(December 2007): 235–38. <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5284970%5Cnhttp://dl.acm.org/citation.cfm?id=1632300>.
- Graf, Sabine, and Kinshuk. 2009. "Advanced Adaptivity in Learning Management Systems by Considering Learning Styles." In *Proceedings - 2009 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT Workshops 2009*, Ieee, 235–38.
- Gutiérrez García, José Alberto. 2016. "Comenzando Con Weka: Filtrado y Selección de Subconjuntos de Atributos Basada En Su Relevancia Descriptiva Para La Clase." (September).
- Hofmann, M, and R Klinkenberg. 2013. *RapidMiner: Data Mining Use Cases and Business Analytics Applications*. CRC Press.
- Huapaya, Constanza R et al. 2012. "Minería de Datos Educacional En Ambientes Virtuales de Aprendizaje Resumen Contexto Introducción." : 996–1000.
- Jácome León, José G. 2017. "Minería de Datos Para Proponer Un Modelo Didáctico Estructural de Aprendizaje En La Facultad Ciencias Administrativas y Económicas de La Universidad Técnica Del Norte. Maestría En Gestión de Sistemas de Información e Inteligencia de Negocios." Universidad de las Fuerzas Armadas ESPE. <http://repositorio.espe.edu.ec/handle/21000/12791>.
- Kohavi, Ron, and George H. John. 1997. "Wrappers for Feature Subset Selection." *Artificial Intelligence* 97(1–2): 273–324. <http://www.sciencedirect.com/science/article/pii/S000437029700043X>.
- Kritikou, Yiouli et al. 2008. "User Profile Modeling in the Context of Web-Based LearniKritikou, Y.,

- Demestichas, P., Adamopoulou, E., Demestichas, K., Theologou, M., & Paradia, M. (2008). User Profile Modeling in the Context of Web-Based Learning Management Systems. *Journal of Network and Computer Applications* 31(4): 603–27. <http://www.sciencedirect.com/science/article/pii/S1084804507000720>.
- Langley, Pat (Institute for the Study of Learning and Expertise). 1994. "Selection of Relevant Features in Machine Learning." *In Proceedings of the AAAI Fall Symposium on Relevance*: 140–44.
- Limongelli, Carla, Filippo Sciarrone, and Giulia Vaste. 2011. "Personalized E-Learning in Moodle: The Moodle_LS System." *Journal of E-Learning and Knowledge Society* 7(1): 49–58.
- López-Moratalla, J. 2005. "Desarrollo de La Adaptatividad En Un Sistema de E-Learning Basado En Estándares IMS." *Simposium ...*: 16–18. <http://www.niee.ufrgs.br/eventos/SIIE/2005/PDFs/Comunica%E7%F5es/c275-Moratalla.pdf> (November 24, 2014).
- Luzcando, Diego Riofrío. 2017. "Propuesta de Un Modelo de Comportamiento Colectivo de Estudiantes Para Un Sistema Inteligente de Tutoría Dirigido Al Entrenamiento Procedimental Diego Riofrío Luzcando."
- Martínez, David et al. 2016. "Descubrimiento de Perfiles de Rendimiento Estudiantil. Un Modelo de Integración de Datos Académicos y Socioeconómicos." *Campus Virtuales* 5(2): 70–83. <https://dialnet.unirioja.es/servlet/articulo?codigo=5696008>.
- McClelland, D.C. 1973. "Testing for Competence Rather than for 'Intelligence'." *The American psychologist* 28(1): 1–14.
- McCormick, Keith et al. 2013. *IBM SPSS Modeler Cookbook*. Packt Publishing ©2013. <https://www.ibm.com/es-es/marketplace/spss-modeler>.
- Méndez, ND Duque. 2009. "Modelo Adaptativo Multi-Agente Para La Planificación y Ejecución de Cursos Virtuales Personalizados." <http://www.bdigital.unal.edu.co/2220/> (November 24, 2014).
- Millán Valldeperas, Eva. 2000. System and Computer Thesis "Sistema Bayesiano Para Modelado Del Alumno." Universidad de Málaga.
- Moine, Juan Miguel, Ana Haedo, and Silvia Gordillo. 2011. "Estudio Comparativo de Metodologías Para Minería de Datos." *XIII Workshop de Investigadores en Ciencias de la Computación*: 278–81. <http://sedici.unlp.edu.ar/handle/10915/20034>.
- Molina, José, and Jesús García. 2008. "Técnicas de Minería de Datos Basadas En Aprendizaje Automático." *Técnicas de Análisis de Datos*: 96–266.
- Monsalve, Julián, Fredy Aponte, and Jorge Hoyos. 2013. "Aplicación de Minería de Datos Educativos a Procesos B-Learning." *Conferencias LACLO* 4(1). <http://www.laclo.org/papers/index.php/laclo/article/view/75> (November 23, 2014).
- Morales, Cristóbal Romero, Sebastián Ventura Soto, and Cesar Hervás Martínez. 2005. "Estado Actual de La Aplicación de La Minería de Datos a Los Sistemas de Enseñanza Basada En Web." *III Taller Nacional de Minería de Datos y Aprendizaje*: 49–56. <http://www.lsi.us.es/redmidas/CEDI/papers/189.pdf> (September 5, 2014).
- Morales, P, B Urosa, and A Blanco. 2003. *Construcción de Escalas de Actitudes Tipo Likert. Una Guía Práctica*. Madrid: La muralla.
- Olmos, S. 2008. "Evaluación Formativa y Sumativa de Estudiantes Universitarios: Aplicación de Las Tecnologías a La Evaluación Educativa." Universidad de Salamanca.
- Paramythis, Alexandros, and S Loidl-Reisinger. 2004. "Adaptive Learning Environments and E-Learning Standards." *Electronic Journal of Elearning* 2(1): 181–194. <http://books.google.com/books?hl=en&lr=&id=nthPcVjs2b0C&oi=fnd&pg=PA369&dq=Adaptive>

+Learning+Environments+and+eLearning+Standards&ots=A9Sr-dRGSX&sig=MCJIQB-AsT9Dyau5LBZHTSLE3Wc.

- Petropoulou, Ourania, Katerina Kasimatis, Ioannis Dimopoulos, and Symeon Retalis. 2014. "LAe-R: A New Learning Analytics Tool in Moodle for Assessing Students' Performance." *Bulletin of the IEEE Technical Committee on Learning Technology* 16(1): 1–13. <http://www.ieeetclt.org/issues/january2014/FullIssue.pdf> (October 17, 2014).
- Pushkarev, Val, Henry Neumann, Cihan Varol, and John R Talburt. 2010. "An Overview of Open Source Data Quality Tools." *University of Arkansas at Little Rock*: 370–76.
- Salcines, Enrique García, Cristóbal Romero Morales, Sebastián Ventura, and Carlos De Castro. 2008. "Sistema Recomendador Colaborativo Usando Minería de Datos Distribuida Para La Mejora Continua de Cursos E-Learning." 3: 19–30.
- Schiaffino, Silvia, Patricio Garcia, and Analia Amandi. 2008. "ETeacher: Providing Personalized Assistance to e-Learning Students." *Computers & Education* 51(4): 1744–54.
- Sciences, Organizational, and Jove Ili. 2012. "Providing Adaptivity in Moodle LMS Courses Srdjan Krčić o Adaptive E-Learning Systems." *Educational technology & society* 15(1): 326–38.
- Sloep, Peter B., and Adriana José Berlanga Flores. 2011. "Redes de Aprendizaje, Aprendizaje En Red." *Comunicar: Revista científica iberoamericana de comunicación y educación* 19(37): 55–63. <http://dialnet.unirioja.es/servlet/articulo?codigo=3733909&info=resumen&idioma=ENG>.
- Talavera, Luis, and Elena Gaudioso. 2004. "Mining Student Data to Characterize Similar Behavior Groups in Unstructured Collaboration Spaces." *Proceedings of Workshop on Artificial intelligence in CSCL*: 17–23.
- Tobon, S. (2017). Essential axes of knowledge society and socioformation. Mount Dora: Kresearch. doi: [dx.doi.org/10.24944/isbn.978-1-945721-19-9](https://doi.org/10.24944/isbn.978-1-945721-19-9)
- Velez, Jeimy. 2009. "Entorno de Aprendizaje Virtual Adaptativo Soportado Por Un Modelo de Usuario Integral." *Universitat de Girona*.
- Veléz, Jeimy. 2009. "Entorno de Aprendizaje Virtual Adaptativo Soportado Por Un Modelo de Usuario Integral (Tesis de Doctorado)." *Universitat de Girona*: 164. <http://dugi-doc.udg.edu/handle/10256/4414> (November 24, 2014).
- Witten, Ian H, Eibe Frank, and Mark a Hall. 2011. Complementary literature None *Data Mining: Practical Machine Learning Tools and Techniques (Google EBook)*. <http://books.google.com/books?id=bDtLM8CODsQC&pgis=1>.
- Zaiane, Omar R. 2001. "Web Usage Mining for a Better Web-Based Learning Environment." *Proceedings of Conference on Advanced Technology for Education*: 60–64. http://link.springer.com/chapter/10.1007/978-3-540-72079-9_3.

12. ANEXOS A

Ponencia en eventos

- I Congreso Internacional de Tecnología, Ciencia y Educación para el Desarrollo Sustentable
Ponente con artículo "*Análisis de variables de estudiantes en un curso bajo la metodología de proyectos formativos*", Cuernavaca, Morelos, México Noviembre 2017.

Publicaciones en Revistas

- Osorio, L, Vélez, J, & Cardona, S. (2018). *Análisis de variables de estudiantes en un curso bajo la metodología de proyectos formativos*. Revista Espacios, ISSN 0798-1015 (**En proceso de aceptación**)

13. ANEXOS B

- Anexo 1: Variables consideradas para el análisis
- Anexo 2: Encuesta metodología de proyectos formativos
- Anexo 3: Preparación y selección de atributos
- Anexo 4: Técnicas de minería
- Anexo 5: Análisis estadístico

Anexo 1: Variables consideradas para el análisis

El conjunto de datos de este estudio correspondió a estudiantes de la institución de Administración y Mercadotecnia (EAM), en el programa de Ingeniería de Software, en cursos del área de programación de computadores bajo la metodología de proyectos formativos. El número de individuos del estudio empírico fue de 262, los estudiantes pertenecían al segundo semestre de ingeniería de software, la intervención tuvo una duración de 12 semanas académicas, durante el segundo periodo del 2016. Como se presenta en el anexo 6.

Se analizaron algunas de las características y variables propuestas por Cardona (2017), así mismo, se consideraron variables adicionales que posiblemente pueden tener incidencia sobre el rendimiento académico del estudiante, ver figura No 1.

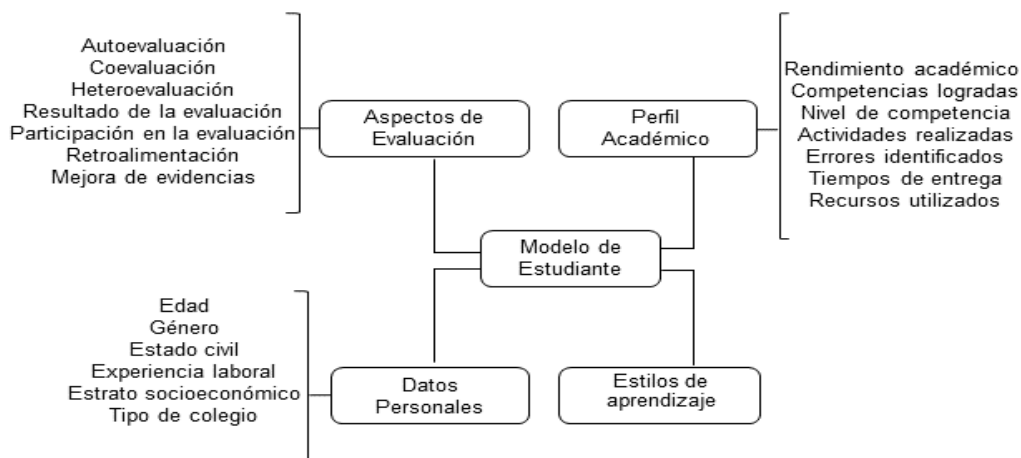


Figura 1: Variables consideradas para el análisis.(Cardona, 2017)

1.1. Dominio de los atributos

Nombre	Descripción	Tipo
Rendimiento académico	Nota definitiva de los estudiantes en el curso	Numérico
Datos Personales		
Edad	Edad	Numérico
Género	Masculino Femenino	Nominal
Estrato	0 - Bajo 1 - Medio bajo 2 - Medio 3 - Medio Alto 4 - Alto 5 - Alto-alto	Nominal
Estado_civil	Soltero Casado/Unión libre Separado / Divorciado Viudo	Nominal
Trabaja	Si No	Nominal
Personas_Cargo	Número de personas a cargo	Numérico

Repitente	Si No	Nominal
Promedio Universidad	Esta variable permite tener en cuenta el promedio que llevan los estudiantes en carreras anteriores o en el transcurso de su carrera.	Numérico
egresado_colegio	Publico Privado	Nominal
promedio_Colegio	Promedio del alumno en el colegio	Numérico
ICFES	Examen	Numérico
Lugar_residencia	Armenia Calarcá Circasia La tebaida Montenegro Pijao	Nominal
Nivel_edcativo_madre	Primaria Secundaria Universidad Especialización Maestría	Nominal
Nivel_eduativo_padre	Primaria Secundaria Universidad Especialización Maestría	Nominal
Numero_hermanos	Numero de hermanos	Numérico
Posición_hermanos	Primero Segundo Tercero Cuarto Quinto Posterior	Nominal
Tipo_vivienda	Arrendada Familiar Propia	Nominal
Cumplimiento de los elementos de proyectos formativos		
Contribución Metodología PF	Selección de 1 – 5 La metodología de proyectos formativos contribuye al desarrollo de las competencias a formar	Numérico
Complejidad del PF	Selección de 1 – 5 El proyecto formativo se enfoca en la solución de un problema del contexto y su nivel de reto está acorde con mi nivel de formación	Numérico
Participación en el PF	Selección de 1 – 5 Participo en la definición del problema del contexto del proyecto formativo	Numérico
Realiza las actividades del PF	Selección de 1 – 5 Realizo las actividades correspondientes a cada una de las fases del proyecto formativo	Numérico
Trabajo colaborativo	Selección de 1 – 5 La metodología de proyectos formativos promueve el trabajo colaborativo y la autogestión	Numérico

Comprensión y participación en las actividades del proceso de evaluación		
Orientan Elaboración de las evidencias	Selección de 1 – 5 Orientan la elaboración de las evidencias del proyecto formativo	Numérico
Elaboración de las evidencias	Selección de 1 – 5 Realiza las evidencias propuestas basado en los criterios definidos en el/los mapa(s) de aprendizaje,	Numérico
Realizo autoevaluación	Selección de 1 – 5 Realizo autoevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s) de aprendizaje	Numérico
Realizo coevaluación	Selección de 1 – 5 Realizo coevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s)	Numérico
Retroalimentación del profesor	Selección de 1 – 5 La retroalimentación del profesor promueve la reflexión sobre las competencias desarrolladas, los logros y los aspectos por mejorar	Numérico
Retroalimentación de los compañeros	Selección de 1 – 5 La retroalimentación de los compañeros muestra objetividad en la valoración de las evidencias y fomenta una cultura democrática en la evaluación	Numérico
Contribución de las actividades del proceso de evaluación al rendimiento académico		
contribución de la autoevaluación al rendimiento académico	Selección de 1 – 5 La autoevaluación de evidencias contribuye al rendimiento académico	Numérico
contribución de la coevaluación al rendimiento académico	Selección de 1 – 5 La coevaluación de evidencias contribuye al rendimiento académico	Numérico
contribución de la Heteroevaluación al rendimiento académico	Selección de 1 – 5 La heteroevaluación de evidencias contribuye al rendimiento académico.	Numérico
contribución de metodología PF al rendimiento académico	Selección de 1 – 5 La metodología de proyectos formativos incentiva a mejorar el rendimiento académico	Numérico
contribución de la evaluación al rendimiento académico	Selección de 1 – 5 La metodología de evaluación contribuye a mejorar mi rendimiento académico	Numérico
Utilidad de los recursos y las actividades de aprendizaje en la plataforma Moodle		
Utilidad de los recursos de aprendizaje	Selección de 1 – 5 Los recursos de aprendizaje ofrecidos en Moodle Utilidad de los recursos de aprendizaje: Los recursos de aprendizaje ofrecidos en Moodle son útiles para la solución del proyecto formativo.	Numérico
coherencia de los recursos para el aprendizaje	Selección de 1 – 5 Los recursos de aprendizaje ofrecidos en Moodle son coherentes con las actividades de aprendizaje del proyecto formativo	Numérico
Coherencia de los recursos con las necesidades	Selección de 1 – 5 Los recursos de aprendizaje ofrecidos en Moodle están acorde con mis necesidades de aprendizaje	Numérico

coherencia de las Actividades con las necesidades	Selección de 1 – 5 Las actividades de aprendizaje ofrecidas en Moodle están acorde con mis necesidades de aprendizaje	Numérico
coherencia con el nivel de competencia	Selección de 1 – 5 Las actividades de aprendizaje ofrecidas en Moodle son coherentes mi nivel de competencia	Numérico
Opinión sobre el diseño del curso en la Plataforma Moodle		
Navegación del curso	Selección de 1 – 5 El diseño gráfico del curso en Moodle permite una navegación adecuada.	Numérico
Diseño gráfico	Selección de 1 – 5 El diseño gráfico del curso en Moodle es estético	Numérico
Experiencia del curso en la plataforma	Selección de 1 – 5 Me siento a gusto con el diseño del curso en la plataforma Moodle	Numérico
Interfaz del curso	Selección de 1 – 5 La interfaz del curso facilita la comprensión de los elementos del proyecto formativo	Numérico
Diseño es estética agradable	Selección de 1 – 5 El diseño de los recursos y las actividades es estéticamente agradable	Numérico
MejoraEvidencia	Selección de 1 – 5 Realiza las sugerencias propuestas por el docente	Numérico
Retroalimentación	Retroalimentación de las evidencias	Numérico
Periodo	Periodo académico	Numérico
rendimiento académico	Nota definitiva de los estudiantes en el curso	Numérico
Aspectos de la Evaluación		
Autoevaluación	Nota de la evaluación realizada por cada estudiante	Numérico
Coevaluación	Evaluación definida por el docente	Numérico
Heteroevaluación	Calificación entre los estudiantes	Numérico
Recomendaciones	Comentarios sobre el curso.	Caracteres

Tabla 11: Dominio de los atributos

Se eligió como clase el siguiente atributo:

1. Rendimiento académico: que corresponde al logro del nivel de competencia, el cual se obtuvo a partir de los resultados de un proceso de evaluación “Nota Definitiva”.

2. Descripción de cada atributo

2.1. Perfil Académico

Como se muestra en la Figura 2 se definen las variables que se consideraron en el modelo de estudiante en la categoría Perfil académico en el cual se tiene:



Figura 2: Perfil académico

- Repitente: Conocer si los estudiantes están repitiendo el curso de formación por competencias.
- Promedio Universidad: Esta variable permite tener en cuenta el promedio que llevan los estudiantes en carreras anteriores o en el transcurso de su carrera.
- Promedio Colegio: Permite conocer el promedio de los estudiantes en su colegio
- ICFES: Resultados de las pruebas ICFES
- Definitiva: Nota definitiva de los estudiantes en el curso
- Periodo: Periodo académico

2.2. Aspectos de la Evaluación

Como se muestra en la Figura 3 se definen las variables que se consideraron en el modelo de estudiante en la categoría Aspecto de evaluación en el cual se tiene:



Figura 3: Aspecto de Evaluación

- Autoevaluación: Nota de la evaluación realizada por cada estudiante
- Coevaluación: Evaluación definida por el docente
- Heteroevaluación : Calificación entre los estudiantes
- Rendimiento Académico: Resultado ítems de cuestionario de nivel de competencia final, se define con el proceso de evaluación.
- Nivel de Competencia: Esta variable se inicializa a través de un cuestionario que presentan los estudiantes al iniciar el curso, en la plataforma Moodle. El nivel de dominio hace referencia a los momentos a través de los cuales se desarrolla el talento y las competencias, y se establece de acuerdo con las categorías definidas por la socioformación: receptivo, resolutivo, autónomo y estratégico, siendo ejes esenciales de la sociedad del conocimiento y la socioformación (Tobón, 2017).

Del mismo modo, en el nivel receptivo la persona actúa ante los problemas con unas nociones básicas. En el nivel resolutivo: La persona actúa ante los problemas comprendiendo los procedimientos elementales, se resuelven problemas sencillos del contexto. En el nivel autónomo: La persona actúa con autonomía y criterio propio ante los problemas. En el nivel estratégico: La persona actúa ante los problemas logrando un impacto, considerando las consecuencias en las diferentes opciones de resolución de problemas en el contexto. El nivel de competencia se actualiza a medida que los estudiantes realizan actividades a través de la plataforma Moodle.

2.3. Datos Personales

En la **¡Error! No se encuentra el origen de la referencia.** se definen las variables que se consideraron en el modelo de estudiante en la categoría A datos Personales en el cual se tiene:



Figura 4: Datos Personales

- Edad: En esta variable se define la edad de los estudiantes que cursan la materia y como puede tener incidencias en la nota definitiva
- Género: esta variable define el género de los estudiantes, los cuales se definen en dos categorías mujer y hombre

- Estrato socioeconómico: el estrato permite categorizar los estudiantes en bajo, medio bajo, medio, medio alto y alto
- Estado civil: En esta variable se categoriza los estudiantes como solteros, casados, separado / divorciados, unión libre, viudos
- Laboral Actualmente: Permite determinar si los estudiantes trabajan o no
- Personas a Cargo: Cuantas personas tiene a cargo los estudiantes o dependen de ellos económicamente
- Tipo de colegio: qué tipo de colegio es egresado el estudiante como público o privado
- Lugar de residencia: Lugar de residencia de los estudiantes
- Nivel educativo madre: grado de escolaridad de la madre
- Nivel educativo padre: Grado de educación del padre
- Tipo de vivienda: Tipo de vivienda si es propia, familiar o alquilada
- Número de hermanos: Numero de hermanos
- Posición entre hermanos: Posición que ubica el estudiante entre hermanos

2.4 Aspectos Didácticos

Se puede determinar el comportamiento de los estudiantes teniendo en cuenta como son las notas de los estudiantes basados en la metodología proyecto formativo.

En la **¡Error! No se encuentra el origen de la referencia.** se definen las variables que se consideraron en el modelo de estudiante en la categoría Aspectos Didácticos en el cual se tiene:

- Comprensión y participación en las actividades del proceso de evaluación
- Cumplimiento de los elementos de proyectos formativos
- Contribución de las actividades del proceso de evaluación al rendimiento académico.
- Utilidad de los recursos y las actividades de aprendizaje en la plataforma Moodle
- Opinión sobre el diseño del curso en la plataforma Moodle

2.4.1 Comprensión y participación en las actividades del proceso de evaluación



Figura 5: Comprensión y participación en las actividades del proceso de evaluación

- Elaboración de las evidencias: Los criterios definidos en el/los mapa(s) de aprendizaje, orientan la elaboración de las evidencias del proyecto formativo
- Realizo autoevaluación: Realizo autoevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s) de aprendizaje.
- Realizo coevaluación: Realizo coevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s)
- Retroalimentación del profesor: La retroalimentación del profesor promueve la reflexión sobre las competencias desarrolladas, los logros y los aspectos por mejorar.
- Retroalimentación de los compañeros: La retroalimentación de los compañeros muestra objetividad en la valoración de las evidencias y fomenta una cultura democrática en la evaluación.

2.4.2 Cumplimiento de los elementos de proyectos formativos

En la Figura 6 se definen las variables que se consideraron en el modelo de estudiante en la categoría Aspectos didácticos en el Cumplimiento de los elementos de proyectos formativos en el cual se tiene:

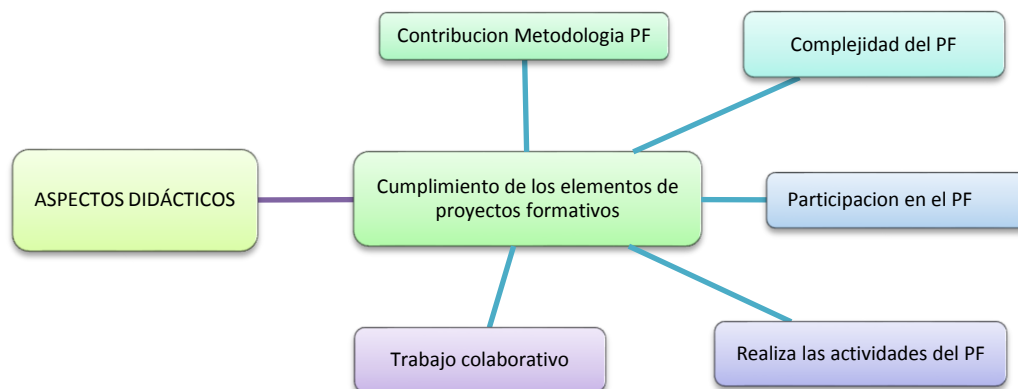


Figura 6: Cumplimiento de los elementos de proyectos formativos

- Contribución Metodología PF: La metodología de proyectos formativos contribuye al desarrollo de las competencias a formar.
- Complejidad del PF: El proyecto formativo se enfoca en la solución de un problema del contexto y su nivel de reto está acorde con mi nivel de formación
- Participación en el PF : Participo en la definición del problema del contexto del proyecto formativo.
- Realiza las actividades del PF : Realizo las actividades correspondientes a cada una de las fases del proyecto formativo.
- Trabajo colaborativo Orientan: La metodología de proyectos formativos promueve el trabajo colaborativo y la autogestión.

2.4.3 Contribución de las actividades del proceso de evaluación al rendimiento académico.

En la **Figura 7** se definen las variables que se consideraron en el modelo de estudiante en la categoría Contribución de las actividades del proceso de evaluación al rendimiento académico en el cual se tiene:

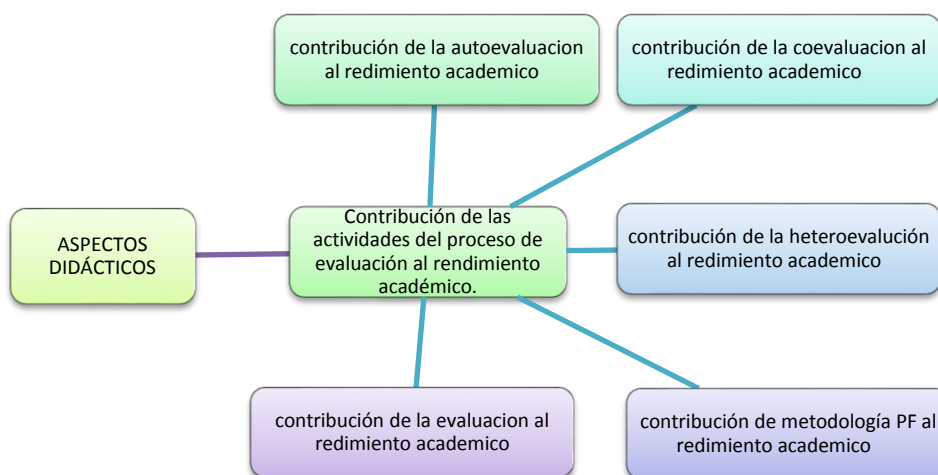


Figura 7: Contribución de las actividades del proceso de evaluación al rendimiento académico

- contribución de la autoevaluación al rendimiento académico: La autoevaluación de evidencias contribuye al rendimiento académico.
- contribución de la coevaluación al rendimiento académico: La coevaluación de evidencias contribuye al rendimiento académico
- contribución de la Heteroevaluación al rendimiento académico: La heteroevaluación de evidencias contribuye al rendimiento académico.
- contribución de metodología PF al rendimiento académico: La metodología de proyectos formativos incentiva a mejorar el rendimiento académico.
- contribución de la evaluación al rendimiento académico: La metodología de evaluación contribuye a mejorar mi rendimiento académico.

2.4.4. Utilidad de los recursos y las actividades de aprendizaje en la plataforma Moodle

En la **Figura 8** se definen las variables que se consideraron en el modelo de estudiante en la categoría Utilidad de los recursos y las actividades de aprendizaje en la plataforma Moodle en el cual se tiene:

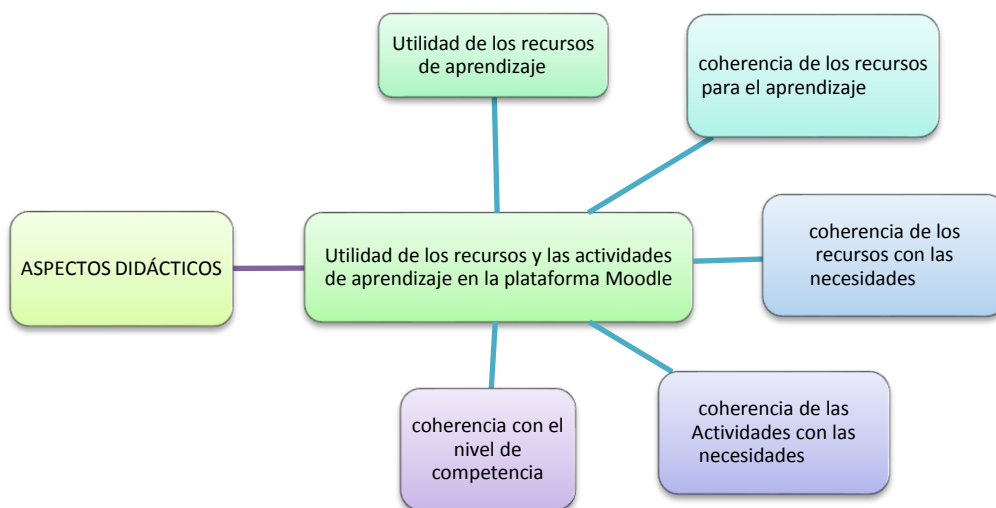


Figura 8: Utilidad de los recursos y las actividades en la plataforma Moodle

- Utilidad de los recursos de aprendizaje: Los recursos de aprendizaje ofrecidos en Moodle son útiles para la solución del proyecto formativo.
- coherencia de los recursos para el aprendizaje: Los recursos de aprendizaje ofrecidos en Moodle son coherentes con las actividades de aprendizaje del proyecto formativo.
- Coherencia de los recursos con las necesidades: Los recursos de aprendizaje ofrecidos en Moodle están acorde con mis necesidades de aprendizaje.
- coherencia de las Actividades con las necesidades: Las actividades de aprendizaje ofrecidas en Moodle están acorde con mis necesidades de aprendizaje.
- coherencia con el nivel de competencia: Las actividades de aprendizaje ofrecidas en Moodle son coherentes mi nivel de competencia.

2.4.5. Opinión sobre el diseño del curso en la Plataforma Moodle

En la **Figura 9** se definen las variables que se consideraron en el modelo de estudiante en la categoría Opinión sobre el diseño del curso en la Plataforma Moodle en el cual se tiene:

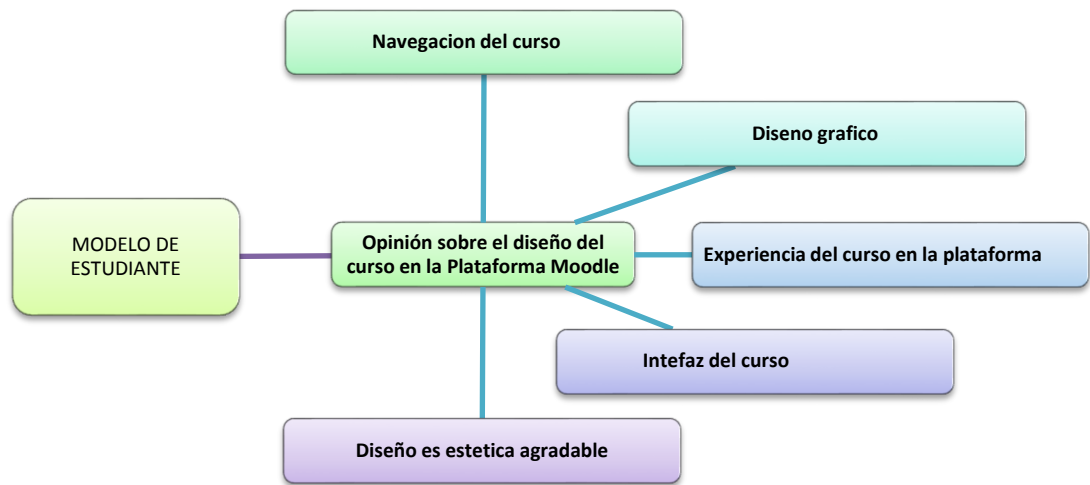


Figura 9: Opinión sobre el diseño del curso en la Plataforma Moodle

- Navegación del curso: El diseño gráfico del curso en Moodle permite una navegación adecuada.
- Diseño gráfico: El diseño gráfico del curso en Moodle es estético.
- Experiencia del curso en la plataforma: Me siento a gusto con el diseño del curso en la plataforma Moodle.
- Interfaz del curso: La interfaz del curso facilita la comprensión de los elementos del proyecto formativo
- Diseño es estética agradable: El diseño de los recursos y las actividades es estéticamente agradable

Anexo 2: Encuesta metodología de proyectos formativos

ENCUESTA METODOLOGÍA DE PROYECTOS FORMATIVOS

Edad: _____

Género: Hombre ___ Mujer ___

Estrato socioeconómico: 1 2 3 4 5 6 (marque con una X)

Estado Civil: Soltero (a) ___ Casado (a) /Unión libre ___ Separado (a) /Divorciado ___ Viudo (a) ___

Trabaja actualmente: Si ___ No ___

Número de personas a cargo: 0 1 2 3 4 5 o más (marque con una X)

Promedio acumulado en la Universidad: _____

Repitente del curso: Si ___ No ___

Egresado Colegio: Público ___ Privado: ___ Promedio Colegio: ___ Puntaje ICFES: ___

Lugar de Residencia: _____

Nivel educativo de la madre: Primaria ___ Secundaria ___ Universitaria ___ Especialización ___
Maestría ___ Doctorado ___

Nivel educativo del padre: Primaria ___ Secundaria ___ Universitaria ___ Especialización ___
Maestría ___ Doctorado ___

Número de hermanos: _____

Posición entre hermanos: Primero ___ Segundo ___ Tercero ___ Cuarto ___ Quinto ___ Posterior ___

Tipo de Vivienda: Propia ___ En alquiler ___

Esta encuesta se realiza a estudiantes de la Escuela de Administración y Mercadotecnia, con el propósito de conocer sus apreciaciones sobre la metodología de proyectos formativos.

Como estudiante del curso, entiendo que esta encuesta será tomada para propósitos de investigación y por lo tanto, el (los) investigador(es) realizarán la recopilación, análisis y reporte de datos. Entiendo también que cuando los datos se recogen y analizan, mi nombre será eliminado y que esta información no será utilizada durante el análisis o cuando se presenten los resultados. Entiendo que participo de esta actividad de manera voluntaria reconociendo que hace un aporte a mi formación profesional.

Las preguntas del cuestionario son clasificadas por categorías.

- Cumplimiento de los elementos esenciales de la metodología de proyectos formativos.

Puntúa las siguientes afirmaciones de 1 a 5 según tu opinión (siendo 1 el valor más bajo y 5 el valor más alto)	1	2	3	4	5
1. La metodología de proyectos formativos contribuye al desarrollo de las competencias a formar.					
2. El proyecto formativo se enfoca en la solución de un problema del contexto y su nivel de reto está acorde con mi nivel de formación.					
3. Participo en la definición del problema del contexto del proyecto formativo.					
4. Realizo las actividades correspondientes a cada una de las fases del proyecto formativo.					
5. La metodología de proyectos formativos promueve el trabajo colaborativo y la autogestión.					

Autor: Sergio Augusto Cardona Torres.

- **Comprensión y participación en las actividades del proceso de evaluación.**

Puntúa las siguientes afirmaciones de 1 a 5 (siendo 1 el valor más bajo y 5 el valor más alto)	1	2	3	4	5
6. Los criterios definidos en el/los mapa(s) de aprendizaje, orientan la elaboración de las evidencias del proyecto formativo.					
7. Realizo autoevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s) de aprendizaje.					
8. Realizo coevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s)					
9. La retroalimentación del profesor promueve la reflexión sobre las competencias desarrolladas, los logros y los aspectos por mejorar.					
10. La retroalimentación de los compañeros muestra objetividad en la valoración de las evidencias y fomenta una cultura democrática en la evaluación.					

- **Contribución de las actividades del proceso de evaluación al rendimiento académico.**

Puntúa las siguientes afirmaciones de 1 a 5 según tu opinión (siendo 1 el valor más bajo y 5 el valor más alto)	1	2	3	4	5
11. La autoevaluación de evidencias contribuye al rendimiento académico.					
12. La coevaluación de evidencias contribuye al rendimiento académico.					
13. La heteroevaluación de evidencias contribuye al rendimiento académico.					
14. La metodología de proyectos formativos incentiva a mejorar el rendimiento académico.					
15. La metodología de evaluación contribuye a mejorar mi rendimiento académico.					

- **Utilidad de los recursos y las actividades de aprendizaje en la plataforma Moodle.**

Puntúa las siguientes afirmaciones de 1 a 5 según tu opinión (siendo 1 el valor más bajo y 5 el valor más alto)	1	2	3	4	5
16. Los recursos de aprendizaje ofrecidos en Moodle son útiles para la solución del proyecto formativo.					
17. Los recursos de aprendizaje ofrecidos en Moodle son coherentes con las actividades de aprendizaje del proyecto formativo.					
18. Los recursos de aprendizaje ofrecidos en Moodle están acorde con mis necesidades de aprendizaje.					
19. Las actividades de aprendizaje ofrecidas en Moodle están acorde con mis necesidades de aprendizaje.					
20. Las actividades de aprendizaje ofrecidas en Moodle son coherentes mi nivel de competencia.					

- **Opinión sobre el diseño del curso en la Plataforma Moodle**

Puntúa las siguientes afirmaciones de 1 a 5 según tu opinión (siendo 1 el valor más bajo y 5 el valor más alto)	1	2	3	4	5
21. El diseño gráfico del curso en Moodle permite una navegación adecuada.					
22. El diseño gráfico del curso en Moodle es estético.					
23. Me siento a gusto con el diseño del curso en la plataforma Moodle.					
24. La interfaz del curso facilita la comprensión de los elementos del proyecto formativo.					
25. El diseño de los recursos y las actividades es estéticamente agradable.					

Recomendaciones u observaciones del instrumento

Gracias por su apoyo

Autor: Sergio Augusto Cardona Torres.

Anexo 3: Preparación y selección de atributos

1. Procesamiento de los datos

Con el fin de conocer la calidad de los datos que se utilizó para el presente trabajo, se realizó un diagnóstico de los mismos utilizando el software DQ Analizar; es una herramienta para perfilamiento de datos y ayuda a identificar problemas de calidad de datos como: datos nulos, datos duplicados, errores en formatos.

A continuación se presenta el perfil de los datos que contiene la base de datos de estudiantes.

1.1. Limpieza de datos

En esta fase se eliminan redundancias y se resuelve problemas de ruido, valores perdidos, se elimina outliers, además de resolver inconsistencias y/o conflictos entre datos. Para este trabajo se revisaron los datos con la herramienta DQ Analizar, para luego realizar la corrección o eliminación de información de acuerdo a su importancia en el modelo de minería.

Edad	INTEGER	enum		133	0	8	19	16	19	43
Genero	STRING	enum	Phone number	133	0	0	2	HOMBRE	HOMBRE	MUJER
Estrato	INTEGER	enum		130	3	1	6	0	3	5
Estado_civil	STRING	enum patt...	Phone number	133	0	0	2	CASADO/UNIONL...	SOLTERO	SOLTERO
Trabaja	STRING	enum	Country	132	1	0	2	NO	NO	SI
Personas_Cargo	INTEGER	enum		128	5	1	6	0	0	9
Repitente	STRING	enum	Country	132	1	0	2	NO	NO	SI
promedio_universidad	STRING	float enu...	Number	133	0	5	19	0	0	4,5
egresado_colegio	STRING	enum	Phone number	90	43	0	2	PRIVADO	PUBLICO	PUBLICO
promedio_Colegio	STRING	float enu...	Number	38	95	7	15	3	4,0	96
ICFES	INTEGER	enum		41	92	26	33	40	243	417
Lugar_residencia	STRING	enum patt...	Phone number	84	49	11	15	ALAMOS	ARMENIA	SORRENTO
Nivel_edcativo_madre	STRING	enum	Phone number	88	45	0	4	ESPECIALIZACION	SECUNDARIA	UNIVERSIDAD
Nivel_eduativo_padre	STRING	enum	Phone number	81	52	1	5	ESPECIALIZACION	SECUNDARIA	UNIVERSIDAD
Numero_hermanos	INTEGER	enum		85	48	1	7	0	1	6
Posicion_hermanos	STRING	enum	Phone number	83	50	0	6	CUARTO	PRIMERO	TERCERO
Tipo_vivienda	STRING	enum	Phone number	88	45	0	3	ARRENDADA	PROPIA	PROPIA
Contribucion_Metodologia_PF	INTEGER	enum		133	0	0	4	2	4	5
Complejidad_del_PF	INTEGER	enum		133	0	0	4	2	4	5
Participacion_en_el_PF	INTEGER	enum		131	2	1	5	1	4	5
Realiza_las_actividades_del_PF	INTEGER	enum		132	1	1	5	1	4	5
Trabajo_colaborativo	INTEGER	enum		133	0	0	4	2	4	5
Orientan_Elaboracion_de_las_eviden...	INTEGER	enum		132	1	1	5	1	4	5
Realizo_autoevaluacion	INTEGER	enum		133	0	1	5	1	4	5
Realizo_coevaluacion	INTEGER	enum		133	0	0	4	2	4	5
Retroalimentacion_del_profesor	INTEGER	enum		133	0	1	5	1	4	5
Retroalimentacion_de_los_companer...	INTEGER	enum		131	2	1	5	1	4	5
contribución_de_la_autoevaluacion_...	INTEGER	enum		133	0	0	5	1	4	5
contribución_de_la_autoevaluacion_...	INTEGER	enum		133	0	0	4	2	4	5
contribución_de_la_heteroevaluación_...	INTEGER	enum		133	0	0	4	2	4	5
contribución_de_metodología_PF_al...	INTEGER	enum		133	0	0	4	2	4	5
contribución_de_la_evaluacion_al_re...	INTEGER	enum		133	0	1	5	1	4	5
Utilidad_de_los_recursos_de_aprendi...	INTEGER	enum		133	0	0	4	2	4	5
coherencia_de_los_recursos_para_el_...	INTEGER	enum		133	0	0	5	1	4	5
coherencia_de_los_recursos_con_las_...	INTEGER	enum		133	0	0	4	2	4	5
coherencia_de_las_Actividades_con_l...	INTEGER	enum		133	0	1	5	1	4	5
coherencia_con_el_nivel_de_compet...	INTEGER	enum		133	0	0	4	2	4	5
Navegacion_del_curso	INTEGER	enum		133	0	1	5	1	4	5
Diseño_grafico	INTEGER	enum		133	0	0	4	2	4	5
Experiencia_del_curso_en_la_platafor...	INTEGER	enum		133	0	0	4	2	4	5
Intefaz_del_curso	INTEGER	enum		133	0	1	5	1	4	5
Diseño_es_estetica_agradable	INTEGER	enum		133	0	0	4	2	4	5
MejoraEvidencia	INTEGER	enum		133	0	0	6	0	2	5
Retroalimentacion	INTEGER	enum		133	0	0	6	0	3	5
Periodo	INTEGER	enum		133	0	0	4	1	2	4
Definitiva	FLOAT			133	0	3	27	2,4	3,4	6,0
Autoevaluacion	FLOAT	enum		133	0	3	14	2,0	4,0	5,0
Heteroevaluacion	FLOAT			133	0	3	25	2,2	3,2	5,0
coevaluacion	FLOAT	enum		133	0	3	15	2,5	4,4	5,0
RECOMENDACIONES	STRING	enum patt...	Name surname	18	115	18	18	DEMASIADO LAR...	MEJORAR DX ...	VIDEO D EcAPAdtAcION

Figura 10: Datos DQ Analizar

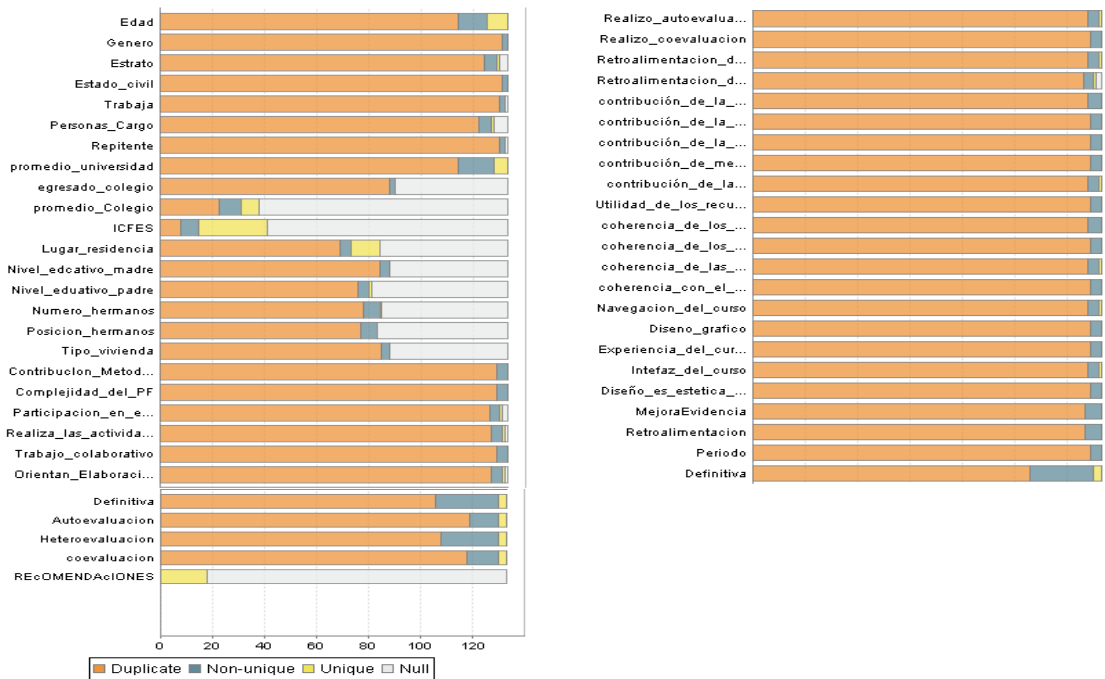


Figura 11: Información DQ Analyzer

Esta herramienta permitió analizar en detalle cada variable considerada, y así conocer la calidad de los datos de cada característica, como se presenta para la variable Repitiente en la siguiente tabla.

➤ **Repitente**

Expresion	Type	Domain	Business Domain	Non-null	Null	Unique	Distinct	Min	Median	Max
Repitente	STRING	enum	Country	132	1	0	2	NO	NO	SI

Tabla 12: DQ Analyzer Detalle variable

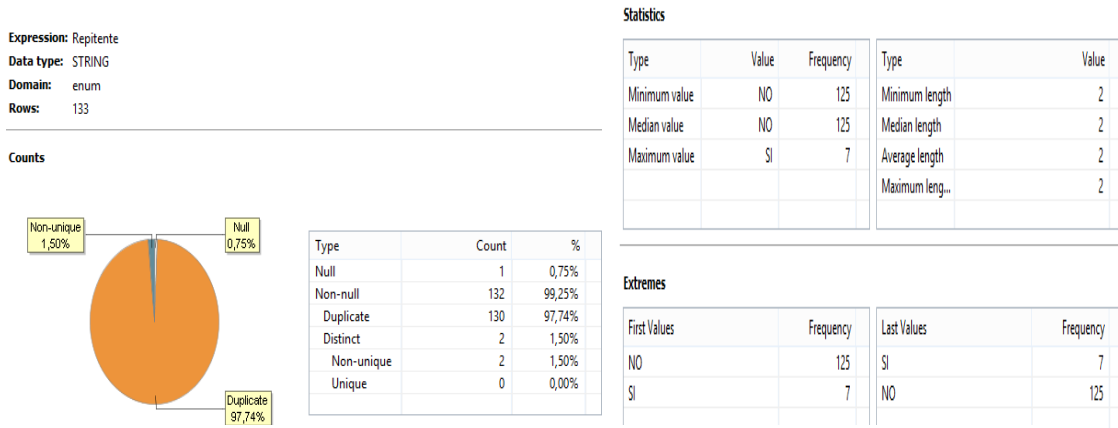


Figura 12: Detalle Variable Grafico

Datos nulos

Se encuentra que la variable “Repitente” tiene 1 registros con valores nulos, los cuales equivalen a un 0.75% de problemas de calidad de datos para efectos del objeto minable.

Este procedimiento se realizó a cada variable, con el objetivo de determinar la calidad de los datos.

Además se encontraron 232 registros sin dato en la columna observaciones, para este caso se suprime la columna.

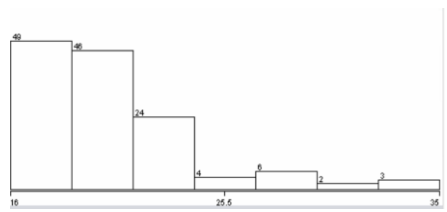
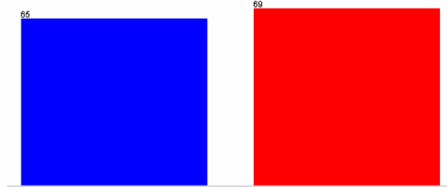
Datos Atípicos:

En la columna Edad se encontró 1 registro con datos atípicos, edad de 35 años dicho campo fue promediado con las demás edades.

En la columna definitiva existía un valor de 6, dicho campo fue promediado con loas demas notas.

1.2. Análisis estadístico

A continuación en la **tabla 2**, se presenta las descripciones estadistas de cada uno de los datos.

Atributo	Descripción												
<p data-bbox="462 940 906 1094">Selected attribute Name: Edad Missing: 0 (0%) Distinct: 18 Type: Numeric Unique: 7 (5%)</p> <table border="1" data-bbox="462 997 906 1087"><thead><tr><th>Statistic</th><th>Value</th></tr></thead><tbody><tr><td>Minimum</td><td>16</td></tr><tr><td>Maximum</td><td>35</td></tr><tr><td>Mean</td><td>20.425</td></tr><tr><td>StdDev</td><td>3.728</td></tr></tbody></table>  <p data-bbox="427 1325 948 1354">Ilustración 22: análisis estadístico - Edad</p>	Statistic	Value	Minimum	16	Maximum	35	Mean	20.425	StdDev	3.728	<p data-bbox="1166 940 1375 1087">El atributo “Edad” contiene una edad mínima de 16 y una máxima de 35 años.</p>		
Statistic	Value												
Minimum	16												
Maximum	35												
Mean	20.425												
StdDev	3.728												
<p data-bbox="462 1371 906 1501">Selected attribute Name: Genero Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)</p> <table border="1" data-bbox="462 1438 906 1501"><thead><tr><th>No.</th><th>Label</th><th>Count</th><th>Weight</th></tr></thead><tbody><tr><td>1</td><td>MUJER</td><td>65</td><td>65.0</td></tr><tr><td>2</td><td>HOMBRE</td><td>69</td><td>69.0</td></tr></tbody></table>  <p data-bbox="412 1711 963 1740">Ilustración 23: análisis estadístico - Género</p>	No.	Label	Count	Weight	1	MUJER	65	65.0	2	HOMBRE	69	69.0	<p data-bbox="1166 1375 1375 1585">Esta variable contiene valores posibles Mujer y Hombre, para un total de 65 Mujeres y 69 hombres.</p>
No.	Label	Count	Weight										
1	MUJER	65	65.0										
2	HOMBRE	69	69.0										

Name: Estrato		Type: Numeric
Missing: 0 (0%)	Distinct: 6	Unique: 1 (1%)
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	2.739	
StdDev	0.988	

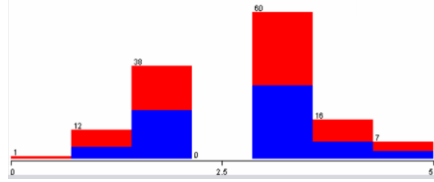


Ilustración 24: Análisis estadístico -Estrato

12 0 - Bajo Mj
 3 1 - Medio bajo
 14 2 - Medio
 5 3 - Medio Alto
 16 4 - Alto
 17 5 - Alto-alto

Name: Estado_civil		Type: Nominal	
Missing: 0 (0%)	Distinct: 4	Unique: 1 (1%)	
No.	Label	Count	Weight
1	SOLTERO	121	121.0
2	CASADO/UNIONLIB...	10	10.0
3	SEPARADO/DIVORC...	2	2.0
4	VIUDO	1	1.0

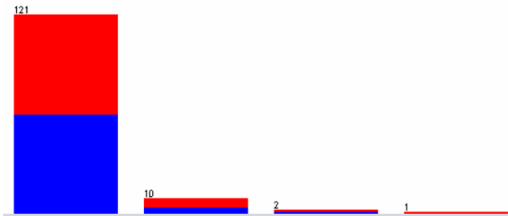


Ilustración 25: Análisis estadístico - Estado_civil

La variable “estado_civil” Tiene 4 categorías, donde la categoría “soltero” presenta el mayor número de registros, en total 121, seguido por la categoría “Casado/Unión libre” con 10 registros y la categoría con menos registros es “Separado / Divorciado” con 2 resgitros y “Viudo”, con 1.

Soltero
 Casado/Unión libre
 Separado / Divorciado
 Viudo

Name: Trabaja		Type: Nominal	
Missing: 0 (0%)	Distinct: 2	Unique: 0 (0%)	
No.	Label	Count	Weight
1	SI	51	51.0
2	NO	83	83.0

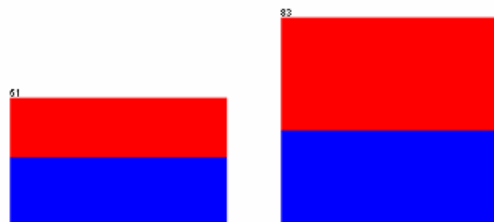
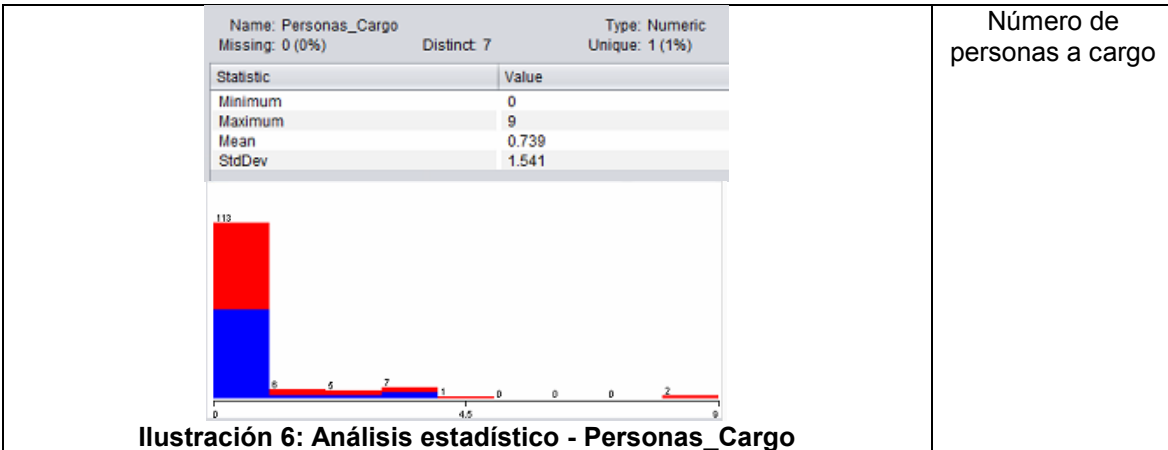


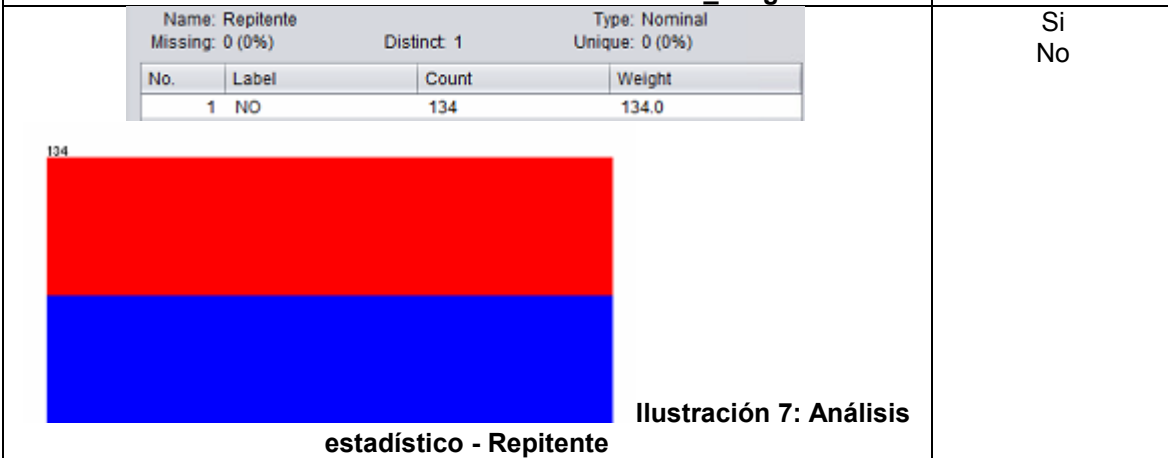
Ilustración 5: Análisis estadístico - Trabaja

Si
 No



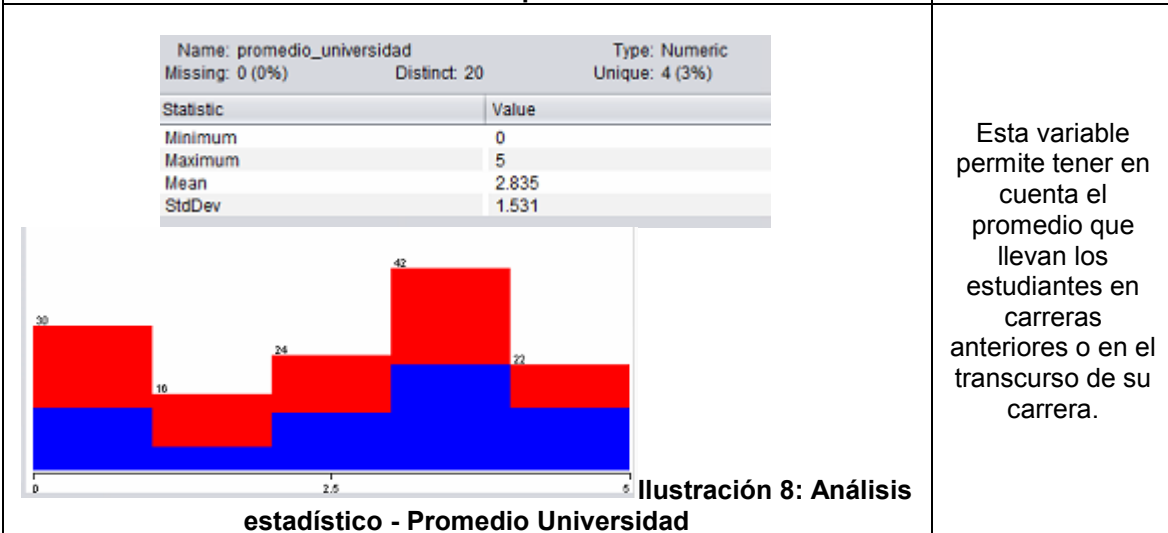
Número de personas a cargo

Ilustración 6: Análisis estadístico - Personas_Cargo



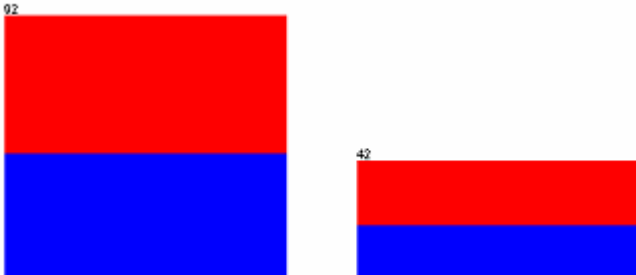
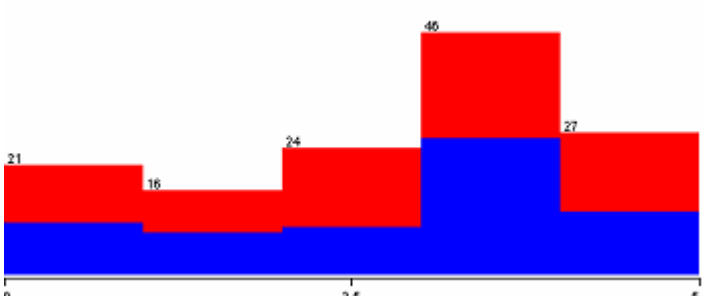
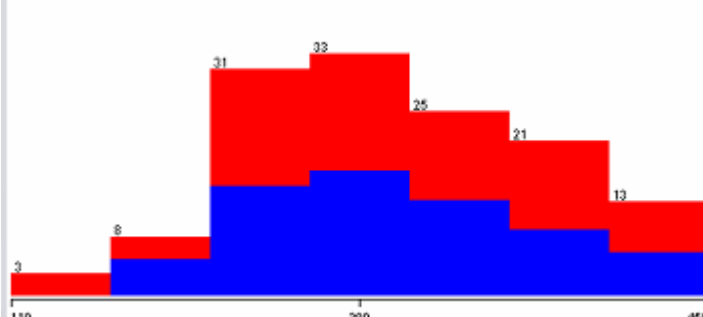
Si
No

Ilustración 7: Análisis estadístico - Repitente



Esta variable permite tener en cuenta el promedio que llevan los estudiantes en carreras anteriores o en el transcurso de su carrera.

Ilustración 8: Análisis estadístico - Promedio Universidad

<table border="1"> <thead> <tr> <th>No.</th> <th>Label</th> <th>Count</th> <th>Weight</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>PUBLICO</td> <td>92</td> <td>92.0</td> </tr> <tr> <td>2</td> <td>PRIVADO</td> <td>42</td> <td>42.0</td> </tr> </tbody> </table>  <p>Ilustración 9: Análisis estadístico - egresado_colegio</p>	No.	Label	Count	Weight	1	PUBLICO	92	92.0	2	PRIVADO	42	42.0	<p>Publico Privado</p>
No.	Label	Count	Weight										
1	PUBLICO	92	92.0										
2	PRIVADO	42	42.0										
<table border="1"> <thead> <tr> <th>Statistic</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Minimum</td> <td>0</td> </tr> <tr> <td>Maximum</td> <td>5</td> </tr> <tr> <td>Mean</td> <td>3.167</td> </tr> <tr> <td>StdDev</td> <td>1.392</td> </tr> </tbody> </table>  <p>Ilustración 10: Análisis estadístico - promedio_Colegio</p>	Statistic	Value	Minimum	0	Maximum	5	Mean	3.167	StdDev	1.392	<p>Promedio del alumno en el colegio</p>		
Statistic	Value												
Minimum	0												
Maximum	5												
Mean	3.167												
StdDev	1.392												
<table border="1"> <thead> <tr> <th>Statistic</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Minimum</td> <td>110</td> </tr> <tr> <td>Maximum</td> <td>450</td> </tr> <tr> <td>Mean</td> <td>299.5</td> </tr> <tr> <td>StdDev</td> <td>71.062</td> </tr> </tbody> </table>  <p>Ilustración 11: Análisis estadístico - ICFES</p>	Statistic	Value	Minimum	110	Maximum	450	Mean	299.5	StdDev	71.062	<p>Examen</p>		
Statistic	Value												
Minimum	110												
Maximum	450												
Mean	299.5												
StdDev	71.062												

Name: Lugar_residencia		Type: Nominal	
Missing: 0 (0%)		Distinct: 6	
		Unique: 2 (1%)	
No.	Label	Count	Weight
1	CALARCA	11	11.0
2	ARMENIA	115	115.0
3	PIJAO	1	1.0
4	MONTENEGRO	2	2.0
5	CIRCASIA	4	4.0
6	LA TEBaida	1	1.0

Armenia
Calarcá
Circasia
La tebaida
Montenegro
Pijao

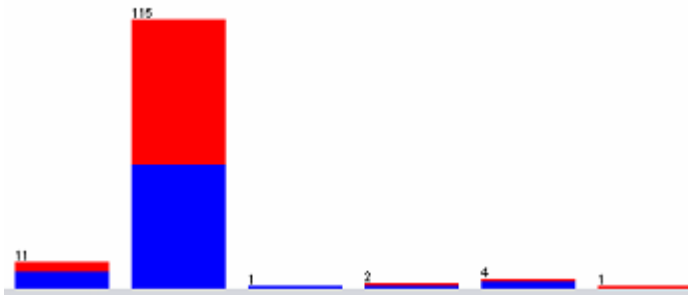


Ilustración 6:
Análisis estadístico - Lugar_residencia

Name: Nivel_edcativo_madre		Type: Nominal	
Missing: 0 (0%)		Distinct: 5	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	UNIVERSIDAD	13	13.0
2	SECUNDARIA	67	67.0
3	PRIMARIA	45	45.0
4	ESPECIALIZACION	7	7.0
5	MAESTRIA	2	2.0

Primaria
Secundaria
Universidad
Especialización
Maestría

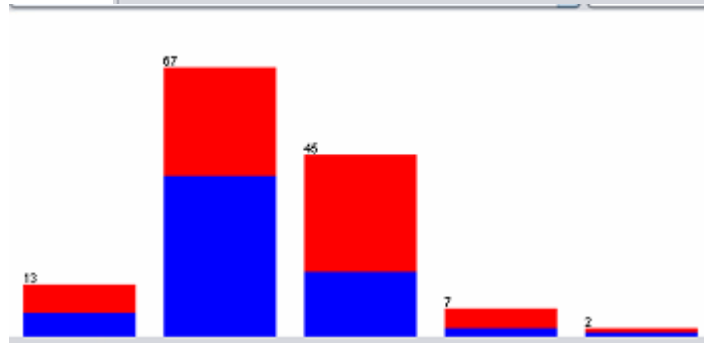
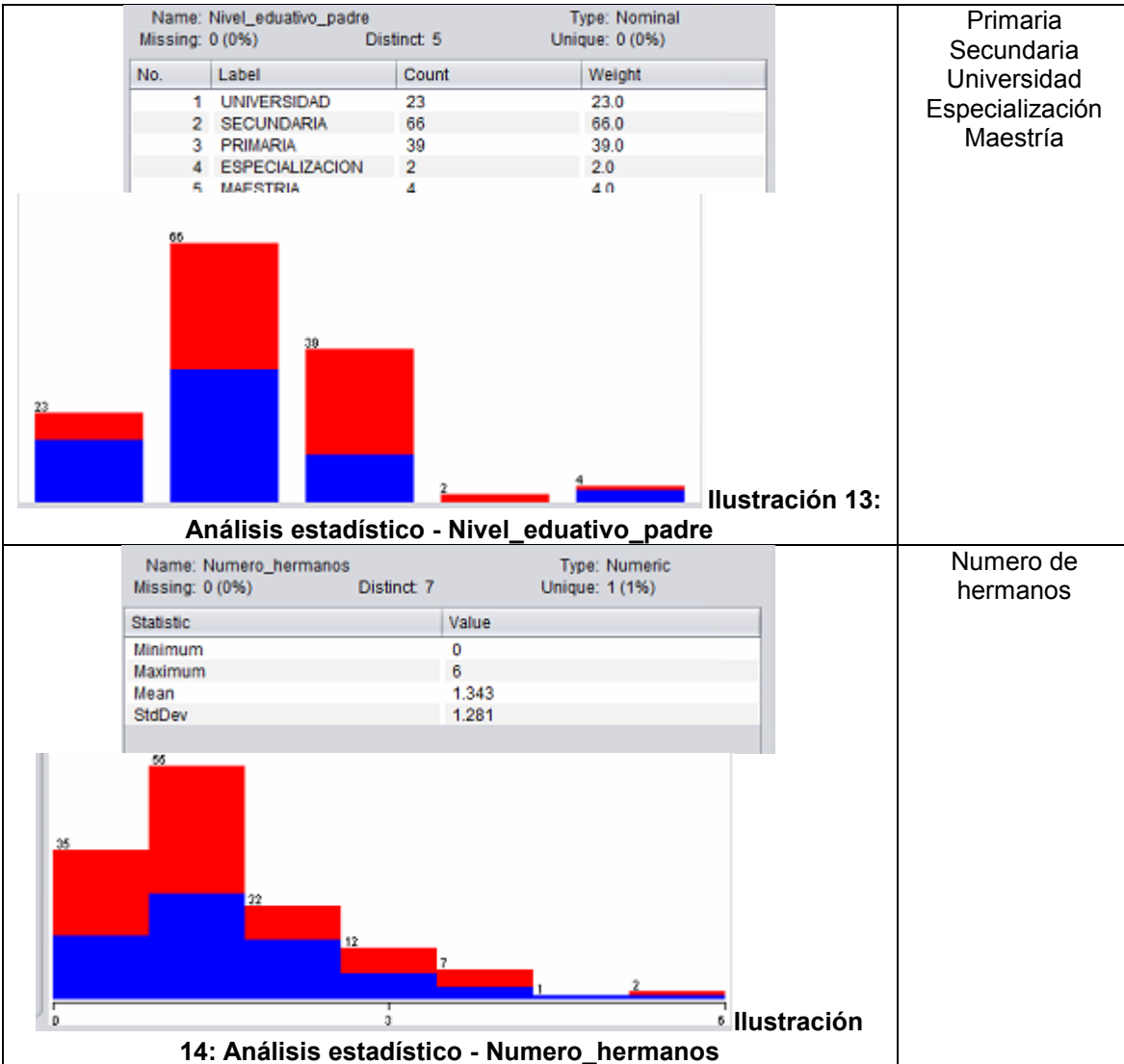


Ilustración 12:
Análisis estadístico - Nivel_edcativo_madre

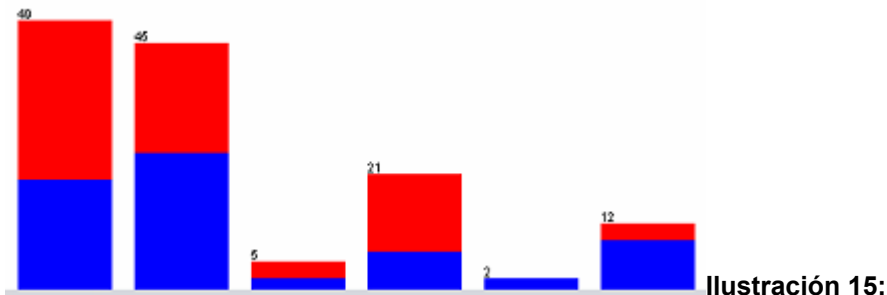


Primaria
Secundaria
Universidad
Especialización
Maestría

Numero de
hermanos

Name: Posicion_hermanos		Type: Nominal	
Missing: 0 (0%)		Distinct: 6	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	SEGUNDO	49	49.0
2	PRIMERO	45	45.0
3	QUINTO	5	5.0
4	CUARTO	21	21.0
5	OTRO	2	2.0
6	TERCERO	12	12.0

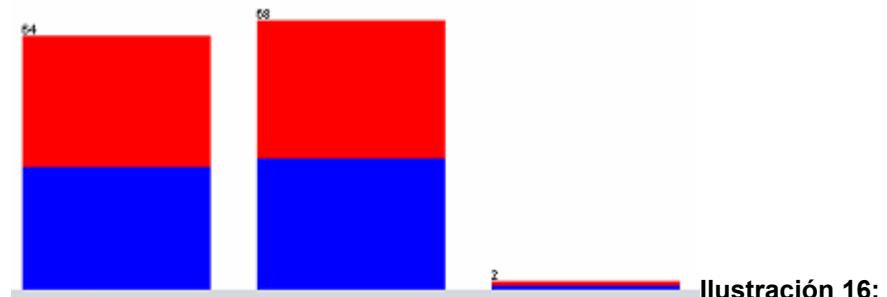
Primero
Segundo
Tercero
Cuarto
Quinto
Posterior



Análisis estadístico - Posicion_hermanos

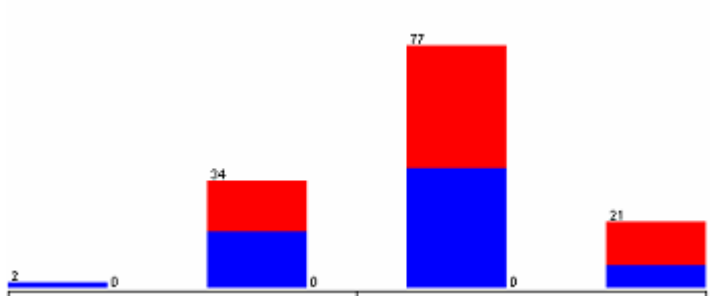
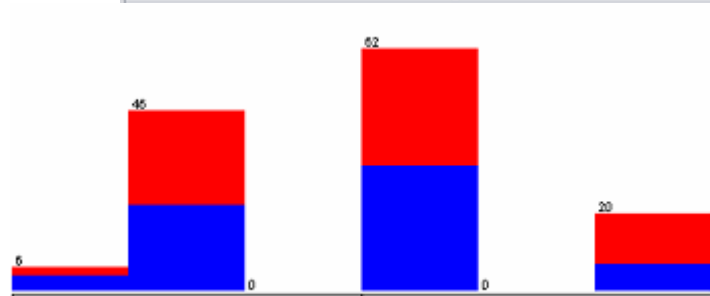
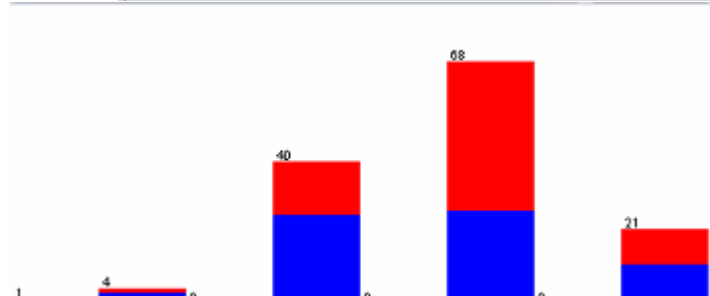
Name: Tipo_vivienda		Type: Nominal	
Missing: 0 (0%)		Distinct: 3	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	ARRENDADA	64	64.0
2	PROPIA	68	68.0
3	FAMILIAR	2	2.0

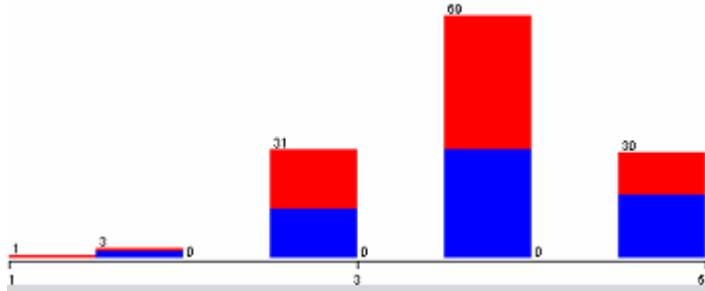
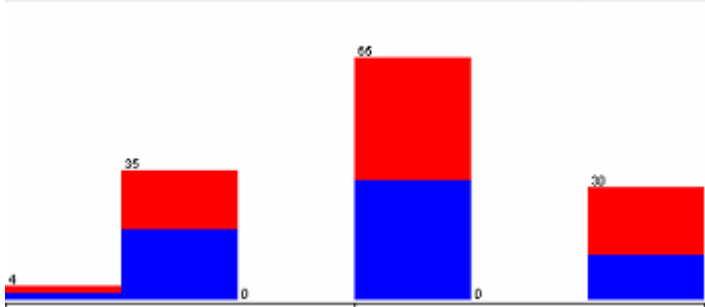
Arrendada
Familiar
Propia



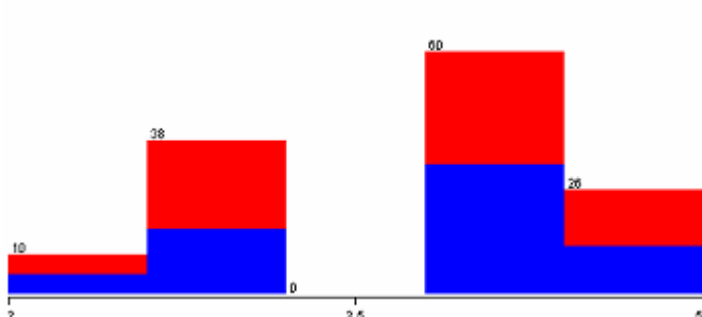
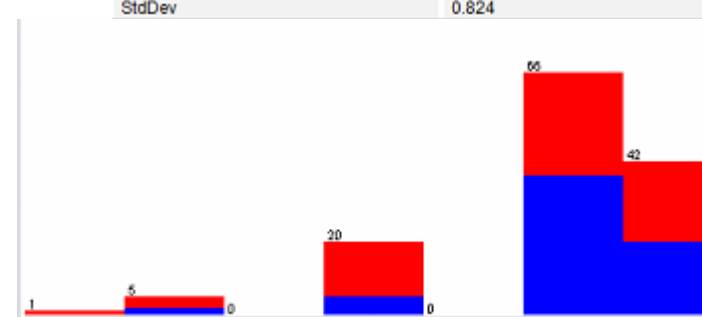
Análisis estadístico - Tipo_vivienda

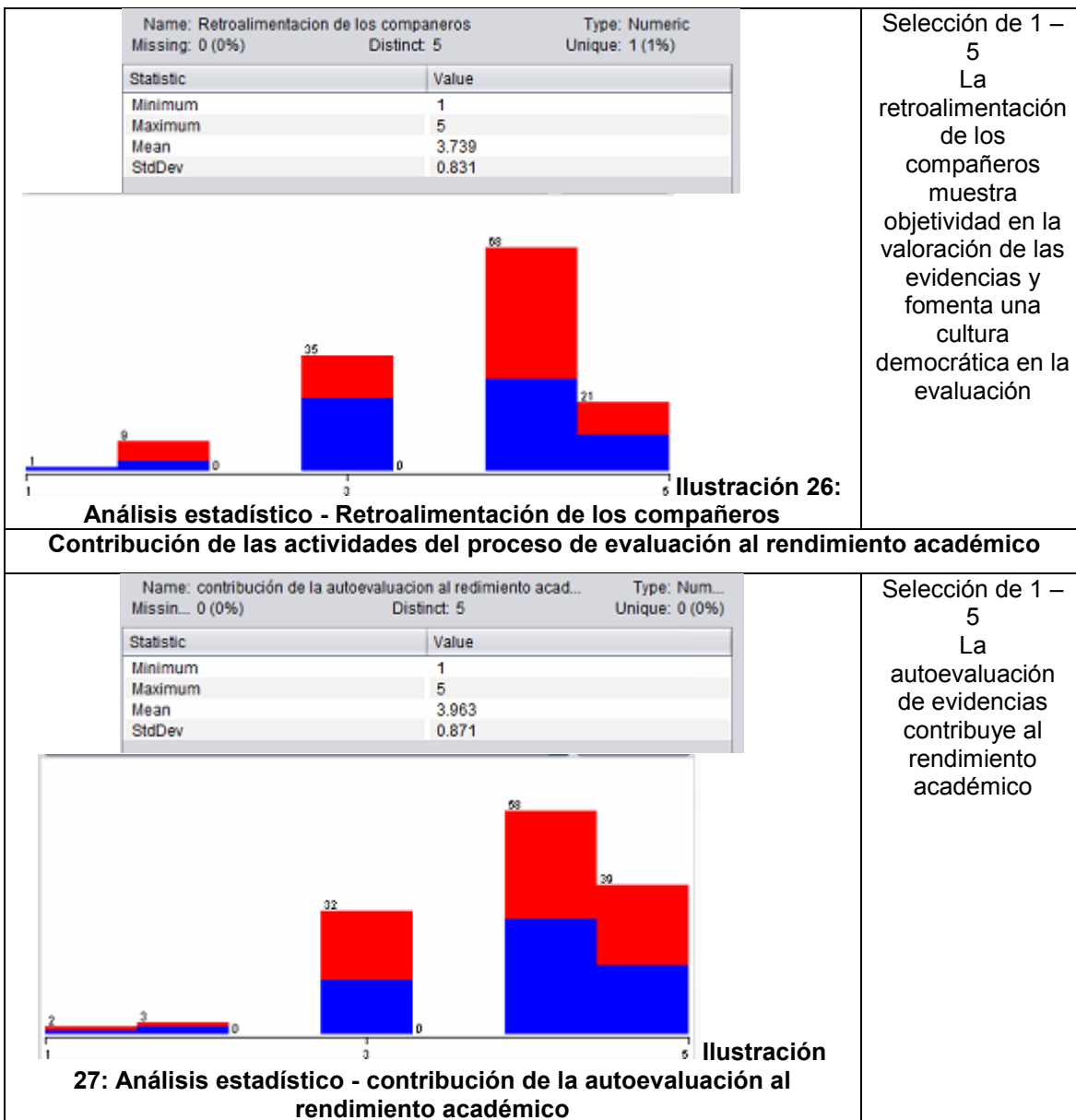
Cumplimiento de los elementos de proyectos formativos

<p>Name: Contribucion Metodologia PF Missing: 0 (0%) Distinct: 4 Type: Numeric Unique: 0 (0%)</p> <table border="1"> <thead> <tr> <th>Statistic</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Minimum</td> <td>2</td> </tr> <tr> <td>Maximum</td> <td>5</td> </tr> <tr> <td>Mean</td> <td>3.873</td> </tr> <tr> <td>StdDev</td> <td>0.676</td> </tr> </tbody> </table>  <p>Ilustración 17: Análisis estadístico - Contribución Metodología PF</p>	Statistic	Value	Minimum	2	Maximum	5	Mean	3.873	StdDev	0.676	<p>Selección de 1 – 5 La metodología de proyectos formativos contribuye al desarrollo de las competencias a formar</p>
Statistic	Value										
Minimum	2										
Maximum	5										
Mean	3.873										
StdDev	0.676										
<p>Name: Complejidad del PF Missing: 0 (0%) Distinct: 4 Type: Numeric Unique: 0 (0%)</p> <table border="1"> <thead> <tr> <th>Statistic</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Minimum</td> <td>2</td> </tr> <tr> <td>Maximum</td> <td>5</td> </tr> <tr> <td>Mean</td> <td>3.716</td> </tr> <tr> <td>StdDev</td> <td>0.772</td> </tr> </tbody> </table>  <p>Ilustración 18: Análisis estadístico - Complejidad del PF</p>	Statistic	Value	Minimum	2	Maximum	5	Mean	3.716	StdDev	0.772	<p>Selección de 1 – 5 El proyecto formativo se enfoca en la solución de un problema del contexto y su nivel de reto está acorde con mi nivel de formación</p>
Statistic	Value										
Minimum	2										
Maximum	5										
Mean	3.716										
StdDev	0.772										
<p>Name: Participacion en el PF Missing: 0 (0%) Distinct: 5 Type: Numeric Unique: 1 (1%)</p> <table border="1"> <thead> <tr> <th>Statistic</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Minimum</td> <td>1</td> </tr> <tr> <td>Maximum</td> <td>5</td> </tr> <tr> <td>Mean</td> <td>3.776</td> </tr> <tr> <td>StdDev</td> <td>0.772</td> </tr> </tbody> </table>  <p>Ilustración 19: Análisis estadístico - Participación en el PF</p>	Statistic	Value	Minimum	1	Maximum	5	Mean	3.776	StdDev	0.772	<p>Selección de 1 – 5 Participo en la definición del problema del contexto del proyecto formativo</p>
Statistic	Value										
Minimum	1										
Maximum	5										
Mean	3.776										
StdDev	0.772										

<table border="1"> <tr> <td colspan="2">Name: Realiza las actividades del PF</td> <td>Type: Numeric</td> </tr> <tr> <td>Missing: 0 (0%)</td> <td>Distinct: 5</td> <td>Unique: 1 (1%)</td> </tr> <tr> <th>Statistic</th> <th>Value</th> <td></td> </tr> <tr> <td>Minimum</td> <td>1</td> <td></td> </tr> <tr> <td>Maximum</td> <td>5</td> <td></td> </tr> <tr> <td>Mean</td> <td>3.925</td> <td></td> </tr> <tr> <td>StdDev</td> <td>0.782</td> <td></td> </tr> </table>  <p>Ilustración 20: Análisis estadístico - Realiza las actividades del PF</p>	Name: Realiza las actividades del PF		Type: Numeric	Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)	Statistic	Value		Minimum	1		Maximum	5		Mean	3.925		StdDev	0.782		<p>Selección de 1 – 5 Realizo las actividades correspondientes a cada una de las fases del proyecto formativo</p>
Name: Realiza las actividades del PF		Type: Numeric																				
Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)																				
Statistic	Value																					
Minimum	1																					
Maximum	5																					
Mean	3.925																					
StdDev	0.782																					
<table border="1"> <tr> <td colspan="2">Name: Trabajo colaborativo</td> <td>Type: Numeric</td> </tr> <tr> <td>Missing: 0 (0%)</td> <td>Distinct: 4</td> <td>Unique: 0 (0%)</td> </tr> <tr> <th>Statistic</th> <th>Value</th> <td></td> </tr> <tr> <td>Minimum</td> <td>2</td> <td></td> </tr> <tr> <td>Maximum</td> <td>5</td> <td></td> </tr> <tr> <td>Mean</td> <td>3.903</td> <td></td> </tr> <tr> <td>StdDev</td> <td>0.774</td> <td></td> </tr> </table>  <p>Ilustración 21: Análisis estadístico - Trabajo colaborativo</p>	Name: Trabajo colaborativo		Type: Numeric	Missing: 0 (0%)	Distinct: 4	Unique: 0 (0%)	Statistic	Value		Minimum	2		Maximum	5		Mean	3.903		StdDev	0.774		<p>Selección de 1 – 5 La metodología de proyectos formativos promueve el trabajo colaborativo y la autogestión</p>
Name: Trabajo colaborativo		Type: Numeric																				
Missing: 0 (0%)	Distinct: 4	Unique: 0 (0%)																				
Statistic	Value																					
Minimum	2																					
Maximum	5																					
Mean	3.903																					
StdDev	0.774																					
<p>Comprensión y participación en las actividades del proceso de evaluación</p>																						

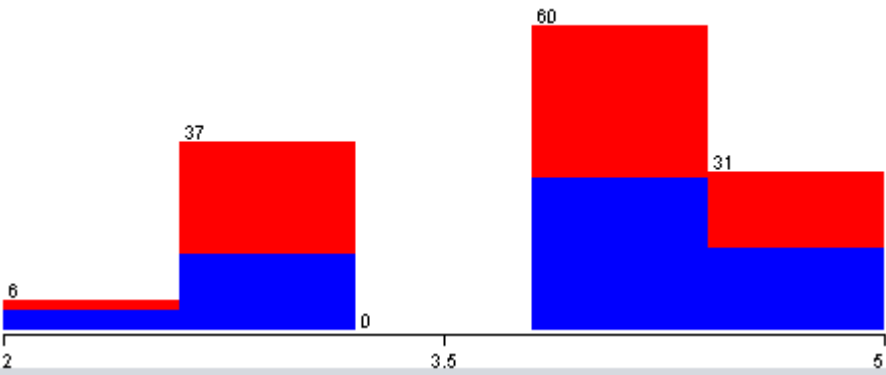
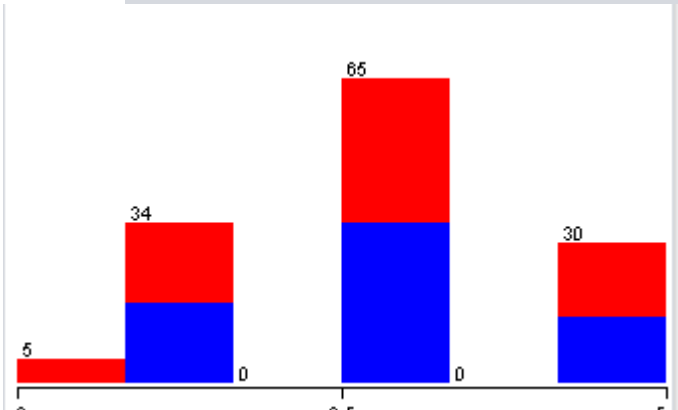
<table border="1"> <tr> <td colspan="2">Name: Orientan Elaboración de las evidencias</td> <td>Type: Numeric</td> </tr> <tr> <td>Missing: 0 (0%)</td> <td>Distinct: 5</td> <td>Unique: 1 (1%)</td> </tr> <tr> <th>Statistic</th> <th>Value</th> <td></td> </tr> <tr> <td>Minimum</td> <td>1</td> <td></td> </tr> <tr> <td>Maximum</td> <td>5</td> <td></td> </tr> <tr> <td>Mean</td> <td>3.948</td> <td></td> </tr> <tr> <td>StdDev</td> <td>0.844</td> <td></td> </tr> </table> <p>Ilustración 22: Análisis estadístico - Orientan Elaboración de las evidencias</p>	Name: Orientan Elaboración de las evidencias		Type: Numeric	Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)	Statistic	Value		Minimum	1		Maximum	5		Mean	3.948		StdDev	0.844		<p>Selección de 1 – 5 Orientan la elaboración de las evidencias del proyecto formativo</p>
Name: Orientan Elaboración de las evidencias		Type: Numeric																				
Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)																				
Statistic	Value																					
Minimum	1																					
Maximum	5																					
Mean	3.948																					
StdDev	0.844																					
<p>Elaboración de las evidencias</p>	<p>Selección de 1 – 5 Realiza las evidencias propuestas basado en los criterios definidos en el/los mapa(s) de aprendizaje,</p>																					
<table border="1"> <tr> <td colspan="2">Name: Realizo autoevaluación</td> <td>Type: Numeric</td> </tr> <tr> <td>Missing: 0 (0%)</td> <td>Distinct: 5</td> <td>Unique: 1 (1%)</td> </tr> <tr> <th>Statistic</th> <th>Value</th> <td></td> </tr> <tr> <td>Minimum</td> <td>1</td> <td></td> </tr> <tr> <td>Maximum</td> <td>5</td> <td></td> </tr> <tr> <td>Mean</td> <td>3.828</td> <td></td> </tr> <tr> <td>StdDev</td> <td>0.889</td> <td></td> </tr> </table> <p>Ilustración 23: Análisis estadístico - Realizo autoevaluación</p>	Name: Realizo autoevaluación		Type: Numeric	Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)	Statistic	Value		Minimum	1		Maximum	5		Mean	3.828		StdDev	0.889		<p>Selección de 1 – 5 Realizo autoevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s) de aprendizaje</p>
Name: Realizo autoevaluación		Type: Numeric																				
Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)																				
Statistic	Value																					
Minimum	1																					
Maximum	5																					
Mean	3.828																					
StdDev	0.889																					

<table border="1"> <tr> <td colspan="2">Name: Realizo coevaluacion</td> <td>Type: Numeric</td> </tr> <tr> <td>Missing: 0 (0%)</td> <td>Distinct: 4</td> <td>Unique: 0 (0%)</td> </tr> <tr> <th>Statistic</th> <th>Value</th> <td></td> </tr> <tr> <td>Minimum</td> <td>2</td> <td></td> </tr> <tr> <td>Maximum</td> <td>5</td> <td></td> </tr> <tr> <td>Mean</td> <td>3.761</td> <td></td> </tr> <tr> <td>StdDev</td> <td>0.851</td> <td></td> </tr> </table>  <p style="text-align: center;">Análisis estadístico - Realizo coevaluación</p>	Name: Realizo coevaluacion		Type: Numeric	Missing: 0 (0%)	Distinct: 4	Unique: 0 (0%)	Statistic	Value		Minimum	2		Maximum	5		Mean	3.761		StdDev	0.851		<p>Selección de 1 – 5</p> <p>Realizo coevaluación de las evidencias, con base en los criterios de evaluación definidos en el/los mapa(s)</p>
Name: Realizo coevaluacion		Type: Numeric																				
Missing: 0 (0%)	Distinct: 4	Unique: 0 (0%)																				
Statistic	Value																					
Minimum	2																					
Maximum	5																					
Mean	3.761																					
StdDev	0.851																					
<table border="1"> <tr> <td colspan="2">Name: Retroalimentacion del profesor</td> <td>Type: Numeric</td> </tr> <tr> <td>Missing: 0 (0%)</td> <td>Distinct: 5</td> <td>Unique: 1 (1%)</td> </tr> <tr> <th>Statistic</th> <th>Value</th> <td></td> </tr> <tr> <td>Minimum</td> <td>1</td> <td></td> </tr> <tr> <td>Maximum</td> <td>5</td> <td></td> </tr> <tr> <td>Mean</td> <td>4.067</td> <td></td> </tr> <tr> <td>StdDev</td> <td>0.824</td> <td></td> </tr> </table>  <p style="text-align: center;">25: Análisis estadístico - Retroalimentación del profesor</p>	Name: Retroalimentacion del profesor		Type: Numeric	Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)	Statistic	Value		Minimum	1		Maximum	5		Mean	4.067		StdDev	0.824		<p>Selección de 1 – 5</p> <p>La retroalimentación del profesor promueve la reflexión sobre las competencias desarrolladas, los logros y los aspectos por mejorar</p>
Name: Retroalimentacion del profesor		Type: Numeric																				
Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)																				
Statistic	Value																					
Minimum	1																					
Maximum	5																					
Mean	4.067																					
StdDev	0.824																					



Selección de 1 – 5
 La retroalimentación de los compañeros muestra objetividad en la valoración de las evidencias y fomenta una cultura democrática en la evaluación

Selección de 1 – 5
 La autoevaluación de evidencias contribuye al rendimiento académico

<p>Name: contribución de la autoevaluación al redimient... Missing: 0 (0%) Distinct: 4 Type: N... Unique: 0...</p> <table border="1"> <thead> <tr> <th>Statistic</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Minimum</td> <td>2</td> </tr> <tr> <td>Maximum</td> <td>5</td> </tr> <tr> <td>Mean</td> <td>3.866</td> </tr> <tr> <td>StdDev</td> <td>0.821</td> </tr> </tbody> </table>  <p>Ilustración 28: Análisis estadístico - contribución de la coevaluación al rendimiento académico</p>	Statistic	Value	Minimum	2	Maximum	5	Mean	3.866	StdDev	0.821	<p>Selección de 1 – 5 La coevaluación de evidencias contribuye al rendimiento académico</p>
Statistic	Value										
Minimum	2										
Maximum	5										
Mean	3.866										
StdDev	0.821										
<p>Name: contribución de la heteroevaluaci... Missi... 0 (0%) Distinct: 4 Type Unique</p> <table border="1"> <thead> <tr> <th>Statistic</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>Minimum</td> <td>2</td> </tr> <tr> <td>Maximum</td> <td>5</td> </tr> <tr> <td>Mean</td> <td>3.896</td> </tr> <tr> <td>StdDev</td> <td>0.788</td> </tr> </tbody> </table>  <p>Ilustración 29: Análisis estadístico - contribución de la Heteroevaluación al rendimiento académico</p>	Statistic	Value	Minimum	2	Maximum	5	Mean	3.896	StdDev	0.788	<p>Selección de 1 – 5 La heteroevaluación de evidencias contribuye al rendimiento académico.</p>
Statistic	Value										
Minimum	2										
Maximum	5										
Mean	3.896										
StdDev	0.788										

Name: contribución de metodología PF...	Type
Missi... 0 (0%)	Distinct: 4
	Unique
Statistic	Value
Minimum	2
Maximum	5
Mean	4.007
StdDev	0.799

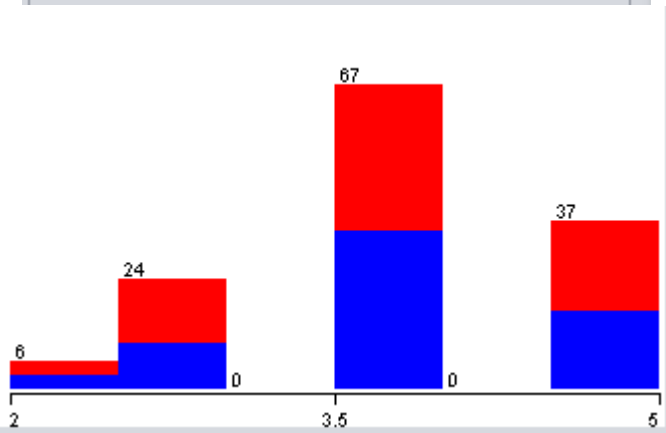


Ilustración 30: Análisis estadístico - contribución de metodología PF al rendimiento académico

Selección de 1 – 5
La metodología de proyectos formativos incentiva a mejorar el rendimiento académico

Name: contribución de la evaluación al red...	Type: ...
Missing: 0 (0%)	Distinct: 6
	Unique: ...
Statistic	Value
Minimum	0
Maximum	5
Mean	3.91
StdDev	0.992

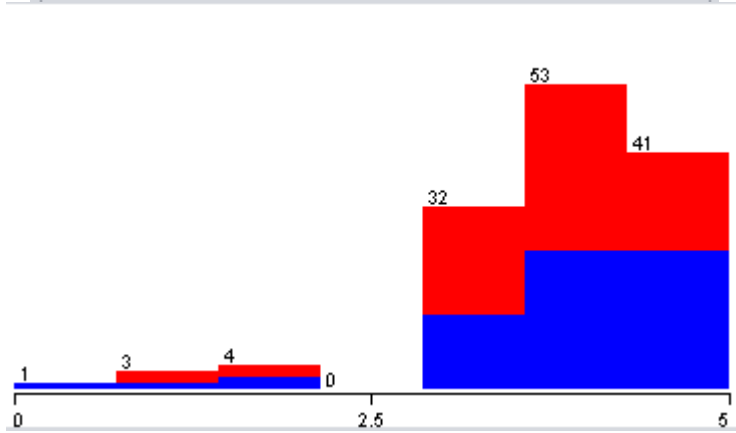


Ilustración 31: Análisis estadístico - contribución de la evaluación al rendimiento académico

Selección de 1 – 5
La metodología de evaluación contribuye a mejorar mi rendimiento académico

Análisis estadístico - Utilidad de los recursos y las actividades de aprendizaje en la plataforma Moodle

Name: Utilidad de los recursos de ap...		Type: Nu...
Missin... 0 (0%)	Distinct: 5	Unique: 0 (...)
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	3.672	
StdDev	0.916	

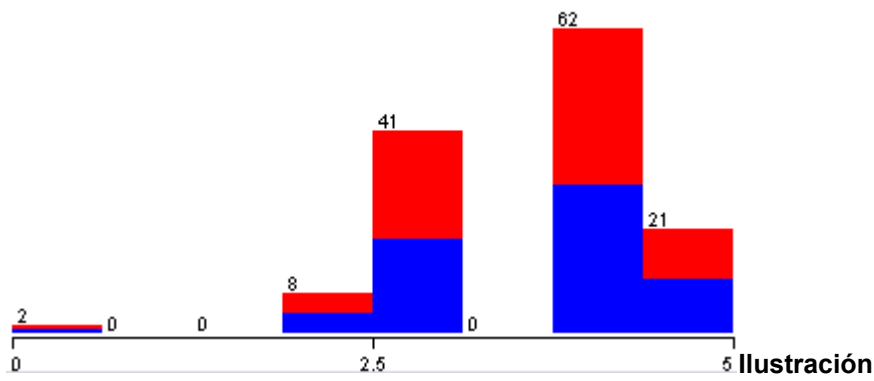


Ilustración 32: Análisis estadístico - Utilidad de los recursos de aprendizaje

Selección de 1 – 5
 Los recursos de aprendizaje ofrecidos en Moodle Utilidad de los recursos de aprendizaje: Los recursos de aprendizaje ofrecidos en Moodle son útiles para la solución del proyecto formativo.

Name: coherencia de los recursos para...		Type: ...
Missing: 0 (0%)	Distinct: 6	Unique: ...
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	3.754	
StdDev	0.945	

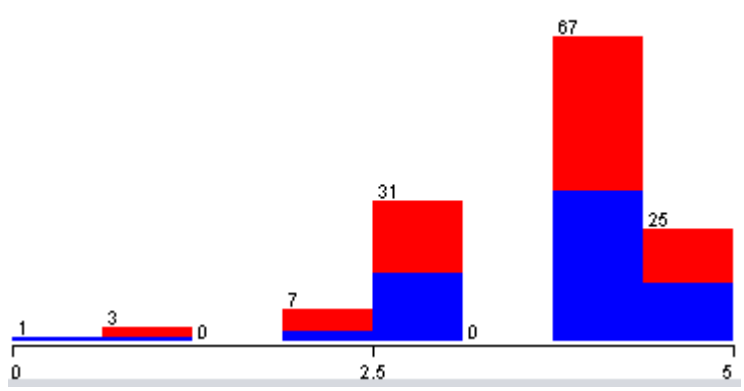


Ilustración 33: Análisis estadístico - coherencia de los recursos para el aprendizaje

Selección de 1 – 5
 Los recursos de aprendizaje ofrecidos en Moodle son coherentes con las actividades de aprendizaje del proyecto formativo

Name: coherencia de los recursos con I...		Type: ...
Missing: 0 (0%)		Distinct: 5
		Unique: ...
Statistic	Value	
Minimum	1	
Maximum	5	
Mean	3.754	
StdDev	0.879	

Selección de 1 – 5
 Los recursos de aprendizaje ofrecidos en Moodle están acorde con mis necesidades de aprendizaje

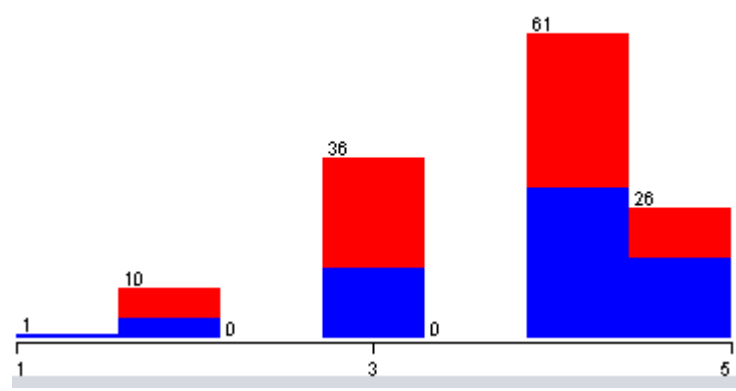


Ilustración 34: Análisis estadístico - Coherencia de los recursos con las necesidades

Name: coherencia de las Actividades con...		Type: ...
Missin... 0 (0%)		Distinct: 6
		Unique: ...
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	3.679	
StdDev	1.001	

Selección de 1 – 5
 Las actividades de aprendizaje ofrecidas en Moodle están acorde con mis necesidades de aprendizaje

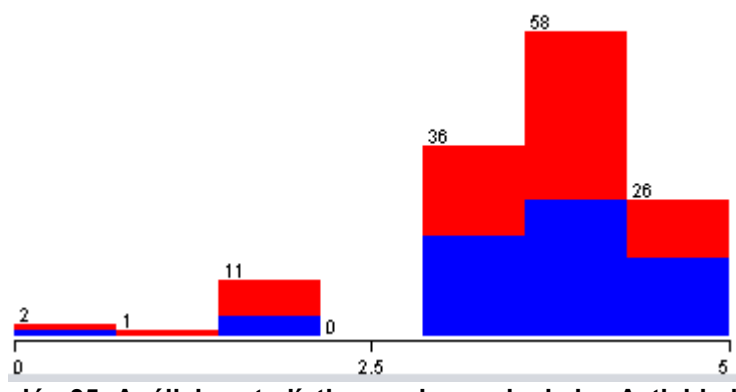


Ilustración 35: Análisis estadístico - coherencia de las Actividades con las necesidades

Name: coherencia con el nivel de co...		Type: Nu...
Missing: 0 (0%)		Distinct: 6
		Unique: 1 (...)
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	3.716	
StdDev	0.978	

Selección de 1 – 5
Las actividades de aprendizaje ofrecidas en Moodle son coherentes mi nivel de competencia

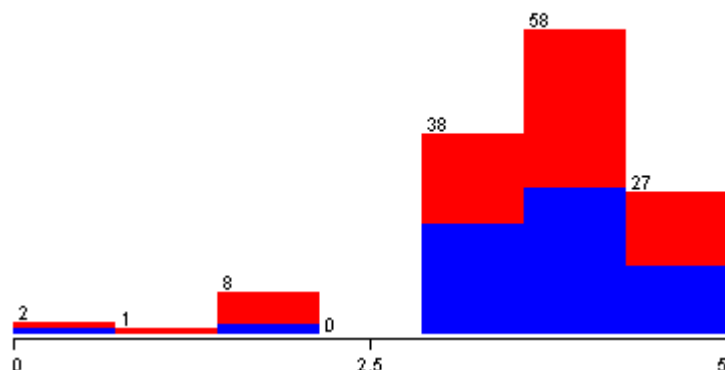


Ilustración 36: Análisis estadístico - coherencia con el nivel de competencia

Análisis estadístico - Opinión sobre el diseño del curso en la Plataforma Moodle

Name: Navegacion del curso		Type: Numeric
Missing: 0 (0%)		Distinct: 6
		Unique: 1 (1%)
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	3.597	
StdDev	0.91	

Selección de 1 – 5
El diseño gráfico del curso en Moodle permite una navegación adecuada.

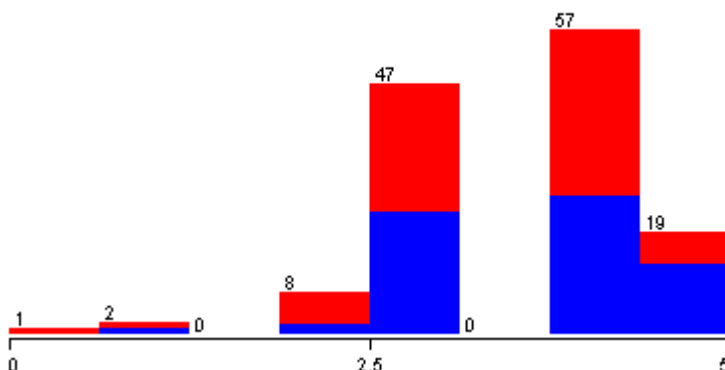


Ilustración 37: Análisis estadístico - Navegación del curso

Name: Diseno grafico		Type: Numeric
Missing: 0 (0%)	Distinct: 5	Unique: 1 (1%)
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	3.582	
StdDev	0.798	

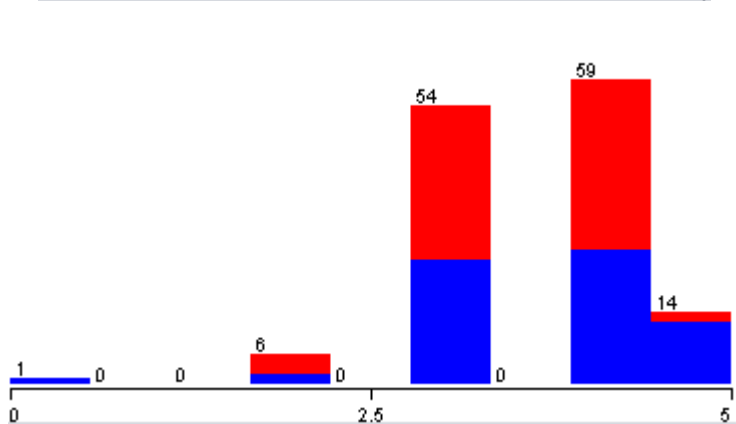


Ilustración 38: Análisis estadístico - Diseño gráfico

Selección de 1 – 5
El diseño gráfico del curso en Moodle es estético

Name: Experiencia del curso en la pl...		Type: Nu...
Missing: 0 (0%)	Distinct: 5	Unique: 1 (...)
Statistic	Value	
Minimum	0	
Maximum	5	
Mean	3.694	
StdDev	0.903	

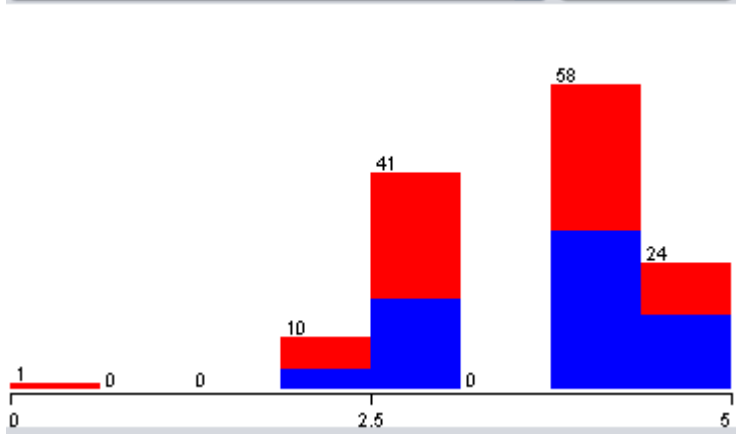
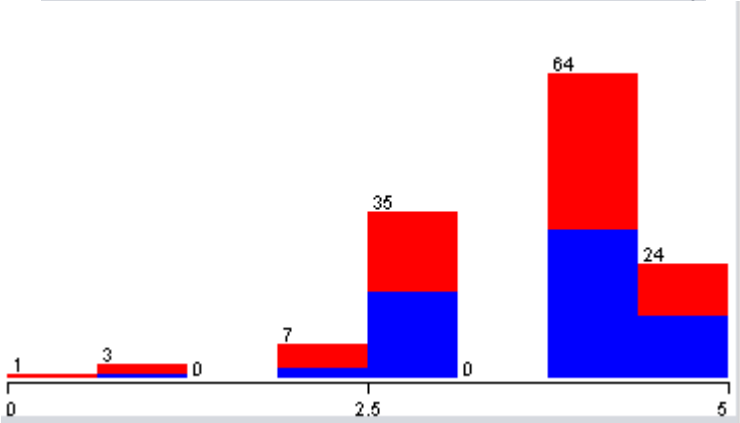
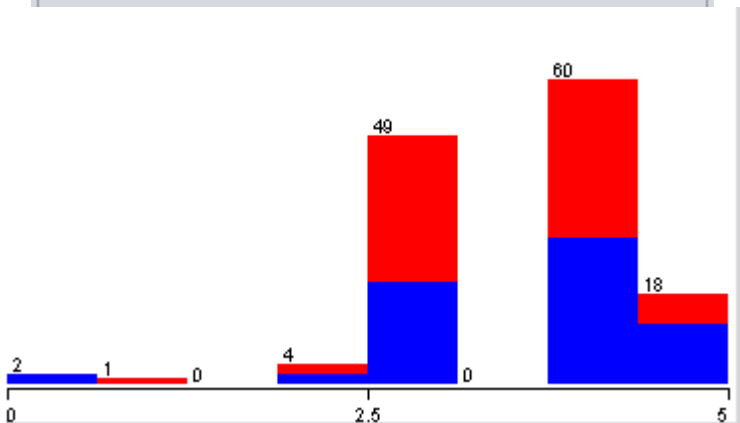
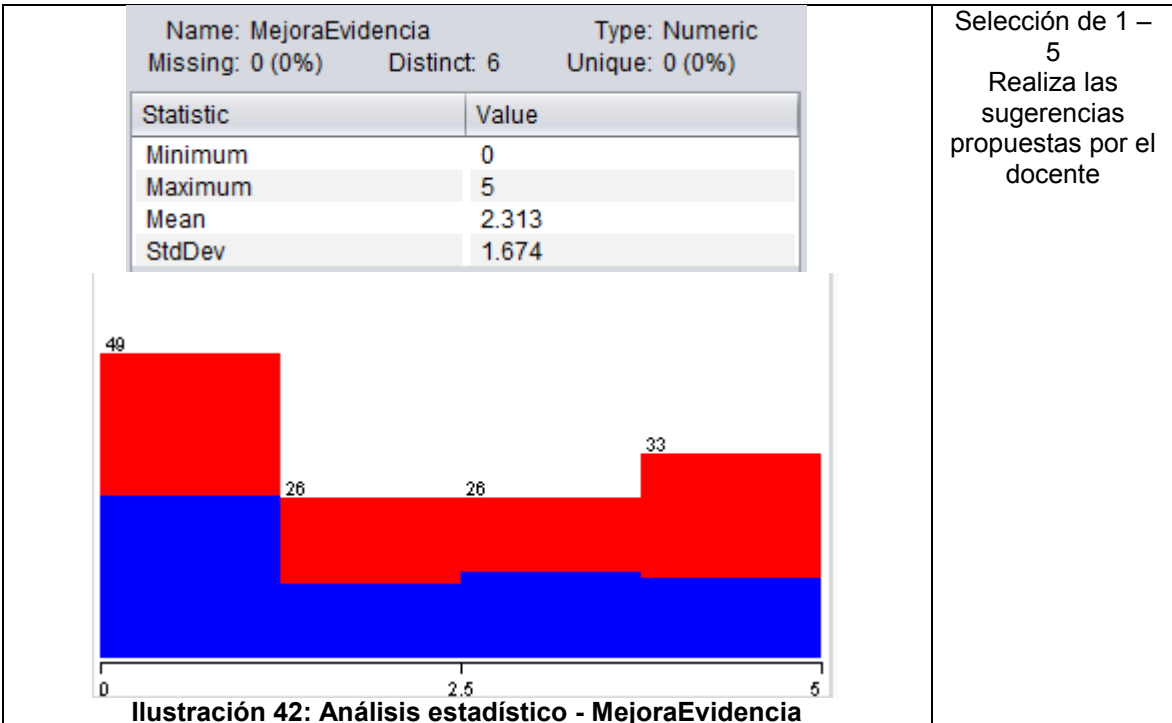


Ilustración 39: Análisis estadístico - Experiencia del curso en la plataforma

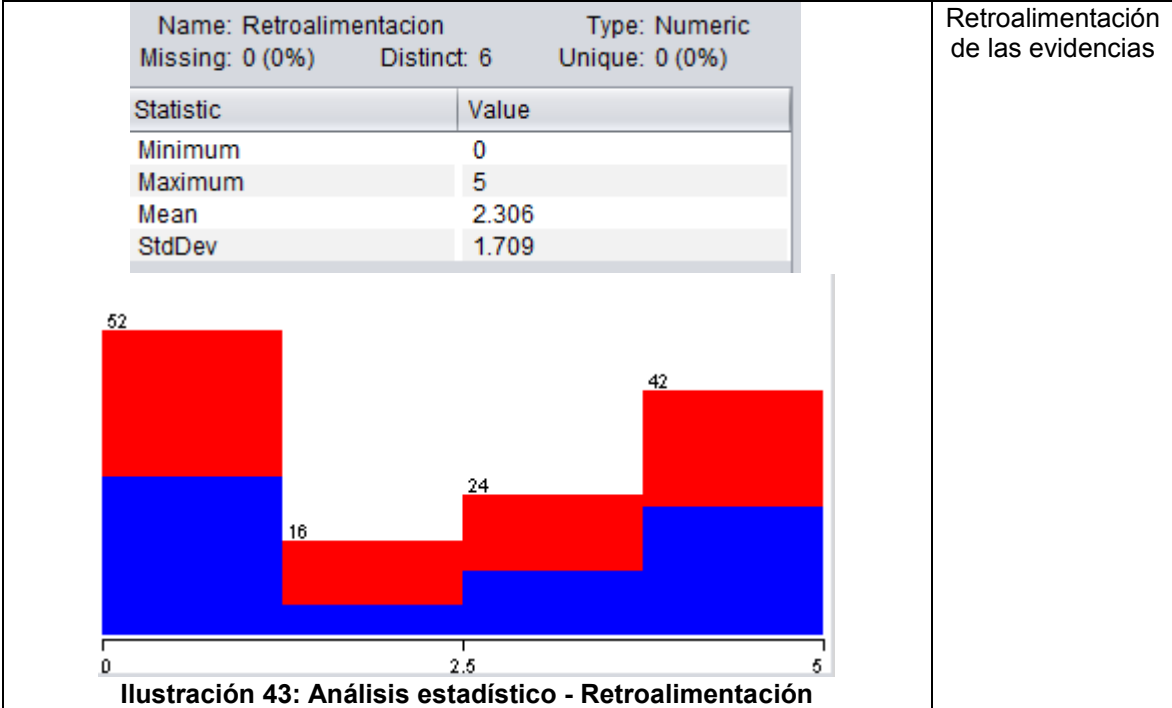
Selección de 1 – 5
Me siento a gusto con el diseño del curso en la plataforma Moodle

<table border="1"> <tr> <td colspan="2">Name: Intefaz del curso</td> <td>Type: Numeric</td> </tr> <tr> <td>Missing: 0 (0%)</td> <td>Distinct: 6</td> <td>Unique: 1 (1%)</td> </tr> <tr> <th>Statistic</th> <th>Value</th> <td></td> </tr> <tr> <td>Minimum</td> <td>0</td> <td></td> </tr> <tr> <td>Maximum</td> <td>5</td> <td></td> </tr> <tr> <td>Mean</td> <td>3.716</td> <td></td> </tr> <tr> <td>StdDev</td> <td>0.947</td> <td></td> </tr> </table>  <p>Ilustración 40: Análisis estadístico - Intefaz del curso</p>	Name: Intefaz del curso		Type: Numeric	Missing: 0 (0%)	Distinct: 6	Unique: 1 (1%)	Statistic	Value		Minimum	0		Maximum	5		Mean	3.716		StdDev	0.947		<p>Selección de 1 – 5</p> <p>La interfaz del curso facilita la comprensión de los elementos del proyecto formativo</p>
Name: Intefaz del curso		Type: Numeric																				
Missing: 0 (0%)	Distinct: 6	Unique: 1 (1%)																				
Statistic	Value																					
Minimum	0																					
Maximum	5																					
Mean	3.716																					
StdDev	0.947																					
<table border="1"> <tr> <td colspan="2">Name: Diseño es estetica agrada...</td> <td>Type: Nume...</td> </tr> <tr> <td>Missing: 0 (0%)</td> <td>Distinct: 6</td> <td>Unique: 1 (1%)</td> </tr> <tr> <th>Statistic</th> <th>Value</th> <td></td> </tr> <tr> <td>Minimum</td> <td>0</td> <td></td> </tr> <tr> <td>Maximum</td> <td>5</td> <td></td> </tr> <tr> <td>Mean</td> <td>3.627</td> <td></td> </tr> <tr> <td>StdDev</td> <td>0.89</td> <td></td> </tr> </table>  <p>Ilustración 41: Análisis estadístico - Diseño es estética agradable</p>	Name: Diseño es estetica agrada...		Type: Nume...	Missing: 0 (0%)	Distinct: 6	Unique: 1 (1%)	Statistic	Value		Minimum	0		Maximum	5		Mean	3.627		StdDev	0.89		<p>Selección de 1 – 5</p> <p>El diseño de los recursos y las actividades es estéticamente agradable</p>
Name: Diseño es estetica agrada...		Type: Nume...																				
Missing: 0 (0%)	Distinct: 6	Unique: 1 (1%)																				
Statistic	Value																					
Minimum	0																					
Maximum	5																					
Mean	3.627																					
StdDev	0.89																					



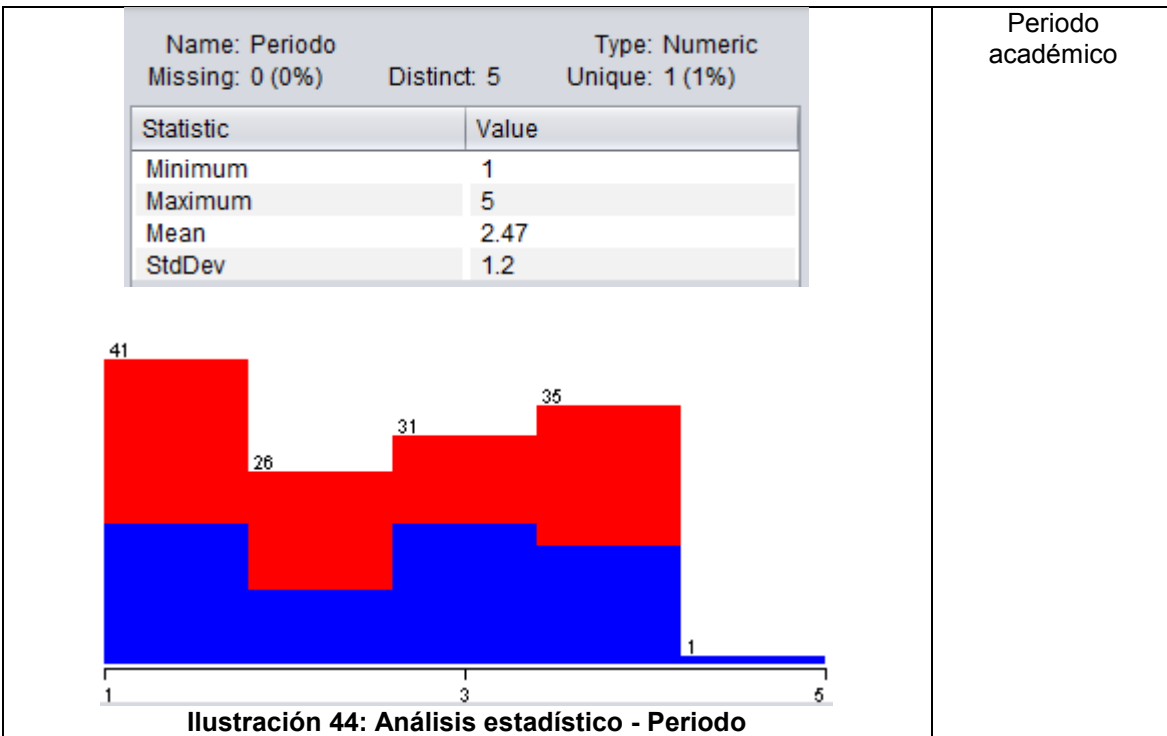
Selección de 1 – 5
Realiza las sugerencias propuestas por el docente

Ilustración 42: Análisis estadístico - MejoraEvidencia

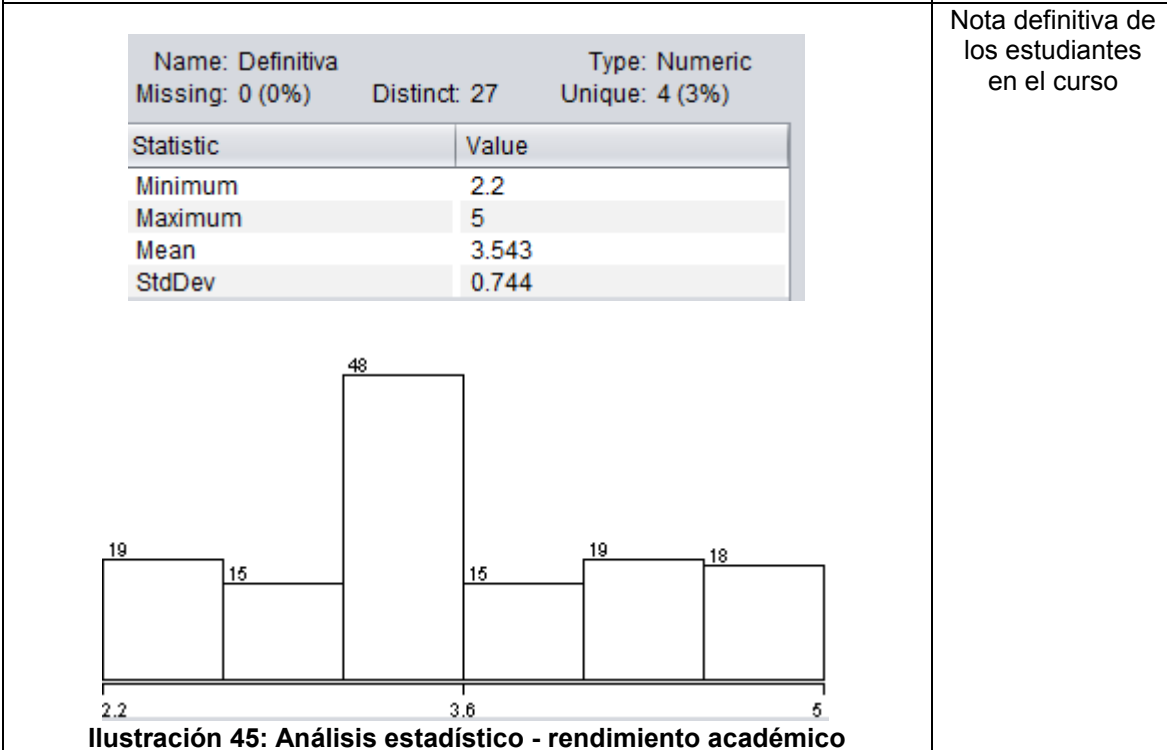


Retroalimentación de las evidencias

Ilustración 43: Análisis estadístico - Retroalimentación

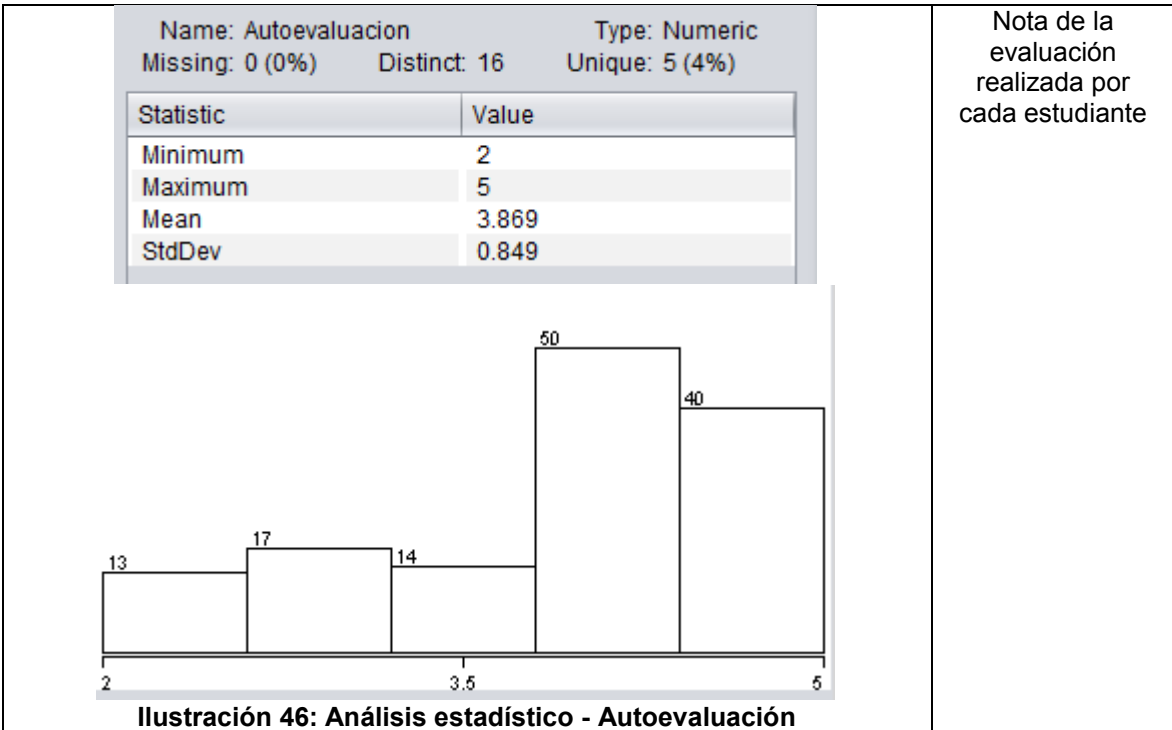


Periodo académico

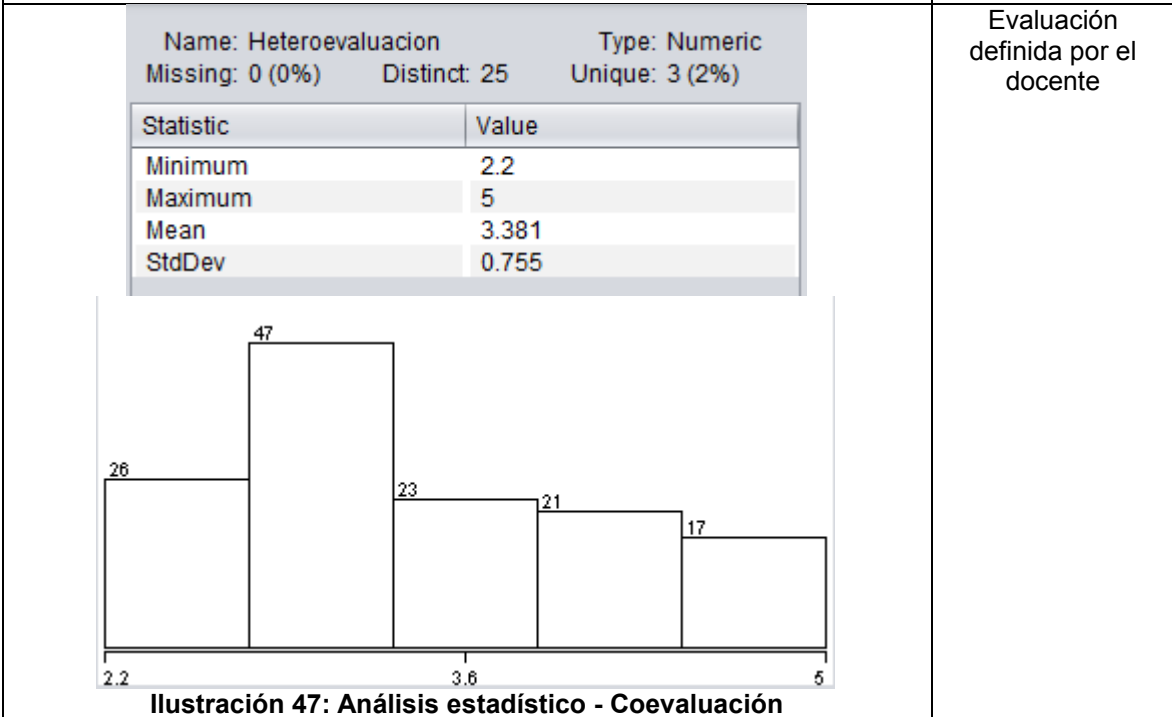


Nota definitiva de los estudiantes en el curso

Aspectos de la Evaluación



Nota de la evaluación realizada por cada estudiante



Evaluación definida por el docente

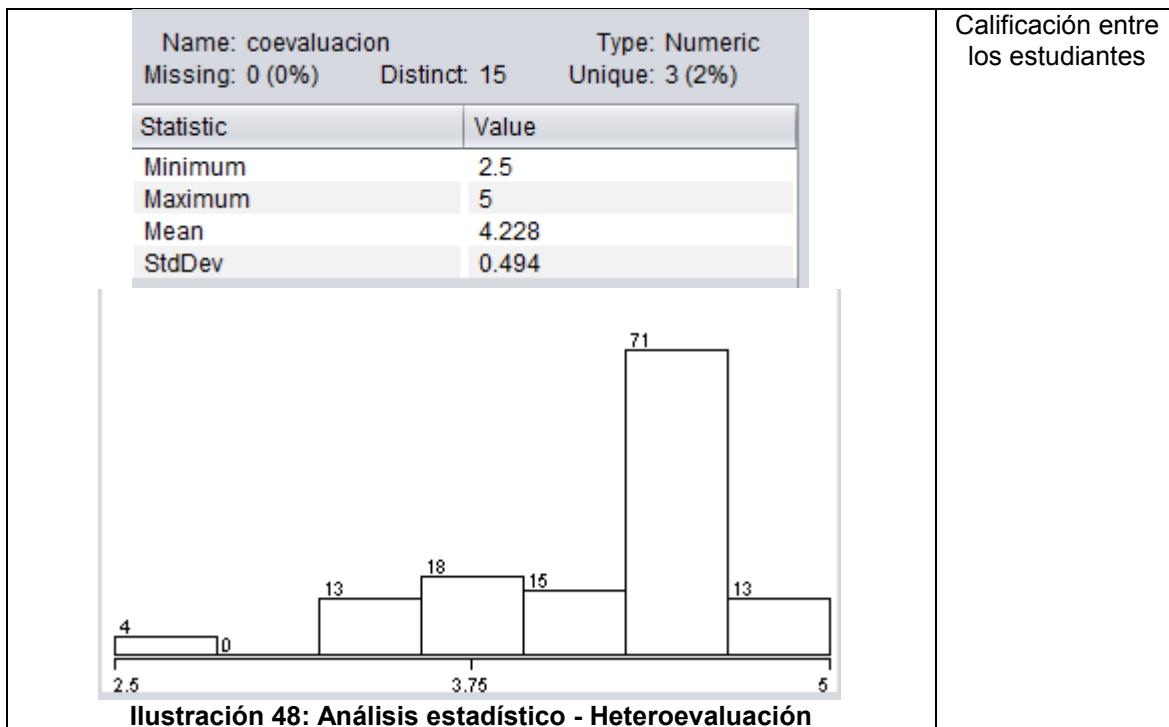


Tabla 13: Descripción Estadística de los Datos

1.3. Transformaciones

Las variables no tuvieron ninguna transformación, a excepción de la utilización del algoritmo “A priori”, el cual no se puede utilizar variables numéricas; para esto se utilizó el método de **DISCRETIZACIÓN**.

2. Selección de Atributos

Un problema en la aplicación de las técnicas de minería de datos, es identificar las variables más representativas para construir el modelo (Kohavi & John, 1997). Ya que en teoría se podría suponer que al tener un número considerable de variables podría generar mejores resultados, sin embargo, se presenta una alta dimensionalidad con mucha información irrelevante que es difícil de manejar en la implementación de estos algoritmos (Gutiérrez García, 2016a).

Langley (Langley, 1994) propone reducir el número de variables antes de aplicar algún algoritmo de clasificación, con el objetivo de generar mayor precisión y menor ruido en los resultados de las técnicas de minería.

Se realiza la reducción del tamaño del conjunto de datos esto permite mejorar eficiencia del proceso de Minería de Datos, además se realiza una selección de datos relevantes en la cual se realiza la selección de características, muestreo o selección de instancias.

Los métodos de filtrado utilizan un conjunto de entrenamiento en los cuales se conoce la clase, para determinar la correlación entre cada característica con dicha clase. La importancia en los atributos

se realiza una vez y con diferentes clasificador(Gutiérrez García, 2016b)es el cual evalúa la misma muestra (Gutiérrez García, 2016).

En la selección de atributos tiene como objetivo realizar un ranking de características que permite disminuir la dimensionalidad permitiendo crear modelos que definen el grupo observado. Se resalta que no existe una técnica que permita seleccionar siempre los mejores resultados, generalmente se requiere la intervención de expertos.

Kohavi y John (Kohavi & John, 1997) presentan tres tipos de relevancia para los atributos: Fuertemente relevantes, Débilmente relevantes e Irrelevante, además Langley (Langley, 1994) propone reducir el número de variables antes de aplicar algún algoritmo de clasificación, con el objetivo de generar mayor precisión y menor ruido en los resultados de las tecnicas de minería.

Se realizó la evaluación de atributos para identificar cuáles eran más relevantes, los métodos que se utilizaron para la selección de estas variables parten del conjunto total de datos, los cuales determinaron la correlación entre atributos y la clase dependiente rendimiento académico “Definitiva”.

- **CfsSubsetEval (CFS):** Este Evaluador genera un ranking de atributos de acuerdo a la correlación con la clase dependiente “Definitiva”, realiza una búsqueda de las variables que presentan una alta correlación con la clase, y una baja intercorrelacion con otras variables; una desventaja de este método es que en algunas ocasiones se pueden eliminar características que no están relacionados con la clase pero si con otros atributos si (Gutiérrez García, 2016a).
- **CorrelationAttributeEval:** Este Evaluador realiza un análisis de relación entre el atributo y la clase dependiente “Definitiva”, los atributos que son categóricos los analiza como indicadores mediante un promedio ponderado (Bouckaert, Frank, & Hall, 2010).
- **GainRatioAttributeEval:** Evalúa el valor de un atributo midiendo la relación de beneficio con respecto a la clase dependiente “Definitiva” (Bouckaert et al., 2010).
- **InfoGainAttributeEval:** Realiza la evaluación de los atributos con el valor de la ganancia con respecto a la clase “Definitiva” (Bouckaert et al., 2010).
- **OneRAttributeEval:** Evalúa el valor de los atributos utilizando el método de clasificación ONER (Bouckaert et al., 2010).
- **ReliefFAttributeEval:** Evalúa las variables mediante un muestreo repetido de una instancia, considerando el valor de un atributo para la instancia más cercana de la misma clase “Definitiva” (Bouckaert et al., 2010).
- **SymmetricalUncertAttributeEval:** Evalúa el valor de un atributo mediante la medición de la incertidumbre simétrica con respecto a la clase “Definitiva” (Bouckaert et al., 2010).

Evaluador de atributos	Algoritmo de clasificación	Método de Búsqueda	Clase	Variables Seleccionadas

CfsSubsetEval (CFS)		GreedyStepwise	Definitiva	<p>promedio_universidad</p> <p>Posicion_hermanos</p> <p>Realizo autoevaluacion</p> <p>Heteroevaluacion</p>
CorrelationAttributeEval		Ranker	Definitiva	<p>0.95855 48 Heteroevaluacion</p> <p>0.18955 2 Genero</p> <p>0.11194 49 coevaluacion</p> <p>0.08546 16 Posicion_hermanos</p> <p>0.06701 1 Edad</p> <p>0.05837 47 Autoevaluacion</p> <p>0.05793 17 Tipo_vivienda</p> <p>0.04454 43 MejoraEvidencia</p> <p>0.04368 12 Lugar_residencia</p> <p>0.04218 14 Nivel_educativo_padre</p> <p>0.0392513 Nivel_educativo_madre</p> <p>0.03753 4 Estado_civil</p> <p>0.01855 3 Estrato</p> <p>0.01557 9 egresado_colegio</p> <p>0.01452 6 Personas_Cargo</p> <p>0.01092 11 ICFES</p> <p>0.00399 5 Trabaja</p>
ReliefAttributeEval		Ranker	Definitiva	<p>0.234965 48 Heteroevaluacion</p> <p>0.027048 2 Genero</p> <p>0.020351 4 Estado_civil</p> <p>0.017905 38 Navegacion del curso</p> <p>0.016203 47 Autoevaluacion</p> <p>0.011976 23Orientan Elaboracion de las evidencias</p> <p>0.010903 8promedio_universidad</p> <p>0.008215-18Contribucion Metodologia PF</p> <p>0.007303 41 Intefaz del curso</p> <p>0.007257 12 Lugar_residencia</p> <p>0.006659 25 Realizo coevaluacion</p> <p>0.00431 19 Complejidad del PF</p> <p>0.003707 39 Diseno grafico</p> <p>0.003382 28 contribución de la autoevaluacion al redimimiento academico</p> <p>0.003051 24 Realizo autoevaluacion</p> <p>0.003018 32 contribución de la evaluacion al redimimiento academico</p>

				<p>0.002875 20 Participacion en el PF</p> <p>0.002407 31 contribución de metodología PF al redimiento academico</p> <p>0.002285 36 coherencia de las Actividades con las necesidades</p> <p>0.002019 26 Retroalimentacion del profesor</p> <p>0.001119 34 coherencia de los recursos para el aprendizaje</p> <p>0.00072 37 coherencia con el nivel de competencia</p> <p>0.000448 1 Edad</p>
ClassifierAttributeEval		Ranker	Definitiva (se discretiza la clase para aplicar árbol j48)	<p>0.63134 48 Heteroevaluacion</p> <p>0.09403 47 Autoevaluacion</p> <p>0.06343 44 Retroalimentacion</p> <p>0.0403 49 coevaluacion</p> <p>0.03358 19 Complejidad del PF</p> <p>0.02239 28 contribución de la autoevaluacion al redimiento academico</p> <p>0.01119 43 MejoraEvidencia</p> <p>0.00746 38 Navegacion del curso</p> <p>0.00448 32 contribución de la evaluacion al redimiento academico</p> <p>0.00373 34 coherencia de los recursos para el aprendizaje</p> <p>0.00373 35 coherencia de los recursos con las necesidades</p> <p>0.00373 27 Retroalimentacion de los companeros</p> <p>0.00373 23 Orientan Elaboracion de las evidencias</p>
Information Gain Ranking Filter		Ranker	Definitiva (se discretiza la clase para aplicar árbol j48)	<p>1.96328 48 Heteroevaluacion</p> <p>0.1389 16 Posicion_hermanos</p> <p>0.10857 14 Nivel_eduativo_padre</p> <p>0.09916 12 Lugar_residencia</p> <p>0.0956 13 Nivel_edcativo_madre</p> <p>0.0607 2 Genero</p> <p>0.04659 4 Estado_civil</p> <p>0.03773 17 Tipo_vivienda</p> <p>0.00529 5 Trabaja</p> <p>0.00257 9 egresado_colegio</p>

Tabla 14: Evaluador de atributos

Como se presenta anteriormente en la tabla 3, los evaluadores de atributos implementados determinando así las variables a considerar.

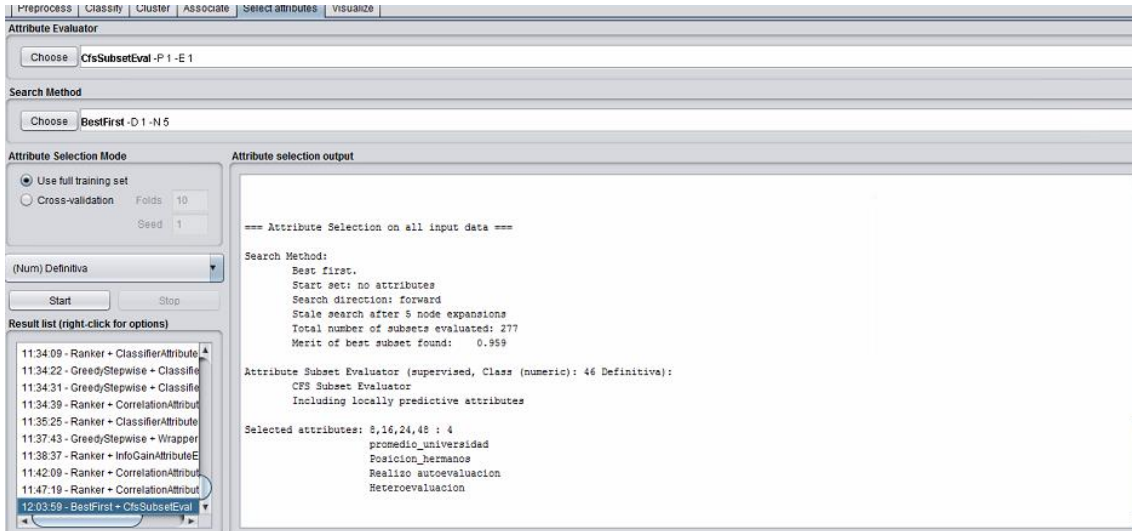


Figura 13: Evaluador de atributos CfsSubsetEval

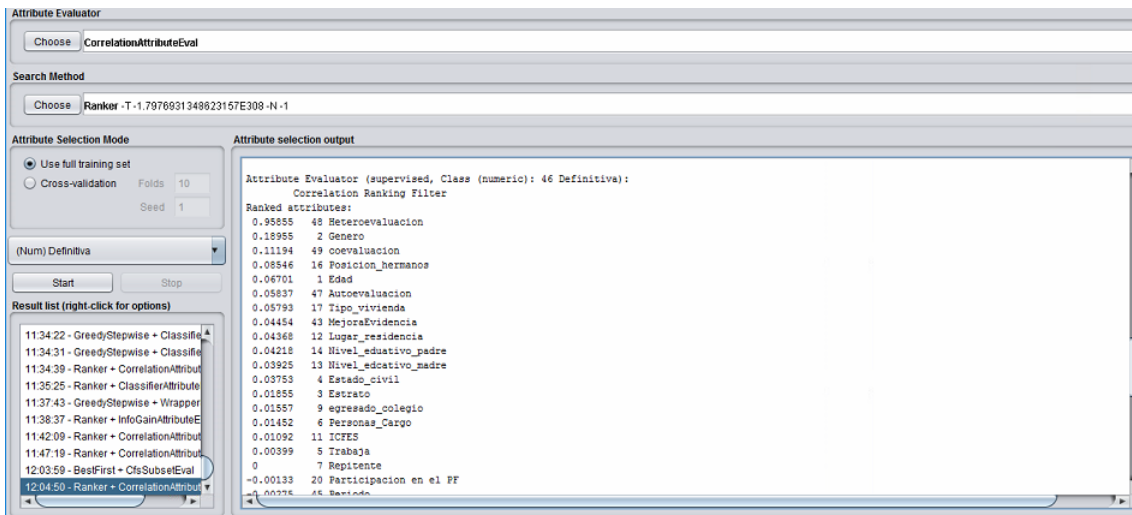


Figura 14: Evaluador de atributos CorrelationAttributeEval

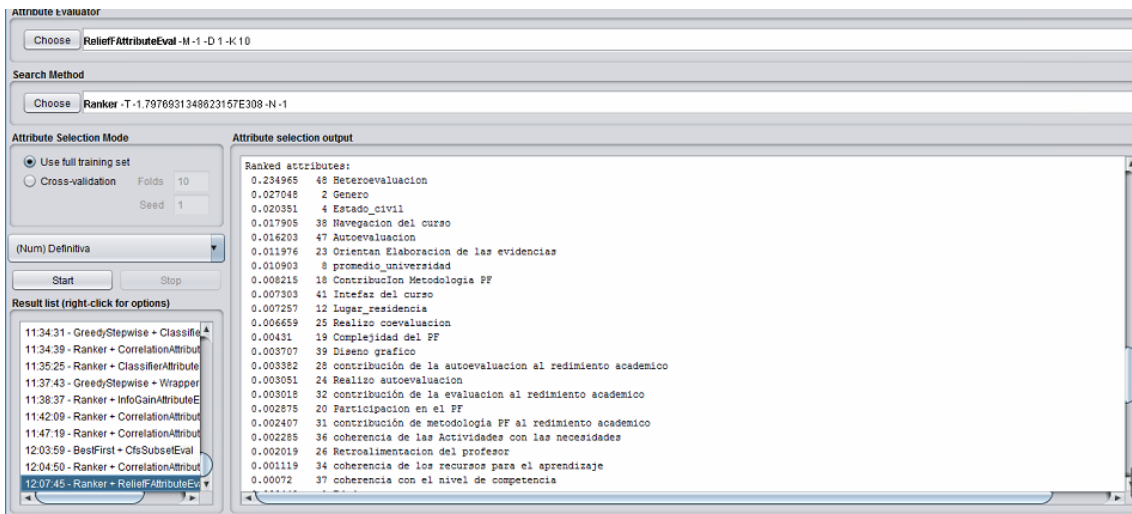


Figura 15: Evaluador de atributos ReliefAttributeEval

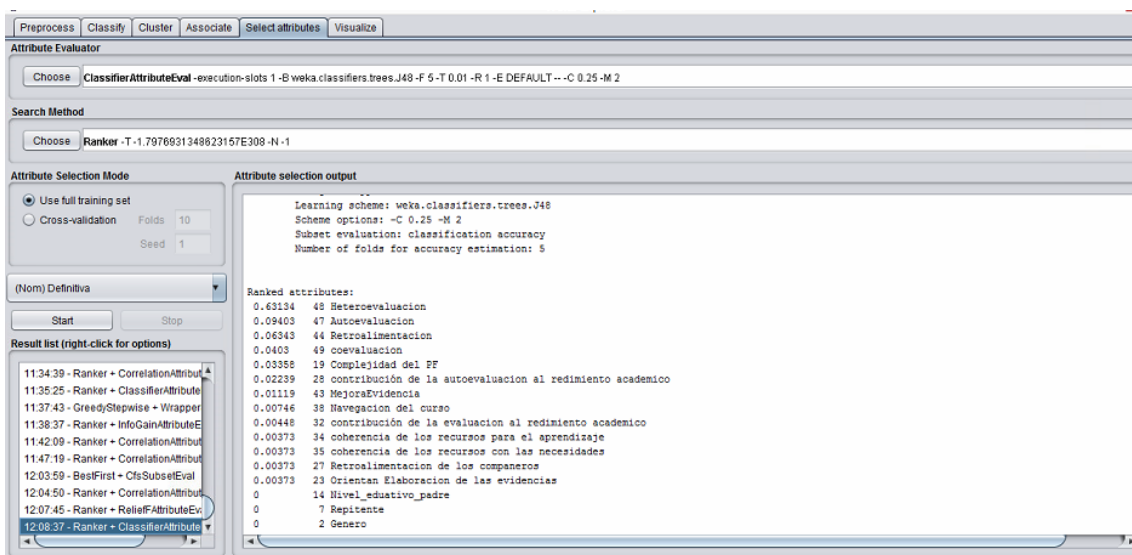


Figura 16: Evaluador de atributos ClassifierAttributeEval

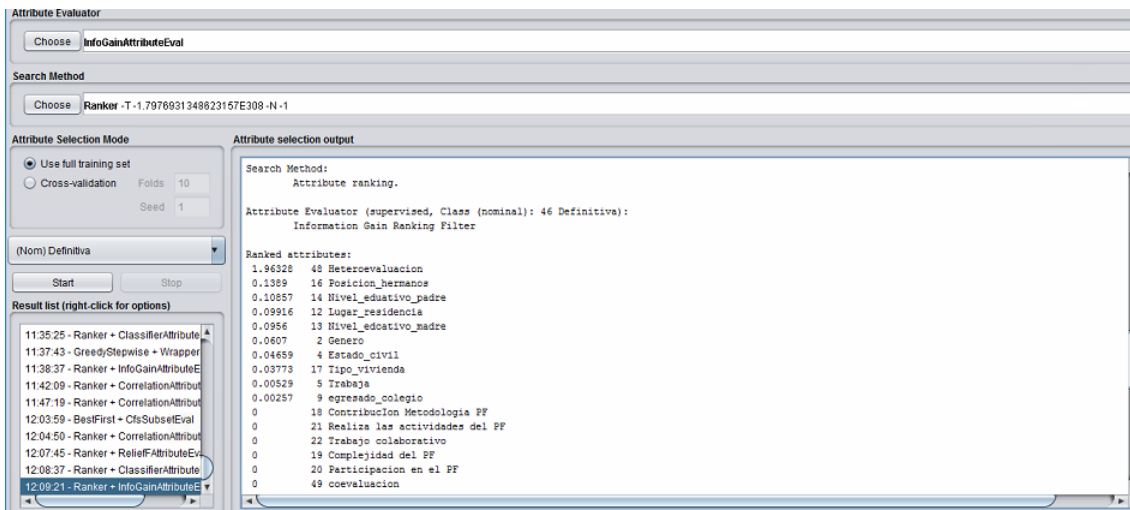


Figura 17: Evaluador de atributos Information Gain Ranking Filter

2.1. Selección de variables y datos que conformarían el conjunto de aprendizaje para el modelo de minería.

Después de realizar la ejecución de los algoritmos para seleccionar los atributos, se procedió a determinar los resultados comunes, es decir qué atributos estaban principalmente en las listas de los rankings en cada ejecución. Como se puede observar la tabla anterior, se resaltó en azul los atributos que son comunes en los métodos, con esta estrategia se puede ver que los atributos de **Correlation Ranking Filter** comparten mayor coincidencia con el resto de los algoritmos, por lo que se puede indicar que las 17 características más relacionadas con el rendimiento académico (definitiva), son:

1. Definitiva (clase)
2. Heteroevaluacion
3. Genero
4. coevaluacion
5. Posicion_hermanos
6. Edad
7. Autoevaluacion
8. Tipo_vivienda
9. MejoraEvidencia
10. Lugar_residencia
11. Nivel_educativo_padre
12. Nivel_educativo_madre
13. Estado_civil
14. Estrato
15. Egresado_colegio
16. Personas_Cargo
17. ICFES
18. Trabaja
19. Autoevaluacion

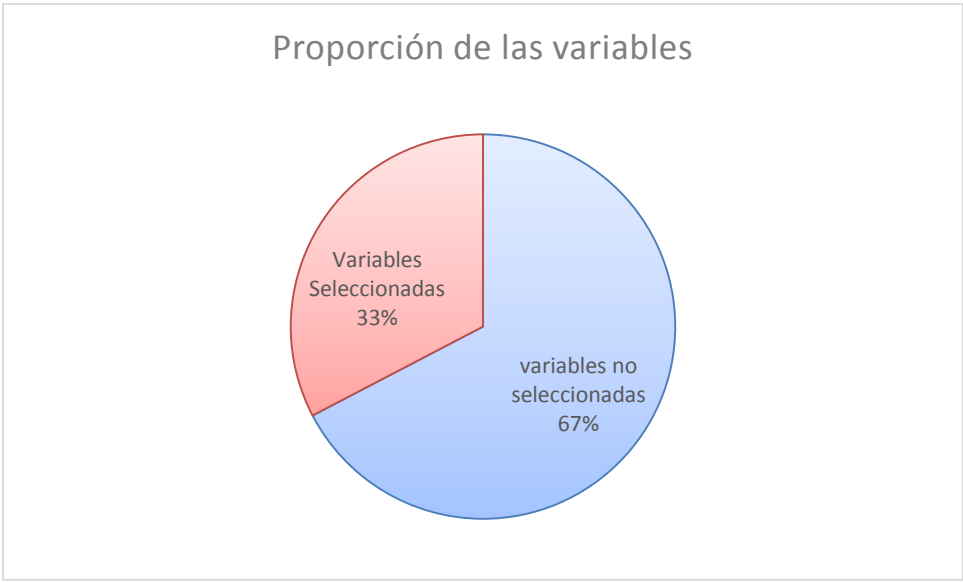


Figura 18: Proporción de las variables

Anexo 4: Técnicas de minería

1. MODELADO

El objetivo de esta fase es aplicar las técnicas de minería en el conjunto de datos obtenido en la fase de pre procesamiento, para detectar información oculta y patrones. Con cada uno de los patrones se pretende encontrar las variables más relevantes de cada uno y finalmente, realizar una votación entre las técnicas utilizadas teniendo en cuenta el número de veces que el factor fue encontrado como el más relevante.

Los algoritmos que se utilizan para determinar los factores que más influyen en la nota definitiva de los cursos bajo la metodología de proyectos formativos son:

- **Experimento 1:** Perfilamiento de los estudiantes. Clústeres
- **Experimento 2:** Selección de factores con Árboles de Decisión
- **Experimento 3:** Selección de factores con PCA
- **Experimento 4:** Selección de factores con regresión logística
- **Experimento 5:** Reglas de asociación

La herramienta que se utilizó para este proyecto en el modelado y procesamiento fue WEKA, los datos analizados se componen de 16 variables seleccionadas definidas en el anexo 5. La cantidad de la muestra es de 133 registros.

Variables			
1	Definitiva (clase)	10	Lugar_residencia
2	Heteroevaluacion	11	Nivel_eduativo_padre
3	Genero	12	Nivel_edcativo_madre
4	coevaluacion	13	Estado_civil
5	Posicion_hermanos	14	Estrato
6	Edad	15	Egresado_colegio
7	Autoevaluacion	16	Personas_Cargo
8	Tipo_vivienda	17	ICFES
9	MejoraEvidencia	18	Trabaja

Tabla 15: Variables para el modelamiento.

1.1. EXPERIMENTO 1: Análisis de perfiles de los estudiantes en cursos bajo la metodología de proyectos formativos

El objetivo es establecer el perfil de los estudiantes que realizaron el curso bajo la metodología de proyecto formativo, permitiendo formar grupos de acuerdo a las características.

Método: K-means y EM

Se debe determinar el número de grupos mediante el algoritmo de EM el cual sugiere 4 clústeres.

Clustered Instances: Log likelihood: -23.78086

0	45 (34%)
1	35 (26%)
2	20 (15%)
3	34 (25%)

Después de obtener los 4 grupos, se aplica el algoritmo de Simple K-means con dos, tres, cuatro y cinco grupos donde se selecciona el que tiene menor cohesión como se presenta a continuación:

- **Con 2 clúster:** Within cluster sum of squared errors: 527.7053244110108
- **Con 3 clúster:** Within cluster sum of squared errors: 497.4589292221793
- **Con 4 clúster:** Within cluster sum of squared errors: 460.9013850414691
- **Con 5 clúster:** Within cluster sum of squared errors: 435.90998566120703

Teniendo en cuenta lo anterior, se concluye que la agrupación más óptima es con 5 clústeres para recorrer el algoritmo, debido a que la cohesión es menor en este caso.

En la ilustración 1 se presentan los resultados generados por Weka de cada uno de los clúster.

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (134.0)	Cluster#				
		0 (28.0)	1 (22.0)	2 (37.0)	3 (26.0)	4 (21.0)
Edad	20.4254	19.5	21.1818	21.7568	19.3462	19.8571
Genero	HOMBRE	HOMBRE	HOMBRE	MUJER	MUJER	HOMBRE
Estrato	2.7388	2.7857	2.8636	2.6757	2.9615	2.381
Estado_civil	SOLTERO	SOLTERO	SOLTERO	SOLTERO	SOLTERO	SOLTERO
Trabaja	NO	NO	NO	SI	NO	NO
Personas_Cargo	0.7388	0.6429	0.8636	0.7027	0.5385	1.0476
egresado_colegio	PUBLICO	PRIVADO	PUBLICO	PUBLICO	PUBLICO	PUBLICO
ICFES	299.5	333.8214	308.1818	308.9459	275.8077	257.3333
Lugar_residencia	ARMENIA	ARMENIA	ARMENIA	ARMENIA	ARMENIA	ARMENIA
Nivel_edcativo_madre	SECUNDARIA	PRIMARIA	SECUNDARIA	SECUNDARIA	SECUNDARIA	PRIMARIA
Nivel_educativo_padre	SECUNDARIA	SECUNDARIA	SECUNDARIA	SECUNDARIA	UNIVERSIDAD	PRIMARIA
Posicion_hermanos	SEGUNDO	SEGUNDO	PRIMERO	SEGUNDO	PRIMERO	PRIMERO
Tipo_vivienda	PROPIA	ARRENDADA	PROPIA	PROPIA	PROPIA	PROPIA
MejoraEvidencia	2.3134	2.1071	2.4545	1.973	3.2308	1.9048
Definitiva	'(2.76-3.32]'	'(2.76-3.32]'	'(3.88-4.44]'	'(-inf-2.76]'	'(2.76-3.32]'	'(3.32-3.88]'
Autoevaluacion	3.8694	3.6036	4.1909	4	3.6192	3.9667
Heteroevaluacion	3.3813	3.3607	4.0045	2.9784	3.4308	3.4048
coevaluacion	4.2284	4.2679	4.4455	4.1892	4.0538	4.2333

Ilustración 26: perfilamiento de los estudiantes WEKA

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

0 28 (21%)
1 22 (16%)
2 37 (28%)
3 26 (19%)
4 21 (16%)

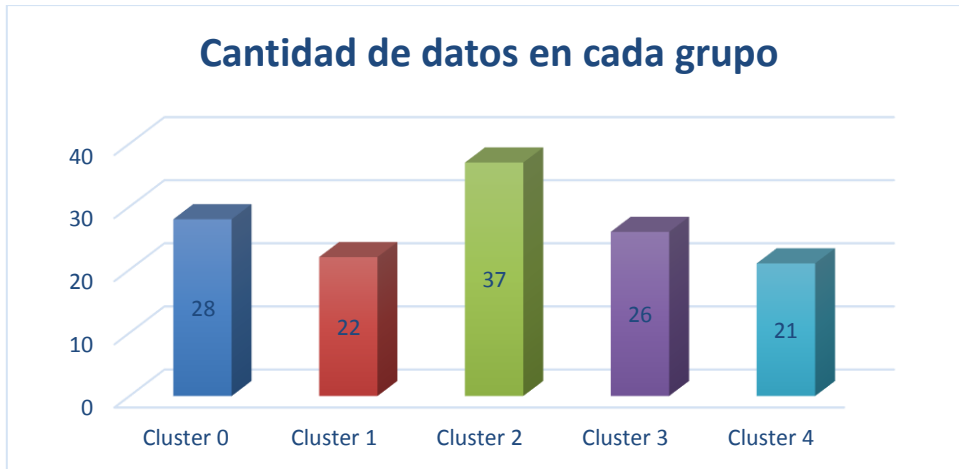


Ilustración 27: Datos por clúster

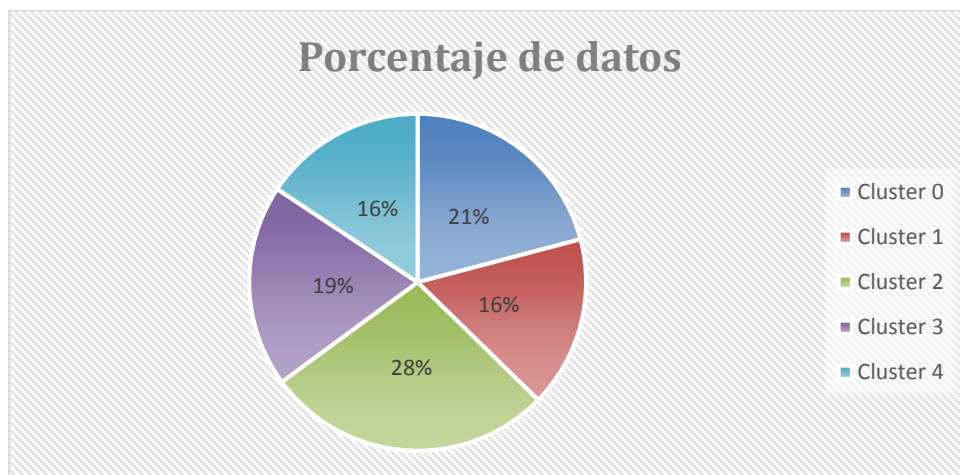


Ilustración 28: porcentaje de datos por Clúster

Clúster	Descripción
---------	-------------

0	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (2.7 – 3.32], siendo el 21 % de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes hombres una edad promedio de 19.5 años, con estrado 2 – 3, hombres solteros que no trabajan y pueden tener entre 0 y 1 persona a cargo. Son egresado de un colegio privado con un promedio de ICFES 333.8, residente en armenia con un nivel educativo de primaria para la madre y secundaria para el padre, es el segundo en la posición de los hermanos y tiene una vivienda arriendo.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que en el ítem mejora la evidencia el promedio fue de 2.1, su autoevaluación de 3.6, su heteroevaluación en 3.3. y la coevaluación en 4.2.</p>
1	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (3.8 – 4.4], siendo el 16% de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes hombres una edad promedio de 21.1 años, con estrado 2 – 3, hombres solteros que no trabajan y pueden tener entre 0 y 1 persona a cargo. Son egresado de un colegio público con un promedio de ICFES 308.9, residente en armenia con un nivel educativo de secundaria para la madre y el padre, es el primero en la posición de los hermanos y tiene una vivienda propia.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que aprobaron el curso, además, en el ítem mejora la evidencia el promedio fue de 2.4, su autoevaluación de 4.1, su heteroevaluación en 4. y la coevaluación en 4.4.</p>
2	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (0 – 2.76], siendo el 28% de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes Mujeres una edad promedio de 21.7 años, con estrado 2 – 3, Mujeres solteras que si trabajan y pueden tener entre 0 y 1 persona a cargo. Son egresados de un colegio público con un promedio de ICFES 308.9, residente en armenia con un nivel educativo de secundaria para la madre y el padre, es el segundo en la posición de los hermanos y tiene una vivienda propia.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que no aprobaron el curso, además, en el ítem mejora la evidencia el promedio fue de 1.9, su autoevaluación de 4, su heteroevaluación en 2.9. y la coevaluación en 4.1.</p>
3	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (2.76– 3.332], siendo el 19% de los estudiantes del curso.</p> <p>Este grupo es conformado por estudiantes Mujeres una edad promedio de 19 años, con estrado 2 – 3, Mujeres solteras que no trabajan y pueden tener entre 0 y 1 persona a cargo. Son egresados de un colegio público con un promedio de ICFES 275.8, residente en armenia con un nivel educativo de secundaria para la madre y universitaria para el padre, es el primero en la posición de los hermanos y tiene una vivienda propia.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que no aprobaron el curso, además, en el ítem mejora la evidencia el promedio fue de 3.2, su autoevaluación de 3.6, su heteroevaluación en 3.4. y la coevaluación en 4.</p>
4	<p>Este clúster está caracterizado por los estudiantes que obtuvieron una nota definitiva entre (3.32 – 3.88], siendo el 16% de los estudiantes del curso.</p>

	<p>Este grupo es conformado por estudiantes hombres una edad promedio de 19.8 años, con estrado 2 – 3, hombres solteros que no trabajan y pueden tener 1 persona a cargo. Son egresados de un colegio público con un promedio de ICFES 257.3, residente en armenia con un nivel educativo de primaria para la madre y el padre, es el primero en la posición de los hermanos y tiene una vivienda propia.</p> <p>Se puede concluir que este grupo está conformado por estudiantes que no aprobaron el curso, además, en el ítem mejora la evidencia el promedio fue de 1.9, su autoevaluación de 3.9, su heteroevaluación en 3.4 y la coevaluación en 4.2.</p>
--	--

Tabla 16: Evaluación de Resultados – Clústeres

Según los resultados presentados en la tabla anterior, se puede interpretar que los estudiantes con un rendimiento académico más bajo son aquellos que:

- Estudiantes mujeres
- Estudiantes que trabajan
- Estudiantes que obtuvieron en su heteroevaluación una nota menor a 3.
- Estudiantes que obtuvieron en su coevaluación una nota menor 4.2.

Ahora, los estudiantes con un rendimiento académico mayor son aquellos que:

- Estudiantes hombres
- Estudiantes que no trabajan
- Estudiantes que son primeros en la posición de sus hermanos
- Estudiantes que obtuvieron en su auto evaluación una nota de 4.1
- Estudiantes que obtuvieron en su heteroevaluación una nota 4.
- Estudiantes que obtuvieron en su coevaluación mayor que 4.2.

Variables que presentan diferencias significativas entre clústeres

- Sexo
- Trabaja
- Heteroevaluación
- Coevaluación
- Posición entre hermanos
- Autoevaluación
- Tipo de vivienda
- Mejora evidencia

1.2. EXPERIMENTO 2: SELECCIÓN DE FACTORES CON ÁRBOLES DE DECISIÓN

En el presente experimento se analizan las variables que más influyen en la nota definitiva mediante arboles de decisión. El modelado se realiza con el algoritmo J48 y su validación cruzada con de grupos y balanceo de los datos para mejorar la calidad de aprendizaje.

Modelado: J48, con nodos binarios y un mínimo numero de instancias de objetos de 400.

Para aplicar este algoritmo se creó una variable categórica con “aprobado” para los estudiantes que su nota definitiva fue mayor o igual a 3 (≥ 3) y “reprobado” para los que obtuvieron una nota menor a 3 (< 3).

DETAILED ACCURACY BY CLASS

TP Rate	FP Rate	Precision	Recall	Class
1.000	0.036	0.991	1.000	aprobado
0.964	0.000	1.000	0.964	reprobado
0.993	0.028	0.993	0.993	Weighted Avg
F-Measure	MCC	ROC Area	PRC Area	Class
0.995	0.977	0.982	0.991	aprobado
0.982	0.977	0.982	0.972	reprobado
0.992	0.977	0.982	0.987	Weighted Avg
CONFUSIÓN MATRIX				
a	b	a = Aprobado		
106	0	b = Reprobado		
1	27			
Correctly Classified Instances 133			99.2537 %	
Incorrectly Classified Instances 1			0.7463 %	

Tabla 17: Evaluación de Resultados - Experimento 1

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	133	99.2537 %
Incorrectly Classified Instances	1	0.7463 %
Kappa statistic	0.9771	
Mean absolute error	0.0075	
Root mean squared error	0.0864	
Relative absolute error	2.2377 %	
Root relative squared error	21.2354 %	
Total Number of Instances	134	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.036	0.991	1.000	0.995	0.977	0.982	0.991	Aprobado
	0.964	0.000	1.000	0.964	0.982	0.977	0.982	0.972	Reprobado
Weighted Avg.	0.993	0.028	0.993	0.993	0.992	0.977	0.982	0.987	

=== Confusion Matrix ===

```

a  b  <-- classified as
106 0 | a = Aprobado
  1 27 | b = Reprobado

```

J48 unpruned tree

```

-----
Heteroevaluacion <= 2.7: Reprobado (26.0)
Heteroevaluacion > 2.7
| Personas_Cargo <= 1.0: Aprobado (91.0)
| Personas_Cargo > 1.0
| | egresado_colegio = PUBLICO: Aprobado (12.0)

```

```

| | egresado_colegio != PUBLICO
| | | Genero = MUJER: Aprobado (2.0)
| | | Genero != MUJER: Reprobado (3.0/1.0)
Number of Leaves : 5
Size of the tree : 9

```

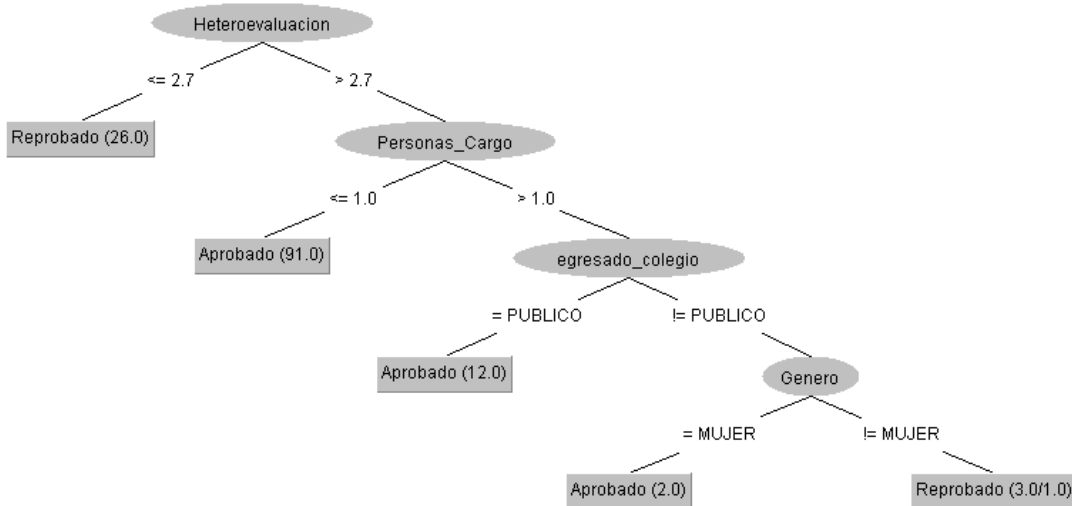


Ilustración 29: Árbol de decisión factores de influencia sobre la nota definitiva

1.2.1. Análisis de resultados

Como se presenta en la tabla 3, el 99.25% de los datos están clasificados de forma correcta. El área ROC para cada grupo es mayor a 0.7 lo que indica que tiene un alto grado de confiabilidad. Además, sobre la diagonal principal de la matriz de confusión se tiene la mayor parte de los datos, esto muestra que el resultado es cercano a la realidad.

De acuerdo con estos resultados, se puede indicar la nota definitiva es influenciada en gran medida por las variables presentadas en la siguiente tabla:

Variables
Heteroevaluación
Personas a cargo
Egresado colegio
Genero

1.3. EXPERIMENTO 3: Análisis de componentes principales PCA

El objetivo de este experimento es observar las variables que más influyen en los el rendimiento académico mediante PCA.

Modelado: Principal Component (PCA)

Ranked attributes:

0.9172	1	0.375Edad+0.347Estado_civil=CASADO/UNIONLIBRE-0.327Lugar_residencia=ARMENIA-0.299Estado_civil=SOLTERO+0.277Posicion_hermanos=PRIMERO...
0.8451	2	-0.371Tipo_vivienda=PROPIA-0.347Estado_civil=SOLTERO+0.341Tipo_vivienda=ARRENDADA-0.299Nivel_educativo_padre=UNIVERSIDAD+0.294Estado_civil=CASADO/UNIONLIBRE...
0.7831	3	-0.4Tipo_vivienda=ARRENDADA+0.366Tipo_vivienda=PROPIA-0.334Nivel_educativo_padre=SECUNDARIA+0.274Nivel_educativo_padre=PRIMARIA-0.24Estado_civil=SOLTERO...
0.7291	4	0.341Nivel_educativo_madre=PRIMARIA-0.316Lugar_residencia=ARMENIA+0.292Posicion_hermanos=SEGUNDO-0.28Nivel_educativo_padre=UNIVERSIDAD-0.267Posicion_hermanos=QUINTO...
0.6778	5	0.407Nivel_educativo_madre=SECUNDARIA-0.333Nivel_educativo_madre=UNIVERSIDAD-0.299Genero=HOMBRE+0.256Tipo_vivienda=FAMILIAR+0.249Nivel_educativo_padre=MAESTRIA...

0.9172	1	0.375Edad+0.347Estado_civil=CASADO/UNIONLIBRE-0.327Lugar_residencia=ARMENIA-0.299Estado_civil=SOLTERO+0.277Posicion_hermanos=PRIMERO...
0.8451	2	-0.371Tipo_vivienda=PROPIA-0.347Estado_civil=SOLTERO+0.341Tipo_vivienda=ARRENDADA-0.299Nivel_educativo_padre=UNIVERSIDAD+0.294Estado_civil=CASADO/UNIONLIBRE...
0.7831	3	-0.4Tipo_vivienda=ARRENDADA+0.366Tipo_vivienda=PROPIA-0.334Nivel_educativo_padre=SECUNDARIA+0.274Nivel_educativo_padre=PRIMARIA-0.24Estado_civil=SOLTERO...
0.7291	4	0.341Nivel_educativo_madre=PRIMARIA-0.316Lugar_residencia=ARMENIA+0.292Posicion_hermanos=SEGUNDO-0.28Nivel_educativo_padre=UNIVERSIDAD-0.267Posicion_hermanos=QUINTO...
0.6778	5	0.407Nivel_educativo_madre=SECUNDARIA-0.333Nivel_educativo_madre=UNIVERSIDAD-0.299Genero=HOMBRE+0.256Tipo_vivienda=FAMILIAR+0.249Nivel_educativo_padre=MAESTRIA...

Variables	
1	Edad
2	Estado_civil
3	Lugar_residencia
4	Posicion_hermanos
5	Trabaja

6	Nivel_edcativo_madre
7	Nivel_eduativo_padre
8	Heteroevaluación
9	egresado_colegio
10	Estrato

Tabla 18: Evaluación de resultados PCA

Seleccionando 10 variables que genero el método PCA en la tabla 10 con un porcentaje de 91.91%.

1.4. Experimento 4: SELECCIÓN DE FACTORES CON REGRSION LOGISTICA

El objetivo de este experimento es encontrar las variables que más influyen en el rendimiento académico mediante regresión logística.

Se realizó una categorización por medio de rangos mediante Dcretize en WEKA de la variable para el modelado.

Modelado: Simple logistic, con un balanceo de datos y una validación cruzada de 10 grupos.

Evaluación de resultados:

SimpleLogistic:

Class '(-inf-2.96]':

$-4.35 + [\text{Personas_Cargo}=(1.8-3.6)] * 5.77 + [\text{Heteroevaluacion}=(\text{-inf}-2.76)] * 10.75$

Class '(2.96-3.72]':

$1.08 + [\text{egresado_colegio}=\text{PRIVADO}] * -0.81 +$
 $[\text{Lugar_residencia}=\text{ARMENIA}] * -0.8 +$
 $[\text{Nivel_eduativo_padre}=\text{SECUNDARIA}] * -0.69 +$
 $[\text{Posicion_hermanos}=\text{CUARTO}] * 0.8 +$
 $[\text{MejoraEvidencia}=(3-4)] * 1.24 +$
 $[\text{Heteroevaluacion}=(2.76-3.32)] * 5.64 +$
 $[\text{Heteroevaluacion}=(3.88-4.44)] * -2.44$

Class '(3.72-4.48]':

$-0.68 + [\text{Edad}=(19.8-23.6)] * -0.97 +$
 $[\text{Estrato}=(4\text{-inf})] * 1.43 +$
 $[\text{Nivel_edcativo_madre}=\text{PRIMARIA}] * -1.34 +$
 $[\text{Posicion_hermanos}=\text{QUINTO}] * 3.35 +$
 $[\text{Posicion_hermanos}=\text{TERCERO}] * -1.03 +$

[MejoraEvidencia='(4-inf)'] * 0.85 +
 [Autoevaluacion='(3.8-4.4)'] * 1.27 +
 [Autoevaluacion='(4.4-inf)'] * -0.68 +
 [Heteroevaluacion='(3.88-4.44)'] * 2.71

Class '(4.48-5.24)' :

-1.77 + [Nivel_educativo_padre=UNIVERSIDAD] * 2.31 +
 [MejoraEvidencia='(2-3)'] * 1.36 +
 [Autoevaluacion='(3.8-4.4)'] * -1.02 +
 [Autoevaluacion='(4.4-inf)'] * 1.33 +
 [Heteroevaluacion='(3.32-3.88)'] * -3.73 +
 [Heteroevaluacion='(4.44-inf)'] * 5.83

Class '(5.24-inf)' :

-9.32 + [Estrato='(3-4)'] * 3.49 +
 [egresado_colegio=PRIVADO] * 0.92 +
 [Posicion_hermanos=CUARTO] * 1.06 +
 [MejoraEvidencia='(1-2)'] * 3.89 +
 [Autoevaluacion='(-inf-2.6)'] * 3.16

Correctly Classified Instances	97	72.3881 %
Incorrectly Classified Instances	37	27.6119 %
Kappa statistic	0.6477	
Mean absolute error	0.1385	
Root mean squared error	0.2786	
Relative absolute error	44.0646 %	
Root relative squared error	70.2632 %	
Total Number of Instances	134	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.833	0.073	0.714	0.833	0.769	0.717	0.924	0.841	'(-inf-2.76)'
	0.756	0.118	0.738	0.756	0.747	0.634	0.904	0.785	'(2.76-3.32)'
	0.462	0.102	0.522	0.462	0.490	0.377	0.809	0.472	'(3.32-3.88)'
	0.739	0.054	0.739	0.739	0.739	0.685	0.969	0.885	'(3.88-4.44)'
	0.850	0.009	0.944	0.850	0.895	0.879	0.997	0.986	'(4.44-inf)'
Weighted Avg.	0.724	0.080	0.723	0.724	0.722	0.644	0.914	0.781	

=== Confusion Matrix ===

```

a b c d e <-- classified as
20 2 1 1 0 | a = '(-inf-2.76)'
```

```

6 31 4 0 0 | b = '(2.76-3.32)'
```

```

2 9 12 2 1 | c = '(3.32-3.88)'
```

```

0 0 6 17 0 | d = '(3.88-4.44)'
```

```

0 0 0 3 17 | e = '(4.44-inf)'
```

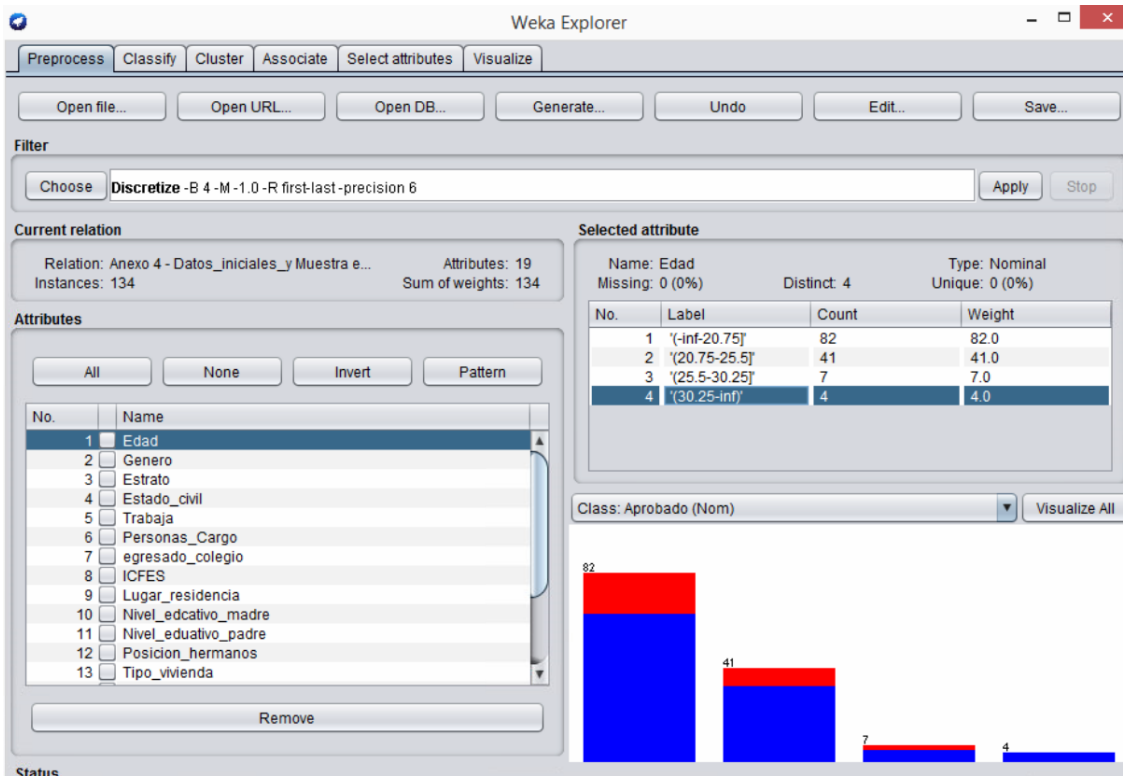
DETAILED ACCURACY BY CLASS				
TP Rate	FP Rate	Precision	Recall	Class
0.833	0.73	0.714	0.833	(-inf : 2.76]
0.756	0.118	0.738	0.756	(2.76 : 3.32]
0.46	0.102	0.522	0.462	(3.32 : 3.88]
0.73	0.054	0.739	0.739	(3.38 : 4.44]
0.85	0.009	0.944	0.850	(4.44 : 5]
0.72	0.080	0.723	0.724	Weighted Avg
F-Measure	MCC	ROC Area	PRC Area	Class
0.769	0.717	0.924	0.841	(-inf : 2.76]
0.747	0.634	0.904	0.785	(2.76 : 3.32]
0.490	0.377	0.809	0.472	(3.32 : 3.88]
0.739	0.685	0.969	0.885	(3.38 : 4.44]
0.895	0.879	0.997	0.986	(4.44 : 5]
0.722	0.644	0.914	0.781	Weighted Avg

Variables relevantes:

- Personas a cargo
- Heteroevaluación
- Posición hermanos
- Mejora evidencia
- Autoevaluación
- Estrato
- Nivel educativo padre
- Nivel educativo madre
- Lugar de residencia
- Edad
- Egresado colegio

1.5. EXPERIMENTO 5: REGLAS DE ASOCIACIÓN A PRIORI

1.5.1. **Método Discretize:** Para las reglas de asociación se discretizan las variables para poder aplicar el algoritmo A priori y se categoriza la clase mediante la herramienta Weka, en aprobado y reprobado mediante el algoritmo “discretize”.



El objetivo de este experimento es seleccionar las variables que más influyen en la nota definitiva mediante la visualización de las reglas.

Método: A priori, 10 reglas con un factor de confiabilidad de 0.5

Apriori

=====

Minimum support: 0.5 (60 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 14

Size of set of large itemsets L(3): 5

Best rules found:

1. Trabaja=NO Personas_Cargo='(-inf-0.9]' 60 ==> Estado_civil=SOLTERO 60 <conf:(1)> lift:(1.11) lev:(0.04) [5] conv:(5.82)
2. Personas_Cargo='(-inf-0.9]' 92 ==> Estado_civil=SOLTERO 91 <conf:(0.99)> lift:(1.1) lev:(0.06) [7] conv:(4.46)
3. Personas_Cargo='(-inf-0.9]' Lugar_residencia=ARMENIA 78 ==> Estado_civil=SOLTERO 77 <conf:(0.99)> lift:(1.09) lev:(0.05) [6] conv:(3.78)
4. Personas_Cargo='(-inf-0.9]' egresado_colegio=PUBLICO 63 ==> Estado_civil=SOLTERO 62 <conf:(0.98)> lift:(1.09) lev:(0.04) [5] conv:(3.06)
5. Trabaja=NO Lugar_residencia=ARMENIA 74 ==> Estado_civil=SOLTERO 71 <conf:(0.96)> lift:(1.06) lev:(0.03) [4] conv:(1.79)
6. Trabaja=NO 83 ==> Estado_civil=SOLTERO 79 <conf:(0.95)> lift:(1.05) lev:(0.03) [4] conv:(1.61)
7. Tipo_vivienda=PROPIA 68 ==> Estado_civil=SOLTERO 64 <conf:(0.94)> lift:(1.04) lev:(0.02) [2] conv:(1.32)
8. Genero=MUJER 65 ==> Estado_civil=SOLTERO 60 <conf:(0.92)> lift:(1.02) lev:(0.01) [1] conv:(1.05)
9. Lugar_residencia=ARMENIA 115 ==> Estado_civil=SOLTERO 105 <conf:(0.91)> lift:(1.01) lev:(0.01) [1] conv:(1.01)
10. Nivel_edcativo_madre=SECUNDARIA 67 ==> Estado_civil=SOLTERO 61 <conf:(0.91)> lift:(1.01) lev:(0) [0] conv:(0.93)

El objetivo de este experimento es observar las variables que más influyen en los el rendimiento académico mediante el algoritmo A priori.

Variables	
1	Trabaja
2	Personas a cargo
3	Estado civil
4	Lugar de residencia
5	Nivel educativo madre
6	Tipo vivienda
7	Egresado colegio

2. Resultado de las técnicas

Después de aplicar los diferentes algoritmos de minería de datos, se realizó una evaluación si con los resultados anteriores se es posible llegar a una conclusión de las variables pueden influir en el rendimiento académico.

En la tabla 4, se presentan todos los resultados que se generaron de cada uno de los algoritmos utilizados en la minería de datos, esto con la finalidad de encontrar patrones que permitan concluir que factores tienen más influencia en el rendimiento académico de cursos bajo la metodología de proyectos formativos.

VARIABLE	EXPERIMENTOS DE MINERÍA DE DATOS					
	CfsSubset Eval	Arboles de decisión	PCA	Regresión logística	Reglas de asociación	Frecuencia
Definitiva (clase)						
Heteroevaluación	x	x	x	x		4
Genero		x				1
coevaluación	x					1
Posicion_hermanos	x		x	x		3
Edad			x	x		2
Autoevaluacion	x			x		2
Tipo_vivienda	x				x	2
MejoraEvidencia	x			x		2
Lugar_residencia			x	x	x	3
Nivel_educativo_padre			x	X		2
Nivel_educativo_madre			x	X	x	3
Estado_civil			x		x	2
Estrato			x	x		2
Egresado_colegio		x	x	x	x	4
Personas_Cargo		x		x	x	3
ICFES						0
Trabaja	x		x		x	3
Sexo	x					1

Tabla 19: Resultado Final.

A continuación se presentan en orden mayor frecuencia los 7 factores más importantes según los experimentos realizados, los que más influyen en el rendimiento académico:

- **Factor 1: Heteroevaluación**
- **Factor 2: Egresado Colegio**
- **Factor 3 Posición hermano**
- **Factor 4: lugar de residencia**
- **Factor 5: nivel educativo madre**
- **Factor 6: trabaja**
- **Factor 7: personas a cargo**

Anexo 5: Análisis estadístico

Pruebas normalidad y homocedasticidad

1. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Edad*

- **Análisis de ANOVA**

Tabla ANOVA para Definitiva por Edad

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Entre grupos	14.2962	17	0.84095	1.44	0.1314
Intra grupos	67.8394	116	0.584822		
Total (Corr.)	82.1355	133			

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 1.43796, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Edad y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para RESIDUOS:

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Estadístico W de Shapiro-Wilk	0.964361	0.0183817

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que RESIDUOS proviene de una distribución normal con 95% de confianza.

Para la variable edad no se puede utilizar ANOVA porque los residuos no tienen distribución normal.

Por tal razón se utiliza el método de Mann-Whitney para esta variable.

2. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Genero.*

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>de Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Entre grupos	3.21024	1	3.21024	5.37	0.0220
Intra grupos	78.9253	132	0.597919		
Total (Corr.)	82.1355	133			

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 5.36902, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de

la prueba-F es menor que 0.05, existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Genero y otro, con un nivel del 95.0% de confianza. Para determinar cuáles medias son significativamente diferentes de otras, seleccione Pruebas de Múltiples Rangos, de la lista de Opciones Tabulares.

Pruebas de Normalidad para RESIDUOS:

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Estadístico W de Shapiro-Wilk	0.955309	0.00136393

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

Para la variable género no se puede utilizar ANOVA porque los residuos no tienen distribución normal.

3. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de estado civil*

Tabla ANOVA para Definitiva por Estado_civil

<i>Fuente</i>	<i>Suma de Cuadrados</i>	<i>Gl</i>	<i>Cuadrado Medio</i>	<i>Razón-F</i>	<i>Valor-P</i>
Entre grupos	2.89495	3	0.964984	1.58	0.1966
Intra grupos	79.2406	130	0.609543		
Total (Corr.)	82.1355	133			

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 1.58313, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Estado_civil y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

<i>Prueba</i>	<i>Estadístico</i>	<i>Valor-P</i>
Estadístico W de Shapiro-Wilk	0.942354	0.0000177618

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

Para la variable Estado civil no se puede utilizar ANOVA porque los residuos no tienen distribución normal.

4. Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de trabaja

Tabla ANOVA para Definitiva por Trabaja

Fuente	Suma de Cuadrados	de Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	0.00572792	1	0.00572792	0.01	0.9237
Intra grupos	82.1298	132	0.622195		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.00920598, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Trabaja y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.942354	0.0000177618

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si B.RESIDUOS puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

Para la variable trabaja no se puede utilizar ANOVA porque los residuos no tienen distribución normal.

5. Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de repitente

Tabla ANOVA para Definitiva por Trabaja

Fuente	Suma de Cuadrados	de Gl	Cuadrado Medio	Razón-F	Valor-P
--------	-------------------	-------	----------------	---------	---------

Entre grupos	0.00572792	1	0.00572792	0.01	0.9237
Intra grupos	82.1298	132	0.622195		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.00920598, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Trabaja y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.945181	0.0000481725

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

6. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de egresado colegio*

Tabla ANOVA para Definitiva por egresado_colegio

Fuente	Suma de Cuadrados	de Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	0.0268009	1	0.0268009	0.04	0.8359
Intra grupos	82.1087	132	0.622036		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.0430857, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de egresado_colegio y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.946153	0.0000674665

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

7. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de lugar de residencia*

Tabla ANOVA para Definitiva por Lugar_residencia

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	2.47588	5	0.495176	0.80	0.5547
Intra grupos	79.6596	128	0.622341		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.795667, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Lugar_residencia y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.948089	0.000130787

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

8. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de nivel educativo madre*

Tabla ANOVA para Definitiva por Nivel_edcativo_madre

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	0.144561	4	0.0361402	0.06	0.9939
Intra grupos	81.991	129	0.635589		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.0568609, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Nivel_edcativo_madre y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.945851	0.0000607961

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza

9. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de nivel educativo del padre*

. Tabla ANOVA para Definitiva por Nivel_eduativo_padre

Fuente	Suma de Cuadrados	de Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	2.33969	4	0.584923	0.95	0.4399
Intra grupos	79.7958	129	0.618572		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.945602, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Nivel_eduativo_padre y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.946608	0.0000789377

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si B.RESIDUOS puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

10. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de posición entre hermanos*

. Tabla ANOVA para Definitiva por Posicion_hermanos

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	2.91747	5	0.583495	0.94	0.4558
Intra grupos	79.218	128	0.618891		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.942807, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Posicion_hermanos y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.945499	0.0000537941

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si B.RESIDUOS puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

11. *Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de tipo de vivienda*

. Tabla ANOVA para Definitiva por Tipo_vivienda

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	0.0931051	2	0.0465526	0.07	0.9284

Intra grupos	82.0424	131	0.626278		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.0743321, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Tipo_vivienda y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.94943	0.000205281

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si B.RESIDUOS puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

12. Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Contribucion Metodologia PF

Tabla ANOVA para Definitiva por Contribucion Metodologia PF

Fuente	Suma de Cuadrados	Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	0.569253	3	0.189751	0.30	0.8236
Intra grupos	81.5663	130	0.627433		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.302424, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Contribucion Metodologia PF y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.94943	0.000205281

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

13. Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Complejidad del PF

Tabla ANOVA para Definitiva por Complejidad del PF

Fuente	Suma de Cuadrados	de Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	4.84487	3	1.61496	2.72	0.0474
Intra grupos	77.2906	130	0.594543		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 2.7163, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la prueba-F es menor que 0.05, existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Complejidad del PF y otro, con un nivel del 95.0% de confianza. Para determinar cuáles medias son significativamente diferentes de otras, seleccione Pruebas de Múltiples Rangos, de la lista de Opciones Tabulares.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.957819	0.00292551

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si B.RESIDUOS puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

14. Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Participacion en el PF

. Tabla ANOVA para Definitiva por Participacion en el PF

Fuente	Suma de Cuadrados	de Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	5.1075	4	1.27688	2.14	0.0797
Intra grupos	77.028	129	0.597116		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 2.1384, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Participacion en el PF y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.959051	0.00420764

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si B.RESIDUOS puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

15. Hipótesis: Existe diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Realiza las actividades del PF

. Tabla ANOVA para Definitiva por Realiza las actividades del PF

Fuente	Suma de Cuadrados	de Gl	Cuadrado Medio	Razón-F	Valor-P
Entre grupos	1.86898	4	0.467245	0.75	0.5591
Intra grupos	80.2665	129	0.622221		
Total (Corr.)	82.1355	133			

El StatAdvisor

La tabla ANOVA descompone la varianza de Definitiva en dos componentes: un componente entre-grupos y un componente dentro-de-grupos. La razón-F, que en este caso es igual a 0.750931, es el cociente entre el estimado entre-grupos y el estimado dentro-de-grupos. Puesto que el valor-P de la razón-F es mayor o igual que 0.05, no existe una diferencia estadísticamente significativa entre la media de Definitiva entre un nivel de Realiza las actividades del PF y otro, con un nivel del 95.0% de confianza.

Pruebas de Normalidad para B.RESIDUOS

Prueba	Estadístico	Valor-P
Estadístico W de Shapiro-Wilk	0.952961	0.000651053

El StatAdvisor

Esta ventana muestra los resultados de diversas pruebas realizadas para determinar si B.RESIDUOS puede modelarse adecuadamente con una distribución normal. La prueba de Shapiro-Wilk está basada en la comparación de los cuartiles de la distribución normal ajustada a los datos.

Debido a que el valor-P más pequeño de las pruebas realizadas es menor a 0.05, se puede rechazar la idea de que B.RESIDUOS proviene de una distribución normal con 95% de confianza.

Análisis factorial

En el trabajo de Cardona (2017) la fiabilidad del instrumento se verificó con base en las apreciaciones de 31 profesores universitarios. Cada uno de los profesores valoró la pertinencia de los ítems del instrumento, en una escala de 1 a 5. Para verificar la consistencia interna de los ítems del instrumento, se realizó un análisis de fiabilidad basado en el coeficiente alfa de Cronbach. Los ítems del instrumento tuvieron discriminación positiva para el coeficiente alfa, y por lo tanto, un alto grado de consistencia interna. El coeficiente resultante fue 0.887, por lo que la fiabilidad puede considerarse aceptable, teniendo como referente que una encuesta de este tipo puede estar desde un 0.70 (Morales, Urosa, and Blanco, 2003), citado en (Olmos 2008). El alfa de Cronbach en caso de eliminar algún elemento sigue teniendo discriminación positiva. Así mismo, la varianza de la escala es homogénea en caso de eliminarse algún elemento (Cardona 2017).

Se realizó la técnica para la validez mediante análisis factorial, el cual se utiliza para identificar propiedades homogéneas de las variables independientes que se están analizando. El objetivo fue establecer si los ítems del instrumento se agrupan en características comunes a ellos. Para el análisis factorial se verificaron los supuestos estadísticos y se analizaron los factores extraídos producto de la rotación de componentes. Los supuestos estadísticos para el análisis factorial fueron:

- El coeficiente de correlaciones entre las variables independientes son altas.
- La prueba de esfericidad Bartlett mostró que el nivel de significación de p-valor = 0.000, permite afirmar que se encontraron relaciones estadísticamente significativas entre ítems del instrumento.
- El valor KMO de 0.702, permite afirmar que es posible realizar el análisis factorial.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.702
Bartlett's Test of Sphericity	Approx. Chi-Square	3065.132
	df	1128
	Sig.	.000

- La diagonal principal de la matriz de correlaciones, tiene coeficientes cercanos a 1.0, a excepción de las variables (p14=0.351 y p26=0.385) y los coeficientes fuera de esa diagonal son bajos. Los resultados obtenidos permiten afirmar que los supuestos estadísticos se cumplen para la extracción de los factores.

Var	Componente				
	1	2	3	4	5
p16: La coevaluación de evidencias contribuye al rendimiento académico	.839				
p13: Utilizo la autoevaluación para la mejora continua de mis evidencias	.823				
p14: Realizo coevaluación y autoevaluación de las evidencias, con base en los criterios de evaluación definidos en la rúbrica.	.762				
p15: La autoevaluación de evidencias contribuye al rendimiento académico.	.621				.381
P1: Las competencias a desarrollar en el proyecto formativo son claras a partir de las evidencias definidas.		.832			
p3: El proyecto formativo se enfoca en la solución de un problema del contexto y su nivel de reto está acorde con la formación de los estudiantes		.798			
P2. La metodología de proyectos formativos contribuye al desarrollo de las competencias del curso		.765			
p10: Los criterios de las rúbricas facilitan la evaluación de las evidencias del proyecto.		.687			
p31: . Me siento a gusto con la metodología de proyectos formativos.	.358	.418			
p25 El diseño gráfico del curso en Moodle permite una navegación adecuada.			.830		
p28: La interfaz del curso facilita la comprensión de los elementos del proyecto formativo.	.326		.814		
p27: Me siento a gusto con el diseño del curso en la plataforma Moodle.			.784		
p26: . El diseño gráfico del curso en Moodle es estético			.706		
p29: . El diseño de los recursos y las actividades es estéticamente agradable.			.686		
p23: Las actividades de aprendizaje ofrecidas en Moodle están acorde con mis necesidades de aprendizaje.				.858	
p22: Los recursos de aprendizaje ofrecidos en Moodle están acorde con mis necesidades de aprendizaje				.835	
p21: Los recursos de aprendizaje ofrecidos en Moodle son coherentes con las actividades de aprendizaje del proyecto formativo				.813	
p24: Las actividades de aprendizaje ofrecidas en Moodle son coherentes mi nivel de competencia				.647	.325

p20: Los recursos de aprendizaje ofrecidos en Moodle son útiles para la solución del proyecto formativo.			.412	.500	
p32: La metodología de evaluación incentiva a mejorar mi rendimiento académico		.355			.770
p34: La metodología de proyectos aumenta mi motivación para trabajar de forma colaborativa.					.695
p35: La autoevaluación me incentiva a mejorar mi rendimiento académico.					.658
p30: La metodología de proyectos formativos incentiva a mejorar el rendimiento académico.	.363				.472

Tabla 20: Matriz de componentes rotados

El método de extracción de los componentes se realizó mediante el análisis de componentes principales. El método de extracción generó 9 factores. Se utilizaron los primeros cinco factores que explican el 70.2% de la variabilidad total. Los factores que se agruparon fueron: metodología de proyectos formativos, proceso de evaluación, diseño del curso en Moodle, adaptación de recursos y actividades de aprendizaje y proceso de aprendizaje, el trabajo de Cardona (2017) permite observar que las variables agrupadas fueron coherente con su investigación.

- El factor 1 está compuesto por variables asociadas al proceso de evaluación (p16, p13, p14 y p15).
- El factor 2 asoció los ítems (p1, p3, p2, p10, p31), estos están relacionados con la metodología de proyectos formativos.
- El factor 3 compuesto por los ítems (p25, p28, p27, p26, p29), los cuales están relacionados con el diseño de la plataforma Moodle.
- El factor 4 compuesto por los ítems (p23, p22, p21, p24, p20), los cuales están relacionados con la utilidad de los recursos y las actividades de aprendizaje de Moodle.
- El factor 5 compuesto por los ítems (p32, p34, p35, p30), los cuales se asocian con la opinión del estudiante sobre el proceso de aprendizaje y el rendimiento académico.