



**UNIVERSIDAD TÉCNICA DE COTOPAXI**  
**FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS**  
**CARRERA DE INGENIERÍA EN INFORMÁTICA Y SISTEMAS**  
**COMPUTACIONALES**

PROYECTO DE INVESTIGACIÓN

**“ANÁLISIS DEL NIVEL DE RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE  
LOS PRIMEROS CICLOS UTILIZANDO ARBOLES DE DECISIÓN COMO  
TÉCNICA DE INTELIGENCIA ARTIFICIAL, PARA DETERMINAR LAS  
POSIBLES CAUSAS DEL BAJO RENDIMIENTO EN LA FACULTAD DE  
CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA  
DE COTOPAXI”**

Proyecto de titulación presentado previo a la obtención del título de Ingeniera en  
Informática y Sistemas Computacionales

**Autora:**

Tomalo Morales Shirley Vanessa

**Tutor:**

MSc. Ing. Cadena Moreano José Augusto

**Latacunga – Ecuador**

Julio 2019



Universidad  
Técnica de  
Cotopaxi



Ingeniería  
Informática Y Sistemas  
Computacionales

### DECLARACIÓN DE AUTORÍA

Yo **TOMALO MORALES SHIRLEY VANESSA** declaro ser autor del presente proyecto de investigación: “ANÁLISIS DEL NIVEL DE RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE LOS PRIMEROS CICLOS UTILIZANDO ARBOLES DE DECISIÓN COMO TÉCNICA DE INTELIGENCIA ARTIFICIAL, PARA DETERMINAR LAS POSIBLES CAUSAS DEL BAJO RENDIMIENTO EN LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI”, siendo el Ing. MSc. Cadena Moreano José Augusto tutor de presente trabajo; y eximo expresamente a la Universidad Técnica de Cotopaxi y a sus representantes legales de posibles reclamos o acciones legales.

Además, certifico que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo de investigativo, son de mi exclusiva responsabilidad.

**TOMALO MORALES SHIRLEY VANESSA**

**C.I: 210080983-5**



## AVAL DEL TUTOR DE PROYECTO DE TITULACIÓN

En calidad de Tutor del Trabajo de Investigación sobre el título: **“Análisis del nivel de rendimiento académico de estudiantes de los primeros ciclos utilizando arboles de decisión como técnica de inteligencia artificial, para determinar las posibles causas del bajo rendimiento en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi”**, de **TOMALO MORALES SHIRLEY VANESSA**, de la Carrera de Ingeniería en Informática y Sistemas Computacionales, considero que dicho Informe Investigativo cumple con los requerimientos metodológicos y aportes científico-técnicos suficientes para ser sometidos a la evaluación del Tribunal de Validación de Proyecto que el Consejo Directivo de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi designe, para su correspondiente estudio y calificación.

Latacunga, julio de 2019

**Ing. MSc. Cadena Moreano José Augusto**

**C.I.: 0501552798**



## APROBACIÓN DEL TRIBUNAL DE TITULACIÓN

En calidad de Tribunal de Lectores, aprueban el presente Informe de Investigación de acuerdo a las disposiciones reglamentarias emitidas por la Universidad Técnica de Cotopaxi, y por la FACULTAD de Ciencias de la Ingeniería y Aplicadas; por cuanto, el postulante: **Shirley Vanessa Tomalo Morales**, con el título de Proyecto de titulación: **“Análisis del nivel de rendimiento académico de estudiantes de los primeros ciclos utilizando arboles de decisión como técnica de inteligencia artificial, para determinar las posibles causas del bajo rendimiento en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi”** han considerado las recomendaciones emitidas oportunamente y reúne los méritos suficientes para ser sometido al acto de Sustentación de Proyecto.

Por lo antes expuesto, se autoriza realizar los empastados correspondientes, según la normativa institucional.

Latacunga, 22 de julio del 2019.

Para constancia firman:

**Lector 1**

Ing. Rodríguez Gustavo  
CC: 1757001357

**Lector 2**

Ing. Edwin Quinaloa Arequipa  
CC: 0502563372

**Lector 3**

Ing. Cantuña Flores Karla  
CC: 0502305113

## **AGRADECIMIENTO**

Agradezco a Dios por brindarme la vida y las capacidades necesarias para salir adelante en una etapa importante en mi crecimiento como persona.

A la Universidad por darme la oportunidad de adquirir nuevos conocimientos y a los docentes por guiarme en mi formación académica, por haberme enseñado sus conocimientos y experiencias en la trayectoria de mi carrera.

Al Ing. José Cadena, por aceptar dirigir mi proyecto y enseñarme sus conocimientos.

Shirley Tomalo

## **DEDICATORIA**

El presente proyecto está dedicado a mis Padres **María Morales** y **Daniel Tomalo** quienes siempre han estado apoyándome en las buenas y en las malas, inculcándome valores de tolerancia y respeto hacia los demás, todo mi amor hacia ellos.

A mi hermano mayor **Willian Tomalo** por ser como mi segundo padre y apoyar mi carrera desde siempre. Danny, Vinicio y Tatiana, muchas gracias por el apoyo incondicional.

A mis hijas **Melanie**, **Doménica** y **Sahily Soto** quienes con su amor, alegría y cariño me impulsaron a seguir adelante gracias mis princesas las amo con todo mi corazón. A mi Abuelita en el cielo eres mi gran ángel de la guarda.

A un gran amigo que al final de mi trayectoria apareció para alentarme en los momentos críticos lejos de mi familia gracias **Luis Banda**.

Shirley Tomalo

## ÍNDICE DE CONTENIDOS

DECLARACIÓN DE AUTORÍA .....	ii
AVAL DEL TUTOR DE PROYECTO DE INVESTIGACIÓN .....	iii
APROBACIÓN DEL TRIBUNAL DE TITULACIÓN.....	iv
AGRADECIMIENTO.....	v
DEDICATORIA.....	vi
ÍNDICE DE CONTENIDOS.....	vii
RESUMEN .....	xii
ABSTRACT .....	xiii
AVAL TRADUCCIÓN.....	xiv
1. INFORMACIÓN GENERAL: .....	1
2. RESUMEN DEL PROYECTO .....	2
3. JUSTIFICACIÓN DEL PROYECTO.....	3
4. BENEFICIARIOS DEL PROYECTO .....	4
5. EL PROBLEMA DE INVESTIGACIÓN .....	4
6. OBJETIVOS.....	5
6.1. Objetivo general: .....	5
6.2. Objetivos específicos:.....	5
7. ACTIVIDADES Y SISTEMAS DE TAREAS EN RELACIÓN A LOS OBJETIVOS .....	6
8. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA.....	8
8.1. Antecedentes.....	8
8.2. Inteligencia Artificial.....	8
8.3. Minería de datos .....	9
8.3.1. Proceso de la minería de datos .....	10
8.3.2. Aplicaciones de la minería de datos.....	12
8.3.3. Retos y tendencias de la minería de datos.....	13
8.3.4. Técnicas de la minería de datos .....	14
8.3.5. Clasificación de las técnicas de minería de datos .....	15
8.3.5.1. Métodos descriptivos.....	15
8.3.5.2. Métodos predictivos .....	15
8.3.6. Algoritmos de clasificación .....	16
8.4. Árbol de decisión.....	17

8.4.1. Elementos de un árbol de decisión.....	18
8.4.2. Clasificación de árboles de decisión .....	18
8.4.3. Características de los árboles de decisión .....	20
8.5. Software Rapidminer Studio .....	20
8.6. Otros software de minería de datos .....	21
8.7. Rendimiento académico .....	22
8.8. Bajo rendimiento académico .....	23
9. PREGUNTAS CIENTÍFICAS O HIPÓTESIS .....	23
9.1. HIPÓTESIS .....	23
10. METODOLOGÍA.....	23
10.1. Tipos de investigación .....	24
10.2. Métodos de investigación.....	24
10.3. Acopio y procesamiento de los datos.....	25
10.4. Diseño experimental .....	25
10.5. Población y muestra .....	25
10.6. Instrumentos para recolectar datos.....	25
10.7. Procesamiento de datos.....	26
11. ANÁLISIS DE RESULTADOS.....	26
11.1. Análisis de la encuesta aplicada.....	26
11.2. Estadística en cuanto al promedio de los estudiantes.....	35
11.3. Caso de estudio del nivel de rendimiento académico .....	36
11.4. Factores del rendimiento académico.....	36
11.5. Software Rapidminer para minería de datos .....	38
11.5.1. Gestionar los datos en Rapidminer.....	40
11.5.2. Analizando los datos recolectados en Rapidminer .....	41
11.5.3. Entrada de datos y muestra de resultados.....	42
11.5.3.1. Resultados obtenidos en arboles de decisión de las variables de identificación .....	44
11.5.3.2. Resultados obtenidos en arboles de decisión variables académicas vs promedio de estudiantes.. .....	47
11.5.3.3. Resultados obtenidos en arboles de decisión variables actitudinales vs promedio de estudiantes.. .....	50
11.6. Análisis de los árboles de decisión .....	52



11.6.1. Análisis de resultados de las variables o atributos de identificación con el promedio del estudiante.....	53
11.6.2. Análisis de resultados de las variables o atributos académicos con el promedio del estudiante.....	53
11.6.3. Análisis de resultados de las variables o atributos actitudinales con el promedio del estudiante.....	54
11.7. Matriz de confusión .....	55
12. IMPACTOS.....	58
12.1. Impacto técnico.....	58
12.2. Impacto social.....	58
12.3. Impacto económico.....	58
13. PRESUPUESTO PARA LA ELABORACIÓN DEL PROYECTO.....	59
14. CONCLUSIONES Y RECOMENDACIONES.....	60
14.1. Conclusiones.....	60
14.2. Recomendaciones.....	61
15. BIBLIOGRAFÍA.....	61
16. ANEXOS.....	64

## ÍNDICE DE TABLAS

Tabla 1: Sistema de tareas en relación a los objetivos planteados.....	6
Tabla 2: Valoración índice Kappa.....	41
Tabla 3: Variables o atributos de identificación del estudiante.....	43
Tabla 4: Matriz de confusión accuracy.....	56
Tabla 5: Matriz de confusión kappa.....	56
Tabla 6: Gastos Directos.....	59
Tabla 7: Gatos Indirectos.....	59
Tabla 8: Presupuesto Total del Proyecto.....	60

## ÍNDICE DE GRÁFICOS

Gráfico 1: Minería de Datos Asociados a distintos campos.....	10
Gráfico 2: Proceso KDD .....	12
Gráfico 3: Estructura de un arbol de decisión .....	20
Gráfico 4: Estadística gráfica – pregunta 1 .....	26
Gráfico 5: Estadística gráfica – pregunta 2 .....	27
Gráfico 6: Estadística gráfica – pregunta 3 .....	27
Gráfico 7: Estadística gráfica – pregunta 4 .....	28
Gráfico 8: Estadística gráfica – pregunta 5 .....	29
Gráfico 9: Estadística gráfica – pregunta 6 .....	29
Gráfico 10: Estadística gráfica – pregunta 7 .....	30
Gráfico 11: Estadística gráfica – pregunta 8 .....	30
Gráfico 12: Estadística gráfica – pregunta 9 .....	31
Gráfico 13: Estadística gráfica – pregunta 10 .....	32
Gráfico 14: Estadística gráfica – pregunta 11 .....	32
Gráfico 15: Estadística gráfica – pregunta 12 .....	33
Gráfico 16: Estadística gráfica – pregunta 13 .....	33
Gráfico 17: Estadística gráfica – pregunta 14 .....	34
Gráfico 18: Estadística gráfica – pregunta 15 .....	34
Gráfico 19: Estadística gráfica – pregunta 16 .....	35
Gráfico 20: Estadística gráfica – Promedio.....	35
Gráfico 21: Rapidminer Studio.....	38
Gráfico 22: Rapidminer Weka extensión .....	39
Gráfico 23: Rapidminer Interfaz principal .....	39
Gráfico 24: Rapidminer: Área de trabajo.....	40
Gráfico 25: Operador read excel .....	42
Gráfico 26: Operadores (read excel, set role y decisión tree) .....	44
Gráfico 27: Árbol gráfico N° 1 variables o atributos de identificación.....	46
Gráfico 28: Árbol gráfico N° 2 variables o atributos académicos.....	48
Gráfico 29: Árbol gráfico N° 3 variables o atributos académicos.....	49
Gráfico 30: Árbol gráfico N° 4 variables o atributos actitudinales .....	51

Gráfico 31: Árbol gráfico N° 5 variables actitudinales vs promedio.....	52
Gráfico 32: Matriz de confusión.....	55

# UNIVERSIDAD TÉCNICA DE COTOPAXI

## INGENIERIA EN INFORMÁTICA Y SISTEMAS COMPUTACIONALES

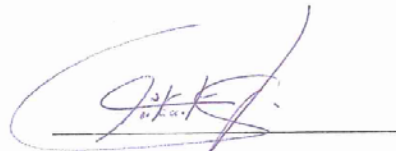
**TÍTULO:** “Análisis del nivel de rendimiento académico de estudiantes de los primeros ciclos utilizando arboles de decisión como técnica de inteligencia artificial, para determinar las posibles causas del bajo rendimiento en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi”.

**Autor:** Tomalo Morales Shirley Vanessa

### RESUMEN

El presente proyecto de investigación trata de determinar las posibles causas del bajo rendimiento desempeño académico, a través de la aplicación de minería de datos, utilizando árboles de decisión como técnica de inteligencia artificial, con el objetivo de predecir su rendimiento académico mediante factores que afectan al estudiante en su desenvolvimiento. Los resultados obtenidos con la aplicación de árboles de decisión realizada a 236 estudiantes mediante una encuesta se pudo verificar los factores que influyeron en el rendimiento del estudiante de acuerdo a su promedio, las variables académicas obtuvieron mayor ganancia de información ya que estas tienen mayor exactitud para predicción, con un 87.33% de los estudiantes se encuentran muy bien en su nivel de desempeño académico. Para el análisis de la información obtenida sobre los factores actitudinales, académicos y de identificación de los estudiantes se aplicó las herramientas decisión tree, set role de Rapidminer que es un software de minería de datos, la utilización de las variables en conjunto correspondientes a un solo aspecto no tiene mayor exactitud que cuando se realizó la combinación de variables de diferentes aspectos.

**Palabras claves:** Arboles de decisión, Inteligencia Artificial, Minería de Datos, set role y Rapidminer.



MSc. Ing. José Cadena Moreano

C.I.: 0501552798

# **TECHNICAL UNIVERSITY OF COTOPAXI**

## **MAJOR ON ENGINEERING ON COMPUTER SCIENCES AND COMPUTERS SYSTEMS**

**THEME:** “Analysis of the academic performance level among students during their first academic period using decision trees as artificial intelligence technique to determine the possible causes of poor performance in the Faculty of Engineering and Applied Sciences at the Technical University of Cotopaxi.”

**Author:** Tomalo Morales Shirley Vanessa

### **ABSTRACT**

This research project tries to determine the possible causes of poor academic performance through the application of data mining, using decision trees as an artificial intelligence technique to predict their academic performance through factors that affect the students in their development. The results obtained with the application of decision trees to 236 students through a survey were able to verify the factors that influenced the students' performance according to their average, the academic variables obtained more significant information gain since they have higher accuracy for prediction, 87.33% of the students are very well in their level of academic performance. For the analysis of the information obtained on the attitudinal, academic, and identification factors, the Rapidminer tree, set role decision tools that are data mining software was applied, the use of the variables together corresponding to a single aspect that has no greater accuracy than the combination of variables of different aspects.

**Keywords:** Decision Trees, Artificial Intelligence, Data Mining, set role, and Rapidminer.



## *AVAL DE TRADUCCIÓN*

En calidad de Docente del Idioma Inglés del Centro de Idiomas de la Universidad Técnica de Cotopaxi; en forma legal **CERTIFICO** que: La traducción del resumen del proyecto de investigación al Idioma Inglés presentado por la señorita Egresada de la Carrera de **INGENIERÍA EN INFORMÁTICA Y SISTEMAS COMPUTACIONALES** de la **FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS TOMALO MORALES SHIRLEY VANESSA**, cuyo título versa **“ANÁLISIS DEL NIVEL DE RENDIMIENTO ACADÉMICO DE ESTUDIANTES DE LOS PRIMEROS CICLOS UTILIZANDO ARBOLES DE DECISIÓN COMO TÉCNICA DE INTELIGENCIA ARTIFICIAL, PARA DETERMINAR LAS POSIBLES CAUSAS DEL BAJO RENDIMIENTO EN LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS DE LA UNIVERSIDAD TÉCNICA DE COTOPAXI”**, lo realizó bajo mi supervisión y cumple con una correcta estructura gramatical del Idioma.

Es todo cuanto puedo certificar en honor a la verdad y autorizo a la peticionaria hacer uso del presente certificado de la manera ética que estimaré conveniente.

Latacunga, Julio 2019

Atentamente,

Lcdo. Collaguazo Vega Wilmer Patricio Mg.  
**DOCENTE CENTRO DE IDIOMAS**  
C.C. 1722417571

## **1. INFORMACIÓN GENERAL:**

### **Título del proyecto:**

Análisis del nivel de rendimiento académico de estudiantes de los primeros ciclos utilizando arboles de decisión como técnica de inteligencia artificial, para determinar las posibles causas del bajo rendimiento en la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

### **Fecha de inicio:**

Abril del 2019

### **Fecha de finalización:**

Agosto del 2019

### **Lugar de ejecución:**

Universidad Técnica de Cotopaxi

### **Facultad que auspicia:**

Facultad de Ciencias de la Ingeniería y Aplicadas

### **Carrera que auspicia:**

Ingeniería en Informática y Sistemas Computacionales

### **Proyecto de investigación vinculado:**

Inteligencia Artificial enfocada a la optimización de algoritmos de control multivariable.

### **Equipo de trabajo:**

#### **Tutor:**

Nombre: José Augusto Cadena Moreano

Cédula de identidad: 050155279-8

Teléfono: 032262854      Celular: 0984059929

Correo electrónico: jose.cadena@utc.edu.ec

Nivel de estudios:

Tercer nivel: Ingeniero en Informática y Sistemas Computacionales; Licenciado en Ciencias de la Educación especialidad Física y Matemáticas.

Cuarto nivel: Magister en Ciencias de la Educación mención Planeamiento y Administración educativa

**Estudiante:**

Nombre: Shirley Vanessa Tomalo Morales

Fecha de nacimiento: 16 de Julio de 1990

Cédula de identidad: 210080983-5

Teléfono: 062362868      Celular: 0987153709

Correo electrónico: shirley\_vanetm1690@hotmail.com

Estudios: Universidad Técnica de Cotopaxi (UTC)

**Área de conocimiento:**

Ciencias

**Línea de investigación:**

Tecnologías de la Información y Comunicación (TICS).

**Sub línea de investigación:**

Robótica e Inteligencia Artificial

## **2. RESUMEN DEL PROYECTO**

El presente proyecto de investigación tiene como finalidad dar a conocer el análisis del nivel de rendimiento académico de los estudiantes de los primeros ciclos utilizando árboles de decisión, pretende determinar el nivel de rendimiento académico de los estudiantes. El método que se utilizara en esta investigación es Árboles de decisión como modelo de predicción basado en reglas utilizado en el ámbito de la inteligencia artificial. La investigación consiste en considerar algunas de las razones que provocan el bajo rendimiento académico de los estudiantes en cuestión a su estudio, el rendimiento académico es una medida de las capacidades que posee el estudiante y que expresa lo que ha aprendido a lo largo del proceso educativo.

El proyecto contiene la justificación del problema como son los factores que inciden en los estudiantes a bajar su rendimiento académico, razón por la cual los estudiantes de los



primeros ciclos de la Facultad de CIYA tienen dificultades para alcanzar los resultados académicos deseados, razón por la cual afecta a su desenvolvimiento académico en la Universidad. A continuación se realiza el problema de investigación a nivel institucional, puesto que de la universidad y de los profesionales que en ella trabajan, depende la evaluación de los conocimientos de los estudiantes y la obtención de altos o bajos rendimientos académicos, los objetivos planteados, la hipótesis y demás metodologías que se utilizaron para la elaboración del proyecto.

### **3. JUSTIFICACIÓN DEL PROYECTO**

En nuestra actualidad el proceso educativo de todo adolescente debe ser entendido como un proceso social en el cual hay que tener en cuenta el contexto en el que se produce, hace muchos años la escuela era el único contexto que se consideraba educativo. Con el pasar de los años se ha comprobado que existen estudiantes que por diversas causas internas, biológicas o socio-ambientales, los estudiantes tienen dificultades para alcanzar los resultados académicos deseados, razón por la cual afecta a su desenvolvimiento académico en la Universidad.

Se justifica el presente trabajo investigativo porque sus resultados aportaran a un mayor conocimiento de los posibles factores que inciden en el desempeño de los estudiantes, esta información permitirá que desde la gestión institucional, se aborden mecanismos correctivos que contribuyan al mejoramiento de los índices de bajo rendimiento académico, abandono y prolongación excesiva de la duración de la carrera, comunes en el ambiente universitario nacional, principalmente notorios en la actuación de los estudiantes en el primer año de carrera.

El rendimiento académico de cada estudiante es un claro indicador del avance exitoso en la carrera de estudios del estudiante en un período particular, a su vez también es un pronosticador de la posibilidad de culminar exitosamente dicha carrera de estudios. Por tal razón las universidades buscan la acreditación de sus programas o áreas que acrediten la calidad de sus servicios y egresados, es por ello que las universidades tanto públicas como privadas inician procesos de autoevaluación institucional.

El siguiente proyecto de investigación se fundamenta en la recolección de información utilizando herramientas de Minería de Datos mediante Árboles de Decisión como técnica de Inteligencia Artificial, el cual sirve para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, esta herramienta nos permitirá conocer el bajo rendimiento

académico de los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

#### **4. BENEFICIARIOS DEL PROYECTO**

##### **Beneficiarios directos:**

- Los beneficiarios directos del presente proyecto son los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

##### **Beneficiarios indirectos:**

- Docentes de la Facultad de Ciencias de la Ingeniería y Aplicadas en la Universidad Técnica de Cotopaxi.

#### **5. EL PROBLEMA DE INVESTIGACIÓN**

Su transcendencia para el individuo y la sociedad es palpable a partir de dos elementos fundamentales: Primero, cuando el bajo rendimiento académico afecta la autorrealización profesional de los educadores; y el segundo elemento, cuando el nivel de conocimientos y habilidades que pueden adquirir les resulta limitado a las exigencias de su práctica profesional. El bajo rendimiento académico en la universidad indican con claridad que en muchos casos la educación anterior ha sido débil, en este sentido un número significativo de estudiantes no ha sabido responder a las exigencias que lo hubieran conducido a logros satisfactorios en la universidad y posterior desempeño en bien de la sociedad.

El bajo rendimiento académico universitario en la actualidad es un problema que enfrentan los estudiantes y docentes en todos los niveles educacionales motivo por el cual constituye un grave problema social, con importantes repercusiones personales y familiares, las repercusiones de alguna manera se podrían amortiguar si se lograra poner los medios adecuados para reducir el fracaso de los estudiantes universitarios.

A nivel institucional, puesto que de la universidad y de los profesionales que en ella trabajan, depende la evaluación de los conocimientos de los estudiantes y la obtención de altos o bajos rendimientos académicos, a fin de cuentas son los que deciden si el estudiante fracasa o no. Por otro lado desde el punto de vista a nivel social, las personas que hayan sufrido esto

durante su proceso educativo se encontraran con problemas mayores al acceder a la vida profesional.

Existe un porcentaje mayor de estudiantes que desapruban asignaturas sobre todo en los primeros ciclos de la carrera, es preocupante saber que cada ciclo este porcentaje se incrementa, diversos indicadores expresan de manera obvia y preocupante que en los estudiantes exista deficiente nivel académico.

## **6. OBJETIVOS**

### **6.1. Objetivo general:**

Análisis sobre las posibles causas del bajo rendimiento académico de los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas utilizando herramientas de minería de datos mediante árboles de decisión.

### **6.2. Objetivos específicos:**

- ✓ Realizar un estado del arte sobre técnicas y herramientas que permiten hacer una minería de datos como parte de Inteligencia Artificial.
- ✓ Recolectar información de estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas.
- ✓ Identificar las variables y operaciones más importantes del proceso de minería de datos de la Facultad.
- ✓ Realizar un análisis sobre los factores demográficos, académicos y actitudinales existentes que conllevan al bajo rendimiento académico que tienen los estudiantes de los primeros ciclos de cada carrera de la facultad CIYA.

## 7. ACTIVIDADES Y SISTEMAS DE TAREAS EN RELACIÓN A LOS OBJETIVOS

**Tabla 1:** Sistema de tareas en relación a los objetivos planteados.

<b>Objetivo</b>	<b>Actividad(tareas)</b>	<b>Resultado de la actividad</b>	<b>Medios de verificación</b>
Realizar un estado del arte sobre técnicas y herramientas que permiten hacer una minería de datos como parte de Inteligencia Artificial.	<p><b>Tarea 1:</b> Recolectar información sobre técnicas de minería de datos.</p> <p><b>Tarea 3:</b> Observar la información recopilada para el análisis.</p> <p><b>Tarea 2:</b> Conocer que tan verídica pueda ser la información recolectada.</p>	Fundamentación teórica del proyecto de investigación.	<ul style="list-style-type: none"> <li>• Investigación Documental</li> <li>• Investigación de campo</li> <li>• Investigación documental</li> <li>• Investigación exploratoria</li> </ul>
Recolectar información de estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas.	<p><b>Tarea 1:</b> Realizar un diagnóstico para conocer si existen causas que llevan al bajo rendimiento académico de los estudiantes.</p> <p><b>Tarea 2:</b> Realizar encuestas a los estudiantes de los primeros ciclos de la facultad CIYA.</p>	Proyecto de investigación.  Análisis del nivel de rendimiento académico mediante arboles de decisión.	<ul style="list-style-type: none"> <li>• Investigación de campo.</li> <li>• Investigación Documental</li> <li>• Encuestas</li> </ul>

<p>Identificar las variables y operaciones más importantes del proceso de minería de datos de la Facultad.</p>	<p><b>Tarea 1:</b> Analizar la información de los porcentajes obtenidos que se encontraron en la Facultad.</p> <p><b>Tarea 2:</b> Desarrollo del análisis del nivel de rendimiento académico mediante arboles de decisión.</p>	<p>Proyecto de Investigación</p>	<ul style="list-style-type: none"> <li>• Investigación Documental</li> <li>• Análisis del nivel de rendimiento académico mediante arboles de decisión.</li> </ul>
<p>Realizar un análisis sobre los factores demográficos, académicos y actitudinales existentes que conllevan al bajo rendimiento académico que tienen los estudiantes de los primeros ciclos de cada carrera de la facultad CIYA.</p>	<p><b>Tarea 1:</b> Analizar la información obtenida del bajo rendimiento que se encontraron en la Facultad.</p> <p><b>Tarea 2:</b> Observar las causas que conllevan a un bajo rendimiento académico.</p>	<p>Proyecto de Investigación. Interpretación de las causas</p>	<ul style="list-style-type: none"> <li>• Investigación de campo</li> <li>• Investigación aplicada</li> </ul>

**Fuente:** El Investigador

## **8. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA**

### **8.1. Antecedentes**

A nivel mundial, diferentes países han direccionado sus políticas de calidad educativa hacia una formación para el desarrollo de competencias. En el modelo educativo por competencias, la evaluación se considera el eje fundamental del proceso de formación, debido a su impacto sobre el proceso de aprendizaje del estudiante, en la cual se valora de forma continua el desempeño del estudiante en contextos específicos, para tomar decisiones formativas en función de los resultados de evaluación (Brahim, Mohammed y Samir, 2010).

La evaluación continua busca superar las prácticas de la evaluación considerada tradicional, la cual en ciertos casos, está direccionada a medir conocimientos, valora aspectos de carácter sumativo y se orienta más a la medición que a la comunicación de la competencia lograda por el estudiante (Tobón, 2010).

La sociedad del conocimiento es el nuevo tipo de sociedad que se comienza a construir en todo el mundo. Consiste en trabajar de manera colaborativa en la resolución de los problemas con apoyo en la tecnología de la información. En esta nueva sociedad, se requiere transformar el proceso de formación (Sánchez, Ruiz y Sánchez, 2011).

Para lo cual se propone la socio formación, un enfoque que consiste en formar a las personas para que resuelvan problemas del contexto y contribuyan a la realización personal, la convivencia, el fortalecimiento del tejido social, el desarrollo socioeconómico y la sustentabilidad ambiental.

En la sociedad del conocimiento, cada vez más se promueve la participación activa del estudiante en el proceso de evaluación, lo cual es posible mediante diferentes tipos de evaluación, como, por ejemplo: la autoevaluación, la coevaluación y la heteroevaluación (Barbosa, 2010).

### **8.2. Inteligencia Artificial**

Es un campo de la ciencia y la ingeniería que se ocupa de la comprensión, desde el punto de vista informático, de lo que se denomina comúnmente comportamiento inteligente, también se ocupa de la creación de artefactos que exhiben este comportamiento. En el seno de la Inteligencia Artificial como ciencia y tecnología se han ido acumulando conocimientos de como emular las diversas capacidades del ser humano para exhibir comportamientos inteligentes y se han desarrollado sistemas cada vez más perfeccionados que reproducen

parcialmente dichas capacidades (Raúl Benítez, Gerard Escudero, Samir Kanaan, David Masip Rodó, 2013).

El estudio de sensores y mecanismos de intercambio de información con el exterior constituyen, por si mismos, pujantes áreas de investigación y aplicaciones prácticas. También la robótica, como especialidad relacionada con máquinas móviles capaces de interactuar inteligentemente con su entorno y manipular objetos, suele considerarse parcialmente incluida en el ámbito de la IA. La Inteligencia Artificial es la ciencia de construir máquinas para que hagan cosas que, si las hicieran los humanos, requerirían inteligencia (Cañas & Novak, 2010).

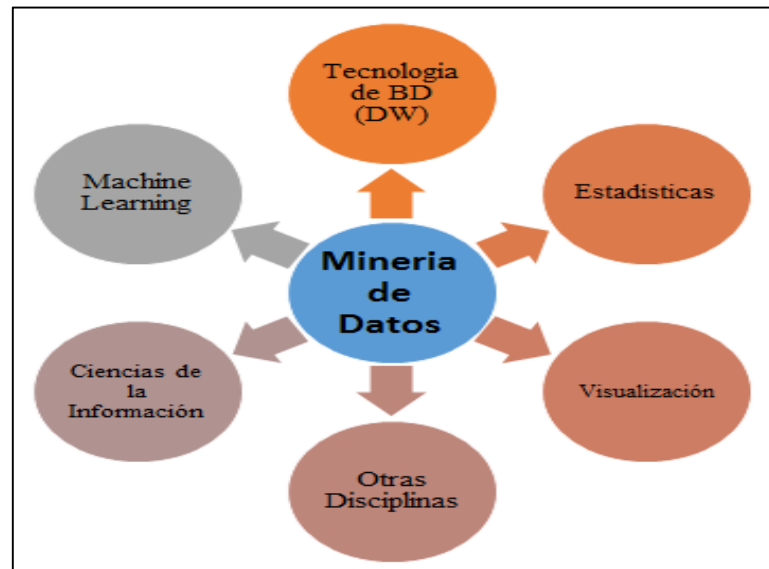
### **8.3. Minería de datos**

La Minería de Datos (MD) como el proceso de descubrir conocimiento útil y entendible, desde grandes bases de datos almacenados en distintos formatos, por medio de modelos inteligibles a partir de los datos, esta utiliza métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos (Icarte Ahumada, 2016).

La Minería de Datos representa una idea que ha venido madurando durante muchos años, en el sentido de recorrer grandes bases de datos para recuperar información conceptual de interés e inferir nueva información útil, permite extraer meta-datos a partir de textos escritos en lenguaje natural (Lugo Reyes, 2014).

La MD es un proceso posterior, destinado a lograr un mejor conocimiento de la información disponible, aumentar beneficios o ventas, y disminuir pérdidas; es decir, tiene un objetivo distinto al que ha motivado la recogida y almacenamiento de información. Esta definición pone de relieve el tratamiento de grandes volúmenes de datos para la generación de conocimiento, siempre que nos encontremos en un escenario donde la extracción de información o conocimiento deba realizarse de forma automática pues su tratamiento manual sería impracticable, entonces podemos hablar de minería de datos (Bedoya et al., 2016).

La minería de datos, es una disciplina de las ciencias e ingenierías de la computación que intenta hallar patrones significativos en conjunto de datos para producir modelos descriptivos, predictivos y clasificadores apoyándose en técnicas de manejo y programación de base de datos, en estadística y aprendizaje automático (Coria, 2016).

**Gráfico 1:** Minería de Datos Asociados a distintos campos

Fuente: (Coria, 2016)

Con todo, la minería de datos es un paso en el proceso de descubrimiento de conocimiento, consiste en la aplicación de algoritmos particulares => métodos que bajo algún objetivo aceptable, para producir una enumeración de patrones => modelos sobre los datos. Se aplican para ello técnicas estadísticas y de inteligencia artificial (algoritmos) para descubrir patrones e irregularidades en los grandes volúmenes de datos. Es, por tanto una tecnología que utiliza técnicas conocidas. Para terminar con la definición de la minería de datos, podemos enumerar también algunos de sus usos:

- Descubre hechos y relaciones de datos.
- Se necesita poca intervención humana.
- Encuentra patrones.
- Determina y establece reglas.
- Almacena y reutiliza reglas.
- Presenta información a los usuarios.
- Puede llevar muchas horas.
- El usuario final debe ser capaz de analizar resultados (Azoumana, 2013).

### 8.3.1. Proceso de la minería de datos

Algunas personas tratan la minería de datos como un sinónimo del término 'Descubrimiento de Conocimiento a partir de datos o KDD', otros en cambio la consideran simplemente un

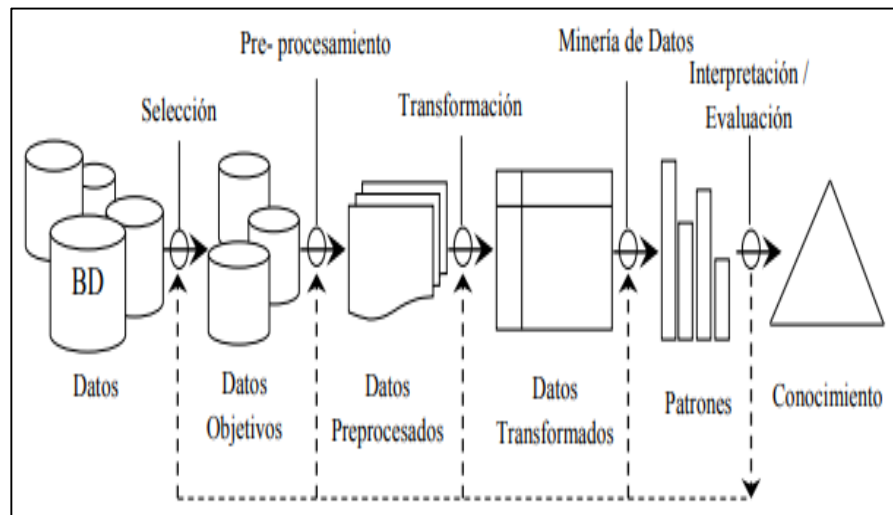


paso esencial en el proceso de descubrimiento de conocimiento que consta de una secuencia interactiva de pasos. La minería de datos emplea una serie de técnicas las cuales son aplicadas para la solución de diversos problemas, sus herramientas predicen futuras tendencias y comportamientos, permitiendo tomar decisiones conducidas por un conocimiento obtenido de los datos. Dentro del proceso interactivo o iterativo del KDD tenemos los siguientes pasos:

- Desarrollo y entendimiento del dominio de la aplicación, este paso requiere cierta dependencia usuario/analista, ya que intervienen factores como=> conocer los cuellos de botella del dominio, saber que partes son susceptibles de un procesado automático y cuáles no, cuales son los objetivos, los criterios de rendimiento exigibles, para que se usaran los resultados que se obtengan, compromisos entre simplicidad y precisión del conocimiento extraído.
- Creación del conjunto de datos objetivo, esto implica consideraciones sobre la homogeneidad de los datos, su variación a lo largo del tiempo, estrategia de muestreo, grados de libertad, etc.
- Procesado de los datos, eliminación de ruido, estrategia para manejar valores ausentes, normalización de los datos, etc.
- Transformación y reducción de los datos, este es un paso critico dentro del proceso global, que requiere un buen conocimiento del problema y una buena intuición, este proceso con frecuencia marca la diferencia entre el éxito o fracaso de la minería de datos.
- Elección del tipo de sistema para minería de datos, este proceso depende de si el objetivo del proceso KDD es la clasificación, regresión, agrupamiento de conceptos (clustering), detección de desviaciones, etc.
- Elección del algoritmo de minería de datos.
- Minería de datos, se realiza la búsqueda de conocimiento con una determinada representación del mismo.
- Interpretación del conocimiento extraído, la obtención de los resultados dependerá de factores como: definición de medidas del interés del conocimiento que permitan filtrarlo de forma automática, existencia de técnicas de visualización para facilitar la valoración de los resultados o búsqueda manual de conocimiento útil entre los resultados obtenidos.
- Consolidación del conocimiento descubierto, incorporándolo al sistema, o simplemente documentándolo y enviándolo a la parte interesada. Este paso incluye la

revisión y resolución de posibles inconsistencias con otro conocimiento extraído previamente (Beltrán Martínez, 2016).

**Gráfico 2:** Proceso KDD



**Fuente:** (Beltrán Martínez, 2016)

### 8.3.2. Aplicaciones de la minería de datos

Algunas de las tareas importantes de la minería de datos incluyen la identificación de aplicaciones para las técnicas existentes, y desarrollar nuevas técnicas para dominios tradicionales o de nueva aplicación, como el comercio electrónico y la bioinformática. Existen numerosas áreas donde la minería de datos se puede aplicar, prácticamente en todas las actividades humanas que generen datos:

- Utilización de árboles de decisión en la construcción de modelos de clasificación de diferentes características del desarrollo del software.
- Comercio y banca: segmentación de clientes, previsión de ventas, análisis de riesgo.
- Medicina y Farmacia: diagnóstico de enfermedades y la efectividad de los tratamientos.
- Seguridad y detección de fraude: reconocimiento facial, identificaciones biométricas, accesos a redes no permitidos, etc.
- Recuperación de información no numérica: minería de texto, minería web, búsqueda e identificación de imagen, video, voz y texto de bases de datos multimedia.
- Astronomía: identificación de nuevas estrellas y galaxias.

- Geología, minería, agricultura y pesca: identificación de áreas de uso para distintos cultivos o de pesca o de explotación minera en bases de datos de imágenes de satélites.
- Ciencias Ambientales: identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales (p.e. plantas depuradoras de aguas residuales) para mejorar su observación, gestión y/o control.
- Ciencias Sociales: Estudio de los flujos de la opinión pública. Planificación de ciudades: identificar barrios con conflicto en función de valores sociodemográficos.

En la actualidad se puede afirmar que la MD ha demostrado la validez de una primera generación de algoritmos mediante diferentes aplicaciones al mundo real. Sin embargo estas técnicas todavía están limitadas por bases de datos simples, donde los datos se describen mediante atributos numéricos o simbólicos, no conteniendo atributos de tipo texto o imágenes, y los datos se preparan con una tarea (Rodríguez Suárez & Amador, 2009).

### 8.3.3. Retos y tendencias de la minería de datos

Existen algunos retos que superar antes de que la minería de datos se convierta en una tecnología de masas. Señalamos algunos de los retos actualmente planteados.

- **Aspectos metodológicos:** Sería muy útil la existencia de una API Standard, de forma que los desarrolladores puedan integrar sin dificultad los resultados de los diversos algoritmos de minería. Esto podría facilitar también la tarea de automatizar y simplificar todo el proceso, integrando aspectos como muestreo, limpieza de datos, minería, visualización, etc... En este mismo sentido sería deseable que los productos de minería de datos estuvieran orientados al programador para fomentar su uso y ampliación. Cuando decimos que una gran decisión se basó en la información disponible, típicamente es una serie de promedios y estimadores estadísticos que presentan una generalización de un gran volumen de datos, donde se hace una inferencia.
- **Escalabilidad:** La escalabilidad de la minería de datos hacia grandes volúmenes de datos es y será siempre una de las tendencias futuras, ya que el volumen de información que se tratara crece de manera exponencial, con lo que los avances en esta área queda siempre superado por las necesidades crecientes. Su tarea consiste en encontrar aquella proyección de los datos, aquella perspectiva donde aparece una

correlación y, lamentablemente, en muchos casos, presentarla como una relación causa-efecto.

- **Simulación, integración en la toma de decisiones y minería de datos:** Los modelos extraídos para un ámbito de interés de una organización. Básicamente se trata de utilizar las salidas de unos modelos como entradas de otros y maximizar el beneficio del conjunto de modelos.
- **Minería para datos con una estructura compleja:** En numerosas ocasiones los datos procedentes de aplicaciones del mundo real no tienen una representación directa en forma de una única tabla, sino que deben ser representadas mediante estructuras jerárquicas (árboles), interrelacionadas (grafos), conjuntos, etc.

Por lo tanto, el reto que se lanza a la comunidad científica que investiga en aprendizaje automático y minería de datos, es el de adaptar o proponer nuevas técnicas que permitan trabajar directamente con este tipo de representaciones. En este campo también entraría la minería de datos distribuida, donde los datos no se encuentran en una única localización sino como es cada vez más habitual en una red de computadores (Beltrán Martínez, 2016).

#### 8.3.4. Técnicas de la minería de datos

La Minería de Datos se podría abstraer como la construcción de un modelo que ajustado a unos datos proporciona un conocimiento. Por tanto podemos distinguir dos pasos en una tarea de MD, por un lado la elección del modelo y por otro el ajuste final de éste a los datos.

La elección del modelo viene determinada básicamente por dos condicionantes: el tipo de los datos y el objetivo que se quiera obtener. Así por ejemplo no sería apropiado aplicar regresión a unos datos constituidos por texto o modelos basados en distancia a datos simbólicos.

En cuanto a la relación modelo-objetivo, la literatura presenta un catálogo de distintos modelos para los diferentes objetivos. Así, si se tiene un problema de clasificación se utilizarán máquinas de vectores soporte o árboles de decisión, si es un problema de regresión se pueden usar árboles de regresión o redes neuronales, si se desea hacer clustering se puede optar por modelos jerárquicos o interrelacionados, etc.

También es importante en esta elección el nivel de comprensibilidad que se quiera obtener del modelo final, ya que hay modelos fáciles de “explicar” al usuario como por ejemplo las reglas de asociación y otros que entrañan claras dificultades como las redes neuronales o los vectores soporte. El segundo paso consiste en realizar una “fase de aprendizaje” con los datos

disponibles para ajustar el modelo anterior a nuestro problema particular. Así si tenemos una red neuronal habrá que definir su arquitectura y ajustar los valores de los pesos de sus conexiones. Si vamos a obtener una recta de regresión hay que hallar los valores de los coeficientes y si usamos los k-vecinos más cercanos necesitamos fijar una métrica y k (Azoumana, 2013).

### **8.3.5. Clasificación de las técnicas de minería de datos**

#### **8.3.5.1. Métodos descriptivos**

Los métodos descriptivos son un modelo para dar mejor comprensión de los datos, sin ninguna variable específica como objetivo único, por lo general se utilizan técnicas descriptivas que incluyen análisis de factores con el fin de extraer dimensiones subyacentes de datos multivariados, análisis de clúster que se refiere a la agrupación de una base de datos de clientes en segmentos y por último el análisis de asociación que se encarga de descubrir relaciones entre tales como productos de ventas al por menor, que son orientados a la interpretación de datos y se enfocan en entender la forma en que los datos subyacentes se refieren a sus partes. Las herramientas de Data Mining barren las bases de datos e identifican modelos previamente escondidos en un solo paso (Azoumana, 2013).

#### **8.3.5.2. Métodos predictivos**

Estos métodos predictivos pretenden construir un modelo de comportamiento, de la cual se obtiene nueva y ocultas muestras y es capaz de predecir valores de una o más variables relacionadas con la muestra, varias técnicas orientadas al descubrimiento están basadas en aprendizaje inductivo donde un modelo es construido implícita o explícitamente mediante la generalización de una cantidad suficiente de muestras de entrenamiento.

Los métodos de verificación incluyen las técnicas comunes de la estadística tradicional, como la bondad de ajuste, escala y técnicas de predicción capaces de manejar grandes cantidades de datos, hay diversas técnicas para la clasificación de datos, tales como:

- Construcción de clasificadores de árboles de decisión
- Clasificadores bayesianas
- Redes bayesianas
- Creencias
- Clasificadores basados en reglas

Los métodos para la predicción incluyendo regresión lineal, regresión no lineal entre otros modelos de basados en la regresión; la clasificación y predicción tienen numerosas aplicaciones incluyendo la detección del fraude, el marketing de destino, la predicción del rendimiento, la fabricación y el diagnóstico médico (Azoumana, 2013).

### 8.3.6. Algoritmos de clasificación

La clasificación de datos se desarrolla una descripción o modelo para una de las clases presentes en la base de datos, existen muchos métodos de clasificación como aquellos basados en los árboles de decisión como el ID3 y el C5.4, los métodos estadísticos, las redes neuronales y los conjuntos difusos entre otros, muchos algoritmos estadísticos han sido utilizados por los analistas para detectar patrones inusuales en los datos y explicar dichos patrones mediante la utilización de los modelos estadísticos como son los modelos lineales (Riquelme, Ruiz, & Gilbert, 2006).

Las técnicas de minería de datos de la inteligencia artificial y más representativa son las siguientes:

- **Redes Neuronales.-** Las redes neuronales imitan la capacidad de la mente humana para encontrar patrones, han sido aplicadas con éxito en aplicaciones que trabajan sobre la clasificación de los datos. Son modelos matemáticos que permiten hacer computación inteligente y llevar a cabo tareas que las computadoras seriales no pueden realizar: reconocimiento de patrones, memorias y aprendizaje asociativo, control adaptivo, predicción de series de tiempo, clasificación de señales y clustering, entre otras. El proceso es “simple”: cada neurona recibe una serie de entradas, que llevarán un peso, emitiendo una salida. La salida viene dada por tres funciones:
  - **La función de propagación**, suele ser el sumatorio de cada entrada multiplicada por el peso asignado.
  - **La función de activación**, tiene como misión modificar a la de propagación. No siempre aparece, coincidiendo en estos casos con la dicha propagación.
  - **La función de transferencia**, se aplica al valor dado por la función de aplicación y se utiliza para acotar la salida de cada neurona según la interpretación que le queramos dar al resultado.

- **Arboles de Decisión.-** En un árbol de decisión cada nodo representa una característica que puede tomar diversos valores, cada uno de los cuales genera una rama, los nodos hojas representan las clasificaciones finales, los arboles de decisión generalmente utilizan una técnica de aprendizaje supervisado que consiste en realizar particiones recursivas en un espacio de datos conformado por un conjunto de casos o ejemplos, la división de cada conjunto se realiza teniendo en cuenta el atributo cuya participación o agrupación por sus valores se parezca más a la partición que produce el atributo clase.

El árbol que se va construyendo con cada partición, sus nodos representan un atributo que sirvió para realizar una partición, los arcos que conectan los nodos entre si representan los diferentes valores que pueden tomar los atributos y las hojas representan subconjuntos de casos pertenecientes a una clase.

- **NaiveBayes.-** Es una de las técnicas de clasificación más usadas basadas en las estadísticas, se trata de un algoritmo de inducción probabilística que representa cada clase como un sumario de probabilidades, el fundamento principal de este clasificador es la suposición de que todos los atributos son independientes conocido el valor de la variable clase.
- **C-Means Clásico.-** Es un algoritmo iterativo que hace parte de las técnicas de agrupamiento no supervisado y tiene como objetivo encontrar patrones o grupos interesantes en un conjunto de datos dados.
- **Fuzzy C-Means.-** Este algoritmo asigna a cada dato un valor de pertenencia dentro de cada clúster y por consiguiente un dato específico puede pertenecer parcialmente a más de un clúster (Lopera, 2018).

#### 8.4. Árbol de decisión

Un árbol de decisión es un modelo de predicción utilizado en diversos ámbitos que van desde la Inteligencia Artificial hasta la Economía, con la aplicación de conjuntos de datos se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema.

Los arboles de decisión se clasifican en instancias, ordenadas a partir de la raíz, recorriendo diversos nodos de decisión, hasta que llegue a un nodo tipo hoja, el cual proporciona la

clasificación de la instancia considerada. El aprendizaje por medio de árboles de decisión es generalmente más apropiado para problemas que poseen las siguientes características:

- Las instancias son representadas por pares atributo – valor.
- La función objetivo posee valores de salida discretos.
- Las descripciones son disyuntivas.
- Los datos de entrenamiento, pueden contener errores.
- Los datos de entrenamiento pueden tener atributos con valores incompletos o faltantes (Banda, 2014).

Los arboles de decisión aplicados a la educación, nos ayuda a determinar el grupo de variables de predicción que estén más relacionadas en nuestra variable final, el índice de fracaso escolar están diseñados para ser capaces de manejar un número muy grande de variables de predicción, en algunos casos, más allá de lo que permitiría el correspondiente modelo paramétrico estadístico (García & Educacional, n.d.).

#### **8.4.1. Elementos de un árbol de decisión**

Los árboles de decisión están formados por nodos, vectores de números, flechas y etiquetas.

- **Nodo.-** Cada nodo se puede definir como el momento en el que se ha de tomar una decisión de entre varias posibles, lo que va haciendo que a medida que aumenta el número de nodos aumente el número de posibles finales a los que puede llegar el individuo. Esto hace que un árbol con muchos nodos sea complicado de dibujar a mano y de analizar debido a la existencia de numerosos caminos que se pueden seguir.
- **Vectores de números.-** Estos vectores serían la solución final a la que se llega en función de las diversas posibilidades que se tienen, dan las utilidades en esa solución.
- **Flechas.-** Las flechas son las uniones entre un nodo y otro y representan cada acción distinta.
- **Etiquetas.-** Las etiquetas se encuentran en cada nodo y cada flecha y dan nombre a cada acción (Rodríguez Suárez & Amador, 2009).

#### **8.4.2. Clasificación de árboles de decisión**

Los árboles de decisión utilizados en la minería de datos son de dos tipos principales:



- Es cuando el resultado predicho es la clase a la que pertenecen los datos.
- Árboles de regresión es cuando el resultado predicho se puede considerar un número real (por ejemplo, el precio de una casa, las notas de un estudiante, etc).

El término Árboles de Clasificación y Regresión (ACR) es un término genérico utilizado para referirse a ambos de los procedimientos anteriores. Los árboles utilizados para la regresión y los árboles utilizados para la clasificación tienen algunas similitudes - pero también algunos diferencias, tales como el procedimiento utilizado para determinar donde dividir.

Algunas técnicas, a menudo llamados métodos conjuntos híbridos, construyen más de un árbol de decisión:

- Bagging, un método de conjunto, construye múltiples árboles de decisión haciendo repetidamente remuestreo de los datos de entrenamiento con sustitución, y votando los árboles para hallar una predicción de consenso.
- Un clasificador Random Forest utiliza una serie de árboles de decisión, con el fin de mejorar la tasa de clasificación.
- Los árboles impulsados se pueden utilizar para problemas de regresión y de clasificación.
- Rotation Forest, en el que cada árbol de decisión es entrenado aplicando primero análisis de componentes principales análisis de componentes principales (ACP) en un subconjunto aleatorio de las características de entrada.

Aprendizaje basado en árboles de decisión es la construcción de un árbol de decisión a partir de tuplas de entrenamiento, cada una etiquetada con su correspondiente clase. Un árbol de decisión es similar a una estructura de diagrama de flujo, donde cada nodo interno (no hoja) denota una prueba en un atributo, cada rama representa el resultado de una prueba, y cada hoja (o terminal) nodo tiene una etiqueta de clase. El nodo superior en un árbol es el nodo raíz.

Hay muchos algoritmos específicos de árbol de decisiones. Entre los más destacados están:

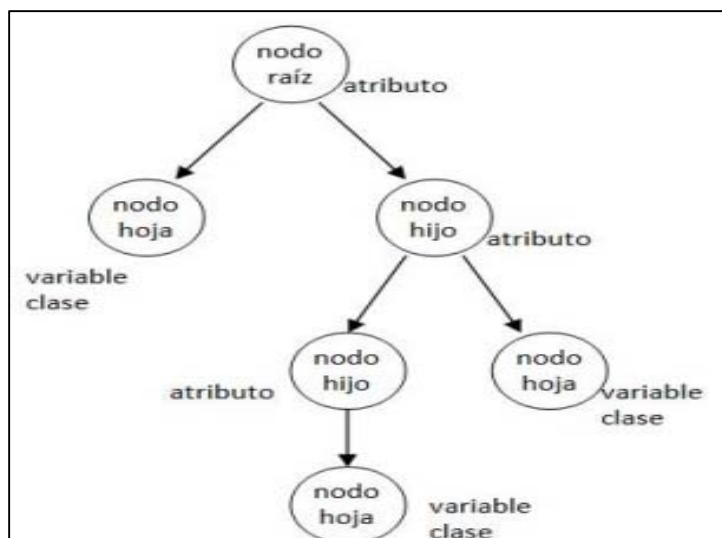
- ID3(Iterative Dichotomiser 3)
- C4.5(Sucesor de ID3)
- ACR(Árboles de Clasificación y Regresión)

- CHAID(Detector automático de Chi-cuadrado de interacción). Realiza divisiones de múltiples niveles al calcular los árboles de clasificación.
- MARS: Extiende los árboles de decisión para manejar mejor datos numéricos.
- Árboles de inferencia condicional: Enfoque que utiliza pruebas no paramétricas como criterios de división, corregidos para múltiples pruebas para evitar el sobreajuste. Este enfoque se traduce en la selección de un predictor imparcial y no requiere poda (Lopera, 2018).

### 8.4.3. Características de los árboles de decisión

Cada nodo de decisión especifica la prueba de algún atributo de la instancia y cada rama descendiente que parte del nodo corresponde a uno de los valores posibles que puede asumir el atributo (Banda, 2014).

**Gráfico 3:** Estructura de un árbol de decisión



Fuente: (Erandi et al., 2009)

## 8.5. Software Rapidminer Studio

Rapidminer es una herramienta perfecta para crear modelos y a posterior la realización de análisis predictivos de grandes volúmenes de datos.

Es una solución que facilita el autoservicio de análisis predictivo permitiendo una avanzada analítica empleando solamente drag and drop y opcionalmente la generación de código.

Se utiliza para realizar análisis de minería de datos (Data Mining) en aplicaciones empresariales, gobierno y academias.

Rapidminer studio permite dar soluciones ya que acelera la creación de prototipos y la validación de modelos así mismo, agiliza la transformación, el desarrollo y la validación de datos. Con esta herramienta tendrá:

- Tener acceso a los datos.
- Crear un conjunto de datos óptimo para el análisis predictivo.
- Limpiar datos de algoritmos avanzados.
- Descubrir patrones o problemas de datos.
- Modelado, entre otras funcionalidades (Banda, 2014).

#### **8.6. Otros software de minería de datos**

- **Orange** es una suite de software para minería de base de datos y aprendizaje automático basado en componentes que cuenta con un fácil y potente, rápido y versátil front-end de programación visual para el análisis exploratorio de datos y visualización, y librerías para Python y secuencias de comando. Contiene un completo juego de componentes para preprocesamiento de datos, característica de puntuación y filtrado, modelado, evaluación del modelo, y técnicas de exploración. Está escrito en C++ y Python, y su interfaz gráfica de usuario se basa en la plataforma cruzada del framework Qt
- **Weka** (Entorno Waikato para el Análisis del Conocimiento) este software es escrito en java por lo tanto es conocida suite de software para máquinas de aprendizaje que soporta varias tareas típicas de minería de datos, especialmente pre procesamiento de datos, agrupamiento, clasificación, regresión, visualización y características de selección. Sus técnicas se basan en la hipótesis de que los datos están disponibles en un único archivo plano o relación, donde cada punto marcado es etiquetado por un número fijo de atributos. WEKA proporciona acceso a bases de datos SQL utilizando conectividad de bases de datos Java y puede procesar el resultado devuelto como una consulta de base de datos. Su interfaz de usuario principal es el Explorer, pero la misma funcionalidad puede ser accedida desde la línea de comandos o a través de la interfaz de flujo de conocimientos basada en componentes.

- **JHepWork** Diseñado para los científicos, ingenieros y estudiantes, jHepWork es un framework para análisis de datos libre y de código abierto que fue creado como un intento de hacer un entorno de análisis de datos usando paquetes de código abierto con una interfaz de usuario comprensible y para crear una herramienta competitiva a los programas comerciales. Esto se hace especialmente para las ploteos científicos interactivos en 2D y 3D y contiene bibliotecas científicas numéricas implementadas en Java para funciones matemáticas, números aleatorios, y otros algoritmos de minería de datos. jHepWork se basa en Jython un lenguaje de programación de alto nivel, pero codificación en Java también puede ser usada para llamar librerías jHepWork numéricas y gráficas.
- **KNIME** (Konstanz Information Miner) es una plataforma de código abierto de fácil uso y comprensible para integración de datos, procesamiento, análisis, y exploración. Ofrece a los usuarios la capacidad de crear de forma visual flujos o tuberías de datos, ejecutar selectivamente algunos o todos los pasos de análisis, y luego estudiar los resultados, modelos y vistas interactivas. KNIME está escrito en Java y está basado en Eclipse y hace uso de sus métodos de extensión para soportar plugins proporcionando así una funcionalidad adicional. A través de plugins, los usuarios pueden añadir módulos de texto, imagen, procesamiento de series de tiempo y la integración de varios proyectos de código abierto, tales como el lenguaje de programación R, WEKA, el kit de desarrollo de Química y LIBSVM (García & Educacional, n.d.).

### **8.7. Rendimiento académico**

El rendimiento es la expresión de capacidades y de características psicológicas del estudiante desarrollado y actualizado a través del proceso de enseñanza-aprendizaje que le posibilita obtener un nivel de funcionamiento y logros académicos a lo largo de un período, año o semestre, que se sintetiza en un calificativo final (cuantitativo en la mayoría de los casos) evaluador del nivel alcanzado (Mora, 2014).

Es un resultado del aprendizaje, suscitado por la intervención pedagógica del profesor o la profesora, y producido en el alumno. No es el producto analítico de una única aptitud, sino más bien el resultado sintético de una suma de elementos que actúan en, y desde la persona que aprende, tales como factores institucionales, pedagógicos, psicosociales y

sociodemográficos. Factores inmediatos como la edad, el sexo, la región geográfica, los ingresos, el número de créditos y la metodología de estudio, este amplio interés ha hecho que el rendimiento académico sea uno de los factores más analizados y de cierta forma una medida de un sistema educativo efectivo y eficaz (Moncada Mora, 2014).

### **8.8. Bajo rendimiento académico**

El bajo rendimiento académico no es un estado permanente y depende tanto del individuo como del entorno familiar. Es uno de los problemas que más preocupa a padres y educadores, pues psicológicamente puede dejar secuelas en la vida.

El bajo rendimiento académico es un problema que enfrentan estudiantes y educadores en todos los niveles educacionales. Su trascendencia para el individuo y la sociedad es palpable a partir de 2 elementos fundamentales: primero, cuando el bajo rendimiento académico afecta la autorrealización profesional de los educandos, y segundo, cuando el nivel de conocimientos y habilidades que pueden adquirir, resulta limitado a las exigencias de su práctica profesional.

En muchos casos el bajo rendimiento académico se da por muchos factores como pueden ser: el fumar a temprana edad afecta la salud del adolescente ya que no solo afecta la salud sino también afecta negativamente al rendimiento académico, otro factor que también afectan es el consumo de bebidas alcohólicas esto a muy corto tiempo hará que los adolescentes pierdan oportunidades de formación y llegar a ser profesionales en el futuro (Giovanny et al., 2017).

## **9. PREGUNTAS CIENTÍFICAS O HIPÓTESIS**

### **9.1. HIPÓTESIS**

Con el análisis del nivel de rendimiento académico de los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi, mediante la utilización de árboles de decisión permitirá determinar las causas del bajo rendimiento académico.

## **10. METODOLOGÍA**

El objetivo de este trabajo es explicar la metodología que sustenta a este estudio. En primer lugar se definió el problema de estudio para poder establecer el tipo de investigación que se va a llevar a cabo, y así poder hacer el diseño de la misma. Seguidamente se definió la recolección y el análisis de los datos que se obtuvieron con la investigación.

## **10.1. Tipos de investigación**

### **Investigación aplicada**

Esta investigación se distingue por tener propósitos prácticos inmediatos bien definidos, es decir se investiga para actuar, transformar modificar o producir cambios en un determinado sector de la realidad. Para realizar investigaciones aplicadas es muy importante contar con el aporte de las teorías científicas que son producidas por la investigación básica y sustantiva.

### **Investigación de campo**

La investigación de campo fue utilizada en el proyecto, ya que los datos e informaciones recolectadas fueron extraídos de forma directa de la realidad y por el propio investigador, a través del uso de instrumentos para recolectar la información, con el fin de dar respuesta a alguna situación o problema planteado.

### **Investigación exploratoria**

Se aplicó la investigación exploratoria, para su respectivo análisis de los antecedentes que puedan causar el bajo rendimiento académico de los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica De Cotopaxi.

## **10.2. Métodos de investigación**

### **Método Hipotético-Deductivo**

El método Hipotético-Deductivo ayudará en la investigación, porque permite encontrar soluciones a los problemas planteados en la investigación y comprobarlos con datos adquiridos, el método hipotético deductivo lo utilizaremos para mejorar el proyecto a desarrollar.

### **Método histórico**

Es aquella orientación que va del pasado al presente para proyectarse al futuro, generalmente la etapa de tiempo proyectada al futuro es equivalente en extensión a la etapa considerada del pasado.

## **Encuesta**

La técnica de la encuesta fue aplicada en el proyecto para recolectar información que sea útil para el investigador. Esta técnica fue realizada para obtener mayor información sobre el nivel de rendimiento académico.

### **10.3. Acopio y procesamiento de los datos**

Data extraída del sistema elcmnc Total matriculados 2018-2019

Data extraída del sistema sistemas Total matriculados 2018-2019

Data extraída del sistema ind Total matriculados 2018-2019

Data extraída del sistema eléctrica Total matriculados 2018-2019

Reporte de notas estudiantes de Facultad CIYA 2018\_I

### **10.4. Diseño experimental**

El diseño experimental, para realizar el proyecto de investigación sobre el análisis del nivel de rendimiento académico mediante arboles de decisión se lo realizó con el software correspondiente para poder realizar minería de datos con los datos recolectados.

### **10.5. Población y muestra**

Para definir la población se tuvo en cuenta todos los Árboles que se pueden formar con la combinación de los atributos (campos) referidos a los registros de los estudiantes de I Ciclo de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi.

La muestra es no probabilística-intencionada, es aquella que el investigador selecciona según su propio criterio, sin ninguna regla matemática o estadística. El investigador procede a seleccionar la muestra en forma intencional, eligiendo aquellos elementos que considera convenientes y cree que son los más representativos.

### **10.6. Instrumentos para recolectar datos**

Los instrumentos que se utilizarán en la investigación consisten en unos cuestionarios que contienen preguntas cerradas, abiertas y de opción múltiple ver Anexo N°1 para la recolección de datos, demográficos, académicos y actitudinales.

## 10.7. Procesamiento de datos

Los datos obtenidos mediante la aplicación de las técnicas e instrumentos antes mencionados; serán incorporados a programas computarizados, tales como el aplicativo de MS Office -Excel y con precisiones porcentuales que son presentados como informaciones en forma de, gráficos y cuadros

## 11. ANÁLISIS DE RESULTADOS

Se aplica una encuesta basado en variables de identificación, académicos y actitudinales a los estudiantes de los primeros ciclos, a continuación tenemos el modelo correspondiente a la encuesta aplicada. Para el siguiente análisis se planteó una encuesta a los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas Ver Anexo N° 1.

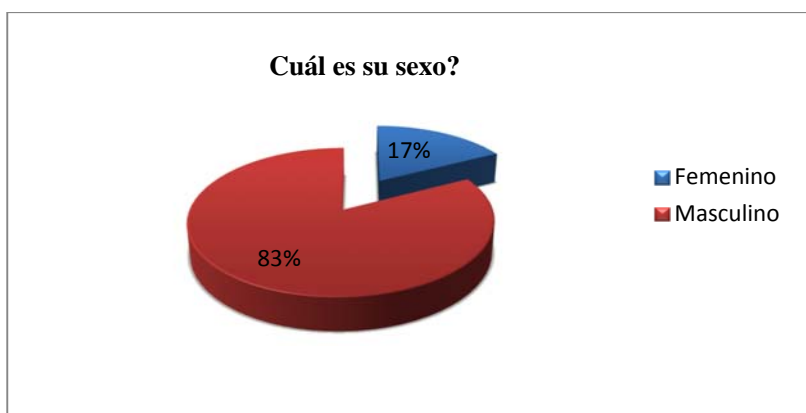
### 11.1. Análisis de la encuesta aplicada

Los datos obtenidos mediante la aplicación de técnicas e instrumentos antes mencionados, fueron incorporados a programas computarizados tales como el aplicativo de Excel y con precisiones porcentuales que son representados como informaciones en forma de gráficos y cuadros, los instrumentos que se utilizarón en la investigación consisten en unos cuestionarios que contienen preguntas abiertas y cerradas y de opción.

Despues de haber recogido la informacion correspondiente de la encuesta emitida, a continuacion nos expone los siguientes resulataados:

#### 1. ¿Cuál es su sexo?

Gráfico 4: Estadística gráfica – pregunta 1



Fuente: Encuesta elaborado por el investigador

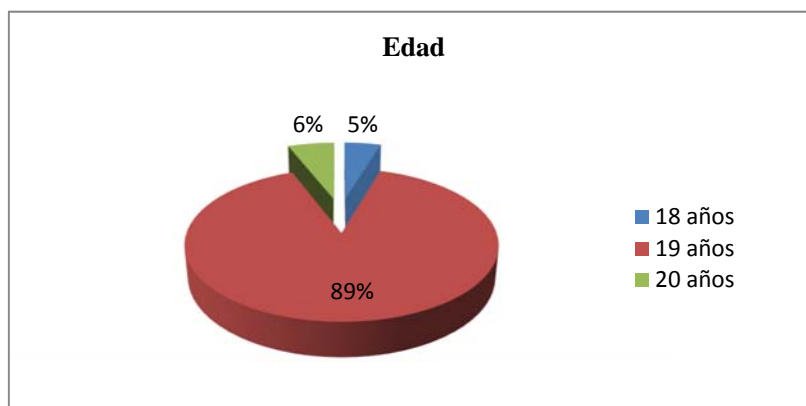


## Interpretación

- En la siguiente pregunta la cual consistía en identificar su sexo, en la cual podemos decir que el 83% de los estudiantes de es sexo masculino mientras que el 17% es de sexo femenino.

## 2. ¿Cuál es su edad?

**Gráfico 5:** Estadística gráfica – pregunta 2



**Fuente:** Encuesta elaborado por el investigador

## Interpretación

- En esta pregunta podemos decir que la edad fluctúa entre los 18 a 19 años de edad de los estudiantes de los primeros ciclos, la cual el 5% está en los 18 años de edad, el 6% tienen 20 años mientras que el 89% de los estudiantes tienen 19 años.

## 3. ¿Estado civil?

**Gráfico 6:** Estadística gráfica – pregunta 3



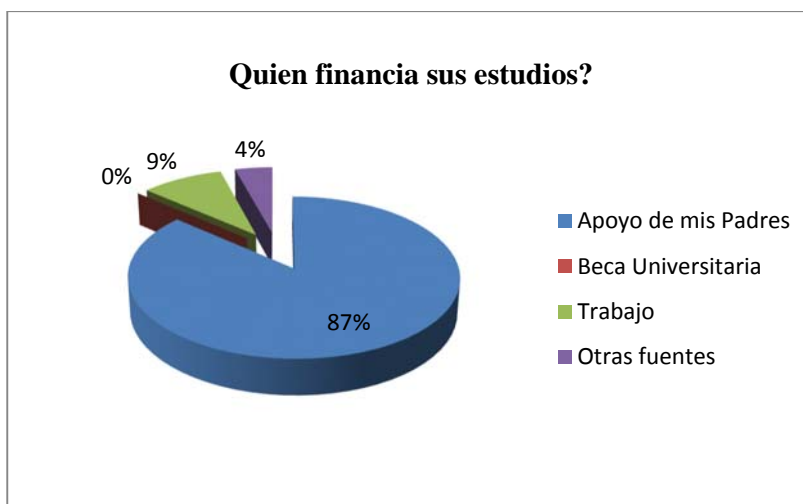
**Fuente:** Encuesta elaborado por el investigador

### Interpretación

- La pregunta 3 demuestra que solo existe el 3% de los estudiantes en unión libre mientras que el 97% de los estudiantes es soltero(a), no tenemos estudiantes que sean casados o separados por lo cual tenemos un 0% en estos ítems.

### 4. ¿Quién financia sus estudios?

**Gráfico 7:** Estadística gráfica – pregunta 4



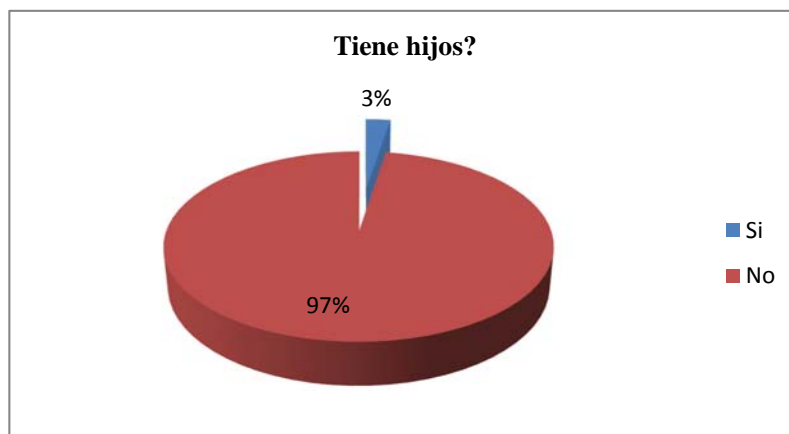
**Fuente:** Encuesta elaborado por el investigador

### Interpretación

- En la pregunta 4 podemos observar que el 87% de los estudiantes tiene el apoyo de sus padres, el 9% de los estudiantes trabajan, el 4% de los estudiantes reciben apoyo de otras fuentes y no existe beca universitaria.

## 5. ¿Tiene usted hijos?

Gráfico 8: Estadística gráfica – pregunta 5



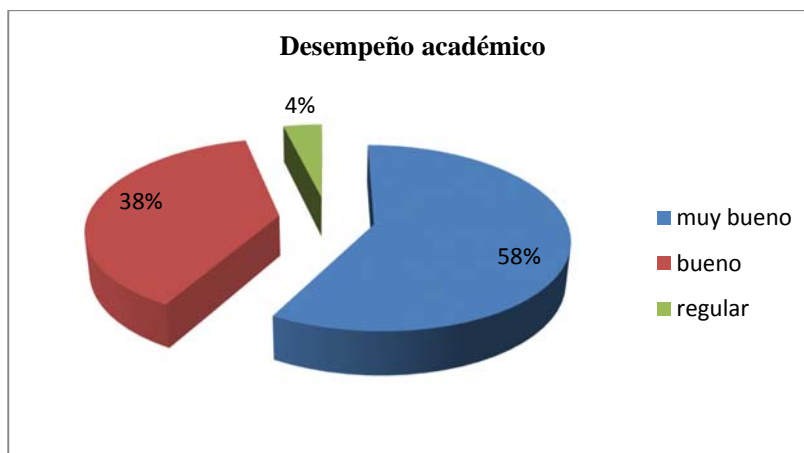
Fuente: Encuesta elaborado por el investigador

## Interpretación

- El 97% de los estudiantes no tiene hijos mientras que el 3% de los estudiantes tienen hijos.

## 6. Cómo es su desempeño académico actual?

Gráfico 9: Estadística gráfica – pregunta 6



Fuente: Encuesta elaborado por el investigador

## Interpretación

- En la pregunta 6 en cuanto al desempeño académico el 58% de los estudiantes es de nivel de desempeño muy bueno, el 38% bueno y el 4% de los estudiantes es regular.

## 7. Usted ha repetido el ciclo académico?

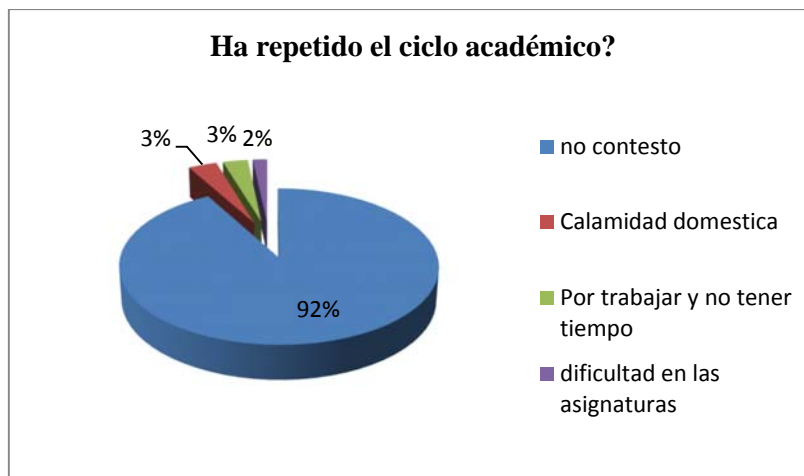
Gráfico 10: Estadística gráfica – pregunta 7



Fuente: Encuesta elaborado por el investigador

## 8. Causa por la cual ha repetido el ciclo académico?

Gráfico 11: Estadística gráfica – pregunta 8



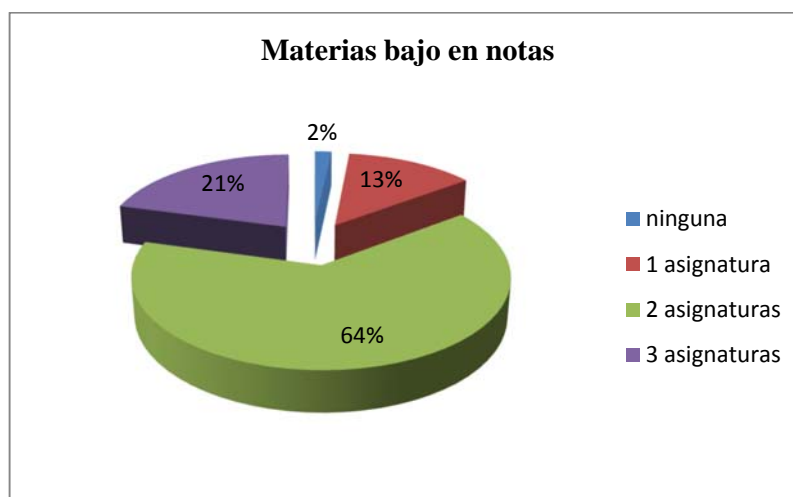
Fuente: Encuesta elaborado por el investigador

### Interpretación

- El 92% de los estudiantes que no contestaron quiere decir que no ha repetido el ciclo académico, mientras que el 8% restante si a repetido el ciclo académico el cual se desglosa a continuación: el 3% ha repetido por calamidad domestica, el 3% por motivos de trabajo y el 2% en dificultad de las asignaturas.

### 9. En cuántas asignaturas usted encuentra bajo en notas?

**Gráfico 12:** Estadística gráfica – pregunta 9



**Fuente:** Encuesta elaborado por el investigador

### Interpretación

- El 2% de los estudiantes no se encuentra bajo en ninguna asignatura, el 13% en 1 asignatura, el 64% en dos asignaturas y el 21% en tres asignaturas.

### 10. ¿En su casa usted dispone de un lugar adecuado para estudiar?

Gráfico 13: Estadística gráfica – pregunta 10



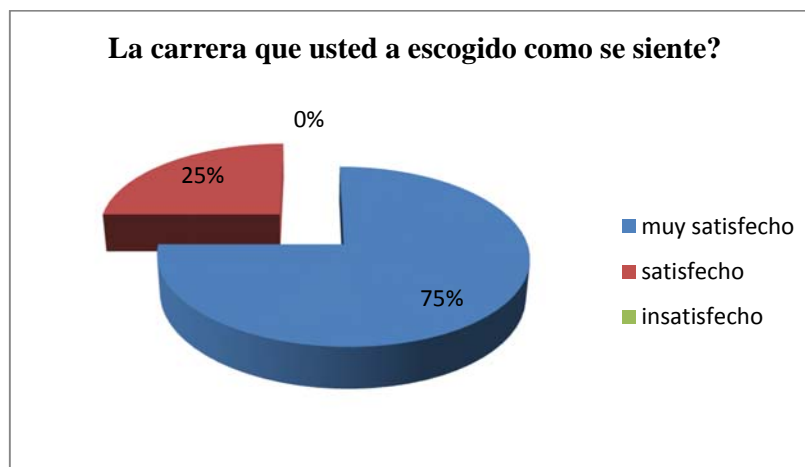
Fuente: Encuesta elaborado por el investigador

### Interpretación

- El 87% de los estudiantes contestaron que si cuentan con un lugar adecuado para estudiar, mientras que el 13% de los estudiantes no posee con el lugar adecuado para estudiar.

### 11. ¿En cuanto a la carrera que usted ha escogido como se siente?

Gráfico 14: Estadística gráfica – pregunta 11



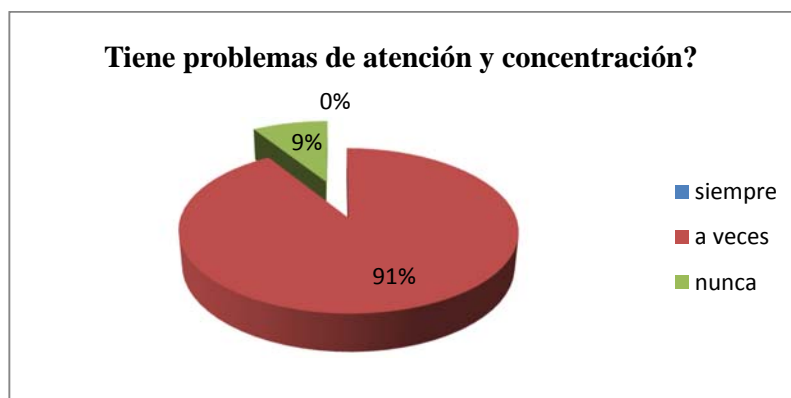
Fuente: Encuesta elaborado por el investigador

## Interpretación

- El 75% de los estuديات contestaron muy satisfactorio en cuanto la carrera que han escogido, el 25% se siente satisfechos y no hay estudiante que se sienta insatisfecho con la eleccion en cuanto a la carrera que han escogido.

## 12. Tiene problemas de atención y concentración en las clases?

**Gráfico 15:** Estadística gráfica – pregunta 12



**Fuente:** Encuesta elaborado por el investigador

## Interpretación

- El 91% de los estudiantes a veces tiene problemas en concentración mientras que el 9% nunca.

## 13. Pone empeño al estudio para superar y sacar buenas notas?

**Gráfico 16:** Estadística gráfica – pregunta 13



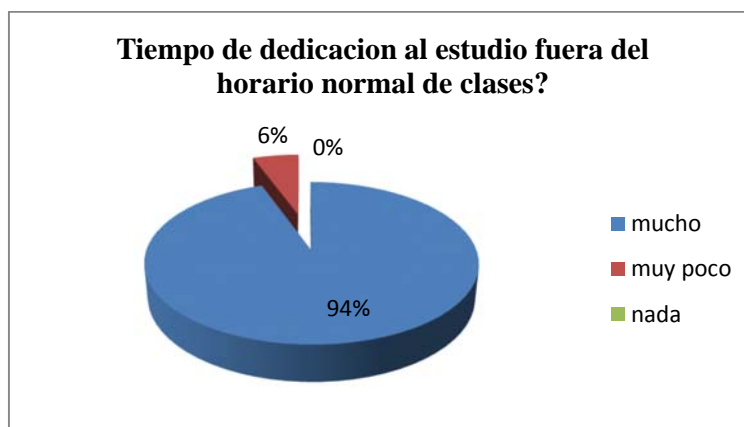
**Fuente:** Encuesta elaborado por el investigador

### Interpretación

- El 6% de los estudiantes respondió a veces se esfuerzan para sacar buenas notas y superarse cada día, mientras que el 94% de los estudiantes se esfuerzan cada día para ser mejores y sacar buenas notas, no existen estudiantes que nunca se esfuercen.

### 14. Cuánto tiempo le dedica usted al estudio fuera del horario normal de clases?

Gráfico 17: Estadística gráfica – pregunta 14



Fuente: Encuesta elaborado por el investigador

### Interpretación

- El 94% de los estudiantes le dan mucho tiempo al estudio despues de sus horas normales de clases mientras que el 6% solo le dedican muy poco al estudio fuera de las horas normales.

### 15. Entrega a tiempo sus trabajos?

Gráfico 18: Estadística gráfica – pregunta 15



Fuente: Encuesta elaborado por el investigador

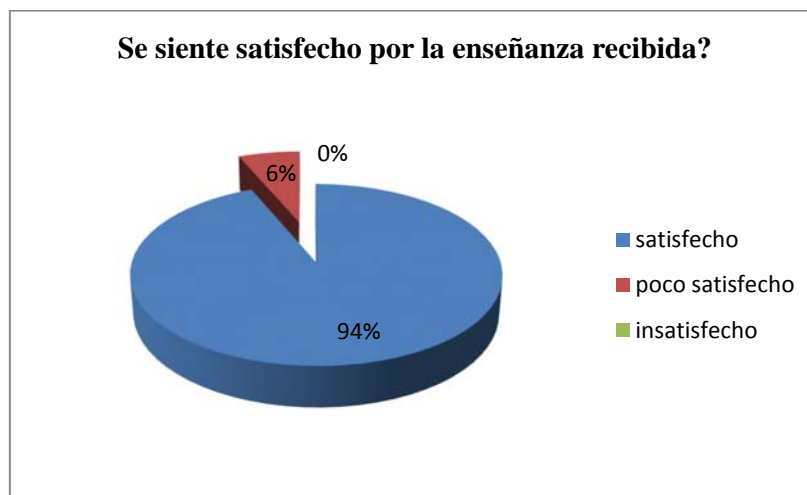


### Interpretación

- El 91% de los estudiantes entrega siempre sus trabajos a tiempo de lo requerido y el 9% a veces entrega sus trabajos a tiempo.

### 16. Se siente satisfecho por la enseñanza recibida?

**Gráfico 19:** Estadística gráfica – pregunta 16



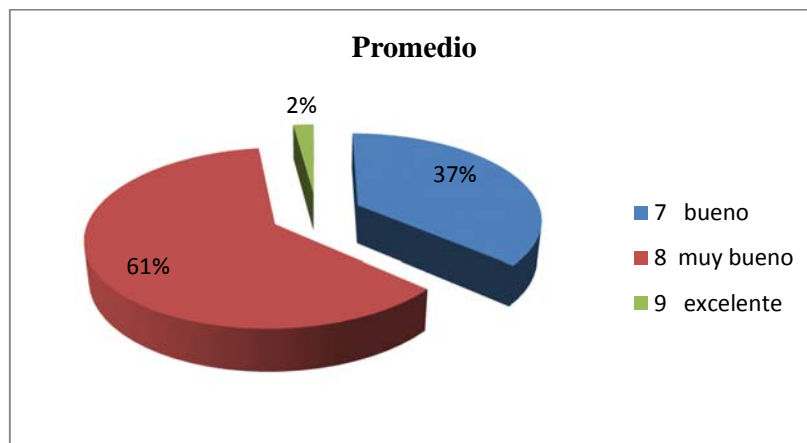
**Fuente:** Encuesta elaborado por el investigador

### Interpretación

- El 94% de los estudiantes se siente satisfecho por la enseñanza recibida y el 6% se siente poco satisfecho.

### 11.2. Estadística en cuanto al promedio de los estudiantes

**Gráfico 20:** Estadística gráfica – Promedio



**Fuente:** Encuesta elaborado por el investigador

### 11.3. Caso de estudio del nivel de rendimiento académico

El nivel de rendimiento académico fue el caso a estudiar para predecir el bajo rendimiento académico que pasan los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas, mediante la aplicación de las técnicas y métodos que posee la minería de datos, para la realización de este estudio se necesitó lo siguiente:

- 1.- Cantidad de datos recolectados
- 2.- Información actual de los estudiantes
- 3.-Accesibilidad

Una vez visto el problema con el cual vamos a trabajar se procedió a elaborar los instrumentos necesarios para la respectiva recolección de datos, mediante cuestionarios que se estructuró para que nos proporcionara la información importante sobre los factores que influyen en el rendimiento académico de los estudiantes para el respectivo análisis, también se trabajó con datos anteriores de los estudiantes para poder realizar mejor el análisis.

Los datos recolectados se analizaron dentro de un software al cual se exportaron los datos con el debido formato que este software admite, la herramienta que se utilizó fue Rapidminer mediante operador Decisión Tree para predecir el rendimiento académico aunque también pudimos hacer un análisis con la extensión de weka con el algoritmo W-J48 que nos permite también predecir el rendimiento académico. Se tuvieron que analizar uno o más atributos para disminuir el margen de error.

### 11.4. Factores del rendimiento académico

Factores	Indicadores
Demográficos o independientes	Cuál es su sexo Edad Estado civil Quién financia sus estudios Tiene usted hijos

Factores	Indicadores
Académicos	<p>Como es su desempeño académico actual</p> <p>Usted ha repetido el ciclo académico</p> <p>Causa por la cual ha repetido el ciclo académico</p> <p>En cuantas asignaturas usted encuentra bajo en notas</p> <p>En su casa usted dispone de un lugar adecuado para estudiar</p>

Factores	Indicadores
Actitudinales	<p>En cuanto a la carrera que usted ha escogido como se siente?</p> <p>Tiene problemas de atención y concentración en las clases?</p> <p>Pone empeño al estudio para sacar buenas notas y superarse cada día?</p> <p>Cuanto tiempo le dedica usted al estudio fuera del horario normal de clases?</p> <p>Entrega a tiempo sus trabajos?</p> <p>Se siente satisfecho por la enseñanza recibida?</p>

**Variable independiente:**

- **Promedio**

### 11.5. Software Rapidminer para minería de datos

El software Rapidminer es un programa informático para el análisis de minería de datos, el cual permite el desarrollo de procesos de análisis de datos mediante encadenamiento de operadores a través de un entorno gráfico, este software se utiliza para la investigación, educación, capacitación, creación rápida de prototipos y en aplicaciones empresariales. En una encuesta realizada por KDnuggets, un periódico de minería de datos, Rapidminer ocupó el segundo lugar en herramientas de analítica y minería de datos utilizadas en proyectos reales en 2009 y fue el primero en el año 2010.

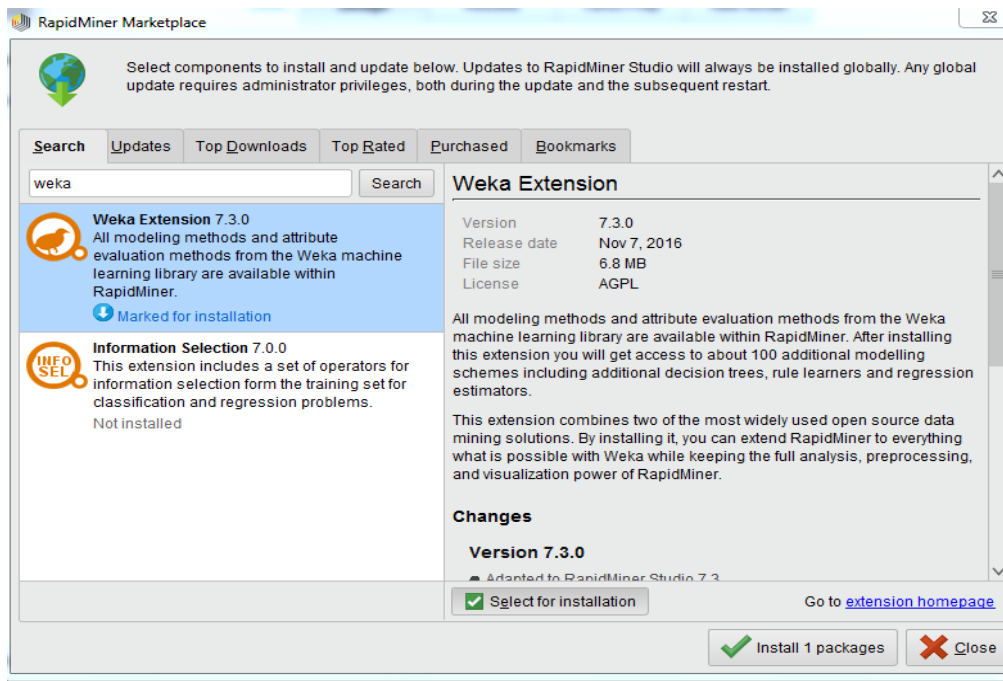
**Gráfico 21:** Rapidminer Studio



**Fuente:** El investigador

Rapidminer proporciona más de 500 operadores orientados al análisis de datos, incluyendo los necesarios para realizar operaciones de entrada y salida, pre procesamiento de datos y visualización, este software también permite utilizar los algoritmos incluidos en Weka.

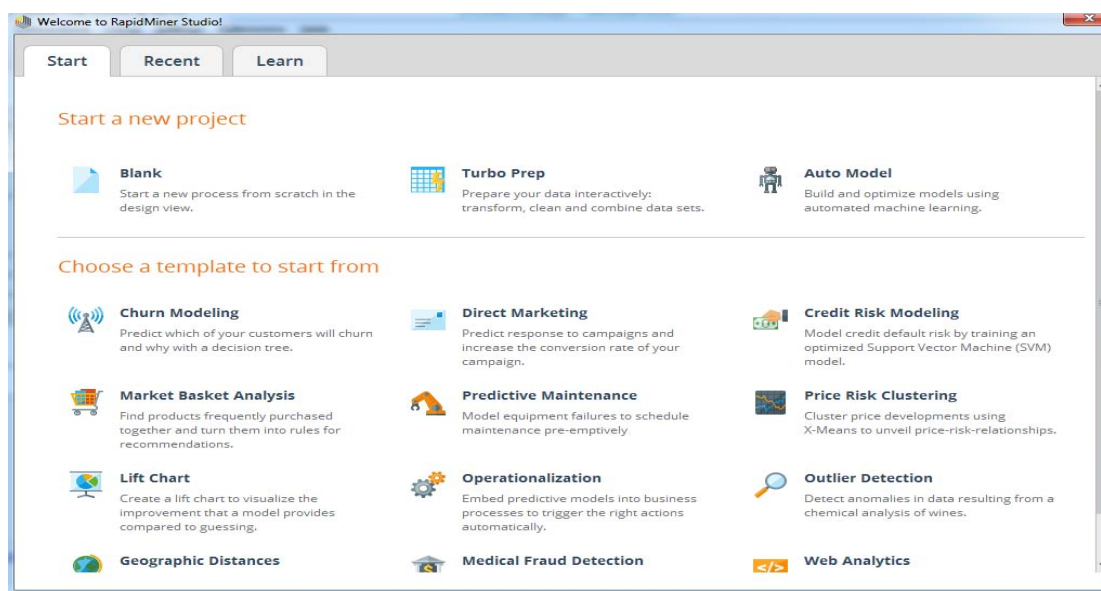
Gráfico 22: Rapidminer Weka extensión



Fuente: Software de rapidminer instalada

Esta herramienta de minería de datos nos muestra una serie de opciones con la cual podemos trabajar en diferentes tipos de casos a analizar.

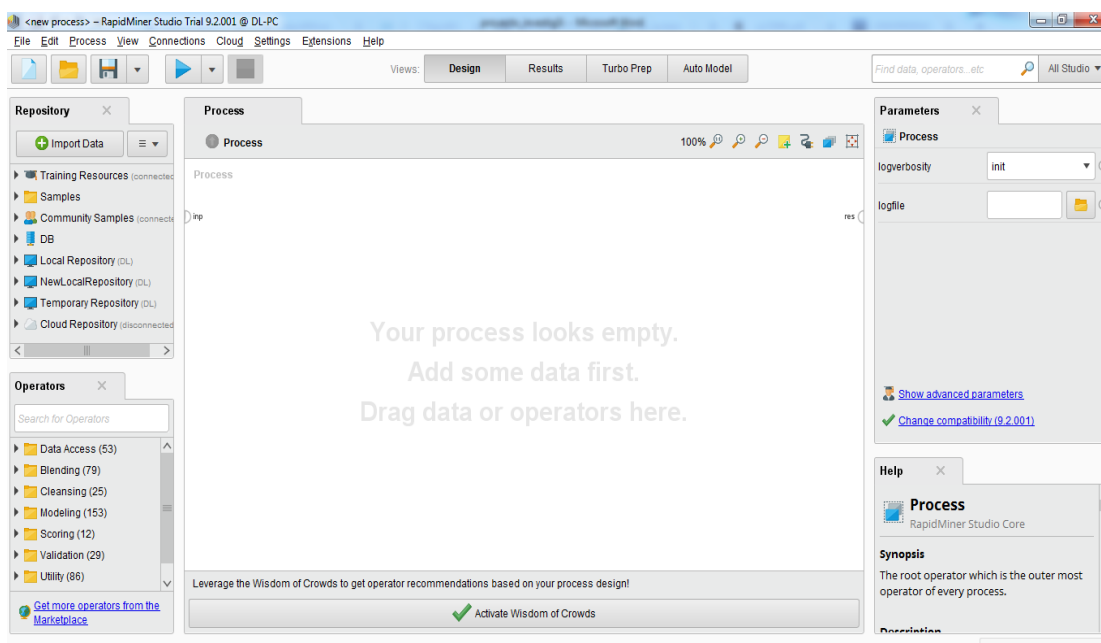
Gráfico 23: Rapidminer Interfaz principal



Fuente: Software de rapidminer instalada

Para el análisis del siguiente proyecto de minería de datos sobre el nivel de rendimiento académico hemos utilizado la opción de Blank el cual es comenzar un nuevo proceso.

**Gráfico 24:** Rapidminer: Área de trabajo



**Fuente:** Software de rapidminer instalada

### 11.5.1. Gestionar los datos en Rapidminer

Para la gestión correcta de los datos se obtuvo la información correspondiente mediante encuesta directamente a los estudiantes e información extraída de los datos académicos sobre los promedios anteriores de los estudiantes. Se pudo crear tablas basadas en los factores que influyen en el rendimiento académico con 236 datos o registros almacenados en las diferentes tablas creadas.

El software rapidminer con el cual se iba a realizar el análisis considero las instancias y los atributos almacenados en las diferentes tablas con los distintos factores académicos colectivamente con los promedios obtenidos de los estudiantes como datos de salida. Esta herramienta nos permite trabajar mediante hojas de cálculo Excel de Microsoft ya puede ser en el formato normal de Excel xlsx o también con el formato csv sea directamente con rapidminer o mediante extensiones de weka con el algoritmo W-J48 que predice arboles de decisión.

En algunos casos se tuvo que eliminar atributos o campos vacíos que no permitían predecir el rendimiento académico de los estudiantes ya que podrían arrojar algún tipo de error al momento de analizar los datos.

**Medida de concordancia:** En cualquier caso de estudio lo primordial es la fiabilidad de los procedimientos empleados, en consecuencia, un objetivo de los estudios de fiabilidad debe consistir en estimar el grado de dicha variabilidad.

**Tabla 2:** Valoración índice Kappa

Valoración del Índice Kappa	
Valor de k	Fuerza de la concordancia
< 0.20	Pobre
0.21 – 0.40	Débil
0.41 – 0.60	Moderada
0.61 – 0.80	Buena
0.81 – 1.00	Muy buena

Fuente: [www.fisterra.com](http://www.fisterra.com)

### 11.5.2. Analizando los datos recolectados en Rapidminer

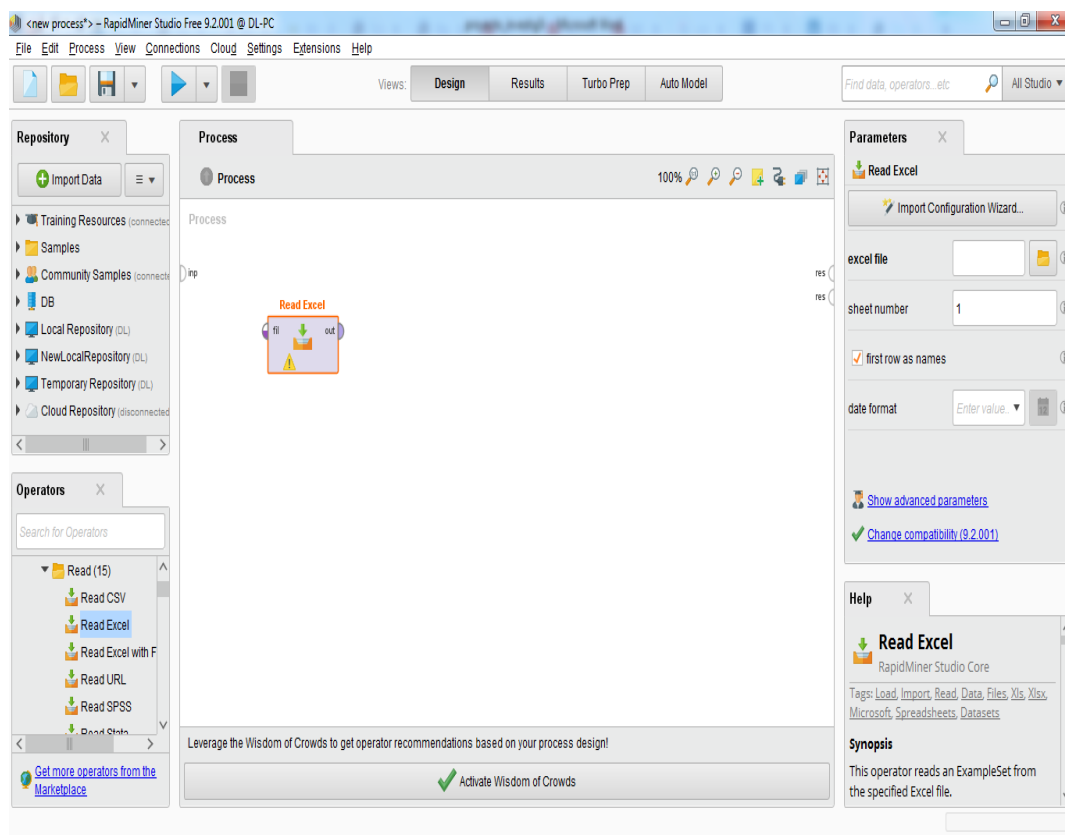
Para el análisis de los datos recolectados se utilizó trabajar con datos almacenados en Excel con la plantilla xlsx, básicamente existen dos formas de importar los datos:

- 1) Mediante una importación directa al proyecto actual.
- 2) Mediante la inclusión de los datos al repositorio.

A continuación importaremos los datos a través de la pestaña Operators Import =>Data Access hacer clic en Files => Read dentro de esta carpeta encontramos la opción de Read Excel el cual lo arrastramos hacia el interior del área=> Main Process. Una vez aparecido este

operador Read Excel en el área del diseño del proceso procederemos hacer clic para llamar al archivo correspondiente para ser analizado más adelante, mediante la pestaña parameters exploramos los archivos que están guardados en el computador.

**Gráfico 25:** Operador read excel



**Fuente:** Elaboración propia mediante software instalado

### 11.5.3. Entrada de datos y muestra de resultados

Importaremos el archivo para realizar el análisis de datos: en este caso analizaremos los atributos que identifican al estudiante.



**Tabla 3:** Variables o atributos de identificación del estudiante

Import Data - Format your columns.

**Format your columns.**

Replace errors with missing values ⓘ

	Cuál es su sexo? * <i>polynomial</i>	¿Edad? * <i>integer</i>	Estado civil? * <i>polynomial</i>	Quién financia ... * <i>polynomial</i>	Tiene usted hij... * <i>polynomial</i>
1	Femenino	19	Soltero(a)	trabajo	no
2	Femenino	18	Soltero(a)	Apoyo de mis Padres	no
3	Femenino	19	Soltero(a)	Apoyo de mis Padres	no
4	Femenino	18	Soltero(a)	Apoyo de mis Padres	no
5	Femenino	20	Soltero(a)	Apoyo de mis Padres	no
6	Femenino	19	Soltero(a)	Apoyo de mis Padres	no
7	Femenino	20	Soltero(a)	Apoyo de mis Padres	no
8	Femenino	20	Soltero(a)	Apoyo de mis Padres	no
9	Femenino	19	Soltero(a)	Apoyo de mis Padres	no
10	Femenino	18	Soltero(a)	trabajo	no
11	Femenino	19	Soltero(a)	Apoyo de mis Padres	no
12	Femenino	19	Soltero(a)	trabajo	no
13	Femenino	18	Soltero(a)	Apoyo de mis Padres	no

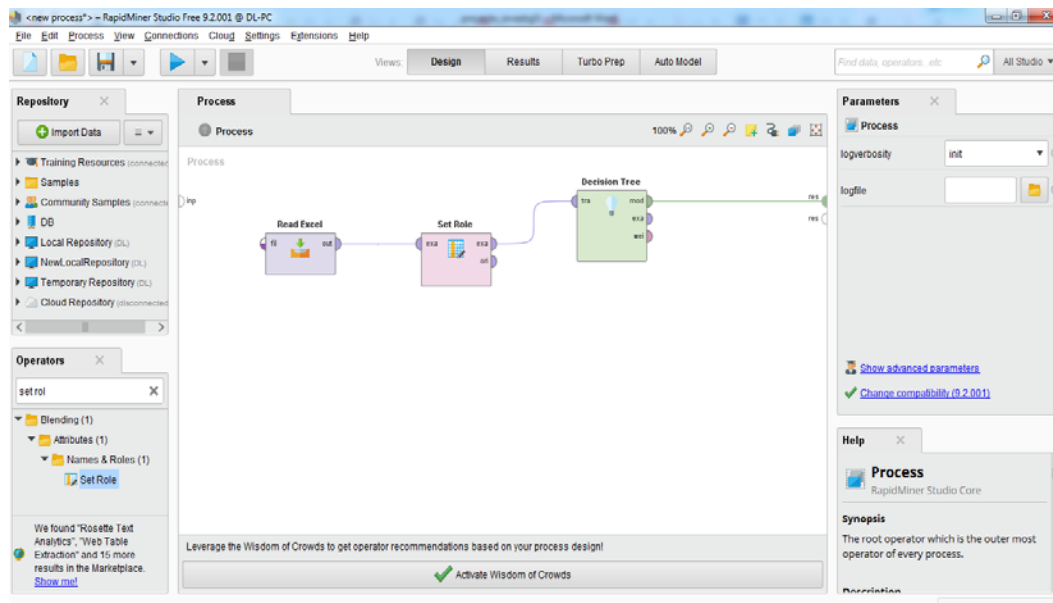
no problems.

← Previous Finish Cancel

**Fuente:** Elaboración propia mediante software instalado

Una vez importados los datos al software se procede a escoger el método de predicción Decisión Tree, acompañado del operador Set Role este operador nos permite escoger la columna a la cual vamos analizar para después mostrar el resultado en un árbol de decisión.

**Gráfico 26:** Operadores (read excel, set role y decisión tree)



**Fuente:** Elaboración propia mediante software instalado

### 11.5.3.1. Resultados obtenidos en arboles de decisión de las variables de identificación

#### Decisión Tree

**Criterion:** gain\_ratio

**Instances:** 236

**Atributes:** 6

Cuál es su sexo?

Edad?

Estado civil?

Quién financia sus estudios?

Tiene usted hijos?

Promedio

#### Resultados Decisión Tree:

=====

Estado civil? = Soltero(a)

| Tiene usted hijos? = no

| | promedio > 7

| | | promedio > 8: Apoyo de mis Padres {trabajo=0, Apoyo de mis Padres=4, otras fuentes=0}

| | | promedio ≤ 8

| | | |Cuál es su sexo? = Femenino

| | | | | ¿Edad? > 19: Apoyo de mis Padres {trabajo=0, Apoyo de mis Padres=8, otras fuentes=0}

| | | | | ¿Edad? ≤ 19

| | | | | | ¿Edad? > 18: Apoyo de mis Padres {trabajo=2, Apoyo de mis Padres=20, otras fuentes=0}

| | | | | | ¿Edad? ≤ 18: Apoyo de mis Padres {trabajo=1, Apoyo de mis Padres=4, otras fuentes=0}

| | | |Cuál es su sexo? = Masculino

| | | | | ¿Edad? > 19: Apoyo de mis Padres {trabajo=1, Apoyo de mis Padres=5, otras fuentes=0}

| | | | | ¿Edad? ≤ 19

| | | | | | ¿Edad? > 18: Apoyo de mis Padres {trabajo=11, Apoyo de mis Padres=70, otras fuentes=8}

| | | | | | ¿Edad? ≤ 18: Apoyo de mis Padres {trabajo=1, Apoyo de mis Padres=5, otras fuentes=0}

| | promedio ≤ 7: Apoyo de mis Padres {trabajo=4, Apoyo de mis Padres=81, otras fuentes=0}

| Tiene usted hijos? = si: Apoyo de mis Padres {trabajo=2, Apoyo de mis Padres=3, otras fuentes=0}

Estado civil? = Unión libre

|Cuál es su sexo? = Femenino: otras fuentes {trabajo=0, Apoyo de mis Padres=0, otras fuentes=1}

|Cuál es su sexo? = Masculino

| | Tiene usted hijos? = no

| | | promedio > 7

| | | | promedio > 8: Apoyo de mis Padres {trabajo=0, Apoyo de mis Padres=1, otras fuentes=0}

| | | | promedio ≤ 8: Apoyo de mis Padres {trabajo=0, Apoyo de mis Padres=1, otras fuentes=1}

| | | promedio ≤ 7: Apoyo de mis Padres {trabajo=0, Apoyo de mis Padres=1, otras fuentes=0}

| | Tiene usted hijos? = si: Apoyo de mis Padres {trabajo=0, Apoyo de mis Padres=1, otras fuentes=0}

236 items in subtree

Ratio of total: 97.43%

Kappa statistic: 0.494

**Gráfico 27:** Árbol gráfico N° 1 variables o atributos de identificación



**Fuente:** Elaboración propia mediante software instalado

### 11.5.3.2. Resultados obtenidos en arboles de decisión variables académicas vs promedio de estudiantes

Decisión Tree

Criterion: gain\_ratio

Instances: 236

Attributes: 6

Como es su desempeño académico actual?

Usted ha repetido el ciclo académico?

Causa por la cual ha repetido el ciclo académico?

En cuantas asignaturas usted encuentra bajo en notas?

En su casa usted dispone de un lugar adecuado para estudiar?

Promedio

#### Resultados Decisión Tree:

=====

Como es su desempeño académico actual? = bueno

| En cuantas asignaturas usted encuentra bajo en notas? > 2: 7 {count=34}

| En cuantas asignaturas usted encuentra bajo en notas? ≤ 2

| | Usted ha repetido el ciclo académico? = no

| | | Causa por la cual ha repetido el ciclo académico? = no contesto

| | | | En su casa usted dispone de un lugar adecuado para estudiar? = si: 7 {count=47}

Como es su desempeño académico actual? = muy bueno

| En cuantas asignaturas usted encuentra bajo en notas? > 2: 7 {count=15}

| En cuantas asignaturas usted encuentra bajo en notas? ≤ 2

| | En cuantas asignaturas usted encuentra bajo en notas? > 0

| | | En cuantas asignaturas usted encuentra bajo en notas? > 1: 8 {count=96}

| | | En cuantas asignaturas usted encuentra bajo en notas? ≤ 1

| | | | En su casa usted dispone de un lugar adecuado para estudiar? = no: 8 {count=30}

| | | | En su casa usted dispone de un lugar adecuado para estudiar? = si: 9 {count=1}

| | En cuantas asignaturas usted encuentra bajo en notas?  $\leq 0$ : 9 {count=4}

Como es su desempeño académico actual? = regular

| Usted ha repetido el ciclo académico? = no

| | Causa por la cual ha repetido el ciclo académico? = no contesto

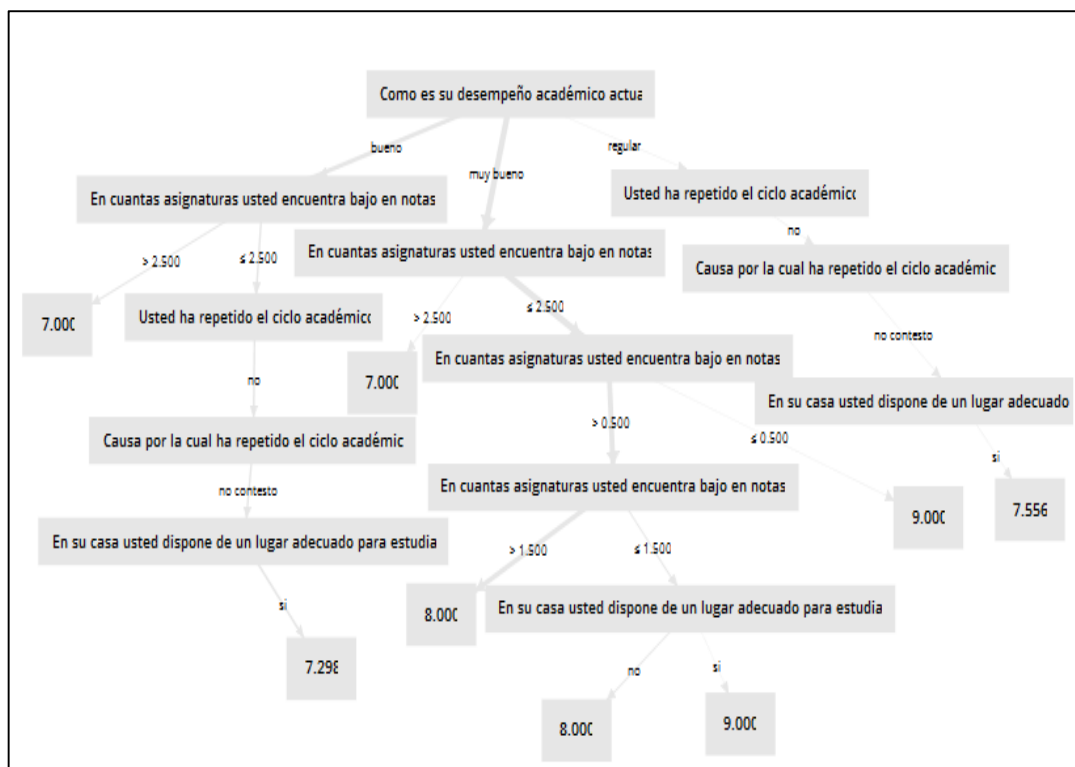
| | | En su casa usted dispone de un lugar adecuado para estudiar? = si: 7 {count=9}

236 items in subtree

Ratio of total: 61.86%

Kappa statistic: 0.66

**Gráfico 28:** Árbol gráfico N° 2 variables o atributos académicos



**Fuente:** Elaboración propia mediante software instalado

### Resultados Decisión Tree:

=====

promedio > 7

| En cuantas asignaturas usted encuentra bajo en notas? > 1: muy bueno {muy bueno=96, bueno=14, regular=5}

| En cuantas asignaturas usted encuentra bajo en notas? ≤ 1: muy bueno {muy bueno=35, bueno=0, regular=0}

promedio ≤ 7

| En cuantas asignaturas usted encuentra bajo en notas? > 2: bueno {muy bueno=15, bueno=34, regular=0}

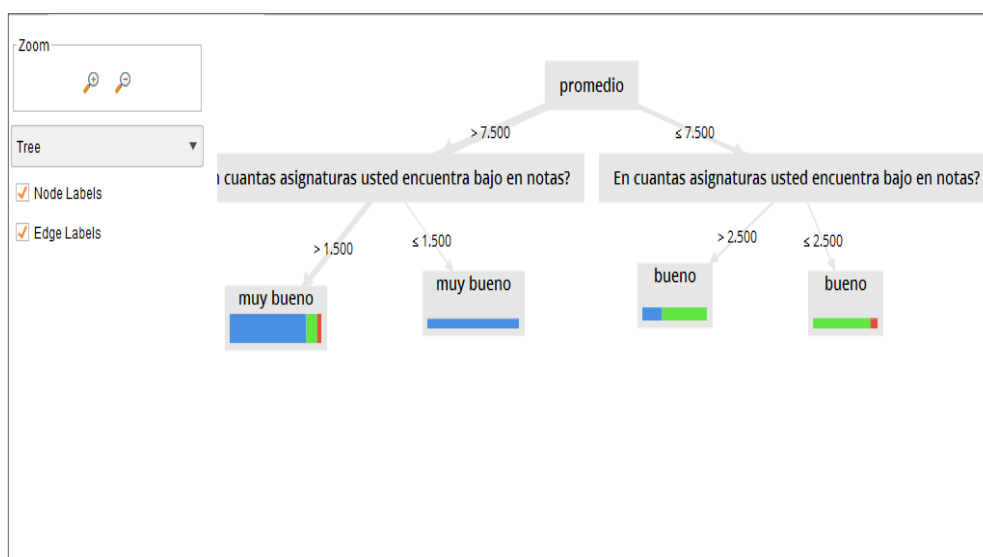
| En cuantas asignaturas usted encuentra bajo en notas? ≤ 2: bueno {muy bueno=0, bueno=33, regular=4}

236 items in subtree

Ratio of total: 89.73%

Kappa statistic: 0.66

**Gráfico 29:** Árbol gráfico N° 3 variables o atributos académicos



**Fuente:** Elaboración propia mediante software instalado

### 11.5.3.3. Resultados obtenidos en arboles de decisión variables actitudinales vs promedio de estudiantes

En cuanto a la carrera que usted ha escogido como se siente?

Tiene problemas de atención y concentración en las clases?

Pone empeño al estudio para sacar buenas notas y superarse cada día?

Cuanto tiempo le dedica usted al estudio fuera del horario normal de clases?

Entrega a tiempo sus trabajos?

Se siente satisfecho por la enseñanza recibida?

Promedio

Decisión Tree

Criterion: gain\_ratio

Instances: 236

Attributes: 7

#### Resultados Decisión Tree:

=====

En cuanto a la carrera que usted ha escogido como se siente? = Satisfecho: satisfecho  
{satisfecho=59, poco satisfecho=0}

En cuanto a la carrera que usted ha escogido como se siente? = muy satisfecho

| promedio > 7

| | Cuanto tiempo le dedica usted al estudio fuera del horario normal de clases? = mucho

| | | promedio > 8: satisfecho {satisfecho=5, poco satisfecho=0}

| | | promedio ≤ 8: satisfecho {satisfecho=125, poco satisfecho=7}

| | Cuanto tiempo le dedica usted al estudio fuera del horario normal de clases? = muy poco:  
satisfecho {satisfecho=13, poco satisfecho=0}

| promedio ≤ 7

| | Tiene problemas de atención y concentración en las clases? = a veces



| | | Entrega a tiempo sus trabajos? = a veces: satisfecho {satisfecho=5, poco satisfecho=0}

| | | Entrega a tiempo sus trabajos? = siempre: satisfecho {satisfecho=9, poco satisfecho=8}

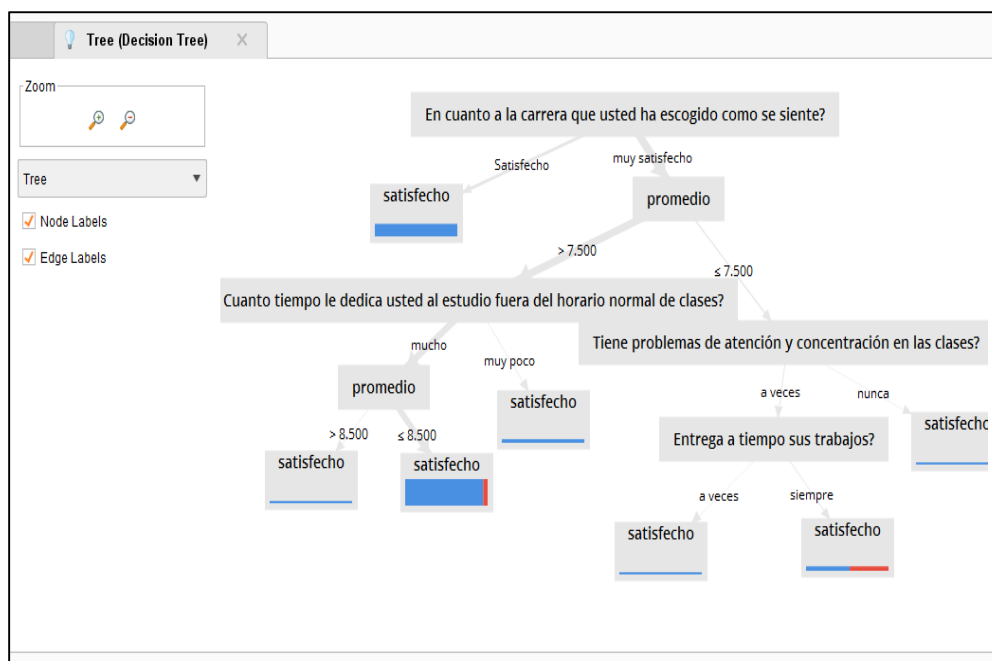
| | Tiene problemas de atención y concentración en las clases? = nunca: satisfecho {satisfecho=5, poco satisfecho=0}

236 items in subtree

Ratio of total: 92.09%

Kappa statistic: 0.80

**Gráfico 30:** Árbol gráfico N° 4 variables o atributos actitudinales



**Fuente:** Elaboración propia mediante software instalado

**Atributo:** Tiene problemas de atención y concentración en las clases?

Decisión Tree

Criterion: gain\_ratio

Instances: 236

Atributes: 3

**Resultados Decisión Tree:**

Tiene problemas de atención y concentración en las clases? = a veces: siempre {siempre=215, a veces=0}

Tiene problemas de atención y concentración en las clases? = nunca

| promedio > 8: siempre {siempre=5, a veces=0}

| promedio ≤ 8

| | En cuanto a la carrera que usted ha escogido como se siente? = Satisfecho: a veces {siempre=2, a veces=9}

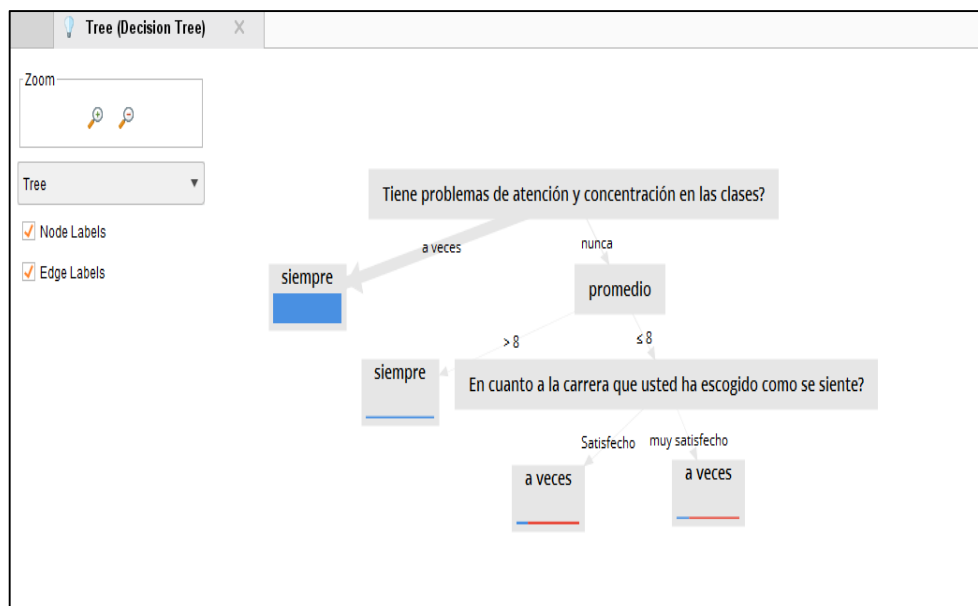
| | En cuanto a la carrera que usted ha escogido como se siente? = muy satisfecho: a veces {siempre=1, a veces=4}

236 items in subtree

Ratio of total: 92.09%

Kappa statistic: 0.80

**Gráfico 31:** Árbol gráfico N° 5 variables actitudinales vs promedio



**Fuente:** Elaboración propia mediante software instalado

## 11.6. Análisis de los árboles de decisión

A continuación es dar a conocer los resultados que se obtuvieron mediante los árboles de decisión.

### **11.6.1. Análisis de resultados de las variables o atributos de identificación con el promedio del estudiante**

El análisis generado por el algoritmo en el árbol gráfico (Ver Gráfico N° 27 Árbol N° 1), se puede decir que la variable estado civil es precisa, por lo siguiente tenemos los siguientes resultados:

Aquellos estudiantes de estado civil soltero(a) y de los cuales no tienen hijos de sexo femenino están entre 19 y 20 años los cuales reciben el apoyo de sus padres se predice que pueden rendir académicamente un promedio que fluctúa entre 7 a 8.

Aquellos estudiantes de estado civil soltero(a) que no poseen hijos de sexo masculino cuya edad es mayor a 19 años los cuales reciben el mayor apoyo de sus padres se predice que pueden rendir académicamente un promedio que fluctúa entre 7 a 8.

Aquellos estudiantes de estado civil unión libre que no poseen hijos de sexo masculino que reciben el apoyo de sus padres pueden rendir académicamente un promedio menor o igual a 7, mientras que aquellos estudiantes que no reciben apoyo de sus padres sino de otras fuentes están en un promedio de 7 a 8.

Aquellos estudiantes de estado civil soltero(a) de sexo femenino cuya edad es menor a 19 años que no tienen hijos pero que reciben apoyo de sus padres como del trabajo pueden rendir académicamente un promedio de 8.

Aquellos estudiantes de estado civil soltero(a) de sexo masculino que no tienen hijos cuya edad es menor de 19 años cuyo financiamiento es de sus padres como también de sus trabajos pueden rendir académicamente un promedio de 7.

### **11.6.2. Análisis de resultados de las variables o atributos académicos con el promedio del estudiante**

El análisis generado por el algoritmo en el árbol gráfico (Ver Gráfico N° 28 Árbol N° 2) se puede decir que la variable desempeño académico es la más precisa, por lo siguiente tenemos los siguientes resultados:

Aquellos estudiantes que tienen el desempeño académico bueno en las cuales están mal en más de dos asignaturas fluctúan un promedio de 7, mientras que aquellos estudiantes que están mal en dos materias o menos que no han repetido el ciclo académico y que cuentan con un lugar adecuado para estudiar fluctúan un promedio de igual a 7.

Aquellos estudiantes un desempeño académico regular que no han repetido el ciclo académico y disponen de un lugar adecuado para estudiar fluctúan un promedio de 7.

Aquellos estudiantes con el desempeño académico muy bueno y que se encuentran bajos en una o ninguna materia los cuales poseen un lugar adecuado para estudiar fluctúan un promedio entre 8 a 9.

El análisis generado por el algoritmo en el árbol gráfico (Ver Gráfico N° 29 Árbol N° 3) tenemos los siguientes resultados:

Aquellos estudiantes que están mal en una materia o ninguna poseen un rendimiento académico muy bueno fluctúan con un promedio mayor a 7, mientras que aquellos estudiantes que están mal en más de una poseen un rendimiento académico entre regular, bueno y muy bueno con un promedio mayor a 7.

Aquellos estudiantes que están mal en más de dos materias poseen un rendimiento académico entre bueno y muy bueno con un promedio menor o igual a 7, mientras que aquellos estudiantes que están mal en dos materias poseen un rendimiento académico entre bueno y regular con un promedio menor o igual a 7.

### **11.6.3. Análisis de resultados de las variables o atributos actitudinales con el promedio del estudiante**

El análisis generado por el algoritmo en el árbol gráfico (Ver Gráfico N° 30 Árbol N° 4) se puede decir que la variable en cuanto a la carrera que usted ha escogido como se siente es la más precisa, por lo siguiente tenemos los siguientes resultados:

Aquellos estudiantes que se sienten muy satisfechos por la carrera que han escogido y dedican mucho al estudio fuera del horario normal de clases se puede decir que su promedio es mayor a 8, mientras que aquellos estudiantes que están entre satisfechos y poco satisfechos poseen un promedio menor o igual a 8.

Aquellos estudiantes que se sienten muy satisfechos por la carrera que han escogido pero tienen a veces problemas de atención y concentración en las clases los cuales a veces entregan sus trabajos a tiempo poseen un promedio de menor o igual a 7.

Aquellos estudiantes que se sienten muy satisfechos por la carrera que han escogido pero tienen a veces problemas de atención y concentración en las clases los cuales siempre

entregan sus trabajos a tiempo aunque se podría decir que también se sienten poco satisfechos por la carrera que han escogido poseen un promedio de menor o igual a 7.

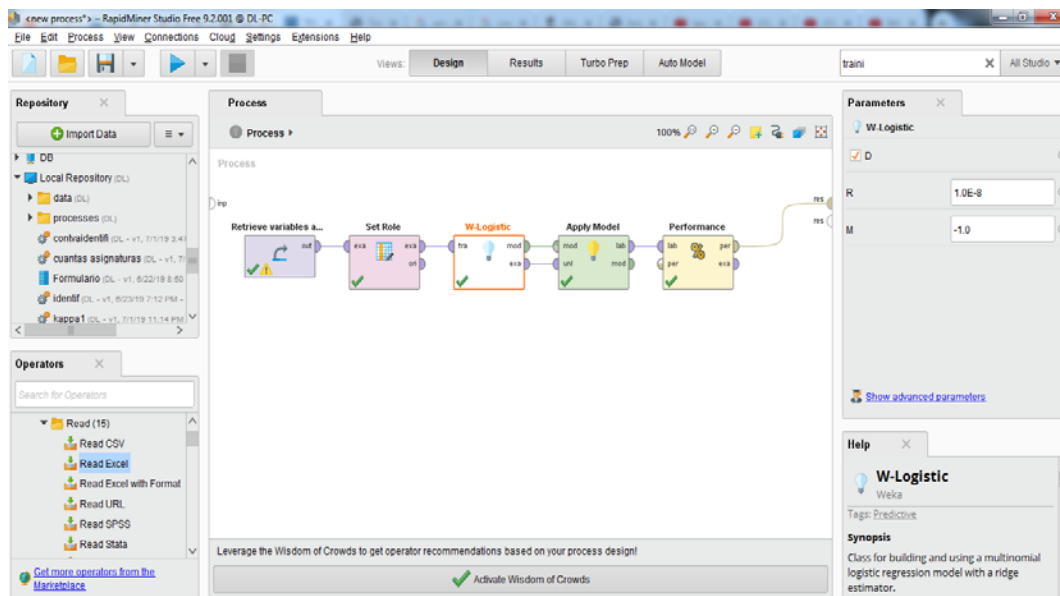
Aquellos estudiantes que se sienten muy satisfechos por la carrera que han escogido pero no tienen problemas de atención y concentración en las clases poseen un promedio de menor o igual a 7.

En el árbol N° 5 del gráfico N° 15 de aquellos estudiantes que no tienen problemas de atención y concentración en las horas de clases y se sienten muy satisfechos fluctúan un promedio entre 8 y 9.

### 11.7. Matriz de confusión

Se construye un modelo a través del operador de validación cruzada nominal, con un grupo de entrenamiento y otro de testeo, para este modelo empleamos la extensión weka para Rapidminer, ya que la propia herramienta no ofrece por sí misma un algoritmo de regresión logística (W-Logistic).

**Gráfico 32:** Matriz de confusión



**Fuente:** Elaboración propia mediante software instalado

La capacidad predictiva del modelo a través de lo que se denomina **Matriz de Confusión**.

PerformanceVector:

accuracy: 83.90%

ConfusionMatrix:

True:	muy bueno	bueno	regular
muy bueno:	131	14	5
bueno:	15	67	4
regular:	0	0	0

kappa: 0.666

ConfusionMatrix:

True:	muy bueno	bueno	regular
muy bueno:	131	14	5
bueno:	15	67	4
regular:	0	0	0

**Tabla 4:** Matriz de confusión accuracy

accuracy: 83.90%				
	true muy bueno	true bueno	true regular	class precision
pred. muy bueno	131	14	5	87.33%
pred. bueno	15	67	4	77.91%
pred. regular	0	0	0	0.00%
class recall	89.73%	82.72%	0.00%	

**Fuente:** Elaboración propia mediante software instalado

**Tabla 5:** Matriz de confusión kappa

kappa: 0.666				
	true muy bueno	true bueno	true regular	class precision
pred. muy bueno	131	14	5	87.33%
pred. bueno	15	67	4	77.91%
pred. regular	0	0	0	0.00%
class recall	89.73%	82.72%	0.00%	

**Fuente:** Elaboración propia mediante software instalado

A continuación la información contenida en la diagonal que se encuentra de color más oscuro representan aquellos elementos que se han clasificado correctamente.

Si se analiza los resultados verticalmente, el total establecido como muy bueno 131 fueron establecidos como tal, 67 como bueno y no tenemos nada como regular. El porcentaje que se obtuvo entre los correctamente clasificados y los predeterminados o también conocidos como falsos negativos se denominan **cobertura class recall**.

$$Cobertura_{clase}(\%) = \frac{Acierto_{clase}}{Falsosnegativos} \times 100$$

Resultados de las coberturas de clase son los siguientes:

$$Cobertura_{muybueno} = \frac{131}{146} \times 100 = 89.73\%$$

$$Cobertura_{bueno} = \frac{67}{81} \times 100 = 82.72\%$$

$$Cobertura_{muybueno} = \frac{0}{9} \times 100 = 0.00\%$$

También se puede horizontalmente, en la cual tenemos 131 muy buenos que son verdaderos, 14 verdaderos como bueno y 5 verdaderos como regular. El porcentaje calculado a partir de la relación entre los correctos clasificados y los falsos positivos se denomina precisión **class precisión**.

$$Cobertura_{clase}(\%) = \frac{Acierto_{clase}}{Falsospositivos} \times 100$$

Resultados de la precisión de clase son los siguientes:

$$Cobertura_{muybueno} = \frac{131}{150} \times 100 = 87.33\%$$

$$Cobertura_{bueno} = \frac{67}{86} \times 100 = 77.91\%$$

$$Cobertura_{muybueno} = \frac{0}{0} \times 100 = 0.00\%$$

Realizado esta matriz de confusión se puede observar en aquellos resultados de muy bueno donde hay un mayor porcentaje de efectividad, esto se debe uno o varios elementos del conjunto de los atributos o variables independientes permiten discernir en una forma clara el tipo de clasificación, lo cual se puede medir el nivel de rendimiento académico de que exista un bajo nivel de rendimiento académico.

## **12. IMPACTOS**

### **12.1. Impacto técnico**

Se pudo alcanzar la predicción del rendimiento académico a un 67.33% a través del modelo matriz de confusión, esto genera grandes expectativas en cuanto al nivel de rendimiento académico ya que no existe bajo rendimiento académico, se pudo analizar el rendimiento mediante el modelo de regresión logística.

### **12.2. Impacto social**

Con el análisis a través de minería de datos mediante arboles de decisión realizada a los estudiantes para medir su rendimiento académico ya que mediante la ejecución del software utilizado sobre los factores que puedan afectar al estudiante en su desempeño.

### **12.3. Impacto económico**

El análisis de minería de datos a través de rapidminer evita el pago de ciertas licencias propietarias, el tiempo de ejecución se lo hizo posible en 5 horas diarias de lunes a miércoles generando un costo de 7 dólares por mes durante 4 meses, los cuales se reflejan en la Tabla N° 6.



### 13. PRESUPUESTO PARA LA ELABORACIÓN DEL PROYECTO

#### 13.1. Gastos directos

**Tabla 6:** Gastos Directos.

<b>Descripción:</b>	<b>Cantidad:</b>	<b>Valor Unitario (\$):</b>	<b>Valor Total (\$):</b>
Hojas de papel Bond	1 Resmas	4	4,00
Impresiones a color	50	0,10	5,00
Impresiones a B/N	80	0,05	4,00
Copias	470	0,02	9,40
Esferos	2	0,35	0,70
Anillados	3	7,00	21,00
Conexión Internet	4 Meses	7,00	28,00
USB/Flash	1	8,50	8,50
Computador	1	Uso diario	100,00
<b>Total Gastos Directos:</b>			<b>\$ 180,60</b>

**Fuente:** El investigador

#### 13.2. Gastos indirectos

**Tabla 7:** Gatos Indirectos.

<b>Descripción:</b>	<b>Valor:</b>
Movilidad	\$ 90,00
Alimentación	\$ 50,00
<b>Total:</b>	<b>\$140,00</b>

**Fuente:** El investigador

### 13.3. Gastos totales del proyecto

**Tabla 8:** Presupuesto Total del Proyecto.

<b>Gastos:</b>	<b>Total:</b>
Gastos Directos	\$ 180,60
Gastos Indirectos	\$ 140,00
10% de Imprevistos	\$ 20,00
<b>Total Presupuesto:</b>	<b>\$ 340,60</b>

**Fuente:** El investigador

## 14. CONCLUSIONES Y RECOMENDACIONES

### 14.1. Conclusiones

- Se ha analizado los resultados y discernido entre atributos influyentes y no influyentes, se han filtrado instancias de los datos obtenidos y se han creado nuevos modelos de minería de datos para nuestro conjunto de datos de muestra.
- Las variables académicas fueron las que resultaron con más ganancia de información por lo que con estas variables la predicción tiene mayor exactitud y definen el rendimiento académico del estudiante.
- El desarrollo de este trabajo permitió validar que la técnica de árboles de decisión aplicados a los factores demográficos, socioculturales, académicos, institucionales predicen el rendimiento académico de los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas.
- Se encuentra mejores resultados sobre todo en el nivel de concordancia cuando la cantidad de intervalos en las salidas del árbol de decisión son menores. Se pudo obtener mejores resultados cuando se utilizó el muy bueno, bueno y regular a cambio de colocar varios intervalos de notas para su rendimiento académico

## 14.2. Recomendaciones

- Se recomienda que los datos del estudiante estén bien organizados para poder predecir mejor el rendimiento académico de los estudiantes, razón por la cual la minería de datos es una herramienta potente y aplicable a un sin número de proyectos y finalidades que estas requieran.
- Se recomienda profundizar más las bondades que nos da las herramientas de rapidminer ya que es una herramienta potente para realizar minería de datos.
- Para investigaciones futuras, se recomienda que al utilizar la herramienta W-Logistic de la extensión de weka se definan bien los operadores que se vincularan a esta herramienta para no tener errores más adelante.

## 15. BIBLIOGRAFÍA

- Azoumana, K. (2013). *Analysis of Student desertion at Universidad Simón Bolívar, Faculty of Systems Engineering, with data mining techniques*. 6(10), 41–51.
- Banda, H. (2014). *Inteligencia Artificial: Principios y Aplicaciones*. (April).
- Barbosa, H. (2010). *Generador de pruebas objetivas adaptadas a las preferencias de presentación de los usuarios*. Universidad de Salamanca.
- Bedoya, O. M., Trujillo, L., Echeverry, M., Bedoya, O. M., Trujillo, M. L., & Eduardo, C. (2016). Minería de datos en egresados de la Universidad de Caldas. *Revista Virtual Universidad Católica Del Norte*, 49, 110–124. Retrieved from <http://revistavirtual.ucn.edu.co/index.php/RevistaUCN/article/view/800/1320>
- Beltrán Martínez, B. (2016). *Notas De Minería De Datos*. 67. Retrieved from <http://bbeltran.cs.buap.mx/NotasMD.pdf>
- Brahim, E. F., Mohammed, K. I. y Samir, B. (2010). A formative assessment model within the competency-based-approach for an individualized e-learning path *World Academy of Science, Engineering and Technology*, 40, 208-212.

- Cañas, A. J., & Novak, J. D. (2010). *Cmc 2010*.
- Coria, S. R. (2016). *Introducción a la Minería de Datos y el Data Warehousing Introducción a la Minería de Datos y el Data Warehousing*. (March).
- Erandi, R., Martínez, B., Ramírez, N. C., Gabriel, H., Mesa, A., Suárez, I. R., ... Morales, S. L. B. (2009). *Árboles de decisión como herramienta en el diagnóstico médico*.
- García, H. Á., & Educacional, I. (n.d.). *Minería de Datos en la Educación*.
- Giovanny, I., Arias, B., Santiago, H., Chávez, T., Colón, I., Ballesteros, A., ... Espinoza, S. (2017). Estilos de vida y su relación con las calificaciones escolares: estudio en Ecuador. *Revista Cubana de Investigaciones Biomédicas*, 36(4), 1–14. Retrieved from <http://scielo.sld.cu>
- Icarte Ahumada, G. A. (2016). Aplicaciones de inteligencia artificial en procesos de cadenas de suministros: una revisión sistemática. *Ingeniare. Revista Chilena de Ingeniería*, 24(4), 663–679. <https://doi.org/10.4067/s0718-33052016000400011>
- Lopera, M. (2018). *Los árboles de decisión como herramienta para el análisis de riesgos de los proyectos*.
- Lugo Reyes, S. O. (2014). Artificial intelligence to assist clinical diagnosis in medicine [Inteligencia artificial para asistir el diagnóstico clínico en medicina]. *Revista Alergia Mexico*, 61(2), 110–120. Retrieved from <http://revistaalergia.mx/ojs/index.php/ram/article/view/33/46%5Cnhttp://www.scopus.com/inward/record.url?eid=2-s2.0-84905817987&partnerID=40&md5=0481b6efab57e92eab71b97be71e3b4d>
- Moncada Mora, L. F. (2014). Determinantes Inmediatos Del Rendimiento Académico En Los Nuevos Estudiantes Matriculados En El Sistema De Educación Superior a Distancia Del Ecuador: Caso Universidad Técnica Particular De Loja. *RIED. Revista Iberoamericana de Educación a Distancia*, 14(2), 77–95. <https://doi.org/10.5944/ried.2.14.790>
- Mora, L. (2014). Como Factor Determinante Del Abandono De Corto Plazo. Un Análisis En El Sistema De Educación Superior a Distancia Del. ... . *Revista Iberoamericana de Educación a Distancia*, 173–196. Retrieved from <http://revistas.uned.es/index.php/ried/article/view/12683>
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). *Minería de datos: Conceptos y tendencias*.

*Inteligencia Artificial*, 10(29), 11–18.

Rodríguez Suárez, Y., & Amador, A. D. (2009). *Herramientas de Minería de Datos Data Mining Tools*. 3(3), 73–80. Retrieved from

[https://rcci.uci.cu/?journal=rcci&page=article&op=viewFile&path\[\]=78&path\[\]=70](https://rcci.uci.cu/?journal=rcci&page=article&op=viewFile&path[]=78&path[]=70)

Sánchez, J., Ruiz, J. y Sánchez, E. (2011). Análisis comparativo de evaluación entre pares con la

del profesorado. Un caso práctico. *Docencia e Investigación*, 36 (21), 11-24.

Tobón, S. (2010). *Formación integral y competencias* (3.<sup>a</sup> ed.). Bogotá: ECOE Ediciones.- (2013a).

*La evaluación de las competencias en la educación básica* (2.<sup>a</sup> ed.). México: Santillana.

# ANEXOS

**Anexo1.-** Encuesta aplicada a los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas.



**Encuesta dirigida a los estudiantes de los primeros ciclos de la Facultad de Ciencias de la Ingeniería y Aplicadas.**

Las respuestas que usted coloque serán anónimas. Serán utilizadas únicamente con fines académicos. Se agradece de antemano su colaboración.

1) **¿Cuál es su sexo?**

Masculino  Femenino

2) **¿Cuál es su edad?** \_\_\_\_\_ años

3) **¿Cuál es su estado civil?**

Soltero(a)  Casado(a)  Unión libre  Separado(a)

4) **¿Quién financia sus estudios?**

Apoyo de mis Padres  Beca Universitaria  Trabajo  Otras fuentes

5) **¿Tiene usted hijos?**

Sí  No

6) **Cómo es su desempeño académico actual?**

Muy bueno  Bueno  Regular

7) **Usted ha repetido el ciclo académico?**

Sí  No

8) **Causa por la cual ha repetido el ciclo académico?**

Dificultad en las asignaturas  Calamidad doméstica

Por trabajar y no tener tiempo

9) **En cuántas asignaturas usted encuentra bajo en notas?**

Asignaturas.

10) **¿En su casa usted dispone de un lugar adecuado para estudiar?**

Sí  No

11) **¿En cuanto a la carrera que usted ha escogido como se siente?**

Satisfecho  Muy poco satisfecho  Insatisfecho

12) **Tiene problemas de atención y concentración en las clases?**

Siempre  A veces  Nunca

13) **Pone empeño al estudio para sacar buenas notas y superarse cada día?**

Siempre  A veces  Nunca

14) **Cuánto tiempo le dedica usted al estudio fuera del horario normal de clases?**

Mucho  Muy poco  Nada

15) **Entrega a tiempo sus trabajos?**

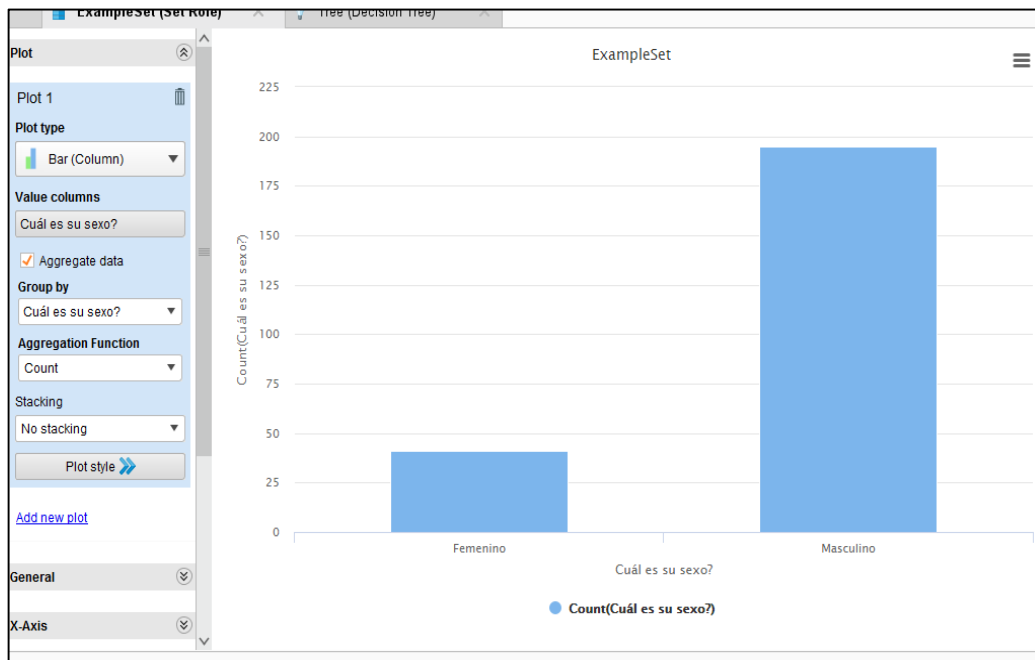
Siempre  A veces  Nunca

16) **Se siente satisfecho por la enseñanza recibida?**

Satisfecho  Poco satisfecho  Insatisfecho

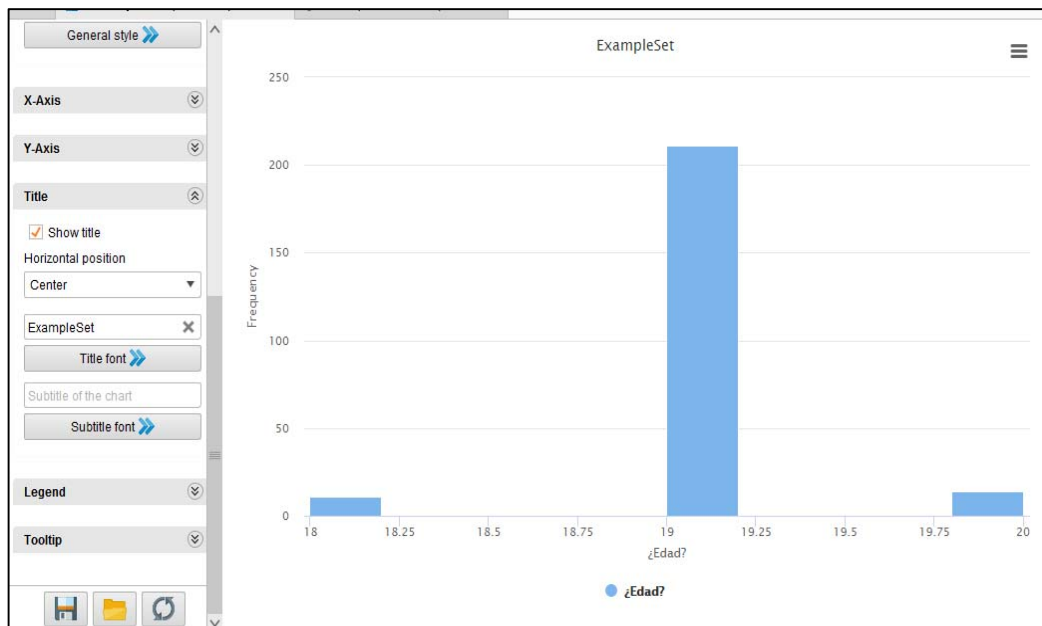


## Anexo2.- Muestra de resultados de rapidminer



Fuente: Elaboración propia mediante software instalado

## Anexo3.- Muestra de resultados de rapidminer



Fuente: Elaboración propia mediante software instalado