



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

TEMA:

Análisis de retención de estudiantes en la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG, durante los dos primeros años de la carrera, mediante técnica de Minería de Datos

AUTOR:

Zambrano Soriano, Pedro Eddison

**Trabajo de titulación previo a la obtención del título de
INGENIERO EN SISTEMAS COMPUTACIONALES**

TUTORA:

Ing. Zurita Fabre, Adela María Eugenia, Ph.D.

Guayaquil, Ecuador

12 de marzo del 2019



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

CERTIFICACIÓN

Certificamos que el presente trabajo de titulación **Análisis de retención de estudiantes en la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG, durante los dos primeros años de la carrera, mediante técnica de Minería de Datos**, fue realizado en su totalidad por **Zambrano Soriano, Pedro Eddison**, como requerimiento para la obtención del título de **Ingeniero en Sistemas Computacionales**.

TUTORA

Ing. Adela María Eugenia Zurita Fabre, Ph.D.

DIRECTORA (e) DE LA CARRERA

Ing. Ana Isabel Camacho Coronel, Mgs.

Guayaquil, 12 de marzo del 2019



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

DECLARACIÓN DE RESPONSABILIDAD

Yo, Zambrano Soriano, Pedro Eddison

DECLARO QUE:

El Trabajo de Titulación, **Análisis de retención de estudiantes en la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG, durante los dos primeros años de la carrera, mediante técnica de Minería de Datos** previo a la obtención del título de **Ingeniero en Sistemas Computacionales**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, 12 de marzo del 2019

EL AUTOR

Zambrano Soriano, Pedro Eddison



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

AUTORIZACIÓN

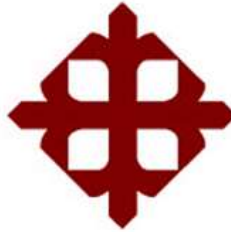
Yo, Zambrano Soriano, Pedro Eddison

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Análisis de retención de estudiantes en la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG, durante los dos primeros años de la carrera, mediante técnica de Minería de Datos**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, 12 de marzo del 2019

EL AUTOR:

Zambrano Soriano, Pedro Eddison



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

REPORTE URKUND

URKUND

Documento: [ZarifaMiro-GISELWALD-Feuerrig-21.docx](#) (14416000)

Presentado: 2019-02-21 12:04 (-05:00)

Presentado por: adela.zarifa@ucsg.edu.ec

Recibido: adela.zarifa@ucsg.edu.ec

Mensaje: [Mensaje de notificación](#)

2% de estas 41 páginas, se componen de texto presente en 6 fuentes.

Categoría	Enlace/nombre de archivo
	COSCOSUA COSCOSA A-sistema y A-sistema-21.docx
	URSA LETAMENDIARIZAL.docx
	URSA LETAMENDIARIZAL-DC.docx
	URSA LETAMENDIARIZAL-URSA-21.docx
	http://www.ucsg.edu.ec/informatica/analisis-de-datos/urca/urca-21.pdf
	Yarifa_21.docx

UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL, FACULTAD DE INGENIERÍA, CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

TÍTULO:

ANÁLISIS DE PRESENCIA DE ESTUDIANTES EN LA CARRERA DE NEGOCIOS INTERNACIONALES DE LA FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS DE LA UCUG, DURANTE LOS DOS PRIMEROS AÑOS DE LA CARRERA, MEDIANTE TÉCNICA DE MUESTRA DE BARRIO.

AUTOR: Zambrano Soriano, Pedro Esteban

Tipo de trabajo de grado: previo a la obtención del título de INGENIERO EN SISTEMAS COMPUTACIONALES

TUTOR(A): Ing. Durbis Fabriz, Adela Hana Espinosa, Ph.D.

Asesor(a): Escobar

UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL, FACULTAD DE INGENIERÍA, CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

Urkund Report - E...pdf

Mostrar todo

12:15 21/02/2019

DEDICATORIA

A mi familia, amigos, y al amor de mi vida, que me impulsaron y apoyaron para alcanzar este nuevo logro.

AGRADECIMIENTO

A mis profesores, y tutores por su gran ayuda y guía durante este trabajo de investigación y toda mi carrera.

A las autoridades de la Universidad Católica de Santiago de Guayaquil por su apoyo y colaboración.

A mi familia y al amor de mi vida por ayudarme a superar las adversidades y a levantarme cada día.



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

TRIBUNAL DE SUSTENTACIÓN

Ing. Adela María Eugenia, Zurita Fabre, Ph.D
TUTORA

Ing. María Paulina, Ching Correa, Mgs.
OPONENTE

Ing. Galo Enrique, Cornejo Gómez, Mgs.
DELEGADO DEL ÁREA

Ing. Ana Isabel, Camacho Coronel, Mgs.
DELEGADO DE LA DIRECCIÓN DE CARRERA

ÍNDICE GENERAL

RESUMEN.....	xiv
INTRODUCCIÓN	2
CAPÍTULO I: PROBLEMÁTICA DE INVESTIGACIÓN.....	5
1.1. Planteamiento del problema	5
1.2. Pregunta de investigación.....	8
1.3. Objetivos	8
1.3.1. Objetivo General.....	9
1.3.2. Objetivos Específicos	9
1.4. Justificación y alcance.....	9
CAPÍTULO II: MARCO TEÓRICO, CONCEPTUAL Y LEGAL.....	11
2.1. Elementos teóricos	11
2.1.1. Deserción Estudiantil vs Rendimiento Académico	11
2.1.2. Predicción del Índice de Permanencia Estudiantil.....	17
2.1.3. Teorías y principios relacionados con la Minería de Datos.....	19
2.1.3.1. Proceso de KDD	20
2.1.3.2. Minería de Datos.....	23
2.1.3.3. Técnicas de Minería de Datos.....	24
2.1.4. Infraestructura tecnológica	27
2.1.5. Herramientas de desarrollo	30
2.1.5.1. Interfaz gráfica.....	30
2.1.5.2. Herramientas.....	30
2.2. Sustento Legal	33
2.3. Ámbito de Aplicación	34
2.3.1. Gestión Empresarial Internacional (Negocios Internacionales)	34
2.3.1.1. Proceso de acreditación	36
2.3.1.2. Tasas de retención.....	38
CAPÍTULO III: METODOLOGÍA Y RESULTADOS	40
3.1. Metodología de la Investigación	40
3.2. Metodología de desarrollo.....	42
3.3. Metodología de Minería de Datos	43
3.3.1. Integración y selección	44
3.3.2. Exploración y procesamiento.....	44
3.3.3. Limpieza	44

3.3.4. Transformación/reducción de datos.....	45
3.3.5. Minería de datos.....	45
3.3.6. Evaluación e interpretación de resultados	45
3.3.7. Difusión y uso de modelos.....	46
3.4. Análisis de Resultados	46
CAPÍTULO IV: PROPUESTA.....	48
4.1. Introducción	48
4.2. Objetivo.....	48
4.3. Responsables	48
4.4. Descripción de la solución	49
4.4.1. Proceso.....	49
4.4.1.1. Fase 1: Minería de Datos	50
4.4.1.2. Fase 2: Implementación de las reglas	61
4.4.2. Requerimientos de Hardware y Software	62
4.4.3. Resultados esperados del aplicativo	63
CONCLUSIONES	64
RECOMENDACIONES	65
REFERENCIAS.....	66
APÉNDICES.....	76
ANEXOS	85

ÍNDICE DE TABLAS

Tabla 1: Comparación de los lenguajes de programación	32
Tabla 3: Clasificación de las variables	51
Tabla 4: Rangos de valor.....	53
Tabla 5: Reglas simplificadas	57
Tabla 6: Reglas simplificadas	57
Tabla 7: Descripción de los nodos	58

ÍNDICE DE FIGURAS

Figura 1: Criterios, subcriterios e indicadores.	7
Figura 2: Modelo de deserción. Tomado de Himmel (2018).....	12
Figura 3: La organización educativa como sistema social y académico.....	15
Figura 4: Fases del proceso KDD	21
Figura 5: Proceso KDD.....	22
<i>Figura 6: Ejemplo del concepto de Minería de Datos.....</i>	<i>24</i>
Figura 7: Algunas técnicas de Minería de Datos	27
Figura 8: Herramientas punteras en Data Science y Machine learning	30
Figura 9: Organigrama	36
Figura 10: Datos informativos del rediseño de la Carrera	38
Figura 12: El paradigma de hacer prototipos	43
Figura 13: Proceso actual	49
Figura 14: Solución	50
Figura 155: Porcentaje de datos faltantes por columnas	52
Figura 16: Vista minable	54
Figura 17: Estructura jerárquica del modelo	54
Figura 18: Comparación con algoritmo J48 (1)	55
Figura 19: Comparación con algoritmo J48 (2)	55
Figura 20: Teorema de Bayes	56
Figura 21: Conjunto de reglas	56
Figura 21: Solución	61

ÍNDICE DE APÉNDICES

Apéndice A. Formato para entrevistas a: Directora de Carrera, Coordinadora Académica, y Asesoría Pedagógica Estudiantil.	76
Apéndice B. Manuales	77

ÍNDICE DE ANEXOS

Anexo 1. Malla curricular rediseñada, Carrera de Gestión Empresarial Internacional (Negocios Internacionales).....	85
--	----

RESUMEN

La presente investigación tuvo el propósito de generar un modelo predictivo de retención-deserción de estudiantes de la cohorte 2016-2018 de la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas, que permita determinar la probabilidad de deserción estudiantil y ayudar en predicciones futuras; este estudio tuvo un enfoque cualitativo, del tipo analítica, exploratoria y descriptiva porque estuvo orientada a realizar un análisis de procesos y aplicativos tecnológicos que permitan determinar los niveles de retención estudiantil, utilizando la minería de datos como herramienta de predicción; dado el ámbito del estudio, no fue necesario establecer población ni muestra, solamente se identificó las personas clave de este proceso y los interesados en los resultados y se los denominó *informantes* a quienes se aplicó la técnica de la *entrevista*. Como resultado de la información recogida se pudo concluir que actualmente no existe un protocolo establecido para identificar a los estudiantes que tienen menor probabilidad de continuar sus estudios, y la mayoría de las selecciones que se realizan se hacen con referencia a la parte académica del alumno. Luego de finalizar el desarrollo del proyecto, se pudo comprobar que el modelo predictivo basado en técnicas de Minería de Datos, permitió la evaluación e interpretación de toda la información de los estudiantes de la cohorte 2016-2018 de la Carrera de Gestión Empresarial Internacional, a través de la cual se determinó un modelo de la retención-deserción y es una herramienta de apoyo en la toma de decisiones de futuras proyecciones sobre este indicador.

Palabras Clave: MINERÍA DE DATOS; RETENCIÓN ESTUDIANTIL; MODELO PREDICTIVO; ÁRBOL DE DECISIÓN; ALGORITMO C4.5; TEOREMA DE BAYES.

INTRODUCCIÓN

Las reformas educativas impulsadas para analizar los niveles en la calidad de la educación ecuatoriana, se han dirigido hacia la evaluación regular de las carreras universitarias para la toma de decisiones que se traduzca en la mejora de la calidad de la educación.

Por motivos de acreditación de las carreras universitarias y evaluación de las mismas por parte de los organismos competentes, los centros de educación superior se encuentran realizando procesos para la implementación de modelos para la mejora de los servicios educativos que en ellos se brindan (Grijalva, Freire, Real, Arellano, & Cornejo, 2018). Las entidades gubernamentales exigen que en las universidades se brinde a los estudiantes una educación de calidad, para lo cual es indispensable la evaluación constante de los distintos indicadores, los mismos que son valorados mediante los procesos internos que en cada institución educativa se realizan para cumplir con este cometido.

Es, a través del análisis de indicadores, que se obtendrían resultados más específicos que sugieran la toma de decisiones, orientadas a proponer gestiones destinadas a que se mejore el rendimiento de una unidad académica, mejora que no sólo trate el aspecto económico en beneficio de la institución de estudios superiores, sino a determinar con anticipación cuáles son los contenidos medulares que deberán ser reforzados para obtener la calidad y así cumplir con la planificación de cada unidad (M. Martínez & Coronado, 2003).

Dentro de los indicadores considerados de importancia para la evaluación de las carreras se encuentra el de Eficiencia Académica, el mismo que permite, a partir de un contexto cuantitativo, presentar características de orden cualitativo tales como calidad de los sistemas educativos anteriores, métodos para la admisión a estudios superiores, incluyendo un abanico elementos que son parte del proceso educativo y las políticas propias del centro de estudios que avalan la retención de los estudiantes en las diversas unidades académicas (K. Rodríguez, Gutiérrez, Wong, & López, 2015).

La carrera Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la Universidad Católica de Santiago de Guayaquil (UCSG) ha dado

impulso a la ejecución de los procesos internos para evaluar el servicio educativo prestado y, para colaborar con estas acciones, se debe realizar el análisis del indicador de Eficiencia Académica, para lo cual se podría tomar como base para el estudio el grupo de estudiantes de la cohorte 2016-2018, con el fin de generar un modelo predictivo basado en patrones, que permita determinar las causas de la retención-deserción de los estudiantes y contribuir con la toma de decisiones en beneficio de la carrera y de los mismos estudiantes, enmarcado siempre en las exigencias de calidad y acreditación determinadas por los entes reguladores de la educación en el país.

Para generar modelos predictivos de evaluación existen algunos métodos y técnicas, siendo la Minería de Datos un proceso que permite descubrir relaciones, patrones y tendencias cuando se realiza un examen a volúmenes de información, los mismos que, gracias a diversidad de herramientas informáticas, se los puede transformar mediante el uso de técnicas. Estas técnicas buscan el descubrimiento automático del conocimiento que se encuentra en los datos almacenados ordenadamente en bases de datos y su objetivo principal es “descubrir patrones, perfiles, tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas de análisis de datos” (Pérez, 2007, p. 1).

Entre las técnicas de minería de datos o *data mining* se encuentran: 1) predictivas en donde las variables iniciales pueden ser dependientes o independientes, 2) descriptivas, en donde las variables pueden tener el mismo estatus al inicio, y 3) auxiliares. Dependiendo de los datos que se obtengan se establecerá el tratamiento de los mismos y la aplicación de las técnicas de minería de datos más adecuadas para el estudio (Cortés Campos, Zapata González, Menéndez Domínguez, & Canto Herrera, 2015; Pérez, 2007).

Con el fin de comprender de mejor forma el desarrollo de este proyecto, se lo ha estructurado en cuatro capítulos. En el capítulo 1 se hace referencia al planteamiento de problema, en donde se presentan los argumentos necesarios para establecer la importancia de su realización, la pregunta de investigación, los objetivos, general y específica, y la justificación y alcance del proyecto.

En el capítulo II se toma en consideración el marco teórico, conceptual y legal, con los elementos teóricos, conceptos relacionados al tema, leyes que sustentan la aplicabilidad de los medios utilizados en el desarrollo del proyecto; además del contexto en donde se aplicará el estudio.

En el capítulo III se analiza la metodología de investigación, en donde se propone el tipo y método de investigación, la población, muestra y muestreo, las técnicas e instrumentos de recolección de datos, la herramienta para el procesamiento de la información, y finalmente el análisis de los resultados de la investigación.

En el capítulo IV se presenta la propuesta tecnológica para la solución del problema, que contiene la arquitectura de la solución, su descripción, los actores del sistema, los requerimientos del sistema, y el plan de entrega de la solución.

Por último, las conclusiones y recomendaciones resultantes del desarrollo del proyecto.

CAPÍTULO I:

PROBLEMÁTICA DE INVESTIGACIÓN

Todo proyecto a proponerse y posteriormente desarrollarse requiere de una estructura que permita comprender cada una de sus partes. Lo inicial en una investigación es determinar por qué se origina el problema y plantear una solución, y luego de esto se necesita establecer los objetivos que permitirán la ejecución del trabajo. Todos estos aspectos se tratan en los párrafos siguientes.

1.1. Planteamiento del problema

La deserción estudiantil constituye el retiro forzoso o voluntario de un estudiante, de una carrera universitaria, y es uno de los problemas con los que se enfrentan las instituciones educativas. En el caso de Ecuador las Instituciones de Educación Superior (IES) son las que se ven afectadas por tal realidad, entendiéndose que la deserción en la educación superior es, según López, Marín, y García (2012) citados por Zamora et al., (2016) “la cantidad de estudiantes que abandona el sistema de educación superior entre uno y otro período académico semestre o año” (p. 27). Se considera, además, que existen tres tipos de deserción: 1) precoz, 2) temprana, 3) tardía (Zamora et al., 2016).

Para medir la calidad de la educación, en el país se creó el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES) que se encarga de ejecutar procesos de: 1) autoevaluación, 2) evaluación externa, 3) acreditación, 4) seguimiento, 5) evaluación de estudiantes, 6) habilitación para el ejercicio profesional (CEAACES, 2016, párr. 2). La evaluación institucional se realiza quinquenalmente para la acreditación de “todas las universidades y escuelas politécnicas del Sistema de Educación Superior ecuatoriano, para determinar la categorización de las instituciones según lo determina la Ley Orgánica de Educación Superior (LOES)” (CACES, 2018b, párr. 2).

Para realizar el proceso evaluativo, el CEAACES tiene un modelo de evaluación con el fin de que se cumpla el nivel de calidad mínimo y “su objetivo es generar una cultura de excelencia” (CACES, 2018b, párr. 5).

A partir de agosto del 2018, el CEAACES pasó a denominarse CACES, que corresponde a las siglas del Consejo de Aseguramiento de la Calidad de la Educación Superior, con la misma filosofía y fines orientados a impulsar la calidad en la educación universitaria.

El proceso que se sigue para evaluar una determinada carrera universitaria comprende dos fases: “la evaluación del entorno del aprendizaje y los resultados de aprendizaje (...). El resultado final de los dos procesos determina la acreditación de una carrera, proceso que termina con la entrega y difusión pública de los resultados” (CACES, 2018a). Por lo que es importante para las carreras buscar finalmente la acreditación.

En cuanto al proceso de la Evaluación del Entorno de Aprendizaje (CACES, 2018a):

Está a cargo de la Dirección de Evaluación y Acreditación de Universidades; mediante el cual se evalúan las condiciones académicas y físicas en las que se desarrolla la carrera, las que deben garantizar el cumplimiento de los estándares definidos en el modelo específico, en función del área de conocimiento, estructurados en criterios, sus criterios e indicadores. (párr. 2).

Lo antes mencionado indica que es de prioridad para las IES ecuatorianas emprender los procesos concernientes a conseguir la evaluación y acreditación, aplicando los métodos que lleven a la institución a encontrar la excelencia académica, respondiendo a los distintos indicadores que requiere la acreditación.

Existen indicadores cuantitativos (medidos por medio de cálculos matemáticos para establecer el desempeño institucional) y cualitativos (medidos mediante las cualidades que debería tener la IES y que se comparan con las que presenta la institución a evaluar), dentro de los cuales se encuentra la Eficiencia Académica (CES, 2016).

CRITERIOS	SUB-CRITERIOS	INDICADORES	
Vinculación con la sociedad	Institucionalización	Planificación de la vinculación Gestión de recursos para la vinculación	
	Resultados de la vinculación	Programas y/o proyectos de vinculación con la sociedad	
Recursos e Infraestructura	Infraestructura	Calidad de aulas Espacios de bienestar Oficinas Tiempo Completo Salas Medio Tiempo/Tiempo Completo	
		Tecnologías de Información y Comunicación	Conectividad Plataforma de Gestión Académica
		Biblioteca	Gestión de la Biblioteca Libros por estudiante Espacio estudiante
		Condiciones	Admisión a estudios de postgrado Bienestar estudiantil Acción afirmativa
	Eficiencia Académica	Tasa de retención grado Tasa de titulación grado Tasa de titulación postgrado	

Figura 1: Criterios, subcriterios e indicadores. Tomado de “Propuesta de Indicadores para la evaluación institucional del Sistema de Educación Superior”, CES (2016).

Los indicadores de acreditación y evaluación de las carreras universitarias de las IES del Ecuador que en la actualidad se exigen, están encaminados a conseguir, como se mencionó anteriormente, la excelencia académica. Por tal motivo, uno de los indicadores a evaluarse es el de Eficiencia Académica, el mismo que mide la retención-deserción de los estudiantes de las distintas carreras en los primeros años de estudio, que son críticos para establecer si un estudiante estudia o no una carrera, puesto que es un período en el cual la deserción es factible.

La Eficiencia Académica es muy importante en el momento de evaluar la calidad de la enseñanza en las IES, ya que estima cómo se comporta el grupo estudiantil y la forma de trabajo que se imparte. De forma general este indicador define variables y se analiza el elemento tiempo de los alumnos que están matriculados y los que culminan exitosamente su carrera, elemento que se aplica al período académico, determinada área que forma parte del pensum académico, carrera estudiada, facultad y universidad (K. Rodríguez et al., 2015).

Si en el momento en que se realiza el cálculo se considera la cantidad de alumnos de una promoción que inicia y termina con éxito sus estudios, en relación con la cantidad que se matriculó en el primer período de la mencionada cohorte, se consigue lo que se llama “Eficiencia Académica Terminal, Eficiencia Interna, Tasa de Éxitos” (K. Rodríguez et al., 2015, párr. 7).

De acuerdo a la presencia de las variables que deben estudiarse para determinar si el estudiante se queda o no en una carrera, se considera que el análisis contempla aspectos no sólo académicos sino también socioeconómicos y personales como factores que afectan tanto a estudiantes como a la propia IES. El análisis de las variables y su comportamiento permitirían tener un mejor control de los estudiantes que ingresan a una carrera y establecer con precisión las causas de la retención-deserción de los mismos para tomar los correctivos necesarios.

Con los antes mencionados antecedentes, que indican la obligatoriedad de realizar los procesos de evaluación y acreditación de las carreras, la UCSG se ha comprometido a cumplir con las fases que implican este proceso aplicando las acciones correspondientes en las distintas facultades y carreras (UCSG, 2018a).

En la carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG no existe una herramienta que permita realizar el análisis de retención-deserción de estudiantes, por lo que se propone la generación de un modelo predictivo basado en técnicas de Minería de Datos para analizar la información disponible de la cohorte del año 2016 proyectada hacia 2018, ya que son en los dos primeros años de una carrera en que se puede determinar la permanencia de un alumno, y sirva de base para predicciones futuras.

1.2.Pregunta de investigación

¿El modelo predictivo basado en patrones mediante el uso de la Minería de Datos permitirá evaluar e interpretar la información de los datos obtenidos de la cohorte 2016-2018 de la Carrera de Negocios Internacionales, para determinar el nivel de retención-deserción de estudiantes y servirá para proyecciones futuras?

1.3.Objetivos

Para que el proyecto llegue a su culminación, se requiere el planteamiento de los objetivos, tanto general como específico.

1.3.1. Objetivo General

Generar un modelo predictivo de retención-deserción de estudiantes de la cohorte 2016-2018 de la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas, para determinar la probabilidad de deserción estudiantil y ayudar en predicciones futuras.

1.3.2. Objetivos Específicos

- Efectuar el diagnóstico de la cohorte de estudiantes del período 2016-2018 de la carrera de Negocios Internacionales, a partir de información levantada en la carrera, para determinar las variables que integrarán el modelo predictivo de información académica;
- Realizar el diseño de la vista minable, para determinar sus atributos más óptimos a través de las técnicas de Minería de Datos, para minimizar posibles errores en la producción del modelo predictivo;
- Generar el modelo predictivo basado en patrones, mediante la utilización de técnicas y algoritmos correspondientes, que permita evaluar e interpretar la información de los datos obtenidos.

1.4. Justificación y alcance

Al tomar las variables relacionadas con el ámbito académico, socioeconómico, laboral, etc., que serán consideradas en el diseño del modelo predictivo que se propone, las autoridades y directivos tendrán elementos de juicio para analizar las variables mencionadas y que pudieran ser intervenidas a lo interno de la unidad académica, con el propósito de incrementar la tasa de retención sin afectar la calidad educativa y la excelencia.

El proyecto sugerido responde al dominio Tecnología y Sistemas Productivos, línea de investigación Ciencia Tecnológica.

Es necesario dejar sentado el alcance de este proyecto, considerando que comprende la generación de un modelo predictivo de clasificación en base a técnicas de minería de datos, que permita analizar los factores o variables de incidencia en la retención-deserción de los estudiantes de la Carrera de Negocios Internacionales de la

Facultad de Ciencias Económicas y Administrativas de la UCSG, de la cohorte ingresada en el año 2016, de modo que se conozca su comportamiento; se tomarían en consideración aspectos como promedio de calificación, cantidad de materias aprobadas, si tiene beca, el promedio de días de retraso en el pago de sus pensiones, información que incide en la retención-deserción de estudiantes en una carrera universitaria y que puede ser solicitada y se encuentra en el sistema.

De las pruebas que se realicen a los datos obtenidos se validará el modelo para determinar las reglas que determinen las proyecciones de los próximos años, con el fin de que se convierta en una herramienta de toma de decisiones para la mejora de los servicios educativos de la UCSG.

El modelo predictivo se mostrará en un aplicativo que se implementará únicamente para uso de la Carrera de Negocios Internacionales, para lo cual se propone realizarlo en un equipo informático de la Sala de Docentes a Tiempo Completo, que se encuentra disponible.

La herramienta que se propone sería un apoyo para la toma de decisiones de mejora continua en la excelencia académica y por consiguiente que la misma institución tenga opciones de poder competir en el mercado educativo.

CAPÍTULO II:

MARCO TEÓRICO, CONCEPTUAL Y LEGAL

En este capítulo se discuten fundamentos teóricos que son utilizados para enmarcar la presente investigación, y tener clara la terminología de los temas que están presentes a lo largo del trabajo.

2.1. Elementos teóricos

En este apartado se presentan los fundamentos teóricos sobre los cuales se basa este proyecto, como son Rendimiento Académico, Deserción Estudiantil, Índice de Permanencia, Minería de Datos, Modelos y Técnicas de Minería de Datos, infraestructura tecnológica para la aplicación de la Minería de Datos, herramientas de desarrollo del aplicativo, el sustento legal y ámbito de aplicación.

2.1.1. Deserción Estudiantil vs Rendimiento Académico

Antes de definir la deserción estudiantil, se necesita conocer lo que es la *retención*, la misma que, de acuerdo a Himmel (2018) es “persistencia de los estudiantes en un programa de estudios universitarios hasta lograr su grado o título” (p. 94), considerando que el término se limita al hecho de que el estudiante podría finalizar sus estudios universitarios en el tiempo que dure su carrera o también no teniendo en cuenta lo que pueda demorarse; esto ya debido a factores como materias a repetir, abandono de la carrera por varios períodos o por no matricularse en todas las asignaturas que tiene cada período académico, de forma que se alarga la carrera. No obstante poder distinguir cada situación antes mencionada, se dificulta diferenciar las dos últimas, ya que puede deberse a varias causas como ser el propio alumno que prolonga su carrera o responder a la propia carrera (programa de estudio y sus características) (Himmel, 2018).

De otro lado, se define como *deserción* “al abandono prematuro de un programa de estudios antes de alcanzar el título o grado, y considera un tiempo suficientemente largo como para descartar la posibilidad de que el estudiante se reincorpore” (Himmel, 2018, pp. 94–95). Para Tinto (1975) citado por Canales y De los Ríos (2018) la deserción es “el proceso de abandono voluntario o forzoso de la carrera en que se

matricula un estudiante, por la influencia positiva o negativa de circunstancias internas o externas al alumno” (p. 175).

El antes mencionado Tinto (1975), referenciado por Canales y De los Ríos (2018) plantea que la deserción universitaria puede ser temporal y/o permanente, según el período que el estudiante deja de estudiar. Es transitoria “cuando el alumno decide revertir la situación de abandono, en la medida de que éste no tenga impedimento para retomar sus estudios; (...) permanente cuando existen causas concretas que impiden al alumno volver a cursar o retomar sus estudios superiores” (Canales & De los Ríos, 2018, pp. 175–176).



Figura 2: Modelo de deserción. Tomado de Himmel (2018)

Otros autores señalan distintas causas para la deserción, tales como: aspectos sociales, personales, económicos, culturales del estudiante y su entorno familiar, por un lado, y, aspectos académicos y de la institución de estudios superiores (Beneyto Sánchez, 2015; Jara et al., 2008).

Entre los motivos de deserción destacados por diversos investigadores se encuentran, por un lado, factores personales, culturales, sociales y económicos de los alumnos y sus familias, y, por otro, factores académicos e institucionales, aunque no existe un acuerdo si el factor de deserción es personal, social o propio de la institución. Es así como Canales y De los Ríos (2018) manifestaron que los factores sociales se relacionan con aspectos familiares (nivel de educación y ocupación de sus padres, perspectiva que el estudiante tiene sobre el futuro y los objetivos que quiere cumplir, el aspecto financiero, responsabilidades de la familia); en el mismo tema social, también es importante el dominio que pueda tener la familia y los amigos sobre el estudiante, el mismo que si es positivo será un factor favorable para disminuir la

deserción. Por otro lado, el factor académico es de mayor importancia en estudiantes que forman parte de familias que sí han tenido estudios universitarios.

Otro de los factores de importancia es la motivación, sobre todo cuando se trata la deserción voluntaria, existiendo una considerable relación entre la motivación del alumno y su rendimiento académico, es decir, la disposición que tenga hacia sus trabajos académicos, la importancia que dé a sus estudios en referencia a sus objetivos, el reto que suponga su carrera y otras motivaciones (Kohler Herrera, 2013). La deserción también tiene que ver con el ingreso del alumno a una universidad, entendiendo que el proceso que implica decidirse por determinada carrera o centro de estudios significaría estratificación y poder, aunque la elección depende de ciertas condiciones (favorables o desfavorables). Factores como costos de matrículas, becas, clase social o raza influyen en la decisión de elegir y permanecer en una determinada universidad, impulsados por ubicarse en un medio en el cual encajen según las antes mencionadas condiciones (Canales & De los Ríos, 2018).

En términos más específicos, al decir de Viale (2014), la deserción estudiantil, vista desde la perspectiva de los hispano-hablantes, las distintas causas que mueven a los estudiantes al abandono de sus estudios antes de haberlos terminado son: a) renuncia involuntaria (inobservancia de normas administrativas o estatutos), b) abandono de la carrera por otra en la misma IES, c) abandono de la carrera por una nueva en otra IES, d) abandono de la IES y terminar sus estudios en otra, e) abandono de la carrera universitaria por trabajo o por cursos o seminarios fuera de la institución, f) detener la carrera con miras a estudiar posteriormente, g) otras causas.

Todo lo antes mencionado confirma que existen distintas formas de concebir las causas de la deserción estudiantil en las universidades con las consiguientes consecuencias tanto para el estudiante como para la institución, considerando que analizar las causas de la deserción es de importancia para control de la misma, ya que la tasa de abandono estudiantil se la puede tomar como indicador para medir la calidad de la educación en la IES y su acreditación, entendiendo que si este indicador es alto la calidad de educación en una universidad es baja y se asumiría que dicho centro de estudio no cumplió con modelos educativos apropiados que sirvan para que el estudiante finalice sus estudios universitarios (Viale, 2014).

Se han abordado los términos retención y deserción, ambos están directamente relacionados con el rendimiento académico. El rendimiento académico y el desempeño de los estudiantes durante el primer año de estudios superiores se ha convertido en una preocupación por los negativos indicadores de deserción y rendimiento estudiantil bajo y es motivo de investigación para establecer sus causas y tomar decisiones para mejorar los modelos educativos (Campo, Escorcía, Moreno, & Palacio, 2016).

Específicamente, el rendimiento académico como factor de deserción escolar “estaría en función de determinar el éxito o el fracaso en los estudios, hace referencia a la certificación académica o calificaciones obtenidas por los estudiantes durante su carrera hasta la obtención del título correspondiente” (Murillo, 2016, sec. Introducción).

Actualmente, el rendimiento académico se lo determina por las calificaciones de un estudiante o grupo de estudiantes durante los períodos académico cumplido y eso constituye la calidad de la educación. Las calificaciones, pues, son determinantes para conocer si aprueba o no sus semestres académicos, por lo que el rendimiento académico es una característica muy significativa en la labor educativa, ya que a través de aquel se rige en gran manera el éxito o fracaso de los estudiantes y que se los conoce a través de informes (G. Martínez & Pedroza, 2017).

Los motivos que influyen para que los alumnos dejen sus estudios universitarios repentinamente por su bajo rendimiento académico “obedecen a factores personales, socioeconómicos, institucionales, demográficos y académicos” (Campo et al., 2016).

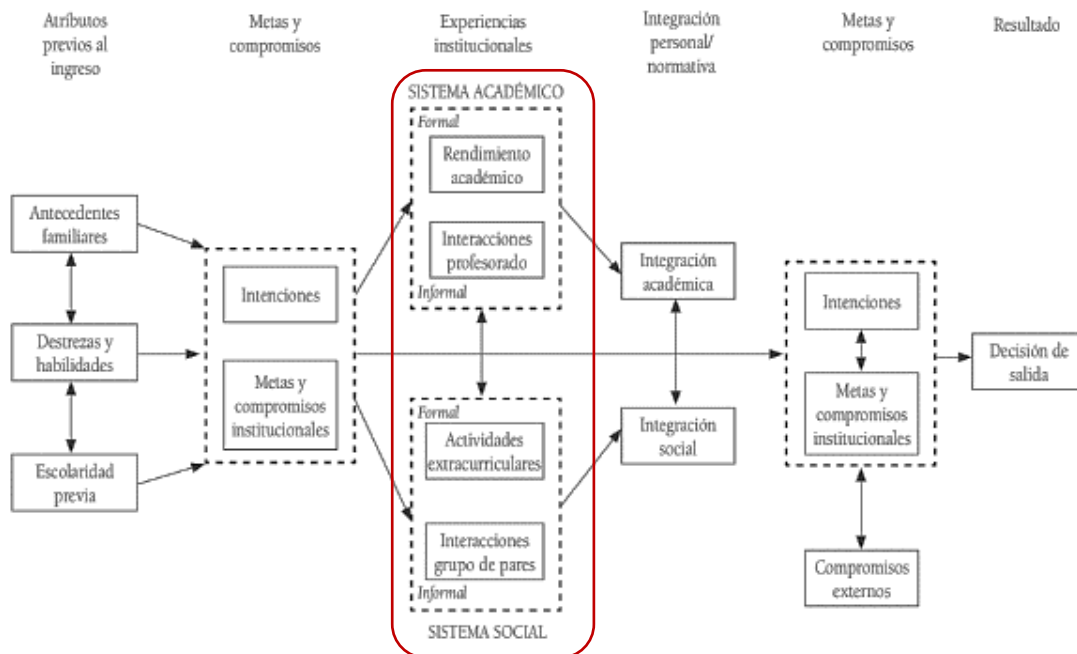


Figura 3: La organización educativa como sistema social y académico. Modelo adaptado de Tinto (1975). Nota: Tomado de Donoso & Cancino, (2018)

De acuerdo a Porcel, Dapozo, y López (2010), la palabra rendimiento tiene varios alcances, si se toma en cuenta que las notas que los estudiantes consiguen son decisivas para determinar el rendimiento académico. Estos datos generarían el concepto de que el rendimiento académico depende únicamente del estudiante, no obstante, el compromiso también recae sobre la institución, ya que es a través de ella que se realiza la evaluación de los parámetros que se consideran como rendimiento. A más de las circunstancias en las cuales las instituciones universitarias se encuentran en cuanto a las condiciones educativas que ofrece y la práctica docente, es indispensable conocer las particularidades que contribuyen las personas que reciben la enseñanza. Con estos datos se podría determinar cuáles serían algunos de los argumentos de la deserción universitaria.

Existen factores que intervienen en el rendimiento de los estudiantes, unos que son parte directa del estudiante (endógenos) y los que se encuentran a su alrededor (exógenos), en donde el rendimiento académico resulta de la confluencia de ambos factores. Dentro de estas características de los alumnos, se encuentra la explicación ofrecida por (Porcel et al., 2010):

Han sido consideradas a la hora de relacionarlas con el rendimiento académico, desde las características aptitudinales, intelectuales y de la

personalidad del alumno hasta los aspectos motivacionales y de percepción personal de los estudiantes durante el transcurso de la carrera, así como también razones de ingreso a la misma... (p. 3).

Al realizar el análisis de los conceptos de deserción estudiantil frente a rendimiento académico, se puede anotar que están directamente relacionadas, porque el rendimiento es un factor que incide en la deserción estudiantil, no son hechos aislados que se tratan de forma separada.

Sobre el mismo tema Vargas, Ramírez, Cortés, Farfán, y Heinze (2011) señalaron el significado de algunas variables que influyen en el rendimiento de los estudiantes:

- **Desempeño académico:** lo que se evalúa del estudiante en el primer año de estudios universitarios, de los promedios de cada materia y promedio final;
- **Aspectos económico-sociales:** los elementos determinantes de su situación social, económica e incluso demográfica del estudiante;
- **Recorrido académico:** lo relacionado con su vida estudiantil, participación en eventos estudiantiles, investigativos y todo lo que se tenga sobre su colaboración, considerando la institución educativa de la que proviene;
- **Atributos de personalidad:** rasgos distintivos propios y adquiridos de su conducta y su reacción frente a situaciones diversas;
- **Pensamiento abstracto:** solución de problemas lógicos y capacidad del estudiante para anticipar y planificar;
- **Pensamiento creativo:** aplicación de los conocimientos adquiridos y experiencias vividas para producir nuevas ideas o responder de forma creativa o novedosa.

Para Kohler (2013) otra variable de influencia en el rendimiento académico es la habilidad que puede tener un alumno para asimilar datos. Según la autora:

Las estrategias de aprendizaje constituyen un proceso de decisión consciente e intencional del sujeto, formado por una secuencia ordenada de operaciones mentales que realiza con el objetivo de organizar y reconstruir la información en su estructura cognitiva de tal manera que logre aprender, y a su vez, le

permita planificar, organizar, ejecutar y evaluar sus actividades de aprendizaje. (Kohler, 2013, p. 279).

De lo anterior se deduce que las estrategias del aprendizaje se constituyen en proceso de selección, segregación, ejecución de un conjunto de características del estudiante, a la par que son factibles de cambio para su adaptación al entorno (F. García, Fonseca, & Concha, 2015; Kohler, 2013)

La deserción universitaria, como se ha visto en párrafos previos, significa el abandono del estudiante de sus estudios por varios motivos, siendo uno de ellos el rendimiento académico, que es una variable que afecta a los estudiantes ya que puede cambiar los motivos y objetivos que tienen los estudiantes por determinada carrera universitaria. Para contrarrestar estos efectos, es necesaria una orientación previa en el momento que el posible estudiante universitario tenga una información clara de la carrera que piensa escoger; esta orientación consiste en la asesoría profesional, todo lo que ella conlleva, cómo acceder a ella, y los avances que se consigan más adelante (Díaz et al., 2014).

En décadas recientes ha tomado importancia mejorar los planes de estudios de las instituciones universitarias, así como modificaciones en el modelo enseñanza-aprendizaje, sobre todo en lo que se relaciona con la forma de adquirir y asimilar los conocimientos, junto con perfeccionamiento y adquisición de habilidades que le servirán en su vida profesional (Díaz et al., 2014).

2.1.2. Predicción del Índice de Permanencia Estudiantil

Los conceptos de permanencia y retención estudiantil, según Velázquez, Posada, & Gómez (2011), citados por García (Julio García, 2018), se encuentran vinculados con “las acciones y estrategias que las diferentes Instituciones de Educación Superior (IES), emprenden para analizar, entender o intervenir las razones y causas de la deserción en la institución” (Julio García, 2018, p. 13). Permanencia significa mantenerse en determinado lugar por un tiempo específico, por lo que permanencia como tema educativo “implica la indagación de las condiciones que puedan afectar la decisión de terminar un proyecto educativo, así estas condiciones sean académicas, familiares, económicas de inclusión u otras” (Julio García, 2018, pp. 13–14).

Para establecer la permanencia de un estudiante en su carrera universitaria se deben calcular los índices de deserción, de forma que se haga un análisis de las posibles causas y se diseñen estrategias que les permitan entender y descubrir los motivos de su incidencia para corregirlos.

Desde otra perspectiva, Torres (2012) asegura que la permanencia de un estudiante está relacionada con la retención, entendiendo que ésta se define como una medición que se realiza a través del contexto institucional, es decir las condiciones y estrategias que aplica la IES para conseguir que los estudiantes se mantengan hasta su graduación, al mismo tiempo que tiene relación con la motivación de los estudiantes para mantenerse en la misma universidad, desde su ingreso hasta la finalización de sus estudios.

El argumento antes mencionado señala, pues, la existencia de algunos factores que se relacionan directamente con la retención estudiantil (Torres, 2012):

- **Generales:** a) preparación académica anterior, b) ambiente universitario, c) compromiso con objetivos educativos y metas institucionales, d) integración social y académica, e) ayuda financiera;
- **Particulares:** a) orientación vocacional, b) bienestar universitario, c) reglamentos estudiantiles en la educación superior.

Para establecer cuántos estudiantes han permanecido en la misma institución universitaria se deberá calcular la tasa de deserción con la información que reposa en los archivos de cada carrera universitaria; estos son los “registros cada uno de los estudiantes matriculados en el programa, discriminando el periodo de matrícula (...), para cada período de estudio. (...) Los registros de cada uno de los estudiantes graduados discriminando el periodo de grado” (Julio García, 2018, p. 15).

Según García (2018), el cálculo de la deserción de una cohorte específica o un período académico determinado requiere especificar el estado del alumno, ya que puede encontrarse en estado *desertor* (si no retoma sus estudios por períodos consecutivos), *estudiante* (matriculado actualmente) o *graduado* (con título universitario). Por lo que la permanencia resultaría de la reconstrucción de los desertores, estudiantes y graduados, cuya información deberá ser procesada.

En concordancia con lo anterior se comprendería que uno de los indicadores relacionados con la gestión del control de la calidad en las IES es la tasa de deserción. Lo mismo sucede con el índice de retención estudiantil, que se considera un indicador de eficiencia interna para medir su desempeño, por lo que las universidades se encuentran creando planes para apoyar la retención para apoyar al estudiante en todo su período universitario para que el estudiante termine de forma satisfactoria sus estudios (Matheu, Ruff, Ruiz, Benites, & Morong, 2018).

El cálculo de los índices antes mencionados se lo puede lograr mediante la implementación de modelos predictivos, en los cuales se determinan las variables de ingreso a la universidad (académica, sociales, familiares y otras de interés) como las de mayor importancia para el análisis que se necesita realizar (Matheu et al., 2018), que utilizan herramientas de análisis estadístico, como lo es la Minería de Datos, que se trata en los apartados siguientes.

2.1.3. Teorías y principios relacionados con la Minería de Datos

Los avances tecnológicos en los distintos ámbitos de la ciencia obligan a que las empresas se actualicen y hagan uso de nuevas herramientas para una mejor generación de información. La era digital ha permitido que todos los datos que se procesan se puedan almacenar, ejecutar, compartir y enviar, por lo que la cantidad de información producida ha aumentado considerablemente y se guarde en bases de datos.

Conocer la abundante cantidad de datos producidos y analizar determinadas características de los mismos es un gran desafío, debido a que la información que se gestiona en una base de datos contiene distintos tipos de datos. Texto, datos numéricos, videos o imágenes se guardan para su posterior estudio, pero no siempre se pueden obtener los resultados esperados si no se aplican técnicas específicas de análisis, como las de Minería de Datos, las mismas que se usan en este proyecto para realizar el análisis del comportamiento de los estudiantes en la Carrera Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG, de la cohorte 2016-2018 y elaborar un modelo predictivo que sirva para una toma de decisiones futuras en cuanto a resultados del análisis del indicador Eficiencia Académica, indispensable para obtener la acreditación de la carrera.

Ya en contexto, la Minería de Datos o *Data Mining* representa una herramienta para analizar gran cantidad de información y descubrir modelos o patrones de ese conjunto de datos (Robles & Sotolongo, 2013), y cuyas técnicas permiten la extracción desconocida de la información que se encuentra contenida en los datos (EcuRed, 2018). Los datos que se desconocen de la gran cantidad de información almacenada resultará de utilidad para descubrir y explicar algún proceso. A través de la Minería de Datos se disponen los datos de manera que se puedan organizar, investigar y explorar para poder descubrir información de importancia que no se aprecia con facilidad (EcuRed, 2018).

Esta herramienta es aplicable en los distintos ambientes, por ejemplo, se puede aplicar Minería de Datos en educación, finanzas, biología, procesos de industria, policial e incluso en la política (Robles & Sotolongo, 2013).

La Minería de Datos se inicia con el proceso de KDD (*Knowledge Discovery in Databases*) que se refiere al procedimiento a través del cual se obtienen los datos y los dispone para su análisis.

2.1.3.1. Proceso de KDD

Knowledge Discovery in Databases o Proceso de Descubrimiento de Conocimientos en Base de Datos KDD, de acuerdo a Pacheco y Fernández (2015) “es el proceso iterativo e interactivo, en el cual se usan, a través de medios automáticos y semiautomáticos, técnicas de aprendizaje inteligente sobre una o varias fuentes de datos” (p. 26) para descubrir y obtener un conjunto de datos válidos o patrones, los mismos que luego de realizar su interpretación y evaluación de acuerdo al contexto al que pertenecen, se transforman en conocimiento que sirve de respaldo en la toma de decisiones, considerando que los pasos previos en la extracción de los datos son categóricas en la confiabilidad e importancia del conocimiento que se ha obtenido.

Las etapas del proceso de KDD son subprocesos que se forman “por una secuencia de etapas específicas” (Pacheco & Fernández, 2015, p. 26) en donde se aplican técnicas que se implementan en las diferentes herramientas para análisis y que consisten en un punto de vista conceptual de dónde se obtiene la información de mayor relevancia de los datos.

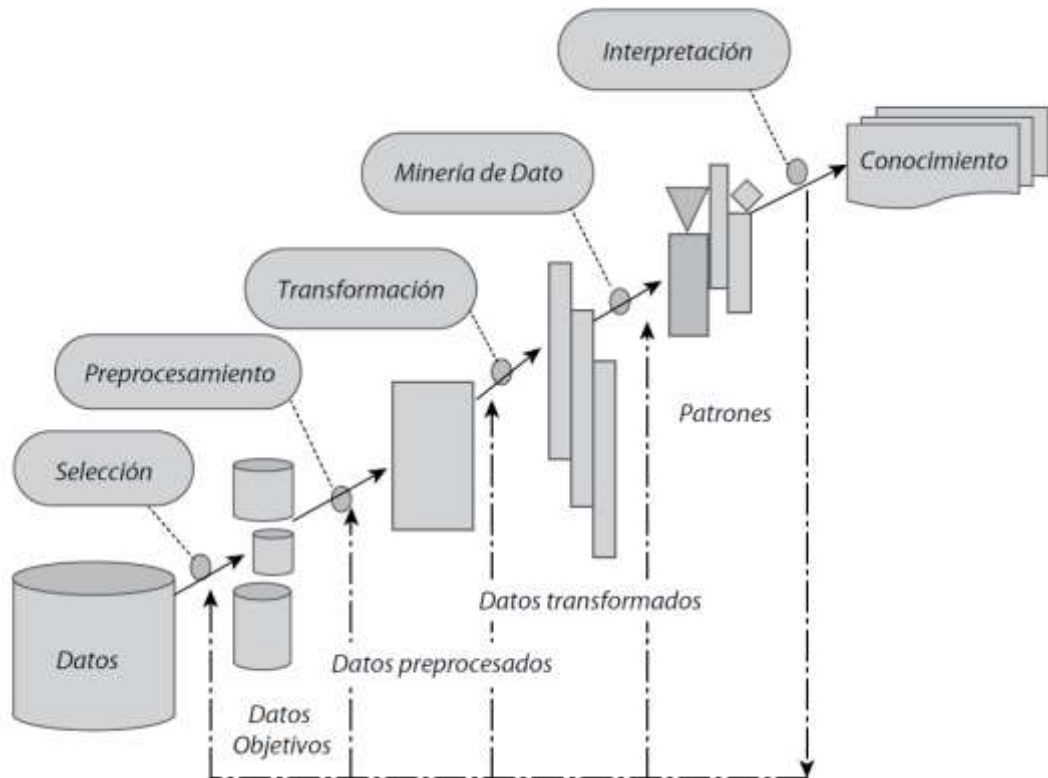


Figura 4: Fases del proceso KDD. Nota: Tomado de Timarán, Hernández, Caicedo, Hidalgo, y Alvarado, (2016)

En las etapas de KDD las técnicas pueden ser catalogadas según dos criterios “la fase del proceso de análisis de datos: a) preprocesado o Minería de datos, b) el objetivo del análisis de los datos: supervisadas o predictivas y no supervisadas o descriptivas” (Pacheco & Fernández, 2015, p. 26). Según Vieira, Ortiz, y Ramírez_(2009) el modelo descriptivo consiste en la recolección de la información, limpieza y análisis de los mismos, mientras que el modelo predictivo tiene que ver con la utilización de los datos para crear el modelo.

Las etapas del proceso KDD se describen en la Figura 5.

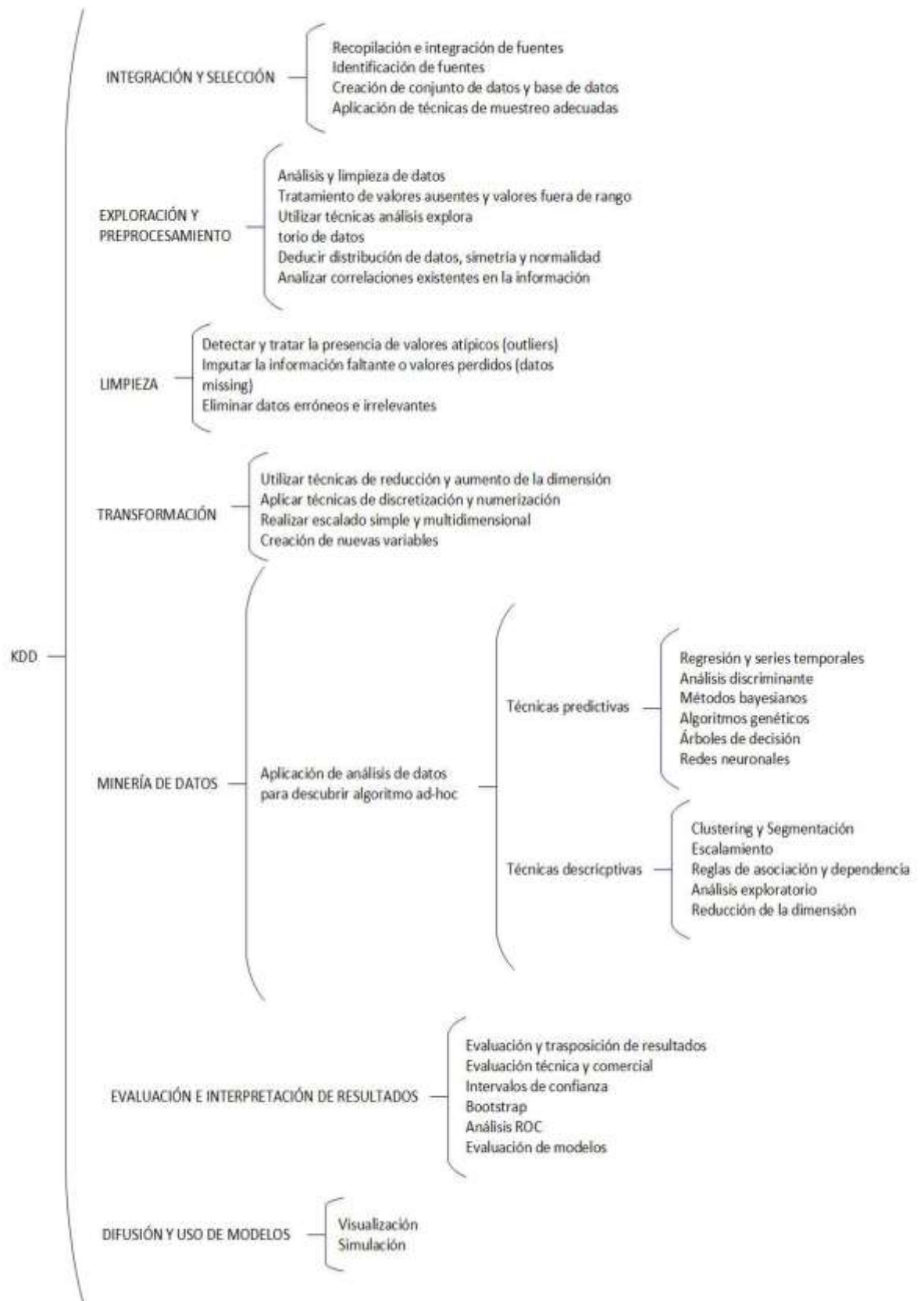


Figura 5: Proceso KDD. Nota: Adaptado de: Barrientos y Ríos (2013), Pérez (2007)

2.1.3.2. Minería de Datos

Según lo manifestado por Molina y García (2006) la Minería de Datos “es un término genérico que engloba resultados de investigación, técnicas y herramientas usadas para extraer información útil de grandes bases de datos” (p. 5). A pesar de que la Minería de Datos se la considera como una fracción del proceso KDD, se los identifica como uno solo. Los mismos autores señalaron que el concepto de Minería de Datos “es usado comúnmente por los estadísticos, analistas de datos, y por la comunidad de administradores de sistemas informáticos como todo el proceso del descubrimiento, mientras que el término KDD es utilizado más por los especialistas en Inteligencia Artificial” (Molina & García, 2006, p. 6).

Esta herramienta de análisis, de acuerdo a Villarino (2015) es de mucha utilidad y el principal medio para la obtención de conclusiones de información que se encuentra en bases de datos y para tal efecto se han creado gran cantidad de “algoritmos de aprendizaje estadístico y computacional como los árboles de clasificación (...) y métodos de mayor complejidad basados en ellos como bagging (...) y random forest (...)” (p. 5).

Sobre la Minería de Datos, Kendall y Kendall (2011) manifestaron:

La minería de datos puede identificar patrones que un humano no puede. O el encargado de tomar decisiones no puede ver un patrón, o tal vez no puede pensar en preguntar si existe ese patrón. Los algoritmos de minería de datos buscan patrones en los almacenes corporativos de datos mediante el uso de algoritmos. (p. 429).

Lo mencionado anteriormente indica que la Minería de Datos es de utilidad para la búsqueda y descubrimiento de modelos que no siempre están visibles.

En la Figura 6 se demuestra el concepto de Minería de Datos para definir un perfil de ventas, en el cual se aprecia que todos los datos externos (listas de correos, historial de compras, información personal de los clientes de una base de datos general), y otro tipo de información, que se recolecta, analiza y determina cuál es la mejor oferta para enviar a sus clientes.

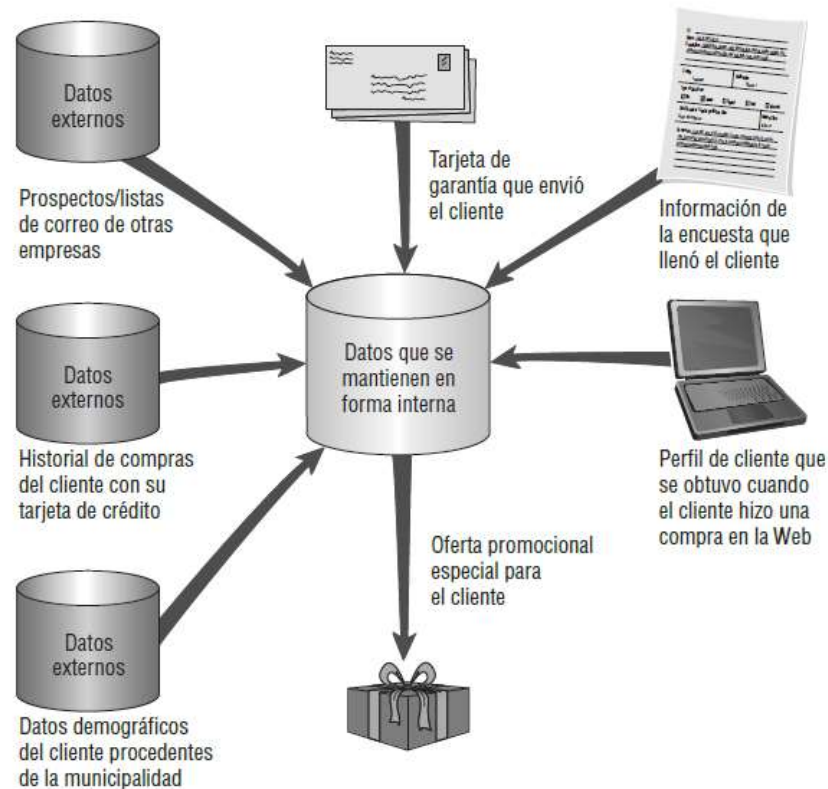


Figura 6: Ejemplo del concepto de Minería de Datos. Tomado de Kendall & Kendall (2011)

Por otro lado, según Febles-Rodríguez & González-Pérez (2002) existen algunos criterios de factibilidad para la aplicación de la Minería de Datos:

Factibilidad económica - organizativa: existe potencialmente un impacto significativo, no se conocen métodos alternativos, se dispone de personal calificado, no existen problemas de legalidad o violación de la información.

Factibilidad técnica: se dispone de suficientes datos, los datos contienen rasgos relevantes, existe poco ruido en los datos y se domina la aplicación de los métodos.

2.1.3.3. Técnicas de Minería de Datos

La herramienta de Minería de Datos ha permitido un progresivo reemplazo del análisis de información “dirigidos a la verificación por un enfoque de análisis de datos dirigido al descubrimiento del conocimiento” (Moreno, Miguel, García, & Polo, 2003, sec. Técnicas de Minería de Datos), diferenciándose ambos enfoques en que el último permite el descubrimiento de datos sin que se requiera plantear una hipótesis. Los algoritmos utilizados en Minería de Datos detectan patrones en la información, lo que

justifica que esta herramienta es de mayor eficiencia que cuando se lleva a cabo un análisis de verificación cuando se quiere examinar datos que vienen de grandes repositorios de datos y de gran diversidad.

Las técnicas de Minería de Datos se clasifican en: *predictivas o supervisadas* y *descriptivas*, las dos técnicas utilizan algoritmos para el descubrimiento del conocimiento (J. Hernández, Ramírez, & Ferri, 2004; Pérez, 2007).

Las *predictivas o supervisadas*, de acuerdo a Moreno et al., (2003):

Predicen el valor de un atributo (*atributos descriptivos*). A partir de datos cuya etiqueta se conoce se induce una relación entre dicha etiqueta y otra serie de atributos. Esas relaciones sirven para realizar la predicción en datos cuya etiqueta es desconocida. Esta forma de trabajar se conoce como aprendizaje supervisado y se desarrolla en dos fases: Entrenamiento (construcción de un modelo usando un subconjunto de datos con etiqueta conocida) y prueba (prueba del modelo sobre el resto de los datos). (sec. Técnicas de Minería de Datos).

Por su parte Hernández, Ramírez, & Ferri, (2004) manifestaron que estas técnicas buscan realizar una estimación de “valores futuros o desconocidos de variables de interés, que denominamos variables objetivo o dependientes, usando otras variables o campos de la base de datos” (p. 12) que se las conoce como variables predictivas o independientes. Las técnicas predictivas consistirían en realizar un clasificación de datos nuevos en base a los que se disponen previamente (Riquelme, Ruiz, & Gilbert, 2006).

Cuando se aplica las técnicas predictivas, se deben pasar las siguientes fases: 1) *identificación objetiva*, en la cual se emplean criterios para la identificación del modelo más apropiado para ajustar los datos, b) *estimación*, en donde se calculan las variables del modelo que se determinó en la fase anterior, c) *diagnosis*, cuando se valida el modelo determinado, d) *predicción*, “proceso de utilización del modelo identificado, estimado y validado para predecir valores futuros de las variables dependientes” (Pérez, 2007, p. 8).

Por otro lado, en las *técnicas descriptivas* “identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos (J. Hernández et al., 2004, p. 12).

El modelo a utilizar permite descubrir y reconocer patrones y directrices en datos presentes sin utilizar los históricos (Moreno et al., 2003).

Lo que se trata de explicar con los dos tipos de técnicas es que son de utilidad para realizar la predicción de las salidas de los datos y descubrir cuáles son las relaciones que existen entre ellos, con el uso de herramientas que utilizan algoritmos que permiten revelar patrones que no se ven a simple vista, agrupaciones, irregularidades o combinaciones por el volumen de datos que se encuentran almacenados en los grandes repositorios; además depuran datos importantes de dichos repositorios (Riquelme et al., 2006).

Pérez (2007) propuso una clasificación sobre las técnicas de Minería de Datos, que se visualizan en la Figura 7.

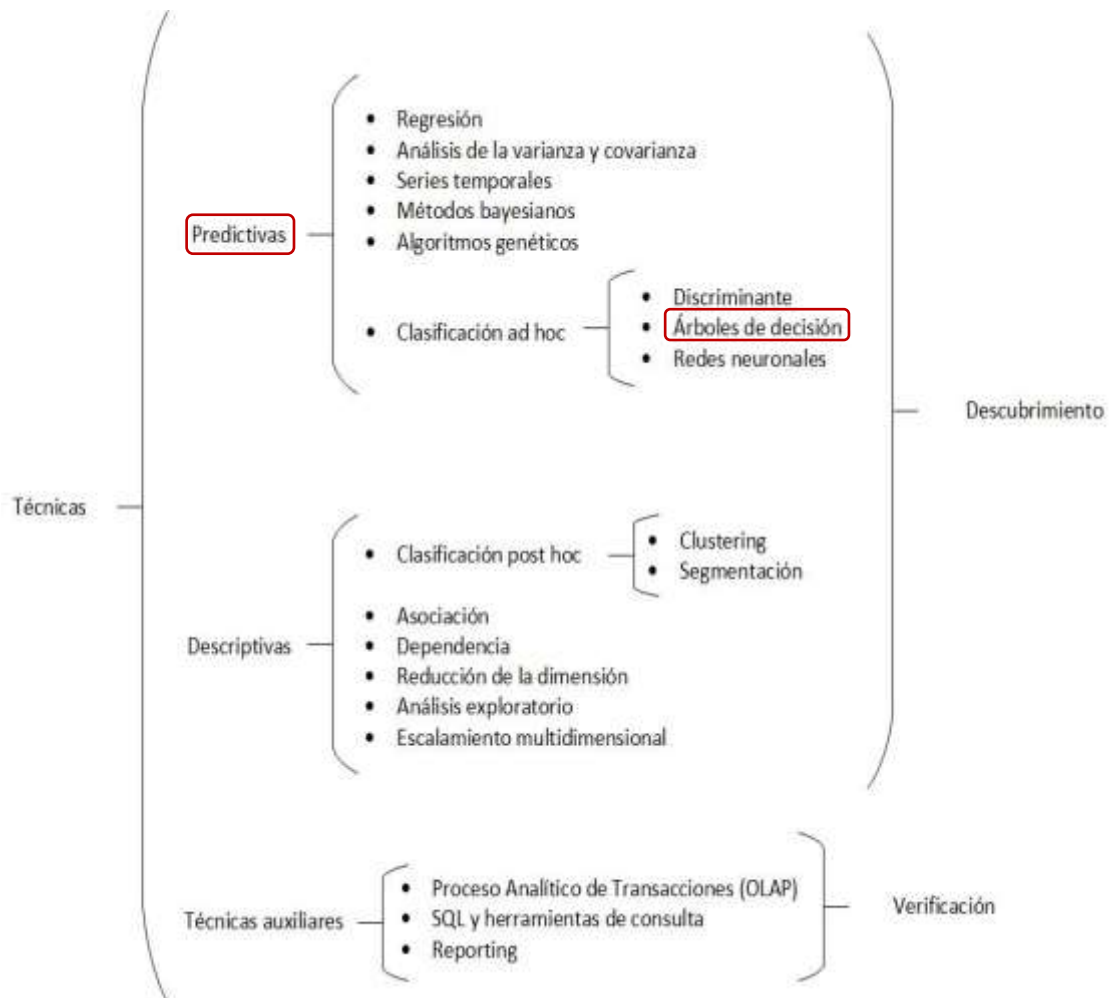


Figura 7: Algunas técnicas de Minería de Datos. Nota: Tomado de Pérez (2007)

Este proyecto emplea las técnicas predictivas, de las cuales se toma los árboles de decisión para los cálculos de retención de estudiantes en la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG, durante los dos primeros años de la carrera.

2.1.4. Infraestructura tecnológica

En cuanto a la infraestructura tecnológica se analizan: los algoritmos C4.5, Cj 48 de Weka, el teorema de Bayes y las herramientas de análisis predictivo.

En cuanto al **Algoritmo C4.5**, de acuerdo a Heredia, Amaya, y Barrientos (2015) el algoritmo C4.5 sirve para la generación de los árboles de decisión:

Es una extensión del algoritmo ID3, el cual trabaja únicamente con valores discretos en los atributos. En cambio, el algoritmo C4.5 permite trabajar con

valores continuos, separando los posibles resultados en dos ramas: una para aquellos $A_i \leq N$ y otra para todos los $A_i > N$. De esta forma, C4.5 genera un árbol de decisión a partir de datos mediante particiones realizadas de manera recursiva. (sec. Metodología).

El algoritmo C4.5 facilita: a) aplicación la concepción razón de ganancia, b) generación de árboles de decisión cuando existan datos con información desconocida para los atributos, c) utilizar atributos de valores continuos, d) poda de árboles de decisión, e) conseguir reglas de clasificación (Molina & García, 2006).

En lo relacionado con el algoritmo **J48 de Weka**, se conoce que es una implementación del C4.5. En este algoritmo, una parte importante de la configuración es “el factor de confianza para la poda, *confidence level*, puesto que influye notoriamente en el tamaño y capacidad de predicción del árbol construido” (Jesús García & Molina, 2012, pp. 196–197).

Los antes mencionados autores explicaron esta forma de construcción de este parámetro se realiza:

Para cada operación de poda, define la probabilidad de error que se permite a la hipótesis de que el empeoramiento debido a esta operación es significativo. Cuanto más baja se haga esa probabilidad, se exigirá que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar. El valor por defecto de este factor es del 25%, y conforme va bajando se permiten más operaciones de poda y por tanto llegar a árboles cada vez más pequeños. Otra forma de variar el tamaño del árbol es a través de un parámetro que especifica el mínimo número de instancias por nodo, si bien es menos elegante puesto que depende del número absoluto de instancias en el conjunto de partida (Jesús García & Molina, 2012, pp. 196–197).

En cuanto a las **redes bayesianas**, García, Martín, Gerassis, Saavedra y Taboada (2017) manifestaron que “modelizan fenómenos aleatorios mediante la representación de un conjunto de factores y las relaciones de dependencia entre ellas. Estas relaciones se representan a través de modelos gráficos, habitualmente grafos acíclicos dirigidos” (pp. 2–3).

El Teorema de Bayes asume, a través de la descomposición, “que un nodo depende solamente de sus padres, entendiendo por nodo padre aquél del cual es descendiente mediante un arco del grafo G. Dicho de otra forma, el nodo padre es el nodo causa y su descendiente es nodo efecto” (J. F. García et al., 2017).

Sobre las **Herramientas de Análisis Predictivo**, según la información que proporciona anualmente el grupo Gartner, sobre el Listado de Plataformas Data Science y Machine Learning punteras en el mercado 2018 (LIS Solutions, 2018), muestra a KNIME se encuentra puntera entre las preferencias en el mercado. Hay que recordar que:

El Cuadrante Mágico de Gartner es una culminación de la investigación en un mercado específico, que le brinda una visión panorámica de las posiciones relativas de los competidores del mercado. Al aplicar un tratamiento gráfico y un conjunto uniforme de criterios de evaluación, un Cuadrante Mágico lo ayuda a determinar rápidamente qué tan bien los proveedores de tecnología están ejecutando sus visiones declaradas y qué tan bien se están desempeñando en relación con la visión del mercado de Gartner. (Gartner, 2019, párr. 1).

Por lo tanto, utilizar la información que proporciona esta herramienta es de utilidad para determinar la herramienta de análisis predictivo que es más óptima para el proyecto.



Figura 8: Herramientas punteras en Data Science y Machine learning. Nota: Tomado de (LIS Solutions, 2018)

2.1.5. Herramientas de desarrollo

Para el desarrollo del proyecto se utilizan algunas herramientas de programación, tales como (1) Interfaz gráfica y (2) Herramientas.

2.1.5.1. Interfaz gráfica

Se entiende como interfaz de usuario al “conjunto de formas y métodos que posibilitan a los usuarios la interacción con un sistema, empleando para esto gráficos e imágenes” (Alegsa, 2016, párrs. 1–2). Estas formas incluyen íconos, fuentes, ventanas y otros que sirven para representar las distintas acciones que realiza el sistema.

2.1.5.2. Herramientas

Una herramienta para desarrollo de software es un programa, aplicación o instrucción que se utiliza para servir de soporte en la realización de diversas tareas y realizarlas de manera más fácil y sencilla (Sistemas, s/f).

Existen diversas herramientas utilizadas por los desarrolladores para crear sistemas de información, entre las cuales se encuentran NetBeans, Eclipse, Visual Studio. En la Tabla 1 se presenta una comparación entre estas herramientas.

NetBeans es un “entorno gratuito para desarrollo de programas Java estándar y applets. Facilita la tarea de desarrollo de aplicaciones empresariales Java EE o J2EE, permitiendo la compilación, empaquetado, distribución y ejecución dentro del mismo entorno” (Garrido Abenza, 2015, p. 6).

Al igual que NetBeans, **Eclipse** “es un entorno de desarrollo gratuito que ayuda en el desarrollo de programas en Java SE y Java EE, así como otros lenguajes por medio de pluggins” (Garrido Abenza, 2015). Facilita la ejecución de “herramientas para la gestión de espacios de trabajo, escribir, desplegar, ejecutar y depurar aplicaciones” (Genbeta, 2014, párr. 2).

Visual Studio

es la herramienta de desarrollo multilenguaje más completa para construir e integrar rápidamente aplicaciones y servicios Web XML. Aumenta de un modo extraordinario la productividad de los desarrolladores y crea nuevas oportunidades de negocio. En su diseño se han integrado a fondo los estándares y protocolos de Internet, como XML y SOAP, por lo que Visual Studio .NET simplifica considerablemente el ciclo de vida del desarrollo de aplicaciones.

En la tabla 1 se presenta la comparación de las herramientas antes mencionadas.

Tabla 1: Comparación de los lenguajes de programación

Características	Visual Studio	NetBeans	Eclipse
Descripción	IDE para sistemas operativos Windows. aplicaciones, sitios y aplicaciones web, así como servicios web en cualquier entorno que soporte la plataforma .NET	IDE desarrollado para lenguaje Java	IDE para desarrollo de aplicaciones de cliente enriquecido, basadas en navegadores. Para entornos integrados. IDE extensible; plataforma ligera para componentes de software
Plataforma	.NET	Multiplataforma	Multiplataforma
Lenguaje	Multilenguaje	Orientado a objetos	Orientado a objetos
Costo		Open source	Open source
Requerimientos	Procesador Pentium II 450 MHz como mínimo, RAM: 64 MB para Windows NT Workstation 4.0; 160 MB para Windows NT Server 4.0; 96 MB para Windows 2000 Professional; 192 MB para Windows 2000 Server; 160 MB para Windows XP Professional; 3,5 GB de espacio de disco duro disponible, que incluye un mínimo de 500 MB en la unidad del sistema	Requerimientos de hardware: espacio en disco 125 Mbytes, memoria: 256 Mbytes, procesador: 500 Mhz, software: JDK 1.3 o superior	Mínimo 1Ghz de procesador, memoria 512 MB, disco de 150 Mb
Ventajas	Construcción de aplicaciones .NET, entorno unificado de desarrollo integrado, Creación de potentes aplicaciones de bases de datos, Versiones completas de tecnologías Servidor, con licencia para desarrollo y pruebas únicamente	Reutilización de módulos, uso de Update Center Module, instalación y actualización simple, incluye templates y wizards, soporte para PHP, menor consumo de memoria, refactorización, soporta JavaScript	Para aplicaciones profesionales, aunque también se utiliza para actividades didácticas por su versión gratuita
Soporte a lenguajes	Visual Basic .NET, Visual C#™ .NET, Visual C++® .NET, Visual J#™ .NET	Java, C/C++ PHP, Groovy, JavaScript, JSP, Ruby, Perl, Go XML, HTML5, CSS, Otros	Java, C/C++, PHP, HTML5, JavaScript, JSP, Perl, Python, SQL, CSS

Nota: Adaptado de Mauleon (2013), Chambi (2016), Cobos (2014), Arias (2015)

2.2.Sustento Legal

El proyecto se sustenta en algunas leyes y normativas. Entre éstas se encuentran la **Constitución de la República** (Asamblea Nacional Constituyente, 2008), Capítulo segundo, Derechos del buen vivir, Sección quinta, Educación.

Art. 26.- La educación es un derecho de las personas a lo largo de su vida y un deber ineludible e inexcusable del Estado. Constituye un área prioritaria de la política pública y de la inversión estatal, garantía de la igualdad e inclusión social y condición indispensable para el buen vivir... (p. 16).

Otro de los documentos legales que enmarcan el proyecto es la **Ley Orgánica de Educación Superior** (LOES), Título I Ámbito, Objeto, Fines y Principios del Sistema de Educación Superior Capítulo 1 Ámbito y Objeto:

Art. 2.- Objeto. - Esta Ley tiene como objeto definir sus principios, garantizar el derecho a la educación superior de calidad que propenda a la excelencia, al acceso universal, permanencia, movilidad y egreso sin discriminación alguna (CES, 2010, p. 7).

Art. 5.- Derechos de las y los estudiantes. - Son derechos de las y los estudiantes los siguientes:

b) Acceder a una educación superior de calidad y pertinente, que permita iniciar una carrera académica y/o profesional en igualdad de oportunidades;

d) Participar en el proceso de evaluación y acreditación de su carrera...(CES, 2010, p. 7).

En cuanto al **Modelo Pedagógico-Educativo de la UCSG**, se considera el Título I, De la Misión y Visión de la Universidad Católica de Santiago de Guayaquil y de la Representación de su Modelo Pedagógico, sobre los ejes que sostienen el Modelo Educativo-Pedagógico.

Art. 4.- Estos ejes, en su integración y complementariedad configuran: tipos de enseñanza, aprendizajes, evaluaciones, formación y los rasgos o características de los

actores centrales: autoridades académicas, docentes, estudiantes, funcionarios, esto es, la comunidad universitaria.

Enseñanza: democrática, participativa, reflexiva, basada en la formulación, análisis y resolución de problemas, con énfasis en la investigación formativa y en el empleo de las nuevas tecnologías de la información y la comunicación.

Aprendizaje: colaborativo, reflexivo, con énfasis en el ejercicio del estudio autónomo, la investigación y las prácticas pre profesionales.

Proceso de aprendizaje: proceso de reflexión en la acción. El proceso propicia y estimula la reflexión sobre la cultura, la investigación entendidos en su connotación más amplia. La transferencia y la co-construcción del conocimiento, desde la reflexión, deben caracterizar al proceso de enseñanza aprendizaje como premisa para la formación sólida e integral de los estudiantes.

Evaluación, mejora y transparencia: acciones sistemáticas pedagógicas, formativas, sumativas de proceso, con medida de resultados del aprendizaje... (UCSG, 2013, pp. 3–4).

2.3.Ámbito de Aplicación

El proyecto se aplica a la Carrera de Gestión Empresarial Internacional, ahora Negocios Internacionales, por su rediseño.

2.3.1. Gestión Empresarial Internacional (Negocios Internacionales)

En el año 1992 se crea la carrera Gestión Empresarial Internacional con la finalidad de formar profesionales en Economía, Comercio Exterior, Administración e Idiomas, profesionales de este tipo, de gran demanda en el mercado laboral. Los conocimientos adquiridos en dichas ramas del conocimiento contribuyen en el cumplimiento de funciones administrativas, ante un panorama ampliamente globalizado, ya que la influencia de fuerzas internacionales está presente, además, en organizaciones nacionales, por lo que es indispensable que existan graduados que tengan formación en técnicas para intercambio comercial (UCSG, 2018b).

Debido a las disposiciones de los entes reguladores educativos, se ha realizado el rediseño de las carreras en la UCSG, por lo que la carrera de Gestión Empresarial Internacional cambio a Negocios Internacionales.

Su misión:

Formar profesionales socialmente responsables en el manejo de los negocios internacionales de las organizaciones a través del desarrollo de competencias en el ámbito económico, administrativo, financiero e investigativo, las nuevas tecnologías de información, técnicas de negociación y el dominio de los idiomas español, francés e inglés acorde con a las exigencias de los diferentes entornos cambiantes y globalizados. (UCSG, 2018c).

Su visión:

Ser una carrera universitaria líder en el campo de los Negocios Internacionales con presencia nacional e internacional, en función de los avances de la tecnología y los cambios económicos administrativos y tecnológicos del mundo contemporáneo. (UCSG, 2018c).

En cuanto a su organización interna, en la Figura 9 se muestra el organigrama de la Facultad de Ciencias Económicas y Administrativas.

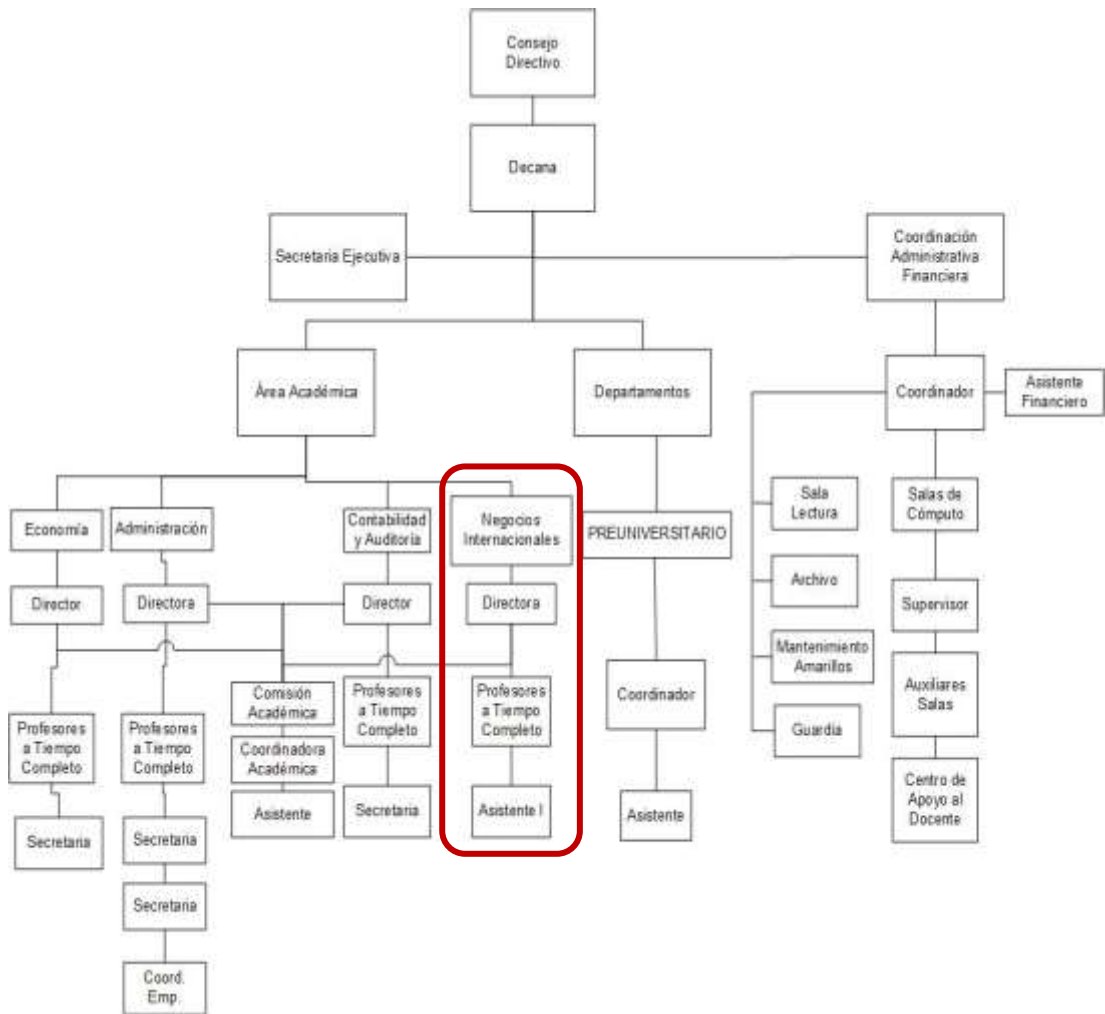


Figura 9: Organigrama de la Facultad de Ciencias Económicas y Administrativas, Carrera de Gestión Empresarial Internacional (Negocios Internacionales)

2.3.1.1. Proceso de acreditación

Con fecha 13 de octubre de 2016, el Consejo de Educación Superior (CES) recibe el proyecto de la Carrera de Negocios Intencionales, presentado por la Universidad Católica de Santiago de Guayaquil (UCSG), por medio de la Plataforma Informática para la presentación de Carreras y Programas. Con fecha 17 de abril de 2017, realiza la evaluación mediante la Plataforma de Presentación de Carreras del CES, y se genera el informe técnico de evaluación de la carrera de Negocios Intencionales realizado por los Facilitadores Académicos Externos.

Mediante plataforma informática del CES el día 24 de abril de 2017, se pone en conocimiento de la UCSG, las observaciones de la evaluación realizada por los Facilitadores Académicos Externos del proyecto de la carrera de Negocios

Internacionales, para que de ser pertinentes se valoren, subsanen y se incluyan las respuestas a las observaciones realizadas en el proyecto académico.

Con fecha 24 de mayo de 2017 se recibe a través de la Plataforma Informática de Presentación de Carreras y Programas, las respuestas a las observaciones realizadas por los Facilitadores Académicos Externos al proyecto de la carrera de Negocios Internacionales, el mismo que incluye la valoración de las observaciones realizadas.

El 13 de septiembre, se recibe el proyecto final de la carrera de Negocios Internacionales, presentado por la Universidad Católica de Santiago de Guayaquil (UCSG).

Código del proyecto	1028-5-6504181A01-3514		
Carrera a rediseñar	<ul style="list-style-type: none"> • GESTION EMPRESARIAL INTERNACIONAL 		
Código de la carrera a rediseñar	<ul style="list-style-type: none"> • 01298 		
Tipo de formación	Licenciaturas		
Tipo de trámite	Re-diseño		
Lugar donde se impartirá el programa	Sede matriz UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL		
Campo amplio	04 - Administración		
Campo específico	041 - Educación comercial y administración		
Campo detallado	04181 - Negocios Internacionales		
Nombre de la carrera	NEGOCIOS INTERNACIONALES		
Título que otorga	LICENCIADO/A EN NEGOCIOS INTERNACIONALES		
Modalidad de aprendizaje	Presencial		
No. de períodos ordinarios	9		
No. de períodos extraordinarios	0		
No. de horas de períodos extraordinarios	0		
No. de semanas por período académico	16		
No. de asignaturas	49		
No. de horas por componente de docencia	2534		
No. de horas por componente de aplicación y experimentación	362		
No. de horas por componente de aprendizaje autónomo	3439		
No. de horas de prácticas pre profesionales	305		
No. de horas de vinculación con la sociedad	160		
No. de horas de trabajo de titulación	400		
Duración de la carrera	7200		
No. de paralelos	2		
No. máximo de estudiantes por paralelo	30		
Jornadas de trabajo	0		
No. Resolución OCAS y fecha de aprobación	065-2016 2016-09-27		
No. Itinerarios Académicos	2		
Itinerarios Académicos	Sostenibilidad de los Negocios Internacionales Relaciones Internacionales		
Proyección de la matrícula	Año	Período I	Período II
	1	60	108
	2	100	90
	3	84	80
	4	72	64
	5	60	60
Convenios de prácticas pre profesionales	Institución	Fecha de suscripción	Estado
	CONVENIO ESPECIFICO DE PASANTIAS GENERALES Y PRÁCTICAS VACACIONALES ENTRE LA UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL Y LA SUPERINTENDENCIA DE TELECOMUNICACIONES	16/05/2013	Vigente
	CONVENIO MARCO DE PASANTIAS ENTRE EL SERVICIO DE RENTAS INTERNAS Y LA UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL	20/08/2015	Vigente
	CONVENIO MARCO DE COOPERACIÓN INSTITUCIONAL ENTRE LA UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL Y LA ASOCIACIÓN COORDINADORA DEL VOLUNTARIADO -ACORVOL	11/09/2014	Vigente
	CONVENIO ESPECÍFICO PARA DESARROLLAR PASANTIAS ENTRE LA UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL Y 3M ECUADOR C.A	13/05/2015	Vigente

Figura 10: Datos informativos del rediseño de la Carrera. Nota: Información proporcionada por la carrera.

2.3.1.2. Tasas de retención

Sobre las tasas de retención de estudiantes en la UCSG, se conoció información desde el año 2013. El cálculo de la tasa de retención se realiza con la cantidad de alumnos correspondiente a una determinada cohorte, es decir, la cantidad que ingresó

dividido para el número que salió de la universidad; la información fue proporcionada por el área de Informática y Estadística de la UCSG.

El porcentaje de las tasas de retención de estudiantes a nivel universitario se encuentra dentro de los parámetros de las IES, y el porcentaje de retención de la Carrera de Negocios Internacionales tiene otro comportamiento, que sería útil se analice y se incluya entre las actividades, métodos para aumentar este porcentaje.

CAPÍTULO III:

METODOLOGÍA Y RESULTADOS

El levantamiento de información realizado para este proyecto permitió realizar el diseño y posterior implementación de la propuesta, y se la obtuvo a través de la metodología de investigación correspondiente y de desarrollo de sistemas. A continuación, se describen, paso a paso, las fases de los procesos antes señalados.

3.1. Metodología de la Investigación

La investigación científica permite construir conocimientos y teorías con base en la ciencia, permitiendo que cada uno de éstos se constituya en una oportunidad para crear nuevos que lleven a dar respuestas y buscar soluciones a problemas de orden social y natural.

En el ámbito de la investigación científica la primera clasificación que debe hacerse es el enfoque que tiene ésta: cuantitativa o cualitativa. En algunos casos una investigación puede ser cuanti-cualitativa, es decir, toma la característica de mixta. Cabe, entonces, establecer algunas diferencias y conceptualizaciones sobre el enfoque de investigación.

A decir de Bernal Torres (2010, p. 60), hay una gran diferencia entre estos dos enfoques:

- La investigación **cuantitativa** “se fundamenta en la *medición* de las características de los fenómenos sociales, lo cual supone derivar de un marco conceptual pertinente al problema analizado, una serie de postulados que expresen relaciones entre las variables estudiadas de forma deductiva. Este método tiende a generalizar y normalizar resultados”.
- Mientras que la investigación **cualitativa** “se orienta a *profundizar* casos específicos y no a generalizar (...) cualificar y describir el fenómeno social a partir de rasgos determinantes, según sean percibidos por los elementos mismos que están dentro de la situación estudiada”.

Para ampliar el campo de estudio de la investigación **cualitativa**, y como complemento a lo aseverado por Bernal, cabe mencionar a Hernández Sampieri, Méndez Valencia, Mendoza Torres, y Cuevas Romo (2017, p. 114) quienes aseguran que este enfoque de investigación propiamente está orientado a analizar procesos, comportamientos, experiencias humanas, aspectos culturales y sociales, historias de vida, entre otros, en donde la intención es llegar a una exploración o descripción de percepciones de un grupo social, ante ello, “la muestra, la recolección de los datos y el análisis pueden sufrir modificaciones de acuerdo con las eventualidades que surjan y la forma en que evolucione el trabajo en el campo” (2017, p. 114).

Con base en lo anterior, la investigación objeto de este estudio tiene un enfoque **cualitativo** ya que está orientada a realizar un análisis de procesos que ocurren dentro de una unidad académica para determinar los niveles de retención de sus estudiantes, con el fin de proponer una técnica utilizando la minería de datos para el diseño de este proceso. Debido a lo mencionado, este estudio es analítico, exploratorio y descriptivo.

Es analítico porque busca revisar los componentes de cada proceso para encontrar sus causas y efectos, así como sus características propias para llegar a una comprensión global del objeto; es exploratorio porque a lo interno del ámbito de estudio no se ha realizado investigaciones relacionadas directamente con la información utilizada para el diseño de nuevos procesos y aplicativos tecnológicos que permitan determinar los niveles de retención estudiantil en la unidad académica beneficiaria del producto final de este proyecto; finalmente, es descriptivo, porque además de identificar el paso a paso de cómo se maneja la información de las variables que intervienen en el proceso de retención estudiantil, se realiza una revisión y actualización de dicho proceso para manejo futuro de directivos y autoridades de la unidad académica.

Todo lo mencionado en el párrafo anterior tiene su base en algunas conceptualizaciones elaboradas por expertos en el tema de la investigación científica: es **analítica** porque está orientada a “organizar los resultados o descubrimientos relacionados con un fenómeno o experiencia humana que está bajo investigación” (R. Hernández, Fernández, & Baptista, 2014, p. 426); es **descriptiva** porque “se reseñan las características o rasgos de la situación o fenómeno objeto de estudio” (César A. Bernal, 2010, p. 113); es **exploratoria** ya que busca “examinar un tema o problema de

investigación poco estudiado, del cual se tienen muchas dudas o no se ha abordado antes (...) o bien, si deseamos indagar sobre temas y áreas desde nuevas perspectivas”.

Por otro lado, en una investigación cualitativa, la población a abordar puede o no ser necesaria, dependiendo del contexto que se indaga. Para el caso de este estudio, que se trata de analizar el proceso que se sigue para determinar los niveles de retención estudiantil, para diseñar una vista minable con técnicas de minería de datos que presente resultados en el momento en que sea solicitado, no fue necesario establecer población ni muestra, solamente se identificó las personas clave de este proceso y los interesados en los resultados y se los denominó *informantes* a quienes se aplicó la técnica de la *entrevista* que es considerada una técnica que tiene como base el diálogo entre el entrevistador y el informante acerca de un tema preestablecido y con apoyo de una guía de preguntas que, pueden o no, ser ampliadas según como vaya presentándose la necesidad de más información, siempre dentro del tema de investigación y las variables que intervienen en el estudio, respetando las funciones propias de este tipo de técnicas de recolección de datos de manera que se cumpla con los fines originalmente identificados (F. Arias, 2012; Lara Muñoz, 2013); las preguntas de la entrevista han sido incluidas en el Apéndice B.

El resultado del análisis realizado a las entrevistas, se presenta en el apartado correspondiente de este capítulo.

3.2. Metodología de desarrollo

En cuanto a la metodología de desarrollo, se utilizó el **prototipo** que, de acuerdo a Pressman (2010, p. 37) es común que los clientes especifiquen cuáles son los objetivos que tendrá el software a desarrollar, pero no sepa determinar los requerimientos; o también se puede dar el caso de que el programador no considere que su algoritmo de su sistema sea lo suficientemente eficiente, adaptable o permita la fácil interacción con los usuarios finales. Surge, entonces, la necesidad de crear prototipos para buscar una mejor y rápida solución, la cual ayudará a comprender y clarificar a todos los involucrados en el sistema los requerimientos que aún no son comprensibles. Este proceso se conoce como paradigma de hacer prototipos.

El proceso antes mencionado inicia con una reunión entre todos los involucrados para determinar los objetivos del sistema, definir requerimientos básicos y especificar cuáles serían los requerimientos y descubrir los temas en los cuales se necesita definición mayor. Se proyecta enseguida “una iteración para hacer el prototipo, y se lleva a cabo el modelado (en forma de un ‘diseño rápido’). Éste se centra en la representación de aquellos aspectos del software que serán visibles para los usuarios finales (...)” (Pressman, 2010, p. 37). Mediante el diseño rápido se construye el prototipo que se lo muestra a los usuarios finales para su retroalimentación y se definen nuevamente requisitos no comprendidos previamente.

La idea de desarrollar un prototipo es la identificación de los requisitos que va a tener el sistema, generando de forma rápida un programa que funcione y muestre lo que en realidad se necesita, por lo que es muy útil en el desarrollo de este proyecto ya que, en caso de un posterior incremento de requerimientos de los ya previstos, se podrá descartar el prototipo realizado y desarrollar otro.

En la Figura 12 se presenta el proceso del paradigma de hacer prototipos.

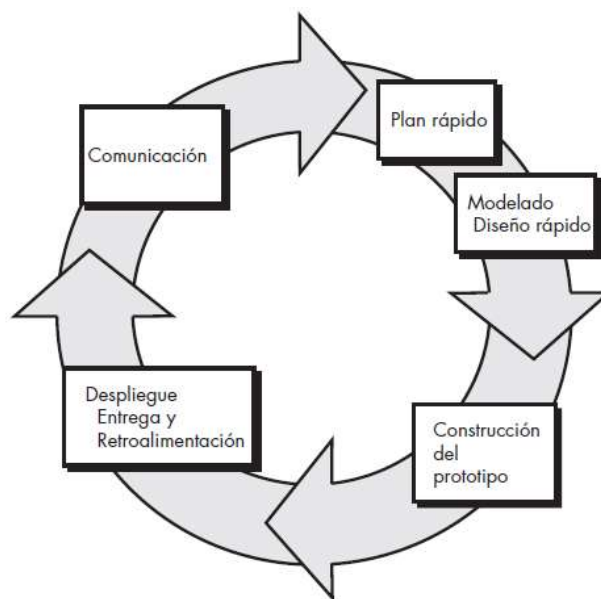


Figura 11: El paradigma de hacer prototipos. Nota: Tomado de: Pressman (2010)

3.3. Metodología de Minería de Datos

Las fases del proceso de Minería de Datos se describen en los párrafos a continuación.

3.3.1. Integración y selección

En esta etapa, luego de que se han determinado los conocimientos más importantes y que se han establecido los objetivos del proceso de KDD visto desde la perspectiva del usuario final, se elabora un grupo “de datos objetivo, seleccionando todo el conjunto de datos o una muestra representativa de este, sobre el cual se realiza el proceso de descubrimiento. La selección de los datos varía de acuerdo con los objetivos del negocio” (Timarán et al., 2016, p. 65).

Esto significa que el grupo de datos seleccionados para el proyecto se los obtuvo del sistema, en donde se encuentra el registro histórico de los estudiantes, con su información, de los cuales se escogió los que se consideran necesarios para la vista minable.

De acuerdo a Pacheco y Fernández (2015, p. 27), en esta etapa se cumplen tres tareas, las mismas que se explican a continuación.

- ***Aprendizaje del tema.***, es necesario descubrir “el proceso detrás de la generación de la información para poder formular las preguntas correctas, seleccionar las variables relevantes a cada pregunta, interpretar los resultados y sugerir el curso de acción después de concluido el análisis”;
- ***Recolección de datos.*** Se deberá encontrar los datos precisos, “cómo fueron obtenidos los datos de varias fuentes, cómo se pueden combinar estos datos y el grado de confianza de cada fuente”; y,
- ***Experiencia en análisis de datos.*** Se deberán conocer técnicas estadísticas.

3.3.2. Exploración y procesamiento

Esta etapa se refiere a la “aplicación de técnicas de análisis exploratorio de datos, buscando la distribución de los datos, simetría, normalidad y correlaciones existentes en la información” (UIAF, 2014, p. 16).

3.3.3. Limpieza

De acuerdo a Pacheco y Fernández (2015) “esta etapa se realiza con el fin de eliminar valores irrelevantes al objetivo, valores redundantes e inconsistencias en los datos de varias fuentes al juntarlos dentro de una sola base de datos” (p. 27).

Es, en esta etapa, en donde se tratan los datos, de manera que de la información recogida se utilice lo necesario, eliminando los datos que no sean relevantes, mediante la aplicación de las técnicas correspondientes.

3.3.4. Transformación/reducción de datos

Según lo manifestado Fayyad et al., (1996) citados por Timarán et al., (2016) en esta etapa se trata de buscar “características útiles para representar los datos dependiendo de la meta del proceso. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables bajo consideración o para encontrar representaciones invariantes de los datos” (p. 66). Dichos métodos de reducción son *horizontales* (suprimir listas ordenadas iguales resultantes del reemplazo del valor de la característica por otra de alto nivel) y *verticales* (suprimir atributos repetidos o intrascendentes como columnas, llaves funcionalmente dependientes) (Timarán et al., 2016).

3.3.5. Minería de datos

En esta fase se determina cuál es la tarea que se va a llevar a cabo, que pueden ser “clasificación, agrupación, otras y se elige la técnica descriptiva (clustering y segmentación, escalamiento, reglas de asociación y dependencia, análisis exploratorio...) o predictiva (regresión y series temporales, análisis discriminante, métodos bayesianos, algoritmos genéricos, árboles de decisión, redes neuronales)” (UIAF, 2014, p. 17).

Lo que se utiliza en el proyecto son los árboles de decisión y el algoritmo C4.5.

3.3.6. Evaluación e interpretación de resultados

Se refiere a cómo se interpretan los resultados que se consiguieron después de utilizar y ejecutar el algoritmo escogido en la fase de minería de datos (Pacheco & Fernández, 2015). Los expertos analizan y evalúan los patrones obtenidos y en caso de que se requiera, se puede regresar a etapas previas y ejecutar otras iteraciones (UIAF, 2014).

3.3.7. Difusión y uso de modelos

Se aplica el nuevo conocimiento y se involucra a todos los posibles usuarios (UIAF, 2014). Luego de que el modelo se construyó y validó, se lo utiliza para dos fines: “para que un analista recomiende acciones basándose en el modelo y en sus resultados, o bien para aplicar el modelo a diferentes conjuntos de datos” (J. Hernández et al., 2004, p. 39). Puede darse el caso de que se necesite asociarse a otras aplicaciones.

Si el modelo es una aplicación automática o manual, se requiere la difusión de la misma, esto significa que comparta con los usuarios finales y el conocimiento adquirido será parte del know-how (saber cómo) empresarial. Por otro lado, es necesario evaluar el buen funcionamiento del modelo, hasta comprobar su verdadera utilidad, ya que es posible que los patrones cambien y el modelo deberá volver a evaluarse, entrenarse o volver a construirse (J. Hernández et al., 2004).

3.4. Análisis de Resultados

De acuerdo a las entrevistas realizadas se pudo concluir que actualmente no existe un protocolo establecido para identificar a los estudiantes que tienen menor probabilidad de continuar sus estudios, y la mayoría de las selecciones que se realizan se hacen con referencia a la parte académica del alumno, y no se analizan sus aspectos socioeconómicos que también influyen en esta decisión.

Para obtener la información de los estudiantes se realizó un oficio dirigido al Rector, con la autorización y firma de la Decana de la Facultad de Ciencias Económicas y Administrativas, la Directora de la Carrera de Negocios Internacionales, y la Coordinadora Académica 1. El Sr. Rector autorizó la disposición de la información siempre y cuando se excluyan los parámetros referentes a la identificación, valores monetarios, y cantidades de estudiantes, motivo por el cual no se detallan en el presente documento.

La información fue proporcionada por el Director del Departamento de Centro de Cómputo de la UCSG, en dos archivos en formato Excel, el primero de los cuales incluía información socio-económica, y el segundo información académica, identificando a cada estudiante con un código único.

Luego de recibida la información se efectuó su análisis, constatando la cantidad de registros, realizando un ordenamiento y reestructuración de los datos, con la finalidad que sean compatibles para su unión y posterior análisis, proceso mediante el cual se determinó que existían columnas que no podrían ser utilizadas para el estudio debido a que tenían ausencia de datos mayores al 20%, y datos que no eran relevantes para el presente estudio.

Como conclusión y según lo expuesto previamente, la Carrera Negocios Internacionales de la UCSG posee información oportuna para la aplicación de procesos de Minería de Datos y determinar patrones y perfiles para mejorar la tasa de retención de sus estudiantes.

CAPÍTULO IV:

PROPUESTA

En los párrafos que se encuentran a continuación, se presenta la propuesta metodológica de utilización de técnicas de Minería de Datos para la determinación de la tasa de retención de estudiantes de la Carrera de Negocios Internacionales (anterior Gestión Empresarial Internacional), cuyos resultados de su aplicación permitirá a las autoridades una toma de decisiones más acertada en cuanto al estudio de las causas de la deserción estudiantil.

4.1. Introducción

Obtener el conocimiento minado de datos a través de las técnicas de minería de datos, se vuelve cada vez más inherente conforme se fue diseñando esta propuesta, ya que nos permite analizar las diferentes variables a considerar y calcular no solo la predicción sino la probabilidad de que esa predicción ocurra.

El proceso de KDD no solo implica la obtención y análisis de los datos sino la fase de corrección de datos erróneos o incompletos, donde se decide el tratamiento de los mismos. El propósito de la fase de minería de datos es decidir que tarea se utiliza para el proceso de extracción de conocimiento, eligiendo el método más adecuado, lo cual permitirá generar un modelo de predicción para analizar la nueva información que se le proporcione y pronosticar una única variable.

4.2. Objetivo

Determinar el nivel de deserción-retención estudiantil de los estudiantes de la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas, mediante la generación de un modelo predictivo con técnicas de Minería de Datos, para ayudar en predicciones futuras.

4.3. Responsables

El requerimiento está a cargo de la directora de la Carrera Negocios Internacionales y la Coordinadora Académica 1.

4.4. Descripción de la solución

La propuesta a presentar tuvo dos fases. En primer lugar, el análisis de la información y creación del modelo predictivo a través de técnicas de Minería de Datos, para extraer sus reglas, y una segunda fase, la implementación de las reglas del modelo a través de una interfaz que identifique los estudiantes con mayor probabilidad de deserción en la Carrera.

4.4.1. Proceso

El proceso actual de reconocimiento de los estudiantes que tienen problemas en su rendimiento realizado por la Directora de la Carrera, se lo realiza de forma manual, el mismo que se lo puede apreciar en la Figura 13.

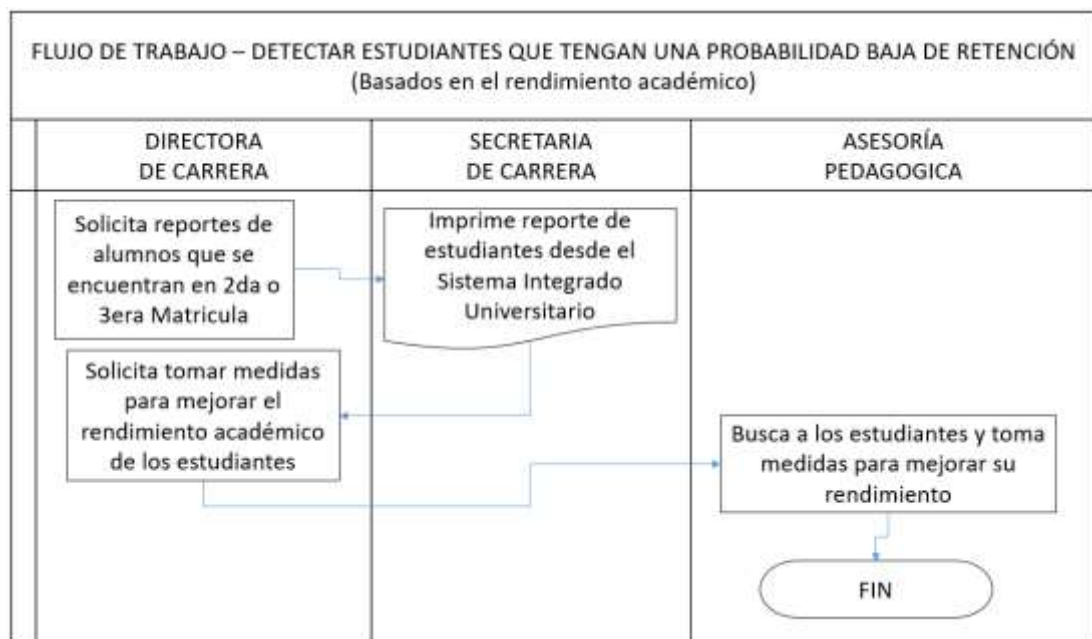


Figura 12: Proceso actual

Para dar una solución automatizada al proceso manual, se propone la implementación de una interfaz en la cual, a pesar de suponerse un proceso un poco más complejo en su ejecución, se integra una solución más amplia en el análisis de la tasa de retención (ver Figura 14).

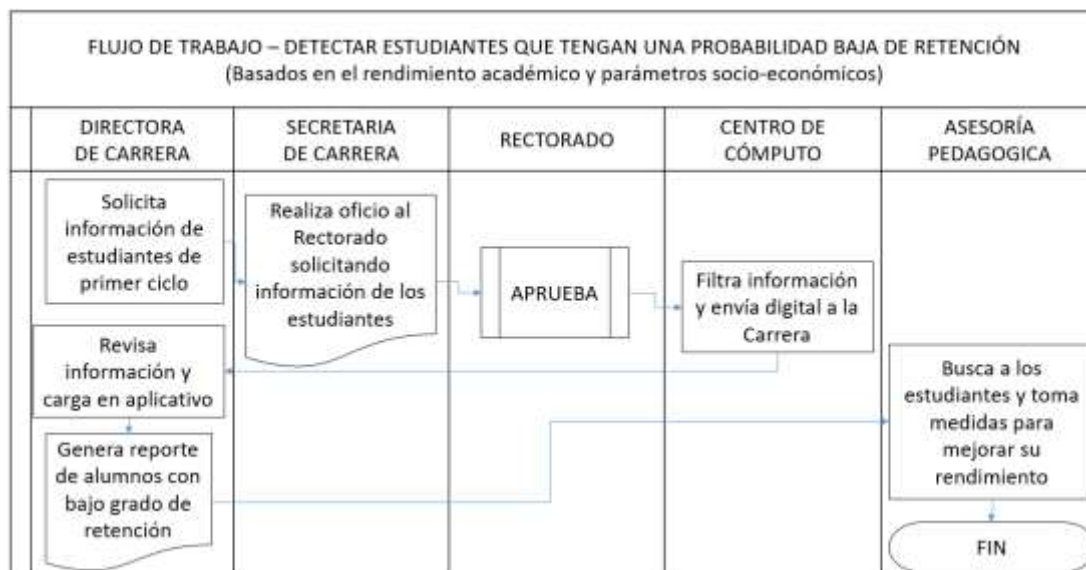


Figura 13: Solución

Para desarrollar la aplicación que contendrá la solución se deberá identificar las reglas del modelo de predicción que se conseguirán a través del proceso de Minería de datos (KDD), el cual se explica a continuación.

4.4.1.1. Fase 1: Minería de Datos

La metodología que se usa para la extracción de conocimiento es “Descubrimiento de conocimiento en base de datos” (KDD), debido a que la minería de datos es una fase dentro de la metodología KDD, que permite la extracción del modelo predictivo que servirá para la toma de decisiones.

Para cumplir con los procesos establecidos para el descubrimiento de la información a través de las bases de datos (KDD), se deberá realizar los procesos de selección, preprocesamiento, limpieza (transformación), minería de datos, e interpretación de los resultados.

La herramienta para **Minería de Datos** que ha quedado mejor posicionada en el Cuadrante Mágico de Gartner como líderes sobre las plataformas de Data Science y Machine Learning para el 2018 ha sido Knime, motivo por el cual es utilizada para este proyecto. Entre las características de Knime se incluyen su código abierto, función multiplataformas y la disponibilidad de diferentes algoritmos de predicción que son útiles para el análisis de la información.

Proceso de selección

El desarrollo de la metodología se basa en el análisis del conjunto de datos obtenidos, y determinar los más esenciales para el estudio del parámetro de retención. Para analizar la retención de estudiantes de los primeros años de estudio en la Carrera Negocios Internacionales, las autoridades de UCSG y la Carrera, proporcionaron la información correspondiente a la cohorte de estudiantes 2016-2018, considerando que existe información confidencial que no fue proporcionada.

De acuerdo al grado de inferencia con respecto a la retención del estudiante se clasifica las variables de acuerdo a la Tabla 3:

Tabla 2: Clasificación de las variables

Variables a considerar	Variables de menor relevancia
COD_ESTUDIANTE	NACIONALIDAD
EDAD	CARRERA
SEXO	TRANSPORTE
ESTADO_CIVIL	TIPO_COLEGIO
UBICACIÓN_COLEGIO	PENSIÓN_COLEGIO
REALIZÓ_ESTUDIOS_ADICIONALES	BECA_OTORGADA_EN_COLEGIO
TRABAJA	VIVE_PADRE
TIENE_HIJOS	ESTADO_CIVIL_PADRE
PORCENTAJE_BECA	INTRUCCIÓN_PADRE
PROM_DIAS_RETRASO	VIVE_MADRE
RETENCIÓN	ESTADO_CIVIL_MADRE
APROBADAS	INTRUCCIÓN_MADRE
REPROBADAS	
PROMEDIO	
SEGUNDA	
TERCERA	

Adicionalmente, después de realizar un análisis de los datos faltantes se descartaron Provincia, Alojamiento y Beca otorgada del colegio a universidad (examen SNNA).

En la Figura 16 se muestran el total de variables y sus porcentajes, en donde se señalan las variables no relevantes por falta de datos, los datos descartados y los datos utilizados a pesar de faltar información.

Porcentaje de datos faltantes por columnas	
COD_ESTUDIANTE	0%
EDAD	0%
NACIONALIDAD	0%
SEXO	0%
CARRERA	0%
ESTADO_CIVIL	0%
PROVINCIA	63%
ALOJAMIENTO	75%
TRANSPORTE	19%
UBICACIÓN_COLEGIO	16%
TIPO_COLEGIO	16%
PENSION_COLEGIO	19%
BECA_OTORGADA_COLEGIO_UNIVERS	100%
BECA_OTORGADA_EN_COLEGIO	19%
REALIZO_ESTUDIOS_ADICIONALES	16%
TRABAJA	97%
TIENE_HIJOS	0%
VIVE_PADRE	16%
ESTADO_CIVIL_PADRE	20%
INSTRUCCIÓN_PADRE	20%
VIVE_MADRE	16%
ESTADO_CIVIL_MADRE	19%
INSTRUCCIÓN_MADRE	19%
PORCENTAJE_BECA	0%
PROM_DIAS_RETRASO_PAG_PEN	0%
RETENCION	0%
	No relevantes por falta de datos
	Datos descartados
	Datos utilizados a pesar de faltar información

Figura 145: Porcentaje de datos faltantes por columnas

Proceso de Pre procesamiento (limpieza)

Debido a que la calidad de los datos no sólo depende del algoritmo de minería de datos sino también de la calidad de los datos minados, se excluyeron las columnas cuyo porcentaje de datos era mayor al 20%, pero para el caso de la variable Trabaja no se la excluyó debido a que es un dato relevante para el estudio.

Como medida adicional y debido a la existencia de datos en blanco que se han presentado se usa la moda para completarlos.

Proceso de transformación

Debido a que existen datos repetidos por cada semestre se reduce la cantidad de registros a un único registro por código de estudiante, pasando a ser éste clave primaria y adicionando el campo de retención que depende si se encuentra o no estudiando en el Semestre B-2018.

Para los datos de la información académica se asignan rangos de valor que permitan el mejor funcionamiento e interpretación del algoritmo en la siguiente etapa de minería de datos, los mismos que se aprecian en la Tabla 4.

Para la clasificación de la edad se seleccionó dos rangos los menores a 19 años y los mayores o iguales a 19 años, debido a que el promedio de los estudiantes que ingresan a la universidad tienen entre 17 y 18 años, considerando como un “adulto” a los estudiantes que tienen más de 18 años.

Tabla 3: Rangos de valor

Clasificación de los datos por rangos		
Columnas	Rangos de valores	Detalle del rango
Edad	<25=1 ; >24=2	<19 standard o >= 19 adulto
Aprobadas	mayor a 31=EXCELENTE	Los estudiantes que ingresaron en el período académico A-2016 deberían tener aprobadas 40 materias para ser alumno regular, y los que ingresaron en el B-2016 deberían tener 32 materias, por este motivo el rango excelente establece con estos valores.
	mayor a 16=MUYBUENO	
	mayor a 8=BUENO	
	mayor a 0=REGULAR	
	0=NINGUNA	
Reprobadas	mayor a 6=ALTO	Para clasificar estos valores se tomó en consideración la mediana sobre los valores mayores a 0 que resultado ser 3.5.
	mayor a 3=MEDIO	
	mayor a 0=BAJO	
	0=NINGUNA	
Promedio	mayor a 9=EXCELENTE	De acuerdo al Instructivo para la aplicación de la evaluación estudiantil, del Ministerio de educación se le asignó los rangos según su escala cualitativa.
	mayor a 6=BUENO	
	mayor a 4=REGULAR	
	mayor a 0=BAJO	
Segunda	0=NINGUNO	Para clasificar estos valores se tomó en consideración la mediana sobre los valores mayores a 0 que resultó ser 1.
	mayor a 3=ALTO	
	mayor a 1=MEDIO	
	mayor a 0=BAJO	
Tercera	0=NINGUNA	Para clasificar estos valores se tomó en consideración la mediana sobre los valores mayores a 0 que resultó ser 1.
	mayor a 3=ALTO	
	mayor a 1=MEDIO	
	mayor a 0=BAJO	
	0=NINGUNA	

Para la aplicación de cualquiera de los algoritmos de minería de datos se debe realizar un análisis de la variable a predecir. Las distintas tareas que ofrece la minería de datos se ordenan en su mayoría en dos grandes grupos, las descriptivas y las predictivas, siendo estas últimas las de interés para el estudio, debido a que se necesita predecir la variable de retención.

Proceso de Minería de Datos

Para el proceso de minería de datos se utilizó dentro de las técnicas predictivas las tareas de clasificación mediante arboles de decisión y los métodos bayesianos, los cuales sirven para realizar comparaciones durante el estudio.

Figura 15: Vista minable

Los árboles de decisión sirven para organizar las variables que se filtraron en la fase de selección, limpieza y transformación, y organizarlas en una estructura jerárquica que determine la probabilidad de retención siguiendo las condiciones que se cumplen desde la raíz hasta las hojas como se muestra en la Figura 17.

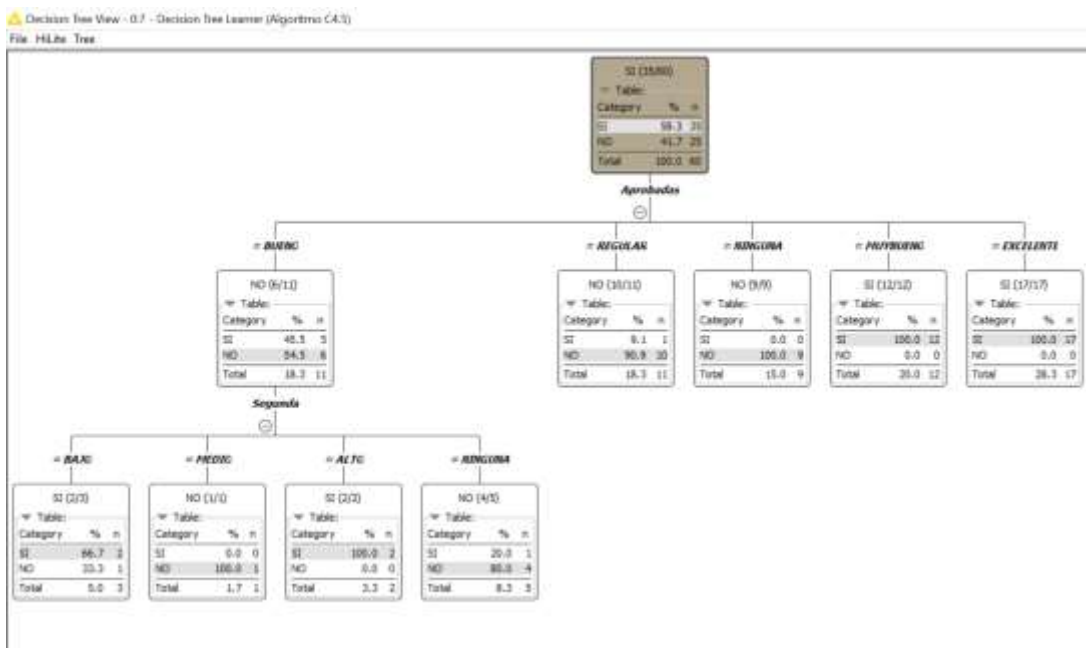


Figura 16: Estructura jerárquica del modelo

Proceso de Interpretación

Para realizar las comparaciones de los algoritmos se utilizaron gráficos circulares (ver Figura 18,19 y 20). Los algoritmos J48 y C4.5 utilizan las técnicas de árboles de

decisión dentro de las tareas de predicción, y ambos arrojan resultados iguales en su predicción, motivo por el cual no se detalla su análisis.

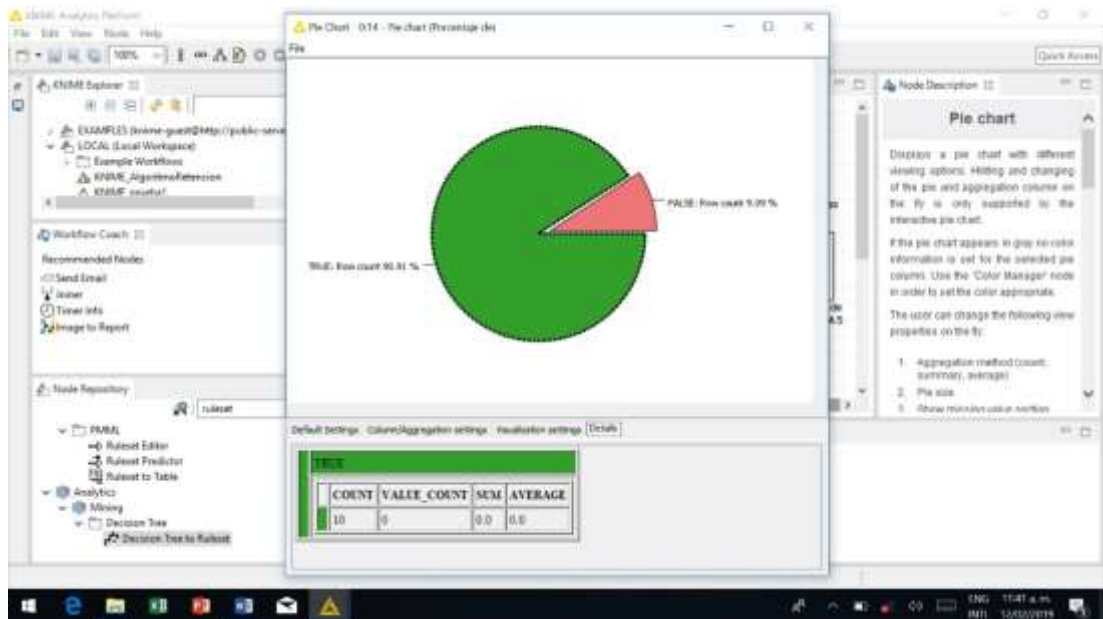


Figura 17: Comparación con algoritmo J48 (1)

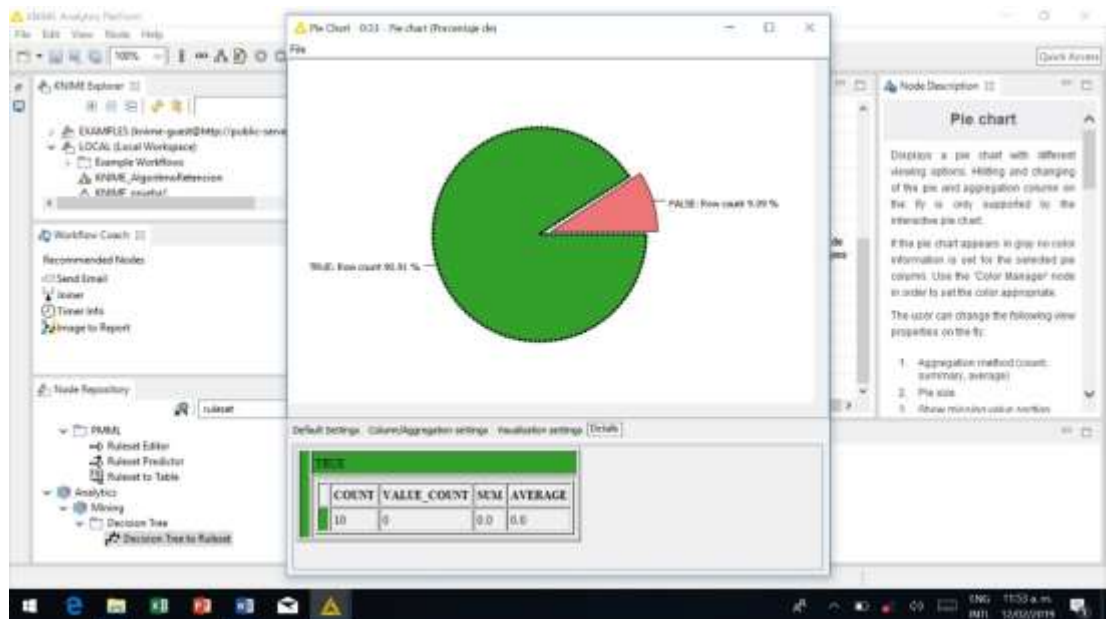


Figura 18: Comparación con algoritmo J48 (2)

Comparación de C4.5 con Teorema de Bayes

El Teorema de Bayes a pesar de que fue un algoritmo muy usado, para este caso de estudio no presenta un resultado confiable ya que su porcentaje de error en la predicción es del 18.18%, siendo inferior al C4.5 por más del 9%, lo cual es representativo para el presente estudio.

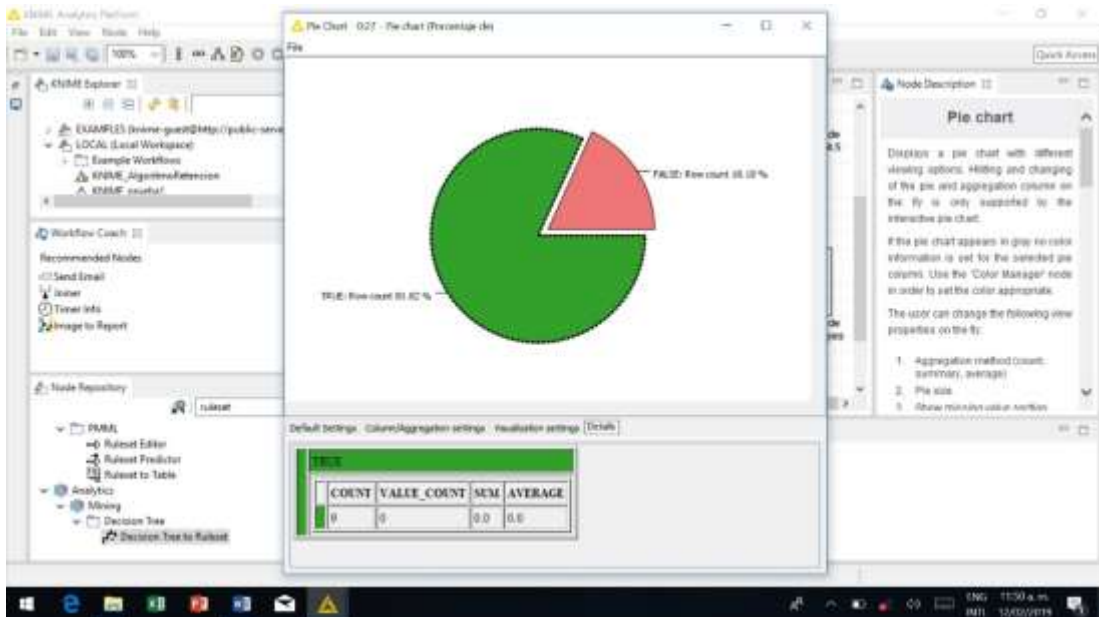


Figura 19: Teorema de Bayes

El resultado generado determino que el algoritmo C4.5 tenía un mayor grado de aciertos en sus predicciones, motivo por el cual fue escogido para extraer las reglas, y su implementación en el aplicativo.

Generación de reglas

Para la preparación de la solución es necesario transformar el árbol de decisión en un sistema de reglas que se deber enfocar en la deserción del estudiante, aislando las reglas que conducen a este resultado y simplificándolas para una mejor comprensión al momento de ingresarlas en el código fuente de la aplicación (ver Figura 21).

File Hilite Navigation View

Table "default" - Rows: 8 Spec - Columns: 3 Properties Flow Variables

Row ID	S Rule	D Record ...	D Number ...
Row1	\$Segunda\$ = "BAJO" AND \$Aprobadas\$ = "BUENO" => "SI"	3	2
Row2	\$Segunda\$ = "MEDIO" AND \$Aprobadas\$ = "BUENO" => "NO"	1	1
Row3	\$Segunda\$ = "ALTO" AND \$Aprobadas\$ = "BUENO" => "SI"	2	2
Row4	\$Segunda\$ = "NINGUNA" AND \$Aprobadas\$ = "BUENO" => "NO"	5	4
Row5	\$Aprobadas\$ = "REGULAR" AND TRUE => "NO"	11	10
Row6	\$Aprobadas\$ = "NINGUNA" AND TRUE => "NO"	9	9
Row7	\$Aprobadas\$ = "MUYBUENO" AND TRUE => "SI"	12	12
Row8	\$Aprobadas\$ = "EXCELENTE" AND TRUE => "SI"	17	17

Figura 20: Conjunto de reglas

Las reglas que se simplifican se las muestra en la Tabla 4.

Tabla 4: Reglas a simplificar

$\$Segunda\$ = "MEDIO" \text{ AND } \$Aprobadas\$ = "BUENO" \Rightarrow "NO"$
$\$Segunda\$ = "NINGUNA" \text{ AND } \$Aprobadas\$ = "BUENO" \Rightarrow "NO"$
$\$Aprobadas\$ = "REGULAR" \text{ AND TRUE} \Rightarrow "NO"$
$\$Aprobadas\$ = "NINGUNA" \text{ AND TRUE} \Rightarrow "NO"$







Debido a que las variables de materias aprobadas se repetían, se simplificó estas reglas en una sola: $\$Segunda\$ = "MEDIO" \parallel "NINGUNA" \text{ AND } \$Aprobadas\$ = "BUENO" \Rightarrow "NO"$. Luego de lo cual el conjunto de reglas simplificadas quedaría como se muestra en la Tabla 5.

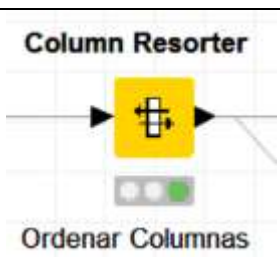

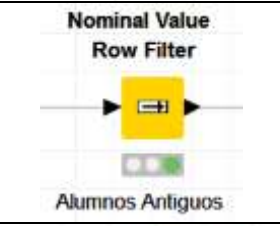
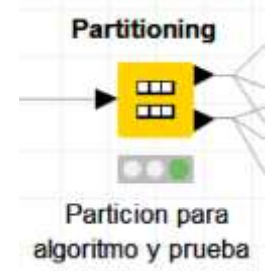



Tabla 5: Reglas simplificadas









Reglas simplificadas enfocadas en la probabilidad de deserción
$\$Segunda\$ = "MEDIO" \parallel "NINGUNA" \text{ AND } \$Aprobadas\$ = "BUENO" \Rightarrow "NO"$
$\$Aprobadas\$ = "REGULAR" \parallel "NINGUNA" \Rightarrow "NO"$

Las reglas mostradas en la Tabla 5 son las que se implementaron en el aplicativo, pero para registro de este proceso de minería de datos se detalla a continuación la ejecución de todos los nodos de manera secuencial, en la Tabla 6.

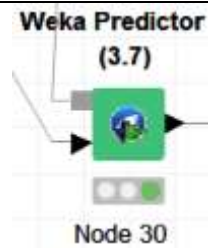
Tabla 6: Descripción de los nodos

Descripción de los nodos usados en Knime (el ordenamiento está directamente relacionado con el orden de procesamiento)		
#	Gráfico de Nodos	Descripción del proceso
1	<p>Excel Reader (XLS)</p>  <p>Lectura de archivo xls FINAL-1</p>	<p>Este nodo lee una de las hojas de cálculo proporcionadas por el Centro de Cómputo y la direcciona a su puerto de salida. Puede leer solo datos numéricos, de fecha, booleanos y de cadena; Establece un tipo de dato para cada columna, para lo cual KNIME actualmente admite String, Time And Date, Double, Boolean e Int. (en el peor de los casos, "String" cubre todo).</p>
2	<p>Column Filter</p>  <p>Filtra columnas no concluyentes</p>	<p>Con este nodo se filtra las columnas que por poseer bajo porcentaje de información no se pueden usar o que son redundantes para el análisis, mientras que solo las columnas restantes se pasan a la tabla de salida. Dentro del cuadro de diálogo, las columnas se pueden mover entre la lista Incluir y Excluir.</p>
3	<p>Missing Value</p>  <p>Agrega la moda</p>	<p>Con este nodo se trata los valores faltantes de las columnas que se incluyeron a pesar de tener valores faltantes, debido a que son muy representativas para el estudio, y para los cuales se está registrando la moda. La primera pestaña del cuadro de diálogo proporciona opciones de manejo predeterminadas para todas las columnas de un tipo determinado. La segunda pestaña permite configuraciones individuales para cada columna disponible (por lo tanto, anulando la predeterminada).</p>
4	<p>Row Filter</p>  <p>Filtro de Ruido</p>	<p>El nodo permite el filtrado de filas según ciertos criterios que generan ruido para el estudio.</p>
5	<p>Joiner</p>  <p>Unir las dos Tablas</p>	<p>Este nodo permite unir dos tablas. La unión se basa en las columnas de unión de ambas tablas, las cuales deben tener el mismo id, caso contrario aumentaría la cantidad de registro y generaría datos faltantes.</p>
5.1	<p>Extract Table Dimension</p>  <p>Tamaño de la Tabla</p>	<p>El nodo muestra el número de filas y columnas de la tabla de entrada, lo cual sirve para confirmar que el proceso anterior se ha realizado de manera exitosa.</p>
5.2	<p>Extract Table Spec</p>  <p>Variables de la Tabla</p>	<p>Este nodo extrae la metainformación de la tabla de entrada (nombres de columna, tipos, etc.) La metainformación se conoce como la especificación de datos de la tabla y contiene información como los tipos de datos que asigno a cada columna para verificar que los mismos sean los adecuados para el estudio.</p>

6	 <p>Column Resorter</p> <p>Ordenar Columnas</p>	<p>Este nodo cambia el orden de las columnas de entrada, en función de la configuración definida por el usuario. Las columnas se desplazan para una mejor comprensión de los datos. La tabla re-ordenada se proporciona en el puerto de salida.</p>
6.1	 <p>Nominal Value Row Filter</p> <p>Alumnos Nuevos</p>	<p>El nodo filtra las filas en función del valor seleccionado, lo cual nos permite identificar a los alumnos que serán descartados por no tener información necesaria para el estudio.</p>
7	 <p>Nominal Value Row Filter</p> <p>Alumnos Antiguos</p>	<p>El nodo filtra a los alumnos que cumplan con los parámetros necesarios para el análisis.</p>
8	 <p>Partitioning</p> <p>Particion para algoritmo y prueba</p>	<p>La tabla de entrada se divide en dos particiones (por número de filas). Las dos particiones están disponibles en los dos puertos de salida. La partición se realiza con muestreo estratificado por retención, es decir, la distribución de valores en la columna seleccionada se conserva (aproximadamente) en las tablas de salida. Las particiones servirán posteriormente para crear el árbol de decisión, y realizar las pruebas al mismo.</p>
9	 <p>Decision Tree Learner</p> <p>Algoritmo C4.5</p>	<p>Este nodo induce un árbol de decisión de clasificación en la memoria principal. El atributo de destino es la retención. Las divisiones nominales pueden ser binarias (dos resultados) o pueden tener tantos resultados como valores nominales. El algoritmo proporciona dos medidas de calidad para el cálculo dividido; El índice de gini y la relación de ganancia. Además, existe un método de poda posterior para reducir el tamaño del árbol y aumentar la precisión de la predicción. El método de poda se basa en el principio de longitud de descripción mínima.</p>
9.1	 <p>Decision Tree to Ruleset</p> <p>Reglas</p>	<p>Convierte el modelo de árbol de decisión a una tabla que contiene las reglas en forma de texto. Las reglas resultantes son independientes entre sí, el orden de las reglas no se especifica.</p>
10	 <p>Decision Tree Predictor</p> <p>Porcentajes de predicción</p>	<p>Este nodo usa un árbol de decisión existente (que se pasa a través del puerto modelo) para predecir el valor de la retención en los patrones que se escogieron para pruebas. El nodo se configura para que se añada columnas con distribución de clases normalizada para cada predicción.</p>

10.1	 <p>Scorer Falsos positivos y falsos negativos</p>	<p>El nodo compara la columna de predicción con la de retención, con el propósito de identificar los falsos positivos que pueden servir para el análisis de la predicción. Los valores de la primera columna seleccionada se representan en las filas de la matriz de confusión y los valores de la segunda columna mediante las columnas de la matriz de confusión. La salida del nodo es la matriz de confusión con el número de coincidencias en cada celda.</p>
10.2	 <p>Missing Value Comprobar salidas sin predicción</p>	<p>Este nodo ayuda a identificar las predicciones faltantes encontradas, lo cual puede ser útil para saber cuáles son las combinaciones que no está tomando en cuenta el algoritmo.</p>
11	 <p>Column Comparator Comparar resultados</p>	<p>Compara la columna de predicción con la de retención, y crea una nueva columna con las variables TRUE o FALSE que servirán para determinar la exactitud del algoritmo.</p>
12	 <p>Color Manager</p>	<p>Se asigna colores para los resultados de la columna generada en el nodo anterior.</p>
13	 <p>Pie chart</p>	<p>Muestra un gráfico circular con el porcentaje de exactitud del algoritmo sobre la partición de prueba. Si el gráfico circular aparece en gris, no se establece información de color para la columna circular seleccionada, esto pasaría en caso de presentarse alguna predicción no disponible.</p>
14	 <p>Naive Bayes Learner Node 16</p>	<p>El nodo crea un modelo bayesiano a partir de los datos de entrenamiento dados. Calcula el número de filas por valor de atributo por clase para atributos nominales y la distribución de Gauss para atributos numéricos. El nodo muestra un mensaje de advertencia si alguna columna se ignora debido a tipos de datos no compatibles.</p>
15	 <p>Naive Bayes Predictor Node 15</p>	<p>Predice la clase por fila según el modelo aprendido. La probabilidad de clase es el producto de la probabilidad por atributo y la probabilidad del propio atributo de clase.</p>
16	 <p>J48 (3.7) Node 31</p>	<p>Clase para generar un árbol de decisión de C4.5 podado o no podado, basado en Weka (herramienta para la minería de datos)</p>

17



El Weka Predictor toma un modelo generado en un nodo Weka y clasifica los datos de prueba en el inport.

El análisis de los algoritmos y la extracción de reglas para posterior implementación en la aplicación de escritorio se aprecia en la Figura 22.

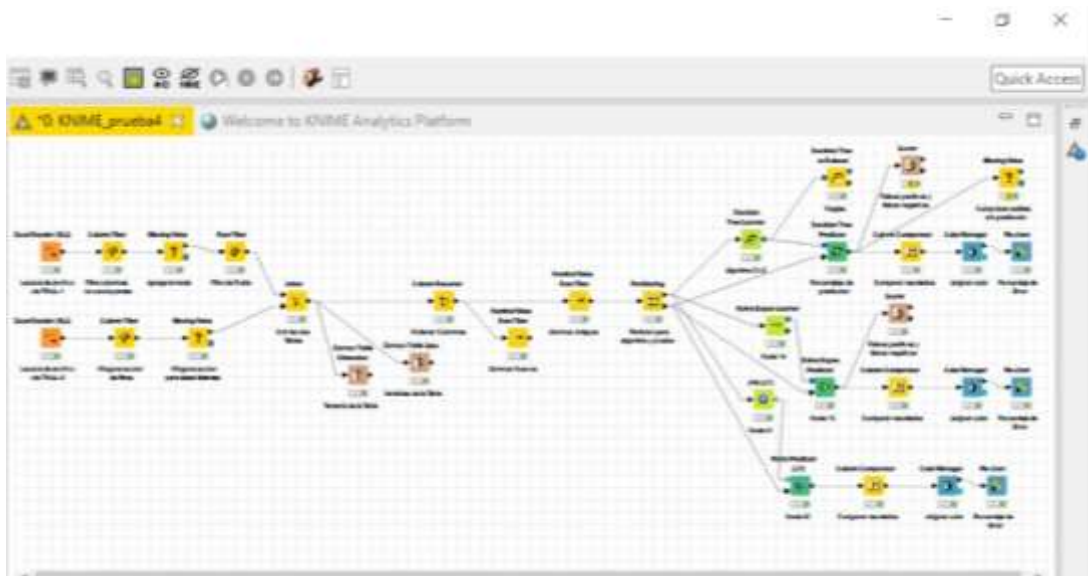


Figura 21: Solución

4.4.1.2. Fase 2: Implementación de las reglas

Sobre las herramientas tecnológicas que utilizamos en la fase de implementación hemos elegido la Interfaz gráfica Windows Forms y Visual Studio como lenguaje de desarrollo, debido a que es una herramienta de código abierto, fácil programación, y reconocida en el ambiente laboral y estudiantil de nuestro país.

El aplicativo consta de dos páginas, que son las de Login y la de ejecución y presentación del archivo con los resultados de los cálculos de la probabilidad de retención procesada mediante técnicas de Minería de Datos, que permite guardarlas en formato PDF, WORD o EXCEL, siendo este último el recomendado para su posterior análisis.

Las páginas del aplicativo están contenidas en un solo módulo que se ejecuta en el equipo asignado para tal actividad. (Ver Manual de Instalación y Ejecución de Aplicación).

4.4.2. Requerimientos de Hardware y Software

La solución informática es un aplicativo de escritorio, que se ejecutará en un ordenador ubicado en las oficinas de la Carrera, y será manejado por la Dirección de la Carrera. Con respecto a los requerimientos técnicos de software y hardware se detallan en la Tabla 7.

Tabla 7: Requerimientos Técnicos de Hardware y Software

Dato	Valor mínimo	Valor recomendado
Procesador	Pentium Celeron	Intel Core I3
Memoria RAM	2GB	4GB
Tamaño Almacenamiento	80MB	100MB
AccessDatabaseEngine	Versión 12.0	Compatible con V12.0
NetFramework	Versión 4.5	Superior a V4.5
Sistema operativo	Windows XP	Windows XP o superior

Debe considerarse que todas las herramientas utilizadas en el proyecto son *open source* por lo que no implica algún costo a la Carrera.

Las características del computador en el cual se encuentra la solución se detallan a continuación, y cabe recalcar que cumple con las solicitadas:

- RAM: 4GB
- Procesador Core i5
- 1 TB de disco
- Sistema operativo: Windows 7

En cuanto al software para el aplicativo, se les proporcionara los siguientes instaladores que son gratuitos y necesarios para la ejecución de la solución:

- NetFramework 4.5
- AccessDatabaseEngine 12

Sobre los requerimientos para KNIME se detallan a continuación a pesar de que en la computadora del usuario no se lo instalará, debido a que no es necesario para el proceso:

- Sistema operativo Linux o Windows
- Java 1.5 JRE o superior
- Memoria RAM: 2GB

4.4.3. Resultados esperados del aplicativo

Se espera proporcionar una herramienta eficiente, capaz de ubicar a los estudiantes de la carrera de Negocios Internacionales que tienen mayor probabilidad de deserción, implementando un modelo predictivo a ejecutarse en la aplicación de escritorio, la cual fue desarrollada e implementada en un práctico archivo ejecutable, el cual no requiere instalación y facilitará la toma de decisiones que ayuden en el aumento de la tasa de retención de la carrera.

El aplicativo presenta los estudiantes de la carrera de Negocios Internacionales que poseen mayor probabilidad de deserción, y genera un reporte en archivo PDF, WORD, o EXCEL, para su fácil manipulación.

Se espera crear un paradigma de uso de los modelos de minería de datos, para la mejora de análisis de datos y toma de decisiones de la educación en la Universidad Católica de Santiago de Guayaquil.

CONCLUSIONES

Luego de finalizar el desarrollo del proyecto, se pudo comprobar que el modelo predictivo basado en técnicas de Minería de Datos, permitió la evaluación e interpretación de toda la información de los estudiantes de la cohorte 2016-2018 de la Carrera de Gestión Empresarial Internacional, a través de la cual se determinó un modelo de la retención-deserción y es una herramienta de apoyo en la toma de decisiones de futuras proyecciones sobre este indicador.

Para la ejecución del proyecto, se levantó la información correspondiente al período académico 2016-2018 en la carrera, en donde se conoció el proceso manual que se sigue para conocer a los estudiantes que tienen problemas de rendimiento, al mismo tiempo se proporcionó las facilidades para que los datos de las tasas de retención de la carrera sean proporcionados por el Centro de Cómputo, de la cual se establecieron las variables que formaron parte del modelo predictivo.

Determinada las reglas de probabilidad de retención de los estudiantes de la carrera mediante técnicas predictivas de Minería de Datos, se diseñó la vista minable, en la cual se utilizaron las variables ya establecidas para el modelo predictivo tales como edad, sexo, estado civil, estudios adicionales, porcentaje de beca, día de retraso en pago de pensiones y otras variables, que ayudaron a reducir errores en el momento del diseño.

Las variables seleccionadas para la generación del modelo predictivo, permitieron establecer los patrones de evaluación e interpretación de la información que se obtuvo de los estudiantes, a través de las técnicas predictivas de Minería de Datos y los algoritmos correspondientes. La evaluación se la realizó a través del árbol de decisión a través del algoritmo C4.5, cuyos resultados se compararon con información resultante del algoritmo J48 y el Teorema de Bayes. Se realizó el modelo predictivo, cuyos resultados se presentaron en el aplicativo desarrollado para tal efecto, herramienta considerada como apoyo en la toma de decisiones para mejorar los servicios educativos de la carrera.

RECOMENDACIONES

Sería conveniente que, para posteriores estudios a realizarse sobre la retención-deserción de los estudiantes, se tome en consideración que la aplicación solo ejecuta las reglas del resultado del presente trabajo, por lo tanto, si se extraen nuevas reglas deberán ser implementadas en otra aplicación o modificadas en la actual.

Se sugiere una actualización de la ficha técnica del estudiante, para tener una información más completa sobre la situación socioeconómica del estudiante, ya que por la falta de datos solicitados se descartaron algunas variables debido a que su porcentaje de ausencia es alto.

En caso de que se realice un estudio posterior con las nuevas variables, o una cohorte más amplia, la aplicación generará nuevas reglas para lograr los resultados esperados.

REFERENCIAS

- Alegsa, L. (2016). Definición de GUI (Interfaz Gráfica de Usuario) [Diccionario tecnología]. Recuperado el 25 de enero de 2019, de <http://www.alegsa.com.ar/Dic/gui.php>
- Almeida, F. (2018). Visual C#.NET: Console Applications and Windows Forms, 63. Recuperado de https://www.researchgate.net/profile/Fernando_Almeida21/publication/326416931_Visual_CNET_Console_Applications_and_Windows_Forms/links/5b4c833f45851519b4c0b8cc/Visual-CNET-Console-Applications-and-Windows-Forms.pdf
- Arias, F. (2012). *El Proyecto de Investigación (Sexta)*. Caracas: Episteme.
- Arias, M. (2015). *Aprende a Programar ASP .NET y C# (Segunda)*. IT Campus Academy. Recuperado de <https://books.google.com.ec/books?id=9WnjCgAAQBAJ&pg=PA17&lpg=PA17&dq=c%23+es+escalable&source=bl&ots=khw719Ohj9&sig=lYEj-4wDzQbHS5qP0BEpARxi2zY&hl=es-419&sa=X&ved=2ahUKEwjkiYD8YLFhVsvFkKHbW8C1MQ6AEwCHoECAAQAQ#v=onepage&q=c%23%20es%20escalable&f=false>
- Asamblea Nacional Constituyente. Constitución de la República del Ecuador (2008). Recuperado de https://www.oas.org/juridico/pdfs/mesicic4_ecu_const.pdf
- Barrientos, F., & Ríos, S. A. (2013). Aplicación de Minería de Datos para Predecir Fuga de Clientes en la Industria de las Telecomunicaciones, *XXVII*, 36. Recuperado de <http://www.dii.uchile.cl/~ris/RIS2013/rios.pdf>
- Beneyto Sánchez, S. (2015). *Entorno familiar y rendimiento académico*. Alicante: 3Ciencias, Área de Innovación y Desarrollo.
- Bernal, César A. (2010). *Metodología de la investigación, administración, economía humanidades y ciencias sociales (Tercera)*. Bogotá: Pearson Educación de Colombia Ltda.

- Bernal, César Augusto. (2010). *Metodología de la investigación: administración, economía, humanidades y ciencias sociales*. Bogotá: Pearson Educación.
- Bot, D., & Cai, S. (2017). Información general sobre formularios Windows Forms. Recuperado el 10 de diciembre de 2018, de <https://docs.microsoft.com/es-es/dotnet/framework/winforms/windows-forms-overview>
- CACES. (2018a). Proceso de Evaluación de Carreras. Recuperado el 4 de noviembre de 2018, de <http://www.caces.gob.ec/web/ceaaces/carreras>
- CACES. (2018b). Proceso de Evaluación Institucional. Recuperado el 4 de noviembre de 2018, de <http://www.caces.gob.ec/web/ceaaces/institucional>
- Campo, K., Escorcía, D., Moreno, M., & Palacio, J. (2016). Metacognición, escritura y rendimiento académico en universitarios de Colombia y Francia. *Avances en Psicología Latinoamericana*, 34(2), 233–252.
<https://doi.org/10.12804/apl34.2.2016.03>
- Canales, A., & De los Ríos, D. (2018). Factores explicativos de la deserción universitaria. *Calidad en la Educación*, 0(26), 173–201.
<https://doi.org/10.31619/caledu.n26.239>
- CEAACES. (2016). CEAACES. Recuperado el 4 de noviembre de 2018, de <https://apps.ceaaces.gob.ec/profesoresColaboradores/paginas/inicio.jsf>
- CES. Ley Orgánica de Educación Superior LOES, Pub. L. No. 0 (2010). Recuperado de http://www.ces.gob.ec/index.php?option=com_phocadownload&view=file&id=651&Itemid=564
- CES. (2016). Propuesta de Indicadores para la evaluación institucional del Sistema de Educación Superior. Recuperado de <http://rauec.gob.ec/images/DIRECTRICESRAUEC/PROPUESTADEINDICADORES.pdf>
- Chambi, R. (2016). IDE's para programadores para crear cualquier tipo de programas. Recuperado el 6 de diciembre de 2018, de <http://www.gitmedio.com/gitmedio/ides-para-programadores/>

- Cobos, C. (2014). *Cuadro comparativo de herramientas de programación eclipse, java*. Educación. Recuperado de <https://es.slideshare.net/CCCRiis/cuadro-comparativo-javi>
- Corso, C., & Gebellini, F. (2012). Uso de herramienta libre para la generación de reglas de asociación, facilitando la gestión eficiente de incidentes e inventarios., 12. Recuperado de http://41jaiio.sadio.org.ar/sites/default/files/16_JSL_2012.pdf
- Cortés Campos, R. L., Zapata González, A., Menéndez Domínguez, V. H., & Canto Herrera, P. J. (2015). El estudio de los hábitos de conexión en redes sociales virtuales, por medio de la minería de datos. *Innovación educativa (México, DF)*, 15(68), 99–114.
- Díaz, I., Machado, G., Chacón, M., Rodríguez, M., Jiménez, O., & Jiménez, A. (2014). Rendimiento académico de estudiantes en asignaturas del área básica en policlínicos universitarios, 17. Recuperado de <http://scielo.sld.cu/pdf/edu/v7n1/edu08115.pdf>
- Donoso, S., & Cancino, V. (2018). Caracterización socioeconómica de los estudiantes de educación superior. *Calidad en la Educación*, 0(26), 205–244. <https://doi.org/10.31619/caledu.n26.240>
- EcuRed. (2018). Minería de Datos. Recuperado el 13 de noviembre de 2018, de https://www.ecured.cu/Miner%C3%ADa_de_Datos
- Febles-Rodríguez, J., & González-Pérez, A. (2002). Aplicación de la minería de datos en la bioinformática. *ACIMED*, 10(2), 69–76. Recuperado de http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1024-94352002000200003&lng=es&nrm=iso&tlng=en
- García, F., Fonseca, G., & Concha, L. (2015). Aprendizaje y rendimiento académico en educación superior: un estudio comparado. *Actualidades Investigativas en Educación*, 15(3). <https://doi.org/10.15517/aie.v15i3.21072>
- García, J. F., Martín, J. E., Gerassis, S., Saavedra, A., & Taboada, J. (2017). Análisis bayesiano de factores de riesgo de accidente en trabajos de movimientos de

tierras. *Informes de la Construcción*, 69(546), 192.

<https://doi.org/10.3989/ic.15.154>

García, Jesús, & Molina, J. (2012). Técnicas de análisis de datos. Aplicaciones prácticas usando Microsoft Excel y Weka, 266. Recuperado de <http://ocw.uc3m.es/ingenieria-informatica/analisis-de-datos/libroDataMiningv5.pdf>

García, Julio. (2018). *Cálculo de indicadores de deserción y permanencia en la Maestría en Administración de las Organizaciones*. UNAD (Maestría). Universidad Nacional Abierta y a Distancia, Colombia. Recuperado de <https://stadium.unad.edu.co/preview/UNAD.php?url=/bitstream/10596/17555/1/2993405.pdf>

Garrido Abenza, P. (2015). *Comenzando a programar con JAVA*. Universidad Miguel Hernández. Recuperado de <https://books.google.es/books?hl=es&lr=&id=4v8QCgAAQBAJ&oi=fnd&pg=PP1&dq=netbeans+vs+eclipse&ots=le0W3qqXiy&sig=UeD1HhHY4zoXPeKJ3wHAdGG2hpE#v=onepage&q=netbeans%20vs%20eclipse&f=false>

Gartner. (2019). Magic Quadrant Research Methodology. Recuperado el 7 de enero de 2019, de <https://www.gartner.com/en/research/methodologies/magic-quadrants-research>

Genbeta. (2014). Eclipse IDE. Recuperado el 25 de enero de 2019, de <https://www.genbeta.com/desarrollo/eclipse-ide>

Grijalva, P., Freire, V., Real, K., Arellano, A., & Cornejo, G. (2018). Aplicación de Técnicas de Minería de Datos para el Análisis de la Eficiencia Académica. *Suplemento especial*, 3, 16. Recuperado de <http://revistas.pucese.edu.ec/hallazgos21/>

Guérin, B.-A. (2016). *ASP.NET en C# con Visual Studio 2015: Diseño y desarrollo de aplicaciones Web*. Ediciones ENI. Recuperado de <https://books.google.es/books?hl=es&lr=&id=sp5v1QwgKGkC&oi=fnd&pg=PA19&dq=visual+studio+que+es&ots=u->

Q9Hu2uFM&sig=o0t8HJxo4EDddRBJIdyU8WX1Ed4#v=onepage&q=visua
l%20studio%20que%20es&f=false

Heredia, D., Amaya, Y., & Barrientos, E. (2015). Student Dropout Predictive Model Using Data Mining Techniques. *IEEE Latin America Transactions*, 13(9), 3127–3134. <https://doi.org/10.1109/TLA.2015.7350068>

Hernández, J., Ramírez, M., & Ferri, C. (2004). *Introducción a la minería de datos*. Madrid: Pearson Educación.

Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (Sexta). México: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.

Hernández Sampieri, R., Méndez Valencia, S., Mendoza Torres, C., & Cuevas Romo, A. (2017). *Fundamentos de investigación* (Primera edición). Ciudad de México: McGraw-Hill/Interamericana.

Himmel, E. (2018). Modelo de análisis de la deserción estudiantil en la educación superior. *Calidad en la Educación*, 0(17), 91–108. <https://doi.org/10.31619/caledu.n17.409>

Hugon, J. (2015). *C# 6: Desarrolle aplicaciones Windows con Visual Studio 2015*. Ediciones ENI. Recuperado de <https://books.google.es/books?hl=es&lr=&id=sp5v1QwgKGkC&oi=fnd&pg=PA19&dq=visual+studio+que+es&ots=u-Q9Hu2uFM&sig=o0t8HJxo4EDddRBJIdyU8WX1Ed4#v=onepage&q=visua>
l%20studio%20que%20es&f=false

Izquierdo, A., Bravo, L., Ceruto, T., & Marín, D. (2015). Nuevos pluggins para la herramienta Knime para el uso de sus flujos de trabajo desde otras aplicaciones. *Ciencias de la Información*, 46(1). Recuperado de <http://cinfo.idict.cu/index.php/cinfo/article/view/673/517>

Jara, D., Velarde, H., Gordillo, G., Guerra, G., León, I., Arroyo, C., & Figueroa, M. (2008). Factores influyentes en el rendimiento académico de estudiantes del primer año de medicina. *Anales de la Facultad de Medicina*, 69(3), 193–197.

- Johnson, D. (2015). Conceptos básicos de las aplicaciones de Windows Forms (Visual Basic). Recuperado el 10 de diciembre de 2018, de [https://docs.microsoft.com/es-es/previous-versions/ms172749\(v%3dvs.120\)](https://docs.microsoft.com/es-es/previous-versions/ms172749(v%3dvs.120))
- Kendall, K. E., & Kendall, J. E. (2011). *Análisis y diseño de sistemas* (Octava). México, DF: Pearson Educación.
- Kohler Herrera, J. L. (2013). Rendimiento académico, habilidades intelectuales y estrategias de aprendizaje en universitarios de Lima. *Liberabit*, 19(2), 277–288.
- Kohler, J. (2013). Rendimiento académico, habilidades intelectuales y estrategias de aprendizaje en universitarios de Lima. Recuperado de <http://www.scielo.org.pe/pdf/liber/v19n2/a13v19n2.pdf>
- Lara Muñoz, E. M. (2013). *Fundamentos de investigación: un enfoque por competencias* (Segunda edición). México, D.F.: Alfaomega.
- LIS Solutions. (2018). Gartner 2018: Listado de Plataformas Data Science y Machine Learning punteras en el mercado. Recuperado el 19 de diciembre de 2018, de <https://www.lis-solutions.es/blog/gartner-2018-listado-de-plataformas-data-science-y-machine-learning-punteras-en-el-mercado/>
- Martínez, G., & Pedroza, M. (2017). Factores incidentes en el rendimiento académico estudiantil de Ingeniería en Sistemas de Información de la FAREM-Matagalpa, 2012-2016. *Revista Científica de FAREM-Estelí*, 0(24), 5–26. <https://doi.org/10.5377/farem.v0i24.5549>
- Martínez, M., & Coronado, G. (2003). Indicadores para la evaluación integral de la productividad académica en la educación superior. *RELIEVE. Revista Electrónica de Investigación y Evaluación Educativa*, 9(1), 45–72. Recuperado de <http://www.redalyc.org/articulo.oa?id=91690103>
- Matheu, A., Ruff, C., Ruiz, M., Benites, L., & Morong, G. (2018). Modelo de predicción de la deserción estudiantil de primer año en la Universidad Bernardo O'Higgins. *Educação e Pesquisa*, 44(0). <https://doi.org/10.1590/s1678-4634201844172094>

- Mauleon, O. (2013). *Tabla comparativa de herramientas case*. Recuperado de <https://es.slideshare.net/oswaldoyuneri/tabla-comparativa-de-herramientas-case-oswaldo-mauleon>
- Molina, J., & García, J. (2006). *TÉCNICAS DE ANÁLISIS DE DATOS*, 266.
- Moreno, M., Miguel, L., García, F., & Polo, J. (2003). *Apoyo a la Decisión en Ingeniería del Software*. España. Recuperado de <http://ceur-ws.org/Vol-84/paper4.pdf>
- Msn. (s/f). ¿Qué es y para qué sirve Visual Studio 2017? Recuperado el 5 de febrero de 2019, de <https://www.msn.com/es-cl/noticias/microsoftstore/%C2%BFqu%C3%A9-es-y-para-qu%C3%A9-sirve-visual-studio-2017/ar-AAAnLZL9>
- Murillo, M. (2016). Riesgo de deserción por rendimiento académico y sus implicaciones para la salud mental desde la perspectiva de los estudiantes. *Congresos CLABES*, 0(0). Recuperado de <http://www.revistas.utp.ac.pa/index.php/clabes/article/view/1114>
- Pacheco, Y., & Fernández, Y. (2015). Aplicación de técnicas de descubrimiento de conocimientos en el proceso de caracterización estudiantil. *Ciencias de la Información*, 46(3), 25–30. Recuperado de <http://www.redalyc.org/articulo.oa?id=181443340004>
- Pérez, C. (2007). *Minería de datos: técnicas y herramientas* (Primera). Editorial Paraninfo. Recuperado de https://books.google.com.ec/books?id=wz-D_8uPFCEC&printsec=frontcover&dq=miner%C3%ADa+de+datos&hl=es-419&sa=X&ved=0ahUKEwj1-4-gvZjeAhXxtlkKHZQHAJYQ6AEIJAA#v=onepage&q=miner%C3%ADa%20de%20datos&f=false
- Porcel, E., Dapozo, G., & López, M. (2010). Predicción del rendimiento académico de alumnos de primer año de la FACENA (UNNE) en función de su caracterización socioeducativa. *Revista electrónica de investigación educativa*, 12(2), 1–21. Recuperado de

http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S1607-40412010000200007&lng=es&nrm=iso&tlng=es

- Pressman, R. S. (2010). *Ingeniería del software: un enfoque práctico* (Séptima). México: McGRAW-HILL INTERAMERICANA EDITORES, S.A. DE C.V.
- Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias, *10*(29), 8. Recuperado de <https://idus.us.es/xmlui/bitstream/handle/11441/43290/Miner%C3%ADa%20de%20datos.pdf?sequence=1&isAllowed=y>
- Robles, Y., & Sotolongo, A. (2013). Integración de los Algoritmos de Minería de Datos 1r, Prism e ID3 a Postgresql. *Journal of Information Systems and Technology Management*, *10*(2), 389–406. <https://doi.org/10.4301/S1807-17752013000200012>
- Rodríguez, K., Gutiérrez, A., Wong, T., & López, D. (2015). Eficiencia académica: un indicador del que se requiere conocer más. *EDUMECENTRO*, *7*(3), 188–194. Recuperado de http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S2077-28742015000300014&lng=es&nrm=iso&tlng=es
- Rodríguez, Y., & Díaz, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas*, *3*(3–4). Recuperado de <http://www.redalyc.org/resumen.oa?id=378343637009>
- Sistemas. (s/f). Definición de Herramienta. Recuperado el 5 de febrero de 2019, de <https://sistemas.com/herramienta.php>
- Timarán, S., Hernández, I., Caicedo, S., Hidalgo, A., & Alvarado, J. (2016). *Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional*. Bogotá: Universidad Cooperativa de Colombia. <https://doi.org/10.16925/9789587600490>
- Torres, L. (2012). *Retención estudiantil en la educación superior. Revisión de la literatura y elementos de un modelo para el contexto colombiano*. (Primera). Pontificia Universidad Javeriana. Recuperado de

<https://www.javeriana.edu.co/documents/15838/273636/Retenci%25C3%25B3nEstudiantil2012.pdf/124fdb5-2318-432a-8e9f-126a2501c229>

UCSG. Reglamento de Régimen Académico y Modelo Educativo-Modelo Pedagógico de la Universidad Católica de Santiago de Guayaquil (2013). Recuperado de http://www.ucsg.edu.ec/wp-content/uploads/transparencia/Reglamento_academico_pedagogicoUCSG.pdf

UCSG. (2018a). Acreditación Internacional. Recuperado el 5 de noviembre de 2018, de <http://www.ucsg.edu.ec/acreditacion-internacional/institucional/compromisos-ucsg/>

UCSG. (2018b). Historia de la Facultad. Recuperado el 23 de noviembre de 2018, de <http://www.ucsg.edu.ec/cea/historia-facultad/>

UCSG. (2018c). Misión, Visión. Recuperado el 23 de noviembre de 2018, de <http://www.ucsg.edu.ec/cea/c004418/mision-vision/>

UIAF. (2014). *Aplicabilidad de la minería de datos y el análisis de redes sociales en la inteligencia financiera* (Primera). Unidad de Información y Análisis Financiero. Recuperado de [http://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Aplicabilidad-de-la-mineria-de-datos-\(1\).pdf](http://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Aplicabilidad-de-la-mineria-de-datos-(1).pdf)

Vargas, I., Ramírez, C., Cortés, J., Farfán, A., & Heinze, G. (2011). Factores asociados al rendimiento académico en alumnos de la Facultad de Medicina: estudio de seguimiento a un año, *34*(4), 8. Recuperado de <http://www.scielo.org.mx/pdf/sm/v34n4/v34n4a2.pdf>

Viale, H. (2014). Una aproximación teórica a la deserción estudiantil universitaria. *Revista Digital de Investigación en Docencia Universitaria*, (1), 59. <https://doi.org/10.19083/ridu.8.366>

Vieira, L., Ortiz, L., & Ramírez, S. (2009). *Introducción a la Minería de Datos*. Río de Janeiro: Editora E-papers. Recuperado de <https://books.google.com.ec/books?id=jJEhHyESFsC&printsec=frontcover&dq=miner%C3%ADa+de+datos&hl=es->

419&sa=X&ved=0ahUKEwiWp5nlidDeAhWx1FkKHYOAC-
gQ6AEILDAB#v=onepage&q=miner%C3%ADa%20de%20datos&f=false

Villarino, G. (2015). Metodología de minería de datos para el estudio de tablas de siniestralidad vial, 91. Recuperado de https://eprints.ucm.es/34870/1/TFM_GuillermoVillarino_Nov_2015.pdf

Zamora, M., Basilio, L., Conde, E., Castellanos, M., Estrada, C., & Lazo, G. (2016). El inevitable ajuste del gobierno de Jimmy Morales, (105), 156. Recuperado de <http://ipn.usac.edu.gt/wp-content/uploads/2016/09/IPN-RD-105.pdf>

APÉNDICES


Apéndice A. Formato para entrevistas a: Directora de Carrera, Coordinadora Académica, y Asesoría Pedagógica Estudiantil.

1. ¿Qué cargo desempeña en la Carrera/Facultad y desde hace qué tiempo?
2. ¿Considera usted importante la tasa de retención de estudiantes de la carrera? ¿Por qué?
3. ¿Se dispone de alguna herramienta para predecir la retención de los estudiantes en la carrera? En caso de ser así ¿Podría describir el proceso para conseguirla?
4. ¿Considera usted importante tener una herramienta que pueda predecir la probabilidad de retención o deserción de los estudiantes en la carrera? ¿Por qué?
5. ¿Qué funcionalidades le gustaría que tenga la herramienta para el cálculo de las probabilidades de retención de los estudiantes en la carrera?

Apéndice B. Manuales

Algoritmo Predicción
Manual de Instalación
Manual de Usuario
Manual Técnico

Queda prohibido cualquier tipo de explotación, reproducción, distribución, comunicación pública y/o transformación, total o parcial, por cualquier medio, de este documento sin el previo consentimiento de la Universidad Católica de Santiago de Guayaquil.

	AlgoritmoPredicción FORMATO MÍNIMO* para la presentación de trabajos titulación Manual de Usuario Manual Técnico	Facultad de Ingeniería Carrera Sistemas Computacionales
---	---	--


Manual de Instalación y Configuración del Software Base y Aplicación

El aplicativo “AlgoritmoPrediccion” no posee ninguna instalación debido a que consta con una carpeta dentro de la cual se encuentra el archivo ejecutable AlgoritmoPrediccion.exe, el cual proporcionará acceso al sistema. Sin embargo, necesita la instalación de dos aplicaciones que se detallan a continuación.

AlgoritmoPrediccion	
Descripción	Carpeta que contiene todos los archivos necesarios para su ejecución
Localización	C:\AlgoritmoPrediccion\
Procedimiento de instalación	
Paso 1	Ejecutar el archivo NetFramework 4.5 y seleccionar las opciones siguientes hasta que finalice su ejecución.
Paso 2	Ejecutar el archivo AccessDatabaseEngine.exe y aceptar su ejecución.
Paso 3	En la ventana “contrato de licencia para el usuario final” marcar la casilla Aceptar los términos de contrato (ver Figura 1), y seleccionar “continuar” e “instalar”(ver Figura 2).
Paso 4	Finalizada la instalación cerrar la ventana e ir a la carpeta de la solución y ejecutar el archivo AlgoritmoPrediccion.exe para su normal funcionamiento.



Figura: Contrato de licencia para usuario final

 <p>UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL</p>	<p>AlgoritmoPredicción FORMATO MÍNIMO* para la presentación de trabajos titulación Manual de Usuario Manual Técnico</p>	<p>Facultad de Ingeniería Carrera Sistemas Computacionales</p>
--	--	---

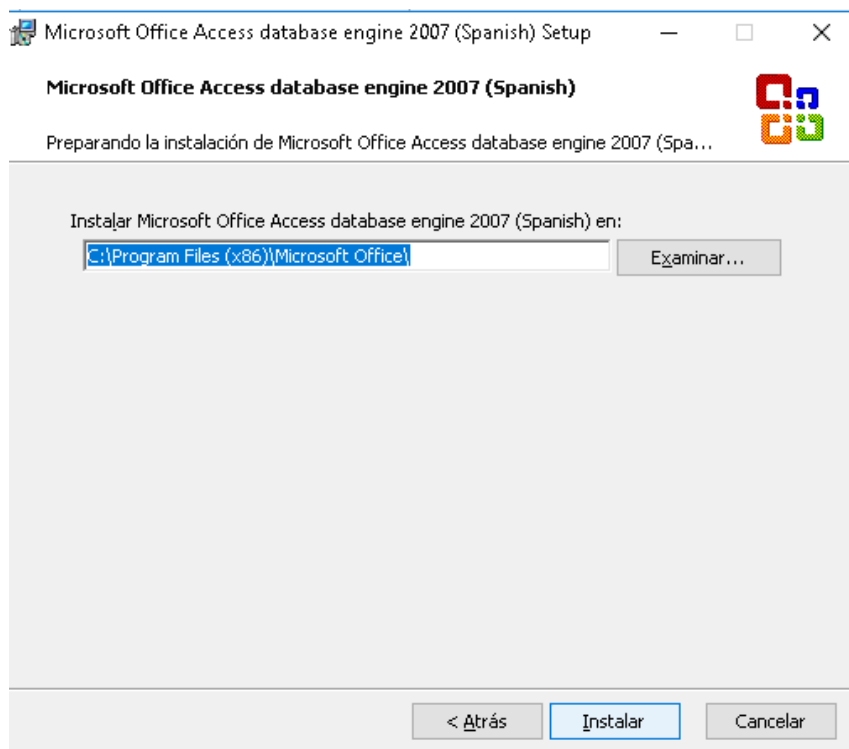



Figura: Instalar Microsoft Office Access database engine 2007 (Spanish)

Desinstalación del Aplicativo

Para su completa eliminación se deberá ingresar a la dirección “C:\” y eliminar la carpeta “AlgoritmoPredicción\” la cual posee todo el aplicativo, así mismo se deberá eliminar cualquier acceso directo que se haya creado durante su uso.

 <p>UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL</p>	<p>AlgoritmoPredicción FORMATO MÍNIMO* para la presentación de trabajos titulación Manual de Usuario Manual Técnico</p>	<p>Facultad de Ingeniería Carrera Sistemas Computacionales</p>
--	--	---

Manual de usuario

Para el uso correcto del aplicativo deberá realizarse los siguientes pasos previos:

Requisitos	
Requisito	Descripción
Fuentes	Deberá solicitarse el archivo de Excel(.xls) con la información proporcionada por el departamento de Centro de Cómputo (Ver Manual Técnico).
Ubicación	Deberá ingresar a la dirección C:\AlgoritmoPrediccion\ o en su defecto al acceso directo ubicado en el escritorio.

La aplicación cuenta con un solo módulo y dos ventanas y su ejecución es presentada a continuación:

Procedimiento de ejecución	
Paso	Descripción
1	Ejecutar el archivo “AlgoritmoPrediccion.exe”.
2	Ingresar el usuario “directora” y contraseña “directora” y seleccionar el botón “login” Ver Figura 3.
3	Seleccionar el botón “Seleccionar Archivo” y buscar el Excel proporcionado por el departamento de Centro de Cómputo de la Universidad. Ver Figura 4.
4	Esperar el mensaje “El archivo se generó exitosamente” y seleccionar el botón “Aceptar”.
5	Se habilitará un botón “azul” el cual se deberá seleccionar el tipo de formato en el cual queremos guardar el archivo. Para una mejor comprensión y uso del mismo se recomienda elegir la opción Excel, aunque también se pueden elegir las opciones PDF y WORD.
6	Se mostrará una ventana donde deberemos seleccionar la ubicación donde se guardará el archivo para su posterior uso, y se deberá elegir un nombre (por defecto se llamará reporte) y seleccionar el botón “Guardar”.
7	Cerrar las ventanas del Aplicativo


 UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL	Algoritmo Predicción FORMATO MÍNIMO* para la presentación de trabajos titulación Manual de Usuario Manual Técnico	Facultad de Ingeniería Carrera Sistemas Computacionales
---	--	--



Figura: Pantalla de Login

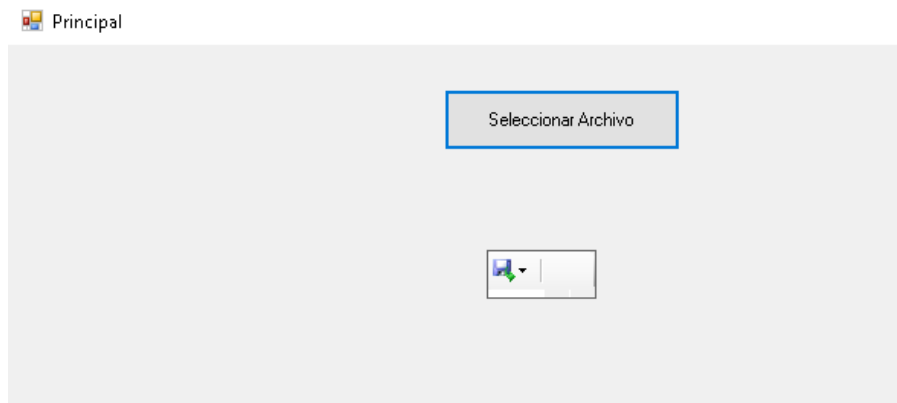



Figura: Carga de archivo y generación de reporte

 <p>UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL</p>	<p align="center">Algoritmo Predicción FORMATO MÍNIMO* para la presentación de trabajos titulación Manual de Usuario Manual Técnico</p>	<p align="center">Facultad de Ingeniería Carrera Sistemas Computacionales</p>
--	--	--

Manual Técnico

Recursos de Hardware y Software


Esta solución no utiliza servidores, debido a que la información es solicitada por la Dirección de Carrera de la Facultad, al departamento de Centro de Cómputo de la Universidad. La estación cliente es la PC ubicada en las oficinas de la Carrera Negocios Internacionales, y para su normal funcionamiento se necesitarán de los siguientes recursos con sus respectivas versiones o versiones posteriores que fueren compatibles, debido a esto la solución incluye un instalador de los dos requerimientos en su versión funcional.

Dato	Valor mínimo	Valor recomendado
Procesador	Pentium Celeron	Intel Core I3
Memoria RAM	2GB	4GB
Tamaño Almacenamiento	80MB	100MB
AccessDatabaseEngine	Versión 12.0	Compatible con V12.0
NetFramework	Versión 4.5	Superior a V4.5
Sistema operativo	Windows XP	Windows XP o superior
AccessDatabaseEngine	Versión 12.0	Compatible con V12.0
NetFramework	Versión 4.5	Superior a V4.5

Debido a que la solución no se conecta a internet no se requiere de ningún tipo de acceso a la red, y debido a que el aplicativo se ejecuta con la biblioteca de “Windows Forms” la única restricción es en cuanto al Sistema Operativo que debe ser Windows.


Requisitos y Relaciones con otros sistemas

Para la ejecución del aplicativo se requiere la información guardada en las bases de datos de la Universidad, la cual debe ser proporcionada en formato .XLS por el Departamento de Centro de Cómputo, la cual incluye información de la parte Académica, Financiera, y Datos de la Ficha Técnica detallados a continuación:

 <p>UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL</p>	<p align="center">Algoritmo Predicción FORMATO MÍNIMO* para la presentación de trabajos titulación Manual de Usuario Manual Técnico</p>	<p align="center">Facultad de Ingeniería Carrera Sistemas Computacionales</p>
--	--	--

Nombre de Columna	Detalle
COD_ESTUDIANTE	Valor entero con identificación del estudiante.
EDAD	Valor entero con la edad del estudiante.
SEXO	Dato tipo texto con el sexo del estudiante con las opciones FEMENINO y MASCULINO.
ESTADO_CIVIL	Datos tipo texto que describe el estado civil del estudiante con las opciones CASADO, SOLTERO, DIVORCIADO, VIUDO.
UBICACION_COLEGIO	Dato tipo texto que describe la ubicación colegio con las opciones DENTRO DE LA CIUDAD o FUERA DE LA CIUDAD.
ESTUDIOS_ADICIONALES	Dato tipo texto que muestra si realizó estudios adicionales con la opción SI o NO.
EMPRESA_TRABAJA	Tiene información tipo texto sobre la empresa donde trabaja.
HIJOS	Muestra en dato tipo texto si el estudiante tiene hijos con las opciones S, y N.
BECA	Muestra en valor flotante el porcentaje de beca del estudiante.
DiasRetrasoPension	Valor entero que muestra el promedio de días de retraso en el pago de pensiones del estudiante.
Aprobadas	Valor entero con el total de materias aprobadas del estudiante.
Reprobadas	Valor entero con el total de materias reprobadas del estudiante.
Promedio	Valor flotante con el promedio de notas del estudiante.
Segunda	Valor entero con la cantidad de materias en segunda matrícula del estudiante.
Tercera	Valor entero con la cantidad de materias en tercera matrícula del estudiante.
Aprobadas	Valor entero con la cantidad de materias aprobadas del estudiante.
Reprobadas	Valor entero con la cantidad de materias reprobadas del estudiante.
Promedio	Valor flotante con el promedio de notas del estudiante.

El archivo que proporciona el departamento de Centro de Cómputo deberá cumplir con los siguientes requisitos:

 <p>UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL</p>	<p>Algoritmo Predicción FORMATO MÍNIMO* para la presentación de trabajos titulación Manual de Usuario Manual Técnico</p>	<p>Facultad de Ingeniería Carrera Sistemas Computacionales</p>
--	---	---

Restricción	Detalle
Nombre de columnas	El normal funcionamiento de la aplicación dependerá de la correcta identificación de las columnas dentro de la tabla, independientemente del orden en el que se encuentren deberán llamarse como se especifica en el manual técnico.
Datos Vacíos	El sistema puede no arrojar información exacta debido a la falta de datos en los campos que se analicen motivo por el cual es necesario que la información este completa para su correcto análisis. (Ver manual técnico).

ANEXOS

Anexo 1. Malla curricular rediseñada, Carrera de Gestión Empresarial Internacional (Negocios Internacionales)

Carrera: NEGOCIOS INTERNACIONALES												
Objeto: La internacionalización de las empresas en contextos de acelerada globalización, a través del intercambio de bienes y servicios, regulado bajo una política exterior vigente y considerando técnicas negociadoras internacionales												
AOC	Núcleo Extracto/tema	Ciclo	Núcleo Problemática	Campo de Formación						Nº. Total de Horas		
				Dinámicas Iniciales	Prácticas profesionales		Epidemiología y metodología de la investigación	Integración de valores, costumbres y cultura	Comunicación y lenguaje			
BÁSICA	Análisis y Operaciones de los Negocios Internacionales	I	Análisis de los Turistas del Comercio Internacional y descripción del flujo de negocios internacionales en Ecuador				Epidemiología de los Negocios Internacionales (100 horas)	Historias I (100 horas)	Lenguaje y Fonética I (100 horas)	100		
								Historia I (100 horas)	Inglés I (100 horas)			
									Ofertinas para la Educación superior (100 horas)			
		II	Adaptación de productos comerciales de acuerdo a preferencias de consumidores internacionales	Marketing Global (100 horas)						Lenguaje y Fonética II (100 horas)	100	
				Marketing (100 horas)				Fundamentos Descriptivos (100 horas)	Inglés II (100 horas)	Fonética II (100 horas)		
		III	Diferenciación de los procesos logísticos de transporte y viabilidad para el flujo de mercancías hacia los mercados internacionales	Matemáticas Aplicadas (100 horas)	Logística Internacional (100 horas)					Inglés III (100 horas)	100	
				Matemáticas (100 horas)						Fonética III (100 horas)		
		IV	Análisis de la política comercial del Ecuador para promover el flujo del comercio internacional	Matemáticas (100 horas)	Logística avanzada (100 horas)				Metodología de la Investigación (100 horas)	Historias II (100 horas)	100	
										Fonética IV (100 horas)		
		PROFESIONAL	Diseño, Desarrollo, Internacionalización y Control de los Negocios	V	Fortalecimiento de propuestas y modelos de grandes empresas a través de los flujos de la internacionalización e inversión para contribuir a la generación de flujos de exportaciones y equilibrio de la balanza comercial	Contabilidad Financiera (100 horas)	Operaciones de los Negocios Internacionales (100 horas)			Investigación de mercados internacionales (100 horas)	Cálculo Comercial (100 horas)	100
						Negocios Externos (100 horas)						
				VI	Diseño de estrategias para administrar los negocios internacionales a través de la implementación de la planeación estratégica	Administración y Comportamiento Organizacional (100 horas)	Administración Estratégica de los Negocios Internacionales (100 horas)	Prácticas Profesionales I (100 horas)				Inglés de Negocios II (100 horas)
Historias I: Análisis de áreas y temas de Negocios (100 horas)												
Historias II: El comercio Internacional (100 horas)											Historia de Negocios II (100 horas)	
Contabilidad Administrativa (100 horas)												
VII	Magnificación de las capacidades del talento humano en términos de conocimientos cultural para afrontar negociaciones vitales en los mercados internacionales			Integración Externas y Comercial (100 horas)	Negocios Internacionales (100 horas)	Prácticas Sociales (100 horas)				Análisis y Viabilidad de Proyectos Internacionales (100 horas)	100	
				Historias I: Integración de Mercados (100 horas)								
				Historias II: Organización (100 horas)								
VIII	Eficiencia para la ejecución y el control de negocios internacionales a través de documentos de planeación			Finanzas Internacionales (100 horas)	Control de los Negocios Internacionales (100 horas)	Prácticas Profesionales II (100 horas)					100	
				Derecho Internacional (100 horas)								
				Historias I: Diversidad y Sostenibilidad (100 horas)								Historias II: Metodología, Análisis y Solución de Casos (100 horas)
TITULACIÓN	Portafolio Inicial de las Relaciones Internacionales	IX	Organización y aplicación de relaciones propiarias y procesos de cooperación de integración regional y mundial	Estadística de Negocios Internacionales (100 horas)	Cooperación Internacional (100 horas)			Trabajo de Titulación (100 horas)	100			



DECLARACIÓN Y AUTORIZACIÓN

Yo, **Zambrano Soriano, Pedro Eddison**, con C.C: # 0927076646 autor del trabajo de titulación: **Análisis de retención de estudiantes en la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG, durante los dos primeros años de la carrera, mediante técnica de Minería de Datos**, previo a la obtención del título de **Ingeniero en Sistemas Computacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, 12 de marzo del 2019

Zambrano Soriano, Pedro Eddison

C.C: 0927076646



REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TEMA Y SUBTEMA:	Análisis de retención de estudiantes en la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG, durante los dos primeros años de la carrera, mediante técnica de Minería de Datos		
AUTOR:	Pedro Eddison, Zambrano Soriano		
TUTOR:	Adela María Eugenia, Zurita Fabre		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Facultad de Ingeniería		
CARRERA:	Ingeniería en Sistemas Computacionales		
TÍTULO OBTENIDO:	Ingeniero en Sistemas Computacionales		
FECHA DE PUBLICACIÓN:	12 de marzo del 2019	No. DE PÁGINAS:	100
ÁREAS TEMÁTICAS:	Sistemas de información, desarrollo de sistemas informáticos		
PALABRAS CLAVES:	minería de datos, retención estudiantil, modelo predictivo, árbol de decisión, algoritmo J48, teorema de Bayes		
RESUMEN	<p>La presente investigación tuvo el propósito de generar un modelo predictivo de retención-deserción de estudiantes de la cohorte 2016-2018 de la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas, que permita determinar la probabilidad de deserción estudiantil y ayudar en predicciones futuras; este estudio tuvo un enfoque cualitativo, del tipo analítica, exploratoria y descriptiva porque estuvo orientada a realizar un análisis de procesos y aplicativos tecnológicos que permitan determinar los niveles de retención estudiantil, utilizando la minería de datos como herramienta de predicción; dado el ámbito del estudio, no fue necesario establecer población ni muestra, solamente se identificó las personas clave de este proceso y los interesados en los resultados y se los denominó <i>informantes</i> a quienes se aplicó la técnica de la <i>entrevista</i>. Como resultado de la información recogida se pudo concluir que actualmente no existe un protocolo establecido para identificar a los estudiantes que tienen menor probabilidad de continuar sus estudios, y la mayoría de las selecciones que se realizan se hacen con referencia a la parte académica del alumno. Luego de finalizar el desarrollo del proyecto, se pudo comprobar que el modelo predictivo basado en técnicas de Minería de Datos, permitió la evaluación e interpretación de toda la información de los estudiantes de la cohorte 2016-2018 de la Carrera de Gestión Empresarial Internacional, a través de la cual se determinó un modelo de la retención-deserción y es una herramienta de apoyo en la toma de decisiones de futuras proyecciones sobre este indicador.</p>		
ADJUNTO PDF:	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
CONTACTO CON AUTOR/ES:	Teléfono: +593-4-2265763 - 0939451997	E-mail: pedroezs@outlook.com	
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE):	Nombre: Toala Quimí, Edison José		
	Teléfono: +593-4-2202763, ext. 1025		
	E-mail: edison.toala@cu.ucsg.edu.ec		
SECCIÓN PARA USO DE BIBLIOTECA			
Nº. DE REGISTRO (en base a datos):			
Nº. DE CLASIFICACIÓN:			
DIRECCIÓN URL (tesis en la web):			