



**UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL**

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

**TEMA:**

**Diseño de un Modelo Predictivo, mediante la técnica de Minería de  
Datos, para identificar el perfil del estudiante en la Unidad de  
Titulación en la Carrera Ingeniería en Sistemas Computacionales de  
la Facultad de Ingeniería de la UCSG**

**AUTOR:**

**Cordovilla Cordovilla, Juan Argenis**

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

**TUTOR:**

**Ing. Salazar Tovar, César Adriano, Mgs.**

**Guayaquil, Ecuador**

**07 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 **Diseño de un Modelo Predictivo, mediante la técnica de Minería de Datos, para identificar el perfil de éxito del estudiante en la Unidad de Titulación en la Carrera Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG**, fue realizado en su totalidad por **Cordovilla Cordovilla, Juan Argenis** como requerimiento para la obtención del Título de **Ingeniero en Sistemas Computacionales**.

**TUTOR**

**Ing. César Adriano Salazar Tovar, Mgs.**

**DIRECTORA (e) DE LA CARRERA**

**Ing. Ana Isabel Camacho Coronel, Mgs.**

Guayaquil, a los 07 días del mes de marzo del año 2019



**UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL**

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

**DECLARACIÓN DE RESPONSABILIDAD**

**Yo, Cordovilla Cordovilla, Juan Argenis**

**DECLARO QUE:**

El Trabajo de Titulación **Diseño de un Modelo Predictivo, mediante la técnica de Minería de Datos, para identificar el perfil de éxito del estudiante en la Unidad de Titulación en la Carrera Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG** 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, a los 07 días del mes de marzo del año 2019

**EL AUTOR**

---

**Cordovilla Cordovilla, Juan Argenis**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERÍA  
CARRERA DE INGENIERÍA EN SISTEMAS  
COMPUTACIONALES

AUTORIZACIÓN

Yo, **Cordovilla Cordovilla, Juan Argenis**

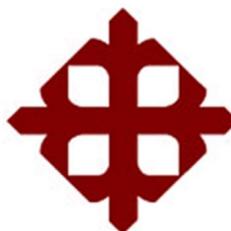
Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación, **Diseño de un Modelo Predictivo, mediante la técnica de Minería de Datos, para identificar el perfil de éxito del estudiante en la Unidad de Titulación en la Carrera Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 07 días del mes de marzo del año 2019

EL AUTOR

---

**Cordovilla Cordovilla, Juan Argenis**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERÍA  
CARRERA DE INGENIERÍA EN SISTEMAS  
COMPUTACIONALES

### REPORTE DE URKUND

URKUND	
<b>Documento</b>	<a href="#">Diseño de un modelo predictivo mediante minería de datos para perfiles de UTE.docx</a> (D48184906)
<b>Presentado</b>	2019-02-21 14:50 (-05:00)
<b>Presentado por</b>	cesar.salazar@cu.ucsg.edu.ec
<b>Recibido</b>	cesar.salazar.ucsg@analysis.orkund.com
<b>Mensaje</b>	Trabajo de titulación B2018 - Juan Cordovilla <a href="#">Mostrar el mensaje completo</a> 0% de estas 29 páginas, se componen de texto presente en 0 fuentes.

## **AGRADECIMIENTO**

Agradezco principalmente a Dios, a mi madre que me apoya en cada momento y decisión que tome. Al Ing. Galo Cornejo G. por el apoyo profesional, que con paciencia guio mi trabajo con su instrucción. A mi tutor Ing. Cesar Salazar por su guía en el trabajo de titulación.

## **DEDICATORIA**

Este trabajo de titulación es dedicado a mi madre que, con su apoyo, amor y confianza me permitió culminar esta etapa de mi carrera profesional.



**UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL**

**FACULTAD DE INGENIERÍA**

**CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES**

**TRIBUNAL DE SUSTENTACIÓN**

**Ing. Ana Isabel Camacho Coronel, Mgs.  
DIRECTORA (e) DE LA CARRERA**

**Ing. Mario Céleri Mujica**  
COORDINADOR DEL ÁREA O DOCENTE DE LA CARRERA

**Ing. Galo Enrique Cornejo Gómez, Mgs.  
OPONENTE**

## ÍNDICE GENERAL

RESUMEN.....	XIV
INTRODUCCIÓN .....	2
CAPÍTULO I EL PROBLEMA .....	4
1.1 Planteamiento del Problema .....	4
1.2 Preguntas de Investigación .....	5
1.3 Objetivos de la Investigación.....	5
1.3.1 Objetivo General .....	6
1.3.2 Objetivos Específicos .....	6
1.4 Justificación y Alcance .....	6
CAPÍTULO II MARCO TEÓRICO, CONCEPTUAL Y LEGAL.....	8
2.1 Elementos teóricos y conceptuales .....	8
2.1.1 El proceso de titulación de grado, correspondencia con su perfil y nivel de empleabilidad.....	8
2.1.2 Minería de Datos: Patrones y Técnicas .....	12
2.1.3 Infraestructura Tecnológica. Técnicas de Minería de Datos a utilizar en el Proyecto .....	16
2.1.4 Herramientas de Desarrollo.....	17
2.1.4.1 C# .....	18
2.1.4.2 Base de datos MySQL.....	19
2.2 Sustento legal .....	21
2.3 Marco contextual .....	23
CAPÍTULO III METODOLOGÍA Y RESULTADOS .....	24
3.1 Metodología de la Investigación.....	24
3.2 Metodología de Desarrollo .....	25
3.3 Metodología de Minería de Datos.....	26
3.3.1 Recopilación de datos.....	26
3.3.2 Pre-procesado o transformación de los datos .....	27
3.3.3 Algoritmos de Minería de Datos .....	28
3.3.4 Interpretación y evaluación de resultados .....	29
3.4 Población y muestra.....	29
3.5 Análisis de Resultados .....	29
CAPÍTULO IV PROPUESTA.....	32

4.1	Introducción .....	32
4.2	Objetivo.....	32
4.3	Responsables.....	33
4.4	Descripción de la solución.....	33
	CONCLUSIONES .....	49
	RECOMENDACIONES .....	50
	REFERENCIAS BIBLIOGRÁFICAS.....	51

## ÍNDICE DE TABLAS

Tabla 1: Características de los estudiantes.....	11
Tabla 2: Comparación de los lenguajes de desarrollo.....	19
Tabla 3: Comparación de bases de datos .....	21
Tabla 4: Resultados de encuesta .....	31
Tabla 5: Porcentaje de estudiantes según área de trabajo .....	31
Tabla 6: Diccionario de datos .....	37

## ÍNDICE DE GRAFICOS

Figura 1: Proceso y etapas de KDD.....	14
Figura 2: Clasificación de las técnicas de Minería de Datos.....	16
Figura 3: Ediciones de SQL Server.....	20
Figura 4: Conjunto de datos.....	30
Figura 5: Promedio del estudiante según la matrícula.....	35
Figura 6: Cálculo del promedio de las materias por área.....	36
Figura 7: Peso del área.....	37
Figura 8: Vista minable.....	38
Figura 9: Workflow de k-Means, para perfiles.....	39
Figura 10: Particionamiento.....	39
Figura 11: Configuración del k-Means.....	40
Figura 12: Clustering.....	40
Figura 13: Gestor de color.....	41
Figura 14: Administrador de formas.....	41
Figura 15: Dispersión de los datos.....	41
Figura 16: Workflow de árbol de decisión, para validación de perfiles.....	42
Figura 17: Configuración del aprendiz de árbol de decisión.....	43
Figura 18: Configuración del árbol de decisión para aplicar el modelo.....	43
Figura 19: Configuración del árbol de decisión. Reglas.....	44
Figura 20: Asignación de registros al clustering.....	44
Figura 21: Conjunto de reglas.....	45
Figura 22: Reglas para desarrollador.....	45
Figura 23: Reglas para infraestructura.....	45
Figura 24: Opciones del aplicativo.....	46

Figura 25: Opción Cargar archivo.....	47
Figura 26: Cálculo del perfil .....	47
Figura 27: Detalle de materias .....	48
Figura 28: Opción Buscar .....	48
Figura 29: Opción Configuración .....	48

## RESUMEN

La asignación de temas para Trabajo de Titulación a los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG es un inconveniente que se presenta en la Unidad de Titulación ya que no se dispone del perfil del estudiante que ha terminado su malla curricular para la elaboración de su trabajo final y esto conlleva a que los temas sugeridos y analizados por la Comisión Académica de la Carrera no se ajusten al conocimiento de las aptitudes del mismo sobre el tema propuesto. Para tal efecto, se propuso el diseño de un modelo predictivo, mediante Minería de Datos, para identificar el perfil del estudiante, de manera que los temas de Trabajo de Titulación estén acordes a sus habilidades. Para realizar la investigación se utilizó el enfoque de investigación cuantitativo, de tipo analítico, descriptivo y exploratorio para estudiar el contexto en donde existe el problema de la identificación del perfil; como técnica de recolección de datos se usó la encuesta, que se realizó a los estudiantes de cohortes A y B 2017 y A 2018, técnicas de Minería de Datos no supervisadas como clustering y árboles de decisión, y la herramienta de Minería de Datos KNIME. Se diseñó la vista minable con las áreas de estudio de la carrera y se generó el modelo predictivo de evaluación para establecer el perfil de acuerdo a las áreas, prácticas y trabajo actual. Al final se plantearon recomendaciones a considerarse como mejoras del modelo predictivo.

***Palabras clave:*** MODELO PREDICTIVO; MINERÍA DE DATOS; VISTA MINABLE, CLUSTERING; ÁRBOL DE DECISIÓN; PATRONES.

## INTRODUCCIÓN

Conocer las habilidades que tienen los estudiantes en los diversos ámbitos de estudio de su carrera universitaria, es de mucha importancia para poder determinar sus fortalezas en el momento de decidir qué es lo más conveniente para él en cuanto se refiere a proponer su tema de investigación para su trabajo de titulación, paso culminante para obtener su título universitario.

Para determinar el perfil de un estudiante confluyen algunas variables a considerar, las que, en la mayoría de los casos, no son tomadas en cuenta ni existe una herramienta que permita analizar las mismas. Estas variables tienen relación, por lo general, con la situación socio-económica, académica, de responsabilidad y profesional, que son las que afectan y determinan las destrezas de un estudiante y que es importante someterlas a análisis.

Como se ha mencionado anteriormente, no se realiza un análisis de variables para determinar el perfil del estudiante. Eso se puede constatar en la Carrera Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la Universidad Católica de Santiago de Guayaquil (UCSG) en donde, en el momento que el estudiante ha terminado su malla curricular e ingresa a la Unidad de Titulación y solicita un tema para desarrollar su Trabajo de Titulación, se encuentra con la novedad de que debe aceptar el proyecto que se le asigne, sin considerar el o las áreas en que tiene mayor pericia o sobresale y, a pesar de que se obliga a trabajar en lo que se le ha destinado, la última palabra para aprobación o rechazo de temas, la tiene la Comisión Académica de la carrera.

Para que exista mayor agilidad en la Unidad de Titulación en el momento de establecer los temas para los trabajos a realizar por los estudiantes, se propone la generación de un modelo predictivo que ayude en la determinación del perfil del estudiante en base a tres variables: situación socio-económica, académica, de responsabilidad y profesional de manera que se optimice la gestión de los docentes en el momento de la aprobación de los temas y se agilice el desarrollo de los proyectos para que los tiempos determinados para la entrega de los avances se cumpla y no

existan retrasos en la revisión de los mismos, lo que perjudicaría al estudiante en su calificación.

Esta investigación pretende generar una interfaz que presente los perfiles del estudiante de la Unidad de Titulación en base a las variables antes señaladas para que sean conocidos por la Comisión Académica. Los resultados de esta investigación han sido incluidos como sigue: en el capítulo I se puede concebir la problemática a resolver, hipótesis, objetivos, justificación, alcance; el capítulo II hace referencia a ciertas teorías y principios relativos al tema en cuestión, así también algunas conceptualizaciones y normativas que sustentan las medidas; en el capítulo III está incluida la metodología de la investigación y al análisis de resultados; el capítulo IV contiene la propuesta objeto de esta investigación; cerrando con algunas conclusiones y recomendaciones.

# **CAPÍTULO I**

## **EL PROBLEMA**

Con el fin de conocer las razones por las cuales se plantea esta investigación, en los apartados que se encuentran a continuación se expone el planteamiento del problema, la pregunta de investigación, el objetivo general y los objetivos específicos y la justificación y alcance del proyecto, de manera que éste sea fácilmente entendido.

### **1.1 Planteamiento del Problema**

En la Carrera Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la Universidad Católica de Santiago de Guayaquil (UCSG), los estudiantes que han terminado su malla curricular ingresan a la Unidad de Titulación para desarrollar su trabajo de investigación y/o implementación de sistemas que los acredite, en el plazo estipulado, como requisito para la obtención del título de Ingenieros en Sistemas Computacionales.

Es en la Unidad de Titulación en donde se receptan los temas sugeridos para los trabajos a realizar por los estudiantes, que luego son analizados en las reuniones de la Comisión Académica de la Carrera para aprobarlos o rechazarlos, teniendo en cuenta que, para su análisis, sólo se dispone del anteproyecto inicial elaborado por el estudiante.

Los temas para los trabajos de titulación a realizarse por los estudiantes son asignados sin tomar en consideración el perfil que cada uno de los estudiantes tiene, es decir, en qué áreas se han destacado durante toda la carrera, su participación en actividades de investigación, eventos realizados, ámbito laboral y otros aspectos que puedan orientar a los docentes a que el tema propuesto sea el apropiado.

A la falta del perfil del estudiante al ingresar a la Unidad de Titulación se limita el conocimiento de las aptitudes del mismo sobre el tema propuesto. La insuficiencia del conocimiento puede formar una cota en el tema y atrasos en las entregas de los avances del trabajo. Uno de los puntos importantes a tener en cuenta es la posterior inserción laboral, ya que el perfil del estudiante no se verá reflejado en su hoja de vida, por

cuanto su trayectoria en la carrera no estuvo orientada a demostrar en lo que se destacó durante sus años de estudio.

Se debe tener en cuenta que se toman en consideración variables académicas, socio-económicas -orientada a una beca- que pueda tener el estudiante por cuanto poseer un incentivo de este tipo le permitiría tener mejor rendimiento estudiantil; y finalmente profesionales. Esto quiere decir nivel de conocimiento que tiene el estudiante en las áreas de mayor impacto de la carrera, entorno social y laboral, todo apoyado en las prácticas y pasantías pre-profesionales obligatorias y que demuestran la participación activa del estudiante al aplicar lo aprendido durante sus años de estudio.

Considerando todas las limitantes que tiene la Unidad de Titulación, se plantea el proyecto para el Diseño de un modelo de evaluación, mediante la técnica de Minería de Datos, para identificar el perfil del estudiante en la Unidad de Titulación en la Carrera de Ingeniería en Sistemas Computacionales para mejorar la asignación de temas y posterior inserción laboral del mismo, cuando llegan ofertas de trabajo a la carrera.

## **1.2 Preguntas de Investigación**

¿Con los datos obtenidos que se utilizan en las variables es posible establecer patrones de perfiles de estudiantes en la Carrera de Ingeniería en Sistemas Computacionales?

¿Los patrones de perfiles de estudiantes en la carrera de Ingeniería en Sistemas Computacionales tienen concordancia con los lineamientos de investigación de la carrera?

## **1.3 Objetivos de la Investigación**

El desarrollo del proyecto se sustenta en el planteamiento de los objetivos, tanto general como específicos.

### 1.3.1 Objetivo General

Definir el perfil del estudiante, mediante un modelo de minería de datos relacionado con los aspectos académicos, socio-económicos y profesionales de los estudiantes en la carrera Ingeniería en Sistemas Computacionales, que sirva de soporte a la asignación o aprobación de temas de Trabajos de Titulación.

### 1.3.2 Objetivos Específicos

- Realizar una revisión bibliográfica de creaciones de perfiles de estudiantes para establecer las variables que inciden en el mismo;
- Diseñar la vista minable, que permita definir los perfiles de estudiantes, para que correspondan al perfil de estudiantes que terminaron la malla curricular de la carrera.
- Generar el modelo de evaluación, con el uso de los algoritmos necesarios respecto a lo obtenido, para detectar la afinidad al perfil que corresponda.

## 1.4 Justificación y Alcance

El propósito de esta investigación es la generación de una herramienta de análisis predictivo para el establecimiento de perfiles de estudiantes de la carrera de Ingeniería en Sistemas Computacionales que ingresan a la Unidad de Titulación

Esta investigación tiene como propósito a través de la minería de datos, tener una herramienta para el análisis de los estudiantes durante la carrera para el ingreso a la Unidad de Titulación. Los resultados podrán ser utilizados para analizar cómo es el perfil del estudiante y tener un mejor desempeño en su tema de titulación.

Cabe recalcar que el estudio se fundamentará en el uso de las técnicas de minería de datos, con el objetivo de aplicarlos en el diseño de la vista minable, que se alimentará de la información proporcionada en la carrera, sobre las notas de los estudiantes, de las asignaturas de la carrera, la misma que representará a una variable **académica**; otra información es la relativa a las asistencias, la misma que representará la variable de **responsabilidad**, además, los perfiles considerados en el momento de realizar las pasantías en las empresas será la variable **profesional**, mientras que la

variable **socioeconómica** se la considerará en caso de que el estudiantes haya contado con una beca. Las mencionadas variables servirán para proceder a la generación de la vista minable propuesta.

En cuanto al alcance del proyecto, el estudio permitirá la generación de una interfaz para la visualización del perfil del estudiante que ingresa a la Unidad de Titulación, tomando en consideración los estudiantes que hayan ingresado en los últimos tres semestres, es decir, A-2017, B-2017 y A-2018. El análisis de la información que alimentará el modelo predictivo dará como resultado el perfil del alumno, considerando las variables: académica, que corresponde a la cartilla histórica del estudiante; de responsabilidad, referente a las asistencias a clases; profesional, que tiene relación con las prácticas pre-profesionales; y socioeconómica, en caso de que tenga beca.

La interfaz que será implementada en el ordenador del docente encargado de la Unidad de Titulación presentará tres pantallas: a) de ingreso como aspirante a la Unidad de Titulación, b) consulta de perfiles, c) reporte de datos.

El encargado de la Unidad de Titulación tendrá la opción de establecer el perfil de los estudiantes, de acuerdo las variables observadas y los perfiles encontrados para asignar los temas de los trabajos de titulación que luego deberán ser aprobados por la Comisión Académica de la carrera. El coordinador de la unidad de titulación podrá generar reportes de los perfiles, de acuerdo a la necesidad.

La investigación se enmarca dentro de la línea investigación y desarrollo de nuevos servicios o productos de la carrera Computación de la Facultad de Ingeniería de la UCSG.

## **CAPÍTULO II**

### **MARCO TEÓRICO, CONCEPTUAL Y LEGAL**

El objeto de este estudio está centrado en la búsqueda de la relación entre algunos elementos socio-económicos y profesionales con los conocimientos y experticia desarrollados por el estudiante de manera que, mediante un modelo de minería de datos, facilite identificar el tema o campo de investigación más cercano a sus competencias para la asignación de temas para sus trabajos de titulación; para abordar este estudio, en este capítulo se hace referencia a los distintos elementos teóricos, conceptos fundamentales, ámbito de aplicación y sustento legal de este trabajo.

#### **2.1 Elementos teóricos y conceptuales**

Para el logro de un producto académico de calidad, es necesario ubicar al estudiante en el ámbito de sus competencias desarrolladas a lo largo de su colegiatura, así como aspectos relacionados con su avance profesional y experiencial. Una de las técnicas que ofrece la tecnología es la aplicación de la minería de datos para hacer proyecciones cuyos resultados pueden facilitar la toma de decisiones ajustada a la realidad, en este caso, del estudiante que ingresa a un proceso de titulación de grado.

En la descripción de los elementos teóricos de este estudio, se consideran algunos aspectos relativos al proceso de titulación en sí mismo, así como el nivel de aplicabilidad de la tecnología y las herramientas de desarrollo, así como la infraestructura requerida para la aplicación de la minería de datos.

##### **2.1.1 El proceso de titulación de grado, correspondencia con su perfil y nivel de empleabilidad**

Un Trabajo de Titulación (TT) tiene un significado estandarizado, pudiendo anotarse que se refiere a la elaboración de un proyecto, realizado por el estudiante en forma escrita, en el cual deberá demostrar absoluto dominio del tema y evidenciar todas las técnicas y aptitudes que ha adquirido y perfeccionado en sus años de preparación académica. Además, el alumno será capaz de plasmar en su trabajo la aplicación de su criterio como profesional en el momento de proponer soluciones al problema, a través de la aplicación del método científico (CIT, 2011).

Un TT es un requisito previo para obtener la titulación en alguna carrera universitaria; antes fue requisito para graduarse de nivel secundario o bachillerato, pero con menos intensidad investigativa. En la actualidad ha sido considerado como una de las opciones para grado universitario; mientras que, el término *tesis*, es asignado para los productos académicos resultantes de un programa de posgrado (San Martín H. & García P, 2006).

En las carreras universitarias se asigna un nivel, al final de la carrera, para desarrollar el TT; su administración y seguimiento está a cargo de la Unidad de Titulación que el Consejo de Educación Superior (CES), en su artículo 21 del Reglamento de Régimen Académico (2017a) la define como una unidad curricular cuyo resultado se ve reflejado en las dos modalidades siguientes: el TT o un examen complejo.

Para realizar un TT digno de un estudiante que va a optar por un título universitario, se requiere disponer de un tiempo considerable para investigar, argumentar, levantar información relacionada con el problema, analizar la información para obtener resultados e interpretarlos, y luego de este proceso, exponer la investigación al tribunal asignado (CIT, 2011).

Los TT favorecen “una experiencia investigativa única durante su proceso formativo (...) pues, desde la perspectiva de la formación investigativa, esta modalidad es la que mayores aportes ofrece” (Romero-Ortega & Sanz-Cabrera, 2017, p. 240), aunque las capacidades que puede tener el estudiante sobre investigación no está supeditado a la elaboración de su TT, sino que se lo adquiere durante su formación académica; todo lo cual y, en base a esa formación y a su experticia en este campo disciplinar, es que se diseña un perfil que responda al ámbito en el que va a desarrollar su investigación.

Según Fingermann (2011, párr. 1) “se habla de perfil del alumno, para referirse a qué características y competitividades” la IES va a orientar sus esfuerzos para formar a sus estudiantes orientado siempre a su aplicabilidad en el contexto laboral y a responder con la filosofía universitaria.

En la actualidad, el estudiante está obligado a cultivar capacidades que le faciliten hacer frente a los nuevos desafíos que se presentan en las distintas esferas. Estas capacidades deben ser profesionales, personales y sociales, y deben incluir nuevos conocimientos, estrategias, herramientas, técnicas de aprendizaje y otras (Bazán & Rojas, 2013), que le permitan ampliar la percepción de todos los problemas a su alrededor y su capacidad de evaluar y solucionar temáticas que se presenten a lo largo de su vida.

El nexo entre los estudios universitarios y el producto académico de titulación con la vida laboral debe estar debidamente diseñado con el fin de garantizar que la IES está formando profesionales que realmente van a tener un espacio en el mercado laboral y/o empresarial, de lo contrario, se produce una sobrepoblación de titulados en un campo disciplinar que no están preparados para hacer frente a las exigencias de la empleabilidad.

De acuerdo a Suriá Martínez, Rosser Limiñana, y Villegas Castrillo (2013) en un estudio realizado sobre el tema de desempleo juvenil y según menciona en el estudio la Organización del Trabajo (OIT) (2011) éste es un problema en todos los países cuando se presenta una crisis económica y son ellos los que pierden su trabajo y los últimos en encontrar otro nuevo. De acuerdo al informe de la OIT, las tasas de desempleo en España son las más altas de los países de la Unión Europea, y se explica por un desajuste entre la oferta y la demanda y, visto desde ese punto de vista, el tema se debe a cambios de orden técnico y desorganización educativa, que provocaría que la juventud no se encuentre óptimamente preparada ni con el perfil correspondiente para desempeñar los nuevo empleos y/o emprendimientos que se exigen en las empresas, aunque sí existan profesionales calificados pero que no tienen un empleo.

En el mismo estudio se menciona que, a pesar de que los jóvenes profesionales no se han visto afectados en sus trabajos y salarios como sucede con otros que tienen menor formación académica, los profesionales acompañan el desgaste del panorama laboral, lo que ha convertido a la preparación superior en obligatoria para conseguir un mejor trabajo, al mismo tiempo que no lo garantizaría. Esto se demostraría en información del Observatorio Joven de Empleo en España de 2012, son los jóvenes españoles graduados los que menos probabilidades de conseguir trabajo estable de

quienes poseen un título de mayor jerarquía, afectando no sólo a jóvenes que tienen aún poca formación profesional o que todavía cursan su carrera, lo que provoca desmoralización, nula motivación y por consiguiente riesgo de apartarse de la inclusión laboral o empresarial y estar inactivo.

Sobre el mismo asunto, existe consenso de varios estudiosos (Martínez Ruiz, 2013; Sánchez-Castañeda, 2014; Suriá Martínez et al., 2013) de que, además de ser importante la oferta de trabajo, debe haber un equilibrio armónico con las competencias desarrolladas durante su período de formación así como con la capacidad que puede tener la persona para conseguir trabajo. Por lo que es necesaria la búsqueda de estrategias de inserción en la juventud profesional, en función de las preferencias y competencias de los futuros graduados.

Dentro de los perfiles de estudiante de una carrera universitaria se deben tener algunas consideraciones. En la Tabla 1 se presentan las características necesarias para un estudiante de Ingeniería en Sistemas Computacionales de tres universidades internacionales, las mismas que permitirían definir el perfil requerido para la implementación del modelo de entrenamiento que se propone.

*Tabla 1:*  
Características de los estudiantes

CARACTERÍSTICAS	UNIVERSIDADES		
	UNIVERSIDAD DEL NORTE	UNIVERSIDAD DE MONTERREY	UNIVERSIDAD LATINA DE COSTA RICA
Investigadores creativos, disciplina de estudio, asimilan los cambios tecnológicos. Aplicación eficiente de la tecnología de información (desarrollador)	X	X	X
Desarrollan creatividad, habilidades analíticas, éticas y de comunicación (analista)		X	X
Toman decisiones libres, autónomas y responsables	X	X	
Comprometido con el proyecto de vida basado en valores y actitudes	X		
Abierto a una dinámica de cambios sociales, políticos, económicos y tecnológicos bajo la óptica de un enfoque humanístico (investigativo)	X	X	
Interés especial por una línea de énfasis de la carrera y enfocan su aprendizaje en materias electivas (infraestructura)		X	X

*Nota:* Tomado de: Universidad del Norte (s. f.); Universidad de Monterrey (2019), Universidad Latina de Costa Rica (s. f.)

### **2.1.2 Minería de Datos: Patrones y Técnicas**

Se entiende como Minería de Datos a la herramienta de análisis que permite sustraer datos importantes que se encuentran dentro de un conglomerado de abundante información que en un principio es desconocida, pero que luego de su estudio resultará de utilidad para un determinado proceso. A través de la Minería de Datos se dispone, ordena, investiga e indaga la información para extraer de la misma datos que no se los observa a simple vista (Morales-Arencibia, Febles-Rodríguez, & Frómeta-Moreno, 2017).

Esta herramienta puede ser utilizada en las siguientes áreas: a) análisis de mercado, b) de empresas, c) gestión de riesgos, d) descubrimiento de hechos fraudulentos, e) producción, f) retención y análisis de clientes, g) investigación científica, h) actividades deportivas, i) estudio astral, j) navegación en internet (Tecnologías Información, 2018); su uso permite a) ahorro de dinero y posibilita nuevos negocios, b) ayuda en la toma de decisiones, c) ofrece a los directivos poder para decisiones, y puede calcular acciones y resultados, d) con modelos descriptivos, se identifican patrones, dependencias de datos y relaciones de importancia en los resultados que se generan, e) con modelos predictivos, se descubren relaciones para orientar el negocio (Rodríguez & Díaz, 2009).

Otra definición de Minería de Datos es la presentada por Solar (2010), quien manifestó que consiste en obtener conocimiento de grandes fuentes de información como lo son las bases de datos, conocimiento que tenga una significación para un determinado proceso, que sea novedoso, que se encuentre contenido en dichos datos y que sea accionable. Un conocimiento accionable consiste en que el conocimiento se lo puede obtener de una determinada base de datos y ejecutar determinadas acciones con él.

Mediante esta herramienta de análisis se consigue la investigación, observación y se hacen visibles el conjunto de datos con un elevado grado de realidad sin que se hayan planteado hipótesis, y la aplicación de sus técnicas se las realiza tomando en cuenta la independencia y distribución en los datos (Solar, 2010).

Según Dávila y Sánchez (2012) la Minería de Datos forma parte de la inteligencia artificial, que se sustenta en distintas ramas de la ciencia, algunas de las cuales son conservadoras, pero que se diferencia de ellas ya que lo que busca es una finalidad antes que el medio, puesto que la finalidad es encontrar patrones, conocer tendencias, pronosticar el comportamiento de variables analizadas, y, en resumen, aprovechar toda la información extraída de los datos guardados en los almacenes de datos que diariamente se genera y que ayuda a los usuarios y empresas a entender y organizar eficientemente y con mayor precisión el entorno en que se desenvuelven para una toma de decisiones adecuada.

La Minería de Datos es una parte de un proceso llamado KDD (por sus siglas en inglés de Knowledge Discovery in Databases), que consiste en obtener información relevante de una gran cantidad de datos; constituye un proceso participativo en el cual se establecen algunas fases, que son: a) recolección o selección de los datos, b) procesado previo, c) minería de datos, d) interpretación y/o evaluación de los datos (Márquez, Romero, & Ventura, 2012; Solar, 2010).

Para ampliar lo relacionado con la *Metodología KDD*, al decir de Troche (2014) es utilizado en el descubrimiento de patrones o rutinas provechosos, de gran utilidad y caracterizado de que se pueden comprender a partir del análisis de los datos. Este proceso busca descubrir conocimiento de los datos que sea importante, valedero y nuevo en relación con ciertas actividades, utilizando herramientas de análisis como son los algoritmos debido a la necesidad de mantener la integridad de la información que se genera permanentemente.

Esta metodología KDD ocurre en cuatro fases: (1) recopilación de datos, (2) reprocesado, (3) aplicación de la minería de datos y (4) interpretación y evaluación de resultados (Figura 1).

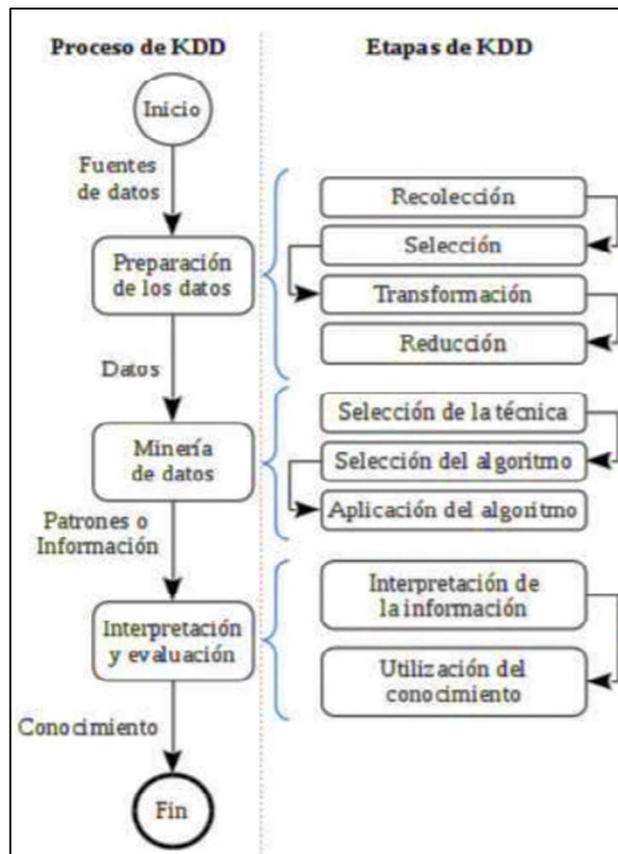


Figura 1: Proceso y etapas de KDD. Nota: Adaptado de Pacheco y Fernández (2015)

Como se ha revisado, la Minería de Datos representa un proceso a través del cual se obtiene información de interés de un conjunto de datos para procesarla de acuerdo a la necesidad del usuario y, por medio de cálculos matemáticos establecer *patrones*, es decir, extraer información que ayude a determinar características, relaciones o propiedades entre los datos que se analizan, además de tendencias que pudieren tenerse entre ellos (Duncan & Cai, 2018; Morales-Arencibia et al., 2017). Por lo general, la detección de patrones no puede ser realizada mediante una revisión habitual de los datos debido a que, por existir gran cantidad de ellos, las conexiones que pueden tener son complejas.

Tanto tendencias y patrones pueden agruparse en un *modelo de minería de datos*, que pueden ser aplicados en: a) pronóstico (predicción de cargas o inactividad de un servidor, cálculo de ventas), b) probabilidad y/o riesgo (selección de clientes, puntos de equilibrio en riesgos, especificación de probabilidad en diagnóstico), c) recomendaciones (especificación de posibles ventas de productos, planteamiento de sugerencias), d) secuencias a buscar (observación de un carrito de compras, predicción

de eventos), e) agrupación (eventos a sucederse a un grupo de clientes) (Duncan & Cai, 2018).

Generar un modelo de minería de datos es sólo un paso de un proceso que incluye: a) elaboración de preguntas que tienen relación con la información contenida en los datos, b) elaboración de un modelo de respuesta a las preguntas, c) implementación de dicho modelo (Duncan & Cai, 2018).

Este proceso de la Minería de Datos tiene sus propias *técnicas* que, según Moreno, Miguel, García, y Polo (2003), se clasifican en dos categorías: 1) predictivas o supervisadas, y b) de descubrimiento del conocimiento o no supervisadas.

Las técnicas *predictivas o supervisadas* o de aprendizaje supervisado “predicen el valor de un atributo (etiqueta) de un conjunto de datos, conocidos otros atributos (atributos descriptivos)” (Moreno et al., 2003, sec. 2). Una vez conocida la información de la etiqueta se provoca una relación de ésta con una o más propiedades, con el fin de que sirva de modelo a aplicarse en predicciones de etiquetas que no son conocidas. El proceso tiene dos etapas: a) entrenamiento, es decir que a partir de la información de una etiqueta que se conoce se crea un modelo a partir de un conjunto menor de datos, b) prueba del modelo creado en la etapa anterior en los otros datos (Moreno et al., 2003).

Una aplicación que aún no es robusta no tendrá las funcionalidades adecuadas para generar un modelo predictivo. Para solucionar ese inconveniente, están los métodos no supervisado, cuya función es establecer tendencias y patrones en los datos que se analizan, sin el uso de datos históricos, para tomar acciones y conseguir provecho, que puede ser para el negocio o para la ciencia (Moreno et al., 2003).

La Figura 2 presenta una clasificación de las técnicas de Minería de Datos.



Figura 2: Clasificación de las técnicas de Minería de Datos. Tomado de Moreno et al., (2003)

Toda la información que se extrae de los datos y su agrupación para procesamiento es realizada a través de la *vista minable*, ya que una datos de distintas fuentes de información y los extrae, escoge, une, limpia y trasforma y los cataloga, para que la modelización de los mismos sea más fácil al desarrollador (J. Hernández, s. f.). De acuerdo a Hasperué (2013, p. 33) “una vista minable consiste en una vista en el sentido más clásico de base de datos; una tabla. La mayoría de métodos de minería de datos son sólo capaces de tratar una tabla en cada tarea”. Esto quiere decir que la mediante la vista minable la información extraída de los datos es agrupada para luego realizar el proceso de minería de datos.

Manejar con datos e información requiere aplicar alguna técnica de análisis, sobre todo cuando se trata de temas como el *big data* que se convierte en “un sistema complejo que lleva a través de distintas fases a estos expertos con la intención de ser capaces de analizar la información, pulirla y obtener los datos más relevantes en cada uno de los contextos” (Universidad Internacional de Valencia, 2018, párr. 1). Una característica de este proceso es que utiliza varios tipos de técnicas con el fin de conseguir determinado objetivo, favoreciendo a que el análisis aplique a cualquier negocio.

### **2.1.3 Infraestructura Tecnológica. Técnicas de Minería de Datos a utilizar en el Proyecto**

Entre las técnicas de Minería de Datos supervisadas o predictivas que son adecuadas para este proyecto se mencionan el agrupamiento o clustering y regresión.

La técnica de *Agrupamiento o Clustering*, de acuerdo con Valenga et al. (2007) “consiste en agrupar un conjunto de datos, sin tener clases predefinidas, basándose en

la similitud de los valores de los atributos de los distintos datos” (p. 263) que no es supervisada por cuanto no existe un conocimiento previo “las clases del conjunto de datos de entrenamiento” (p. 263). El clustering descubre, de un gran grupo de datos, agrupaciones a través de medidas de distancia.

Esta técnica se basa en el principio de “maximizar la similitud interclase y minimizar la similitud interclase” (Valenga et al., 2007, p. 263). El clustering “ayuda a construir particiones significativas de un gran conjunto de objetos basado en la metodología divide y conquista, la cual descompone un sistema de gran escala en pequeños componentes para simplificar el diseño y la implementación” (Timarán, Hernández, Caicedo, Hidalgo, & Alvarado, 2016, p. 71). Esta descomposición o segmentación permite determinar poblaciones similares, con el fin de optimizar los perfiles.

De acuerdo a la estimación que año a año realiza el Cuadrante Mágico de Gartner sobre el avance de las tecnologías en el mundo, para el año 2018 de Plataformas Data Science y Machine Learning punteras en el mercado, la que se encuentra en el top 1 es *KNIME*.

La interfaz gráfica de Minería de Datos más óptima para el proyecto es KNIME que se trata de una herramienta de software libre; “es una plataforma modular de exploración de datos que permite al usuario la creación de flujos de manera visual” (Abet, Carrizo, Corso, & González, 2012, p. 208), datos con los cuales se pueden predecir otros nuevos que ayuden a optimizar el rendimiento del negocio (Ruiz, 2018). Knime, de forma selectiva, efectúa un proceso, que luego es ejecutado. Luego de la ejecución, se analizan los resultados “mediante varias vistas interactivas tanto de los datos como los modelos” (Abet et al., 2012, p. 209).

#### **2.1.4 Herramientas de Desarrollo**

Las herramientas de desarrollo utilizadas en el proyecto son C# y base de datos MySQL. En los siguientes párrafos se describe cada uno de ellos.

#### **2.1.4.1 C#**

Según Ceballos Sierra (2011), C# es considerado un lenguaje de programación para creación de aplicaciones de propósito general, con alcance web, orientado a objetos para desarrollo en Microsoft .NET. C# busca que las aplicaciones a desarrollar en C y C++ sean más fáciles y rápidas de hacer, pero sin dejar de lado las características que tienen estos lenguajes.

El antes mencionado autor señaló que C# “permite trabajar con todo tipo de datos, crear estructuras dinámicas, trabajar con ficheros, atacar a bases de datos, diseñar interfaces gráficas de usuario, etc.” (Ceballos Sierra, 2011, sec. Prólogo). Es un lenguaje portable, ya que se ejecuta independientemente de la plataforma, y como está basado en C, C++, gran parte de su sintaxis se basa en estos lenguajes; su tamaño permite desarrollar las aplicaciones evitando posibles errores, además de tener potencia y flexibilidad.

En la Tabla 1 se realiza una comparación del lenguaje de desarrollo con otros para determinar las características que permitieron su selección.

Tabla 2:  
Comparación de los lenguajes de desarrollo

Características	Java Netbeans	Eclipse	C#
Licencia	GPL2, CDDL	Licencia pública de Eclipse	
Lenguaje	Orientado a objetos	Orientado a objetos	Orientado a objetos
Costo	Sin costo, código abierto	Sin costo, open source	Versión gratuita pero limitada
Utilización	IDE, para cualquier tipo de aplicación (de escritorio, web y móvil); reutilización de módulos, permite el uso de la herramienta Update Center Module; incluye templates y wizards	IDE abierto y extensible	Para desarrollo de aplicaciones web basadas en ASP.NET, formularios de Windows y aplicaciones de escritorio basadas de WPF, también para aplicaciones Windows para teléfonos
Plataforma	Multiplataforma	Multiplataforma	
Soporte para otros lenguajes	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	
Requerimientos de hardware	Espacio en disco: 125 Mbytes, memoria: 256 Mbytes, procesador: 500 Mhz, software: JDK 1.3 o superior		

Nota: Adaptado de Mauleonph (2013), Cobos (2014), Chambi (2016), Dimes (2016), (Perry, 2012)

#### 2.1.4.2 Base de datos MySQL

Esta base de datos MySQL “es un sistema de gestión de base de datos relacional, multihilo y multiusuario (...) como software libre en un esquema de licenciamiento dual” (EcuRed, s. f.-a, párr. 1), es decir licencia GNU GPL y licencia pagada para uso empresarial. El sistema gestor de base de datos SGBD es multiplataforma y se lo utiliza en aplicaciones web. La preferencia por este SGBD es por su relación con PHP, que por lo general se combina con MySQL (EcuRed, s. f.-a).

En cuanto a la velocidad, la arquitectura multihilo permite la conexión concurrente de varios clientes a la vez; es de fácil uso, por la interfaz de comandos y las herramientas gráficas y de escritorio. Tiene soporte multiusuario, esto es, acceso concurrente a bases de datos al mismo tiempo; su sistema de privilegios es potente, portable, cumple estándares, soporte a lenguajes, soporte de aplicaciones, triggers,

vistas, procedimientos almacenados y otras características (Kumar, s. f.) que hacen de MySQL el SGBD el elegido para el desarrollo del proyecto.

Existen algunas versiones de SQL Server, que se muestran en la Figura 4.

Edición de SQL Server	Definición
Empresa	La edición premium de SQL Server Enterprise ofrece una completa gama de capacidades de centros de datos de gama alta con un rendimiento increíblemente rápido, virtualización ilimitada e inteligencia empresarial de extremo a extremo, lo que permite altos niveles de servicio para cargas de trabajo de misión crítica y acceso de usuario final a información de datos.
Estándar	La edición estándar de SQL Server ofrece una base de datos básica de gestión de datos e inteligencia empresarial para que los departamentos y pequeñas organizaciones ejecuten sus aplicaciones y es compatible con herramientas de desarrollo comunes para la nube y las instalaciones, lo que permite una gestión eficaz de la base de datos con un mínimo de recursos de TI.
Web	La edición web de SQL Server es una opción de bajo costo total de propiedad para los hosters web y los VAP web para proporcionar capacidades de escalabilidad, asequibilidad y capacidad de administración para propiedades web de pequeña a gran escala.
Desarrollador	La edición para desarrolladores de SQL Server permite a los desarrolladores construir cualquier tipo de aplicación sobre SQL Server. Incluye toda la funcionalidad de Enterprise Edition, pero tiene licencia para su uso como sistema de desarrollo y prueba, no como servidor de producción. SQL Server Developer es una opción ideal para las personas que crean y prueban aplicaciones.
Ediciones express	Express Edition es la base de datos gratuita de nivel de entrada y es ideal para aprender y crear aplicaciones de escritorio y de servidor pequeño basadas en datos. Es la mejor opción para proveedores de software independientes, desarrolladores y aficionados que crean aplicaciones de clientes. Si necesita funciones de base de datos más avanzadas, SQL Server Express puede actualizarse sin problemas a otras versiones superiores de SQL Server. SQL Server Express LocalDB es una versión liviana de Express que tiene todas sus características de programación, se ejecuta en modo de usuario y tiene una instalación rápida, sin configuración y una breve lista de requisitos previos.

Figura 3: Ediciones de SQL Server. Nota: Tomado de Ray, Milener, Roth, Steen, y Kumar (2017)

En la Tabla 2 se presenta una comparación de MySQL con otros GBD para justificar el uso de esta base de datos frente a las otras.

Tabla 3:  
Comparación de bases de datos

Características	MySQL	SQL Server	Oracle
Tipo de base de datos	Relacional	Relacional	Relacional
Tipo de herramienta	Administración gráfica, apoya el modelo y optimización de modelo de datos	Para aplicaciones inteligentes con SQL Server Machine Learning Services mediante R y Python.	
Leguaje	C, C++, amplio subconjunto del lenguaje SQL		
Plataforma	Multiplataforma	Multiplataforma	Multiplataforma
Licencia	GNU GPL de MySQL		Privada
Modelo	Cliente servidor	Cliente servidor	Soporta Cliente servidor
Soporte a base de datos			De todos los tamaños
Soporte a sistemas operativos	A varios S.O.	Soporte de transacciones	
Escalabilidad	Si	Si	Si
Estabilidad	Si	Si	
Seguridad	Si, conexión segura	Si, protección de datos	
Portabilidad	Gran portabilidad entre distintos sistemas o plataformas		
Arquitectura del sistema	Grupo de hebras en el servidor, gestión de la memoria (grupo de memorias internas, concesiones de muchas memorias, asignación de memoria dinámica, caché de planes y ejecución, api de acceso a datos, comunicación dentro de SQL Server, procesamiento de consultas distribuidas heterogéneas, réplica, modelo de réplica, opciones de réplica		<p>Servidor dedicado: estructuras de memoria (caché de memoria intermedia, memoria interna, registro histórico, rehacer cola, compartida.</p> <p>Servidor dedicado: estructuras de proceso (escritor de las bases de datos, escritor del registro histórico, punto de revisión, monitor del sistema, monitor de procesos, recuperador archivador.</p> <p>Servidor multihenebrado , agrupaciones de aplicaciones reales de Oracle.</p>

Nota: Adaptado de Miranda (2012), Cornejo (2012), Microsoft (s. f.)

## 2.2 Sustento legal

Los documentos legales que sustentan esta investigación son: **la Constitución de la República**, Capítulo segundo, Derechos del buen vivir, Sección quinta, Educación, que, en su artículo 26 establece:

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... (Asamblea Nacional Constituyente, 2008, p. 27)

El **Reglamento de Régimen Académico**, Capítulo III De la Estructura Curricular, artículo 21, Unidades de organización curricular en las carreras técnicas y tecnológicas superiores y equivalentes; y, de grado,

3. Unidad de Titulación. - Es la unidad curricular que incluye las asignaturas, cursos o sus equivalentes, que permiten la validación académica de los conocimientos, habilidades y desempeños adquiridos en la carrera para la resolución de problemas, dilemas o desafíos de una profesión. Su resultado final fundamental es: a) el desarrollo de un trabajo de titulación, basado en procesos de investigación e intervención o, b) la preparación y aprobación de un examen de grado de carácter complejo (CES, 2017a, p. 12).

En cuando al trabajo de titulación, el CES la define como...

Propuesta innovadora que contenga, como mínimo, una investigación exploratoria y diagnóstica, base conceptual, conclusiones y fuentes de consulta. Para garantizar su rigor académico, el trabajo de titulación deberá guardar correspondencia con los aprendizajes adquiridos en la carrera y utilizar un nivel de argumentación coherente con las convenciones del campo del conocimiento. (p. 13).

Sobre el **Modelo Pedagógico-Educativo de la UCSG**, en el Título I, De la Misión y Visión de la Universidad Católica de Santiago de Guayaquil y de la Representación de su Modelo Pedagógico, se toman en cuenta 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 Marco contextual**

El ámbito de aplicación es la Carrera Ingeniería en Sistemas Computacionales, la misma que se encuentra en sus últimas promociones, ya que se ha dado paso a la nueva Carrera que es Ingeniería en Ciencias de la Computación.

Las experiencias logradas mediante la aplicación de este producto de investigación, permitirán establecer de manera más efectiva el ámbito en el que el estudiante ha desarrollado sus conocimientos y cumple con un perfil específico para su etapa de titulación.

Como se ha podido apreciar en todo el desarrollo del capítulo, la presentación de los elementos teóricos permite conocer el contexto en donde se realiza el proyecto, poniendo énfasis en la Minería de Datos que, mediante la aplicación de sus técnicas se puede llegar a resolver el problema de investigación.

## **CAPÍTULO III**

### **METODOLOGÍA Y RESULTADOS**

Este capítulo analiza la Metodología de la Investigación utilizada, la Metodología de Desarrollo de la interfaz para presentación de resultados del proceso de Minería de Datos, la Metodología que sigue el proceso KDD, y el análisis de los resultados de la investigación.

#### **3.1 Metodología de la Investigación**

Para esta investigación se utilizó el enfoque de investigación cuantitativo, de tipo analítico, descriptivo y exploratorio, ya que se requirió estudiar el contexto en donde existe el problema de la identificación del perfil del estudiante, para determinar el modelo predictivo a utilizarse en la Unidad de Titulación y así asignar de forma más conveniente los temas a desarrollar por los estudiantes que optan por su título universitario.

En lo relacionado al enfoque cuantitativo, según lo que manifestado por R. Hernández, Fernández y Baptista (2014) está conformado por un grupo de procesos que se lleva a cabo de forma secuencial y demostrativa, de modo que la idea principal de la investigación va delimitándose, y luego de que la idea ya está definida se especifican los objetivos y preguntas de investigación, se analiza la información existente sobre el tema y se estructura el marco teórico del proyecto; el planteamiento de las preguntas origina hipótesis y variables, que son probadas a través del análisis estadístico y medición numérica. Este proyecto tiene enfoque cuantitativo porque existen datos numéricos que levantarse.

Se dice que el proyecto es analítico ya que se procede a separar las partes que conforman el problema a estudiar, de manera que cada una sea analizada individualmente (Bernal, 2010). A través de este método el problema que existe en la Unidad de Titulación de la Carrera de Ingeniería en Sistemas Computacionales sobre la falta de conocimiento del perfil del estudiante de la Unidad de Titulación para la asignación de un tema para desarrollar, es analizado desde su contexto, entendiendo

que la Carrera no existe ningún medio o método para conocer las habilidades de los estudiantes y las áreas en donde se hayan destacado.

El proyecto es descriptivo, por cuanto se especifican las particularidades o atributos del problema a estudiar, entendiendo que se refiere a lo más importante y distintivo de ese problema. En los estudios descriptivos, además de seleccionar lo más importante del problema, se pueden desarrollar “productos, modelos, prototipos, guías..., pero no se dan explicaciones o razones de las situaciones, los hechos, los fenómenos...” (Bernal, 2010, p. 113). En este proyecto, el análisis de las causas más significativas del problema se traduce en la elaboración del modelo predictivo para determinar el perfil del estudiante y en el desarrollo del prototipo para la presentación de los resultados de la aplicación de las técnicas de Minería de Datos, que es lo utilizado para el análisis de la información.

Por último, el proyecto es exploratorio ya que se parte de un problema de poco estudio y análisis, del que no se tiene una idea clara, cuya investigación se realiza desde un punto de vista innovador, sirviendo de base para posibles futuros estudios. Este tipo de alcance de investigación es utilizado para conocer en profundidad un problema que es de poco conocimiento, conseguir información para conocer si se puede realizar investigaciones más integrales de determinado entorno, analizar nuevos fenómenos, determinar variables, determinar aspectos importantes para tratar posibles investigaciones o referirse a premisas (R. Hernández et al., 2014). En este proyecto, la aplicación de la investigación exploratoria está dado porque el problema del desconocimiento del perfil del estudiante de la Unidad de Titulación es relativamente de poco estudio, y se necesita analizarlo en profundidad para plantear la solución más óptima.

### **3.2 Metodología de Desarrollo**

Para el desarrollo de la interfaz de presentación de resultados del proceso de Minería de Datos se utilizó el prototipo.

#### ***Prototipo***

De acuerdo a lo manifestado por Sommerville (2011):

Un prototipo es una versión inicial de un sistema de software que se usa para demostrar conceptos, tratar opciones de diseño y encontrar más sobre el problema y sus posibles soluciones. El rápido desarrollo iterativo del prototipo es esencial, de modo que se controlen los costos, y los interesados en el sistema experimenten por anticipado con el prototipo durante el proceso de software. (p. 45).

El prototipo se lo utiliza para predecir modificaciones relacionadas con la selección de requisitos o encontrar respuestas concretas en el desarrollo de software y asistir en el diseño de la interfaz de usuario (Sommerville, 2011)

El uso del prototipo en el proyecto es una forma de realizar de forma rápida la interfaz para la presentación de los resultados del proceso de Minería de Datos, y en caso de que se requieran nuevas modificaciones, se podrá desechar ese prototipo y desarrollar otro, de acuerdo a las nuevas necesidades.

### **3.3 Metodología de Minería de Datos**

La Minería de Datos cumple con las siguientes etapas: recopilación de los datos necesarios para la investigación, el procesamiento de los mismos, la aplicación de algoritmos y finalmente la interpretación de los resultados.

#### **3.3.1 Recopilación de datos**

Es el momento en el cual se realiza la recopilación de la información que se encuentra disponible en los repositorios de datos. En primer lugar, se efectúa la selección de un grupo de elementos que aporten al estudio que se realiza y luego se los debe reunir considerando las fuentes de datos en los que la información está disponible. Al final, la información deberá integrarse en un solo grupo de datos para su análisis (Márquez et al., 2012).

Como los datos son la fuente de donde se extrae la información relevante, es en la fase de recopilación que se cumplen algunas tareas, tales como: 1) conseguir los datos, b) dar formato a los datos, c) elaborar las herramientas necesarias y ambientes adecuados, d) validar los datos conseguidos y el formato dado, e) trabajar con muestras

aleatorias elaboradas, f) analizar, calibrar, validar y realizar prueba oculta de los datos (Vieira, Ortiz, & Ramírez, 2009).

La obtención del conocimiento del tema y lo que se quiere conseguir se divide en tres tareas:

La primera tarea, relacionada con el *Aprendizaje del tema*, considerado importante ya que se debe “conocer 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” (Pacheco & Fernández, 2015, p. 27).

Le sigue la *Recolección de datos*, como punto de partida para identificar las necesidades específicas del entorno de estudio. Es importante identificar dónde están los datos precisos, la forma de su obtención, cómo se obtuvieron de las fuentes, su forma de combinación y el nivel de confianza que tienen las fuentes de las cuales provienen los datos, con el propósito de establecer si ofrecen información útil y aplicable al estudio (Pacheco & Fernández, 2015; Planas, Rodríguez, & Lecha, 2004).

Finalmente, es importante tener suficiente *Experiencia en análisis de datos*, ya que de la interpretación e inferencia que se haga de los datos, así como el manejo eficiente de las herramientas estadísticas de soporte, depende la presentación de los resultados y su incidencia en el estudio (Villar-Ledo & Ledo-Ferrer, 2016).

### **3.3.2 Pre-procesado o transformación de los datos**

Se trata de la conversión de datos analizados y transformarlos a un modelo analítico (Moreno et al., 2003; Pacheco & Fernández, 2015).

La eficiencia del tratamiento estadístico a través del procesamiento o transformación de los datos obtenidos, va a permitir presentar resultados idóneos, apegados al objeto de estudio (Ramiro Vásquez, Caballero Núñez, & Herrera Villafranca, 2017). De acuerdo con Martín Pérez (2016) el procesamiento y sistematización de los datos obtenidos de una investigación debe realizarse tomando en consideración las relaciones existentes entre las variables que intervienen en el

objeto de estudio y su interpretación cuantitativa, así como el uso adecuado de las herramientas estadísticas.

En esta fase se disponen todos los datos recopilados para luego utilizar las técnicas más adecuadas de Minería de Datos de procesamiento de la información. El pre-proceso consiste en: a) limpiar los datos, b) transformar las variables, c) realizar la interpretación de los datos. También existen técnicas tales como “la selección de atributos y el re-balanceado de datos para intentar solucionar los problemas de la alta dimensionalidad y desbalanceo que presentan normalmente este tipo de conjuntos de datos” (Márquez et al., 2012, p. 110).

En cuanto a la *selección de los datos*, la UIAF (2014) señaló que esta parte del proceso consiste en la recolección de los datos de fuentes útiles, en donde se identifican las distintas variables contenidas en los datos y a las cuales se les aplica las técnicas más recomendadas para el muestreo.

Sobre la *limpieza*, ésta se refiere al filtrado de los datos en cuanto tiene que ver con información incompleta y/o que no sea la correcta. En cuanto a la *transformación*, se realiza la conversión de los datos, a través de la aplicación de técnicas “de reducción o de aumento de la dimensión y de escalado simple y multidimensional” (UIAF, 2014, p. 16).

### **3.3.3 Algoritmos de Minería de Datos**

Referido al “tratamiento automatizado de los datos seleccionados con una combinación apropiada de algoritmos” (Moreno et al., 2003, sec. Técnicas de Minería de Datos) que, según Ruiz (2018), es la relacionada con el modelado mediante “herramientas de análisis en base a los datos que se tienen” (p. 3) utilizando patrones conocidos o no anteriormente, que sirvan para análisis y proyección de datos, y que “puedan ser útiles, comprensibles y que se encontraban ocultos hasta el momento” (p. 3).

Se determina el trabajo a realizar como puede ser agrupar o clasificar la información, mediante la aplicación de las técnicas (algoritmos) de acuerdo a los modelos de minería de datos (descriptivos o predictivos), cuyos resultados

contribuirán en la toma de decisiones. Al final del proceso, los algoritmos utilizados deberán evaluarse y comparar los resultados obtenidos, para establecer los que tuvieron resultados más fiables (Márquez et al., 2012; UIAF, 2014).

#### **3.3.4 Interpretación y evaluación de resultados**

Se refiere a la evaluación, análisis e interpretación de los patrones encontrados en la fase anterior, que realizan expertos, por lo general, mediante visualización y se determinan modelos y patrones para utilizarlos en el problema a resolver. Si se requiere, se pueden volver a realizar nuevas iteraciones (Márquez et al., 2012; Moreno et al., 2003, sec. Técnicas de minería de datos; UIAF, 2014).

Por último, se socializa la información encontrada a los potenciales usuarios que serán los encargados de analizarlos y tomar las decisiones correspondientes.

### **3.4 Población y muestra**

La población estuvo constituida por los estudiantes de las tres últimas cohortes de los períodos A y B 2017 y A 2018. La muestra estaba conformada por 54 estudiantes, de los cuales se obtuvo información de 47 de los mismos.

### **3.5 Análisis de Resultados**

El levantamiento de información para determinar el perfil de los estudiantes se realizó mediante encuesta a través de la plataforma virtual, a la muestra de los estudiantes de la Unidad de Titulación.

Para la recolección de los datos, se procedió a realizar una carta a la Directora (e) de la Carrera de Computación y Sistemas Computacionales, para que sea autorizado el acceso a los datos de los estudiantes.

Se tomó la información proporcionada por las secretarías de la carrera, es decir, información de la plataforma de la UCSG, el SIU, la misma que fue receptada en formato PDF para su posterior tratamiento. Se transcribió los datos a un archivo de Excel para mejor manipulación de los mismos; además, se recogió la información en

físico de las prácticas tanto de laboratorio o técnicas, sociales y pre profesionales para seleccionar los datos más relevantes.

En la Figura 4 se muestran los datos de los estudiantes, referente a las notas y las prácticas de laboratorio y pre profesionales.

FUNDAMENTOS DE PROGRAMACIÓN	Maticuls	LOGICA COMPUTACIONAL	Maticuls	BASE DE DATOS	Maticuls	ADMINISTRACION DE BASE DE DATOS	Maticuls	ANALISIS Y DISEÑO DE SISTEMAS	Maticuls	FUNDAMENTO DE REDES DE COMPUTADORAS	Maticuls	ENRUTAMIENTO DE REDES	Maticuls	Maticula	TELECOMUNICACIONES	Maticula	PRACTICA PREPROFESIONAL	ASISTENCIA
7.26	1	7.1	2	7.01	1	7.5	1	9.26	1	7.89	1	8.38	1	1	7.25	1	DESARROL	78
7.5	3	6.52	3	8.59	3	8.59	3	6.82	1	7.34	1	7.72	1	1	7.2	1	DESARROL	74
7.38	1	7.02	2	6.79	1	8.43	2	6.77	1	8.53	1	7.46	1	1	7.43	1	DESARROL	93
10	1	8.25	1	7.37	1	8.47	1	8.77	1	8.25	1	8.59	1	1	9.61	1	DESARROL	85
7.19	1	7.88	2	6.52	1	7.14	1	6.79	1	7.22	1	8.8	1	1	7.01	1	DESARROL	88
8.52	1	7	1	7.4	1	7.96	1	8	1	7.74	1	9	1	1	9.11	1	TECNICO TI	95
9	1	7.4	1	8.25	1	8.61	1	8.62	1	7.74	1	7.74	1	1	8.37	1	DESARROL	84
8.5	1	6.5	1	7.69	1	7.94	1	6.53	1	7	1	8.42	2	1	7.18	2	ANALISTA	84.48
8.37	1	6.92	1	7	1	8.67	1	8.99	1	7.1	1	7.3	1	1	7.1	1	DESARROL	88
7	1	7.37	1	8.37	1	9.43	1	8.65	1	9	1	8.4	1	1	8.99	1	ANALISTA	93
7	1	6.75	1	6.53	1	6.53	1	8.3	1	7.06	1	8.7	1	1	7.12	1	DBA	84.27
7.12	1	7.63	2	7.61	1	8.14	1	8.73	1	7.64	1	9.79	1	1	7.27	1	DESARROL	93
9.62	1	7.7	1	7.15	1	9.23	1	7.12	1	7.07	1	8.33	1	1	8.62	1	DBA	78
6.62	1	7.26	1	7	1	8.48	2	8.57	1	8.31	1	7.62	1	1	7.44	1	TECNICO TI	79
8.75	2	7	2	7	1	8.92	3	7.18	2	7	1	7.3	1	1	8.71	2	DESARROL	68
8.25	1	7.63	2	7.95	2	7.97	1	8.03	2	7	1	7.25	1	2	8.04	2	DESARROL	83
6.5	1	7.31	1	7.82	1	7.03	1	8.44	1	7	1	7.91	1	1	7.15	1	ANALISTA	88
6.5	2	6.5	2	6.91	1	7.21	2	6.55	1	7.82	2	7.46	1	1	7	1	TECNICO TI	91
6.69	2	7.25	2	8.15	2	8.52	1	7.24	1	7.92	1	8.95	1	1	8.43	1	TECNICO TI	87
9.67	1	7.25	1	7.75	1	8.63	1	8.91	1	8.01	1	9.49	1	1	7.95	1	DESARROL	86
7	2	6.77	1	7.03	1	7.97	2	8.69	1	7.58	1	7.84	1	1	7.41	1	DESARROL	87
8.87	1	7.37	1	8.13	2	8.64	1	8.84	1	7.08	1	8.25	1	2	7.87	1	DESARROL	80
7.63	1	6.5	1	7.04	1	7.9	1	7	1	8.37	1	8.59	1	1	7.6	1	DESARROL	84
8.25	1	7.73	2	7.02	1	9.57	1	9.04	2	7.56	1	8.06	1	1	7.69	1	DESARROL	80.3

Figura 4: Conjunto de datos

Posteriormente se realizó un análisis de los estudiantes graduados de las promociones A y B 2017 y A 2018, mediante una encuesta vía correo electrónico para conocer si trabajan actualmente y en qué área. La encuesta cubrió algunos aspectos, y se muestra en los párrafos siguientes.

**ENCUESTA PARA DETERMINAR PERFILES DE ESTUDIANTES DE LA UNIDAD DE TITULACIÓN DE LA CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES DE LA FACULTAD DE INGENIERÍA DE LA UCSG**

1. Mientras cursó su carrera universitaria, ¿trabajó en alguna empresa?

SI

NO

2. Si la respuesta anterior fue afirmativa, responda lo siguiente:

¿En qué área de la carrera lo hizo? \_\_\_\_\_

¿Cuánto tiempo trabajó? \_\_\_\_\_

3. Mientras estuvo cursando sus estudios universitarios, ¿participó en algún tipo de evento organizado por la carrera?

SI

NO

4. Si la respuesta anterior fue afirmativa, responda el tipo de evento:

Evento de desarrollo de sistemas \_\_\_\_\_

Evento de incentivo al emprendimiento \_\_\_\_\_

*Tabla 4:*

Resultados de encuesta

Desarrollo o afines	Infraestructura	No trabajan	TOTAL
23	22	2	47

*Tabla 5:*

Porcentaje de estudiantes según área de trabajo

Desarrollo o afines	Infraestructura	No trabajan	TOTAL
49%	47%	4%	100%

Luego de procesar la información recabada se pudo conocer que 23 estudiantes trabajan en el área de desarrollo o afines 22 en infraestructura y 2 no trabajan, esto es 49%, 47% y 4% respectivamente. Esta información sirvió de base para realizar el cálculo del perfil del estudiante de la Carrera de Ingeniería en Sistemas Computacionales.

De lo anotado anteriormente, se puede observar que la determinación de la metodología de investigación, de desarrollo y de Minería de Datos ha permitido recabar información para conocer los perfiles de los estudiantes, de modo que esta información sirva para mejorar la asignación de los temas de los trabajos de titulación.

## **CAPÍTULO IV**

### **PROPUESTA**

La propuesta de un Diseño de un Modelo Predictivo, mediante la técnica de Minería de Datos, para identificar el perfil del estudiante en la Unidad de Titulación en la Carrera Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG, se describe en los párrafos siguientes, a través de los cuales se pretende presentar el tratamiento de cada fase y la correspondiente captura de pantalla.

#### **4.1 Introducción**

El modelo predictivo a aplicarse en la Unidad de Titulación de la Carrera de Ingeniería en Sistemas Computacionales está destinado a determinar el perfil del estudiante para la correcta asignación de los temas para el Trabajo de Titulación. La aplicación de las técnicas de Minería de Datos, a través de las herramientas utilizadas en este proyecto, permiten procesar los datos válidos de los estudiantes relacionados, básicamente, con su experiencia laboral en determinada área y sus actividades estudiantiles previas.

El procesamiento de la información del estudiante permite identificar su perfil en base a los estudiantes que ya han terminado su carrera, de manera que la Unidad de Titulación pueda tener un referente para aplicarlo a las cohortes siguientes y optimizar el desarrollo de los Trabajos de Titulación, por cuanto se facilitará proporcionar al alumno un proyecto ya aprobado por la Comisión Académica de la Carrera, evitando el desperdicio de tiempo en su revisión y se lo trabaje en el período de tiempo establecido.

#### **4.2 Objetivo**

Determinar el perfil del estudiante de la Unidad de Titulación de la Carrera de Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG, con el fin de optimizar el proceso de asignación de temas para Trabajo de Titulación de manera que se reduzcan los tiempos en la aprobación y ejecución de los mismos.

### **4.3 Responsables**

El requerimiento de la determinación de los perfiles de estudiantes está a cargo de la Directora de las carreras de Computación e Ingeniería en Sistemas Computacionales y debe ejecutarse por el Coordinador de la Unidad de Titulación.

### **4.4 Descripción de la solución**

La interfaz que presentará los resultados de la aplicación de las técnicas de Minería de Datos que determinan el perfil del estudiante, se alimentará de un archivo en formato CSV que contiene las notas y las prácticas de los alumnos.

La solución es una aplicación de escritorio, que tiene dos opciones: Cargar datos, que se alimenta de un archivo CSV que contiene el nombre del estudiante, las notas de las materias y las prácticas. Se genera el perfil y se emite un reporte.

#### **Herramienta de Minería de Datos**

La herramienta de Minería de Datos utilizada es KNIME, la misma que es una interfaz que se alimenta de un archivo en formato CSV al cual se aplican las técnicas escogidas (Clustering y Árboles de decisión) para determinar el perfil del estudiante.

#### **Proceso**

En los párrafos siguientes se presenta el proceso de las fases de Minería de Datos para generar el perfil del estudiante.

#### ***Recopilación de datos***

La primera fase consiste en la recolección de la información de los estudiantes de la unidad de titulación conformada por las tres últimas cohortes de los períodos A y B 2017 y A 2018.

En cuanto a las tareas, éstas comprenden:

### ***Aprendizaje del tema***

Conocimiento de la información en base a información recabada a los coordinadores de las áreas de hardware y programación, los mismos que tienen relación directa con los temas de Minería de Datos, utilizada en el proyecto. Se tomaron todas las materias de la malla curricular 2012, para finalmente seleccionar las que corresponden a tres áreas básicas.

En cuanto al área de hardware, se conoció que las asignaturas de Fundamentos de Redes de computadoras y Enrutamiento de redes, Telecomunicaciones y Lógica Computacional son las que influyen para la determinación de los perfiles de los estudiantes, lo mismo ocurre con el área de software y sus asignaturas Administración de Centros de Cómputo, Base de Datos y Administración de base de datos, Ingeniería de Software y Análisis y Diseño de Sistemas orientado a objetos, junto con las asignaturas del área de programación: Fundamentos de Programación, Programación Orientada a Objetos, Programación en Capas, Programación Distribuida y Programación Móvil.

### ***Recolección de datos***

Los datos que se utilizaron en el proyecto fueron solicitados a la Directora (e) de la Carrera, a quien se emitió una carta solicitando los mismos para su revisión en el SIU. La información relacionada con las pasantías fue proporcionada por las Secretarías Administrativas.

### ***Experiencia en análisis de datos***

Se descartan las prácticas sociales que fueron proporcionadas por las Secretarías Administrativas de la Carrera porque son destinadas a la comunidad y no representan un valor específico para el cálculo del perfil del estudiante, y de laboratorio porque van alineadas a las pre profesionales, por lo que la única información relevante para el diseño del modelo predictivo fue los datos de los estudiantes de las tres últimas cohortes de los períodos A y B 2017 y A 2018, en cuanto a las notas de todas las materias de la malla curricular de la carrera y si se encuentran trabajando.

El resultado son los datos presentados en la vista minable. previos a la aplicación de los algoritmos de la Minería de Datos.

### ***Pre-procesado o transformación de los datos***

La fase de pre-procesado permite determinar cuáles son los datos de mayor relevancia, del grupo de información levantada en la etapa de selección.

Se descartó a estudiantes que, a pesar de tener el perfil definido, no trabajan en el área correspondiente y hace que la información se disperse. La calificación de la materia se da en base a la cantidad de veces que se ha tomado la misma, es decir que para a partir de la segunda matrícula, se resta el 25% la calificación (ver Figura 5).

	CJ	CK	CL	CS	CT	CU	DB	DC	DD	DE	DF	DG
INGENIERIA DE SOFTWARE		SS M	Matricula	PROGRAMACION MOVIL	P5 M	Matricula	ADMINISTRACION Y ORGANIZACION DE CENTROS	S6 M	Matricula	ENRUTAMIENTO DE REDES	H3 M	Matricula
	7.94	7.94	1	7.21	7.21	1	7.3	7.3	1	8.38	8.38	1
	7.46	7.46	1	7	7	1	7.06	7.06	1	7.72	7.72	1
	7.4	7.4	1	7.55	7.55	1	7.41	5.5575	2	7.46	7.46	1
	8.03	8.03	1	8.07	8.07	1	7.22	7.22	1	8.59	8.59	1
	7.12	7.12	1	7.27	5.4525	2	7.03	7.03	1	8.8	8.8	1
	7.85	7.85	1	8.26	8.26	1	7.62	7.62	1	9	9	1
	7.22	7.22	1	8	8	1	7.08	5.31	2	7.74	7.74	1

Figura 5: Promedio del estudiante según la matrícula

El promedio de cada área corresponde a la cantidad de materias que tiene cada una. Por ejemplo, el área de programación tiene 5 asignaturas: Fundamentos de Programación, Programación Orientada a Objetos, Programación en Capas, Programación Distribuida y Programación Móvil. Para su cálculo se suman los promedios de cada materia y se divide para 5.

Según la información proporcionada por los docentes coordinadores de cada área, las materias seleccionadas para el cálculo del perfil son las que mayor influencia tienen en la malla curricular. En la Figura 6 se muestra el cálculo del promedio de las materias.

EJ	EK	EL	EM	EN	EO	EP
P (Fundamentos, Objetos, Capas, Distribuida, Movil)	H (Fundamentos redes, Enrutamiento, Telecomunicaciones, Logica)	S1 (Centros de Computo)	S2 (Fundamentos y Administracion de Bases)	S3 (Ing. Software, Analisis y Diseno de Sistemas 1 y 2)		
7.516	7.21125	7.3	7.255	8.15		
5.168	6.38	7.06	4.295	6.6775		
.LA 6.8415	7.17125	5.5575	6.55625	7.186666667		
.LA 9.066	8.675	7.22	7.92	8.423333333		
.LA 6.1445	7.235	7.03	6.83	7.456666667		
8.092	8.2125	7.62	7.68	7.87		
.LA 7.716	7.8125	5.31	8.43	7.67		

Figura 6: Cálculo del promedio de las materias por área

Luego del promedio se asigna el peso al área. La cantidad de materias entre programación, hardware y software de la malla curricular son 23, de las cuales se escogieron 5 para programación, 10 para software y 8 materias para hardware. El cálculo del peso se realiza multiplicando el número de materias de cada área multiplicado por 100 y se divide para el número total de materias de la malla curricular, como se puede apreciar a continuación.

$$\text{Programación} = \frac{5 \times 100}{23} = 22\%$$

$$\text{Hardware} = \frac{8 \times 100}{23} = 35\%$$

$$\text{Software} = \frac{10 \times 100}{23} = 43\%$$

La asignación de la cantidad de materias por área fue realizada por los coordinadores de las áreas para poder realizar los cálculos correspondientes al peso, de acuerdo a los requerimientos del proyecto. En la Figura 7 se presenta el peso del área que se calculó.

B	C	D	E	F
Programación 22%	Hardware 35%	Software1 7.17 %	Software2 14.33 %	Software3 21.5%
1.65	2.52	0.52	1.04	1.75
1.14	2.23	0.51	0.62	1.44
1.51	2.51	0.40	0.94	1.55
1.99	3.04	0.52	1.13	1.81
1.35	2.53	0.50	0.98	1.60
1.78	2.87	0.55	1.10	1.69
1.70	2.73	0.38	1.21	1.65
1.38	2.21	0.25	1.12	1.51
1.67	2.54	0.25	1.12	1.69
1.81	2.95	0.58	1.28	1.88
1.39	2.59	0.43	0.94	1.66
1.65	2.66	0.50	1.13	1.69
1.91	2.78	0.58	1.17	1.44
1.27	2.68	0.53	0.96	1.71
1.53	2.28	0.50	0.82	1.26
1.57	2.28	0.52	1.00	1.61
1.55	2.21	0.58	0.88	1.65
1.45	2.69	0.41	1.05	1.60
1.84	2.86	0.53	1.17	1.72
1.33	2.59	0.52	0.93	1.68
1.80	2.67	0.52	1.06	1.65

Figura 7: Peso del área

La información del peso se descarga en un archivo con formato CSV para su posterior procesamiento en KNIME; el archivo contiene información en donde cada cantidad de registros tiene número, área, prácticas, trabajo, asistencia.

### *Algoritmos de Minería de Datos*

La data del archivo en formato CSV que se ingresa a KNIME es la preprocesada en la fase anterior, que es la información que se genera en la vista minable.

Tabla 6:  
Diccionario de datos

Atributo	Tipo	Valores posibles
Programación	Entera	>1.53 o <=1.53
Hardware	Entera	>2.52 o <=2.52
Software 1	Entera	>0.47 o <=0.47
Software 2	Entera	>0.99 o <=0.99
Software 3	Entera	>1.61 o <=1.61
Trabajo	Nominal	Desarrollador, Administrador, Redes, IT, Soporte, Infraestructura, Atención al cliente, Mesa de Ayuda, Analista
Práctica pre profesional	Nominal	Desarrollador, IT, Analista

Row ID	'D' 'Progra...	'D' 'Hardw...	'D' 'Softwa...	'D' 'Softwa...	'D' 'Softwa...	'I' 'PRACT...	'I' 'PRACT...	'I' 'TRABA...	'I' 'TRABA...	'I' 'TRABA...	'D'
1	1.65	2.52	0.52	1.04	1.75	1	0	1	0	0	78
2	1.14	2.23	0.51	0.62	1.44	1	0	0	1	0	74
3	1.51	2.51	0.4	0.94	1.55	1	0	0	0	1	93
4	1.99	3.04	0.52	1.13	1.81	1	0	0	0	1	85
5	1.35	2.53	0.5	0.98	1.6	1	0	0	0	1	88
6	1.78	2.87	0.55	1.1	1.69	0	1	0	1	0	95
7	1.7	2.73	0.38	1.21	1.65	1	0	0	0	1	84
8	1.38	2.21	0.25	1.12	1.51	1	0	1	0	0	84
9	1.67	2.54	0.25	1.12	1.69	1	0	0	0	1	88
10	1.81	2.95	0.58	1.28	1.88	1	0	0	0	1	93
11	1.39	2.59	0.43	0.94	1.66	0	1	0	1	0	84
12	1.65	2.66	0.5	1.13	1.69	1	0	0	0	1	93
13	1.91	2.78	0.58	1.17	1.44	0	1	0	0	1	78
14	1.27	2.68	0.53	0.96	1.71	0	1	0	1	0	79
15	1.53	2.28	0.5	0.82	1.26	1	0	0	0	1	68
16	1.57	2.28	0.52	1	1.61	1	0	0	1	0	83
17	1.55	2.21	0.58	0.88	1.65	0	0	0	1	0	91
18	1.45	2.69	0.41	1.05	1.6	0	1	0	1	0	87
19	1.84	2.86	0.53	1.17	1.72	0	1	0	0	1	86
20	1.33	2.59	0.52	0.93	1.68	1	0	0	1	0	87
21	1.8	2.67	0.52	1.06	1.65	1	0	0	0	1	80
22	1.7	2.72	0.55	1.07	1.54	1	0	0	0	1	84
23	1.48	2.55	0.31	1.19	1.66	1	0	0	1	0	80
24	1.57	2.33	0.52	0.84	1.49	1	0	0	1	0	87

Figura 8: Vista minable

### ***Descripción de los algoritmos***

Antes de describir los algoritmos utilizados para procesar la información, se debe analizar cada uno de los nodos del workflow tanto de clustering como de árbol de decisión.

En cuanto al workflow para clustering, el nodo de *lectura de datos* realiza la lectura de los datos del archivo CSV de la vista minable de la fase anterior; *valor faltante* permite completar valores faltantes en los campos, en caso de que alguno no exista; *particionamiento*, la tabla que va a definir los clustering: conjunto de entrenamiento y conjunto de pruebas; *k-media* o *k-means*, que genera los clusters; *gestor de color*, para asignar un color a los clusters; *administrador de formas*, da formas a los valores; *gráfico de dispersión*, para la creación de un gráfico de dispersión para mostrar la definición de los grupos. En la Figura 9 se muestra el workflow para clustering.



Figura 9: Workflow de k-Means, para perfiles

En el nodo *particionamiento* se requiere establecer los valores absolutos y relativos (90-10), en donde el 90% de los datos se analicen por el nodo que genera los clusters y el 10% para generar un prototipo o para entrenar al algoritmo.

La data que se envía al conjunto de entrenamiento primero es leída, luego se particiona y finalmente se aplica el algoritmo de clustering.

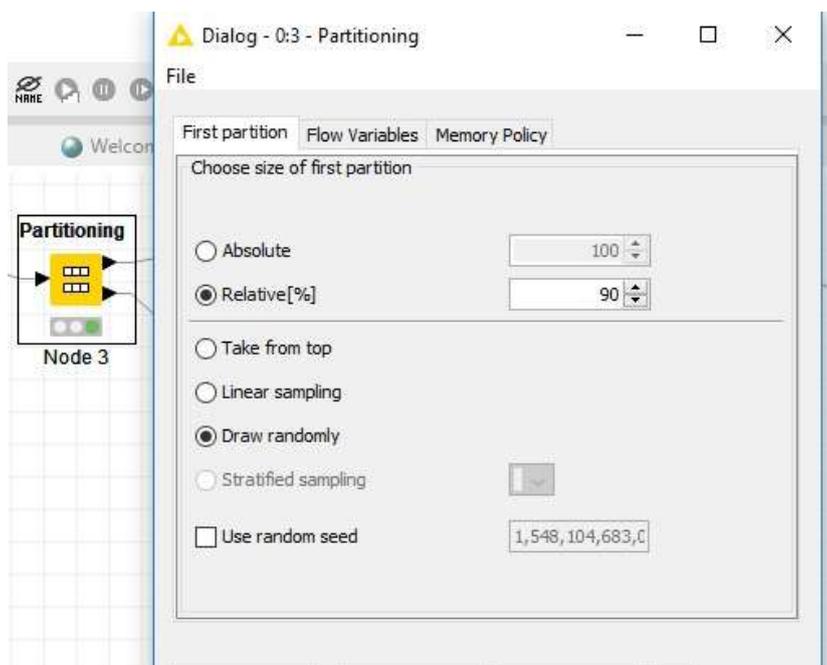


Figura 10: Particionamiento

En cuanto al algoritmo de **Clustering**, se utiliza el método de agrupamiento k-Means con su medida euclidiana, a través de la cual se divide la información en grupos afines: desarrollo e infraestructura. Todos los datos que se recibe de los estudiantes se agrupa de acuerdo a los parámetros de trabajo, prácticas y áreas.

En la Figura 10 se muestran las propiedades de la herramienta para la creación de los dos clusters, en donde el cluster\_0 = desarrollo y el cluster\_1 = infraestructura.

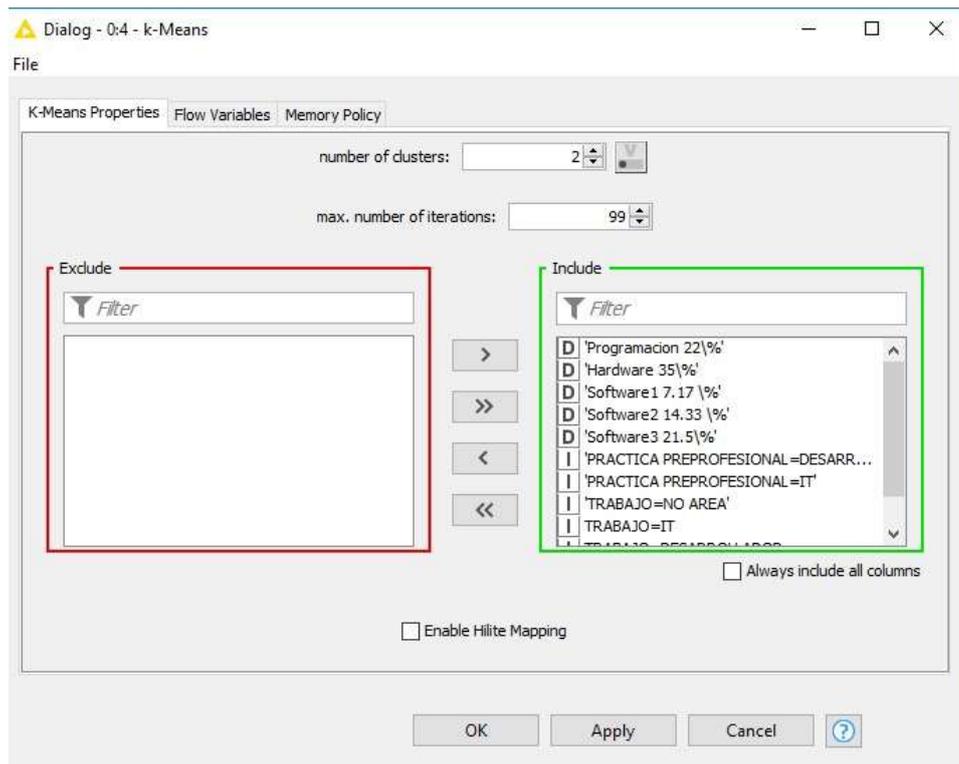


Figura 11: Configuración del k-Means

En la Figura 12 se observan los clusters ya generados y en la figura 13 se muestra la asignación del color a las columnas de los clusters.

Row ID	D 'Progra...	D 'Hardw...	D 'Softwa...	D 'Softwa...	D 'Softwa...	D 'PRACT...	D 'PRACT...	D 'TRABAJO=NO AREA'	D 'TRABAJO=IT'	D 'TRABAJO=DESARROLLAD...	D 'A
cluster_0	1.663	2.692	0.483	1.081	1.676	0.75	0.214	0.036	0.357	0.607	89.76
cluster_1	1.53	2.521	0.47	0.995	1.612	0.571	0.429	0.071	0.571	0.357	77.30

Figura 12: Clustering

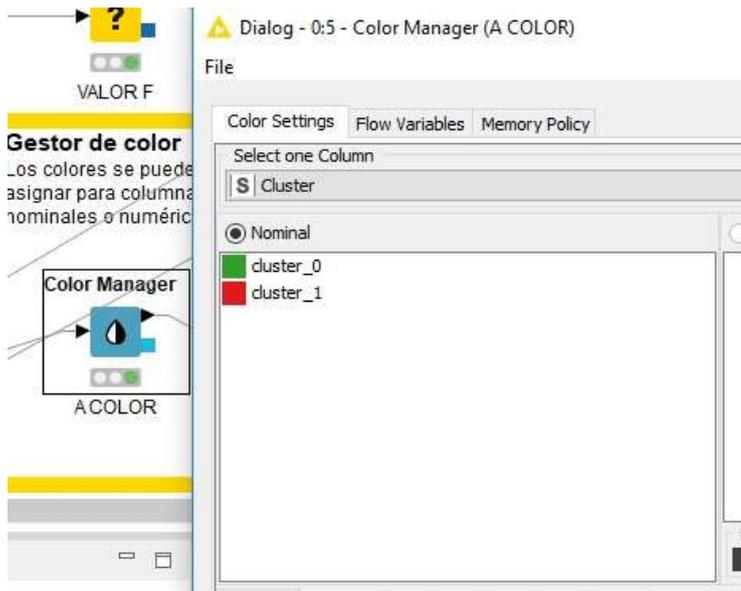


Figura 13: Gestor de color

En la figura 14 se muestra el administrador de formas de los clusters y en la figura 15 se presenta la dispersión de los datos.

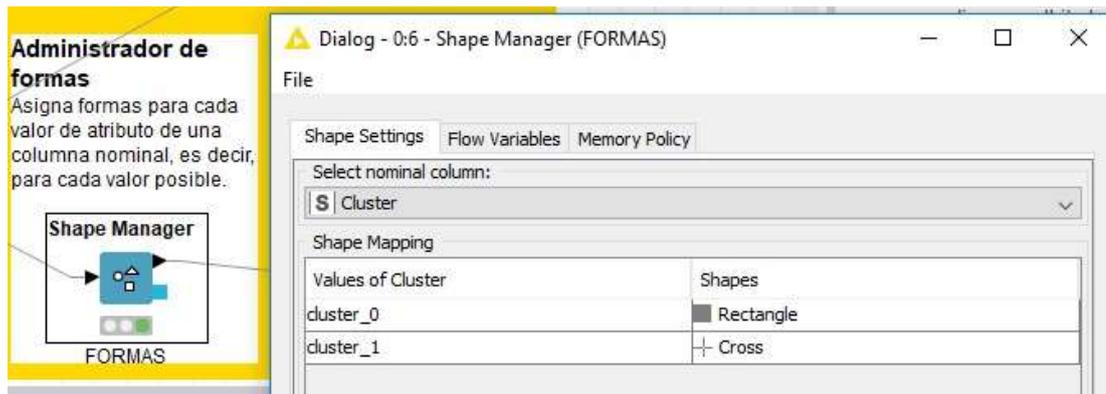


Figura 14: Administrador de formas

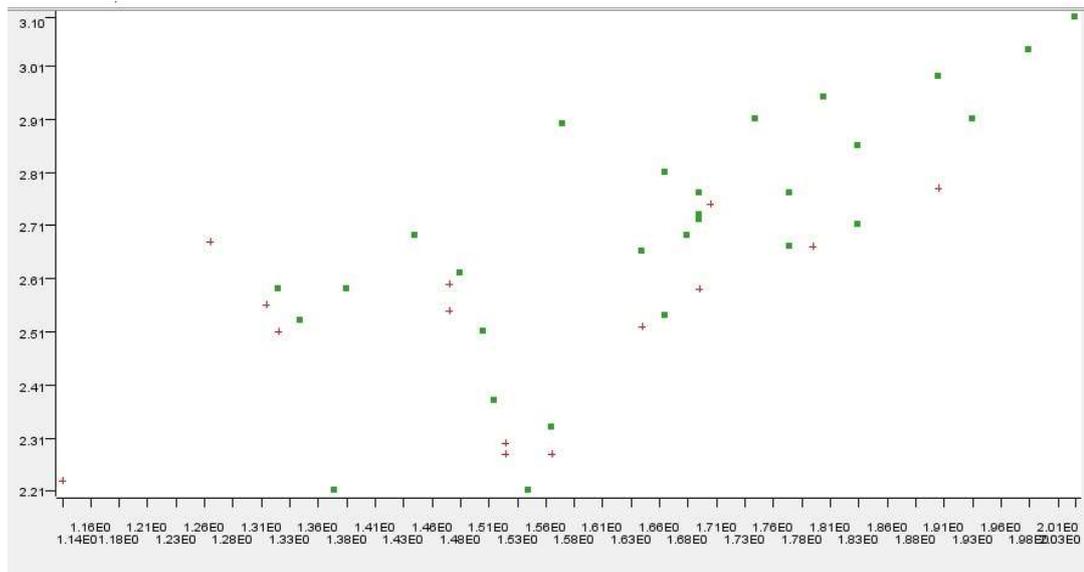


Figura 15: Dispersión de los datos

En el workflow de para el árbol de decisión el nodo *lectura de datos* realiza la lectura del archivo CSV en la vista minable; *gestor de color*, asigna un color a las columnas; *valor faltante*, coloca un valor en caso de que falte para completar; *particionamiento*, asigna el valor relativo 90-10 (90% para entrenamiento, 10% para pruebas); *árbol de decisión (aprendiz)*, los datos enviados para que el algoritmo aprenda; *aplica el modelo*, hace una comparación entre los datos enviados al aprendizaje y predice el valor de los patrones; *árbol de decisiones para el conjunto de reglas*, que genera todo el conjunto de reglas para los perfiles; *resultados del modelo*, que genera la matriz de confusión.

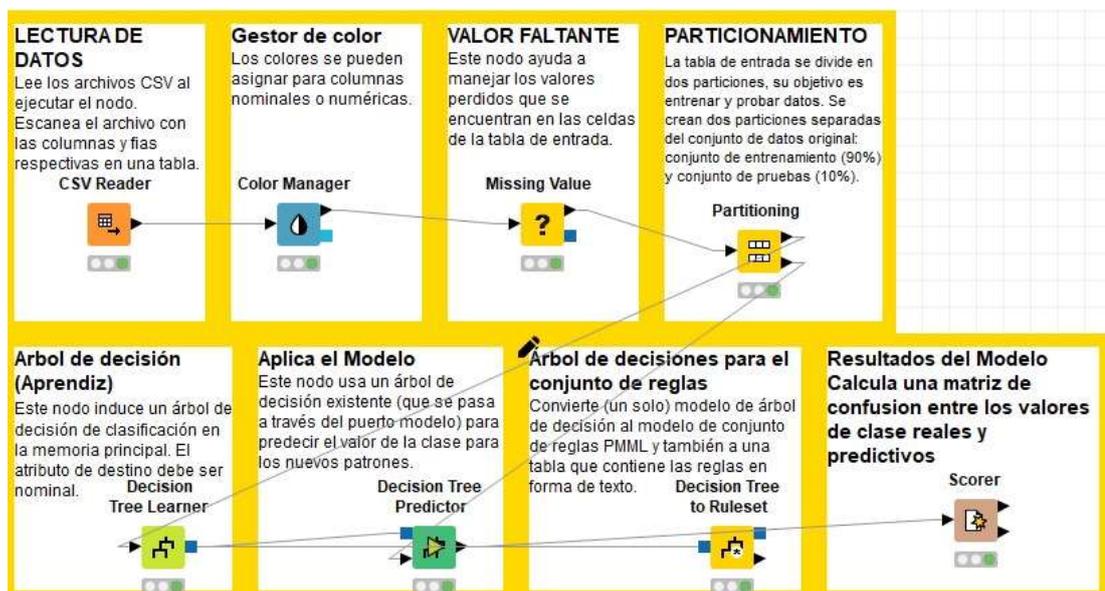


Figura 16: Workflow de árbol de decisión, para validación de perfiles

En la Figura 17 se muestra la configuración del aprendizaje del árbol de decisión, la misma que consiste en el ingreso de información nominal de entrenamiento a una pseudo clasificación.

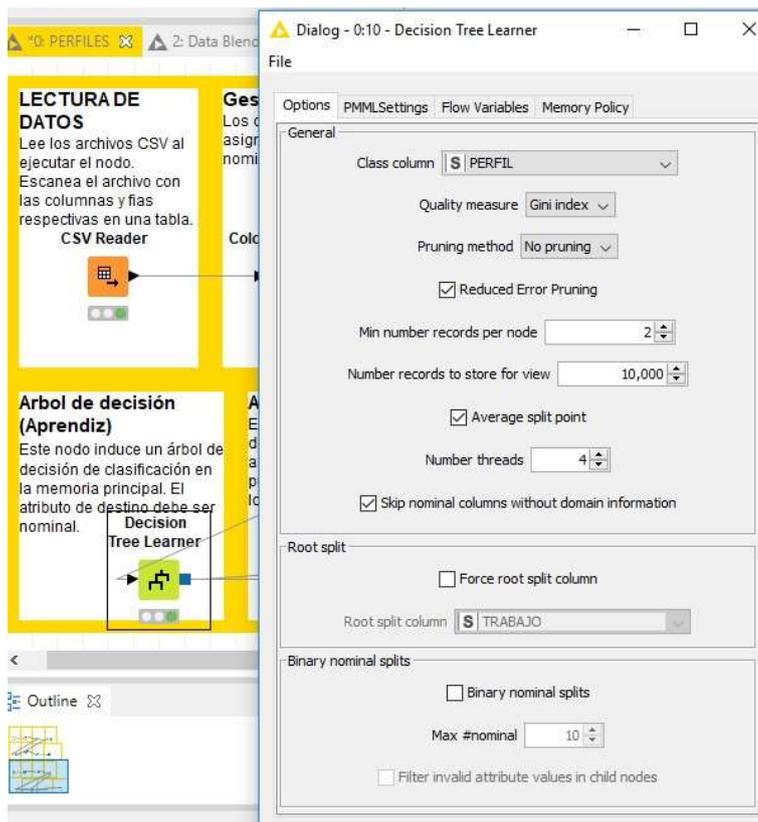


Figura 17: Configuración del aprendizaje de árbol de decisión

La información que resulta del aprendizaje, es decir la pseudo clasificación, se une al conjunto de pruebas para hacer una combinación para el establecimiento de los patrones de perfiles.

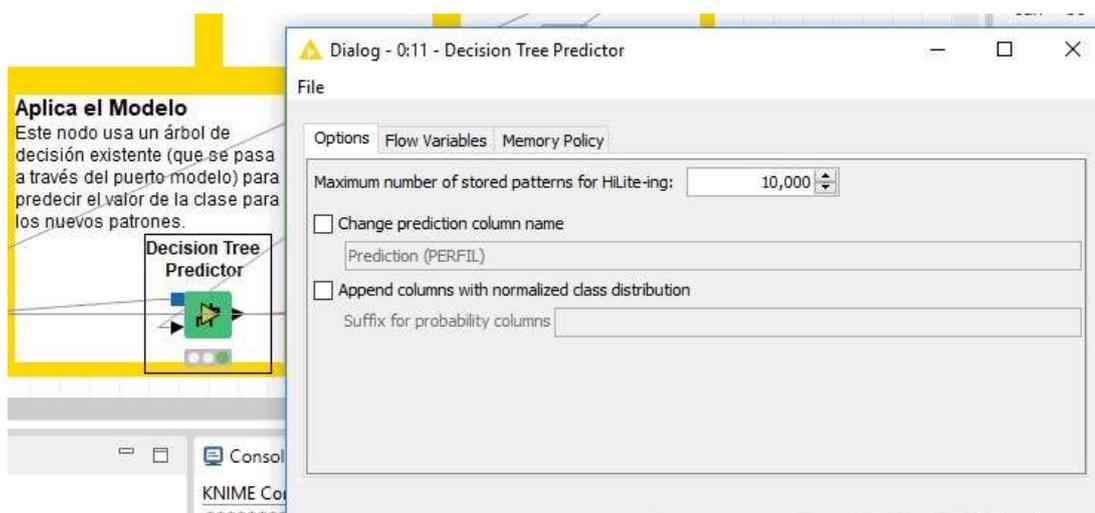


Figura 18: Configuración del árbol de decisión para aplicar el modelo

En la figura 19 se muestra la configuración del conjunto de reglas probables para cada grupo, desarrollador e infraestructura.

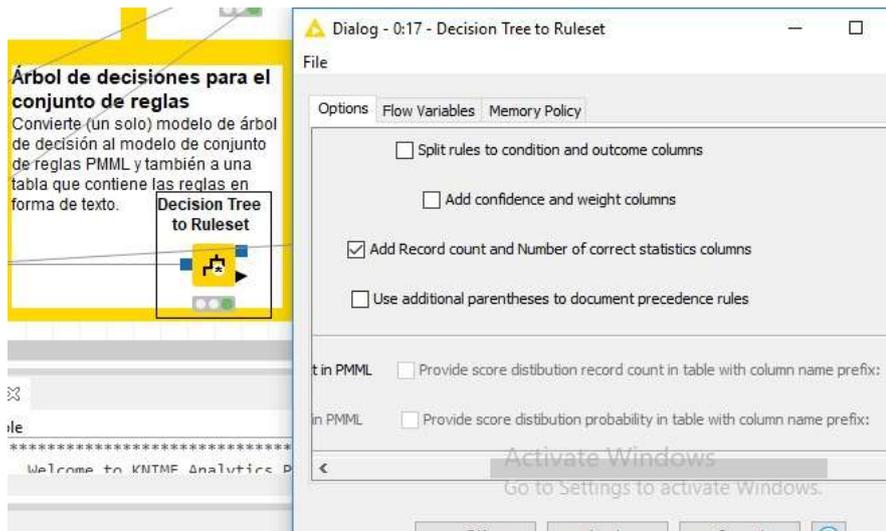


Figura 19: Configuración del árbol de decisión. Reglas

### ***Interpretación y evaluación de resultados***

Cada clustering tiene asignado N cantidad de registros. La información se agrupó de acuerdo a la afinidad que tiene estudiante y se llegó a la conclusión que existen dos grandes grupos. Según la dispersión de los datos se verifica que el cluster 0 está todo un cuadrante, mientras que para cluster 1 está en otra sección del cuadrante.

En la Figura 20 se muestra la asignación de los registros de clustering, lo que permite comprobar la existencia de los dos clusters que se crearon.

Row ID	'Softwa...	'Softwa...	'Softwa...	'PRACT...	'PRACT...	'TRABA...	'TRABA...	'TRABA...	'ASISTE...	'Cluster
1	0.52	1.04	1.75	1	0	1	0	0	78	cluster_1
2	0.51	0.62	1.44	1	0	0	1	0	74	cluster_1
3	0.4	0.94	1.55	1	0	0	0	1	93	cluster_0
4	0.52	1.13	1.81	1	0	0	0	1	85	cluster_0
5	0.5	0.98	1.6	1	0	0	0	1	88	cluster_0
7	0.38	1.21	1.65	1	0	0	0	1	84	cluster_0
8	0.25	1.12	1.51	1	0	1	0	0	84.48	cluster_0
9	0.25	1.12	1.69	1	0	0	0	1	88	cluster_0
10	0.58	1.28	1.88	1	0	0	0	1	93	cluster_0
11	0.43	0.94	1.66	0	1	0	1	0	84.27	cluster_0
12	0.5	1.13	1.69	1	0	0	0	1	93	cluster_0
13	0.58	1.17	1.44	0	1	0	0	1	78	cluster_1
14	0.53	0.96	1.71	0	1	0	1	0	79	cluster_1
15	0.5	0.82	1.26	1	0	0	0	1	68	cluster_1
16	0.52	1	1.61	1	0	0	1	0	83	cluster_1
17	0.58	0.88	1.65	0	0	0	1	0	91	cluster_0
18	0.41	1.05	1.6	0	1	0	1	0	87	cluster_0
19	0.53	1.17	1.72	0	1	0	0	1	86	cluster_0
20	0.52	0.93	1.68	1	0	0	1	0	87	cluster_0
21	0.52	1.06	1.65	1	0	0	0	1	80	cluster_1
22	0.55	1.07	1.54	1	0	0	0	1	84	cluster_0
23	0.31	1.19	1.66	1	0	0	1	0	80.3	cluster_1
24	0.52	0.84	1.49	1	0	0	1	0	87.17	cluster_0
25	0.41	0.95	1.64	1	0	0	1	0	70.94	cluster_1

Figura 20: Asignación de registros al clustering

Por otro lado, se comprueba la generación del conjunto de reglas para desarrollo e infraestructura.

Row ID	Rule	Record ...	Number
Row1	\$TRABAJO\$ = "ADMINISTRADOR" AND TRUE => "I"	3	2
Row2	\$TRABAJO\$ = "REDES" AND TRUE => "I"	1	1
Row3	\$P\$ = ">1.53" AND \$TRABAJO\$ = "DESARROLLA..."	14	14
Row4	\$H\$ = "<=2.521" AND \$P\$ = "<=1.53" AND \$TRA...	3	2
Row5	\$H\$ = ">2.521" AND \$P\$ = "<=1.53" AND \$TRAB...	3	3
Row6	\$TRABAJO\$ = "ANALISTA" AND TRUE => "D"	1	1
Row7	\$TRABAJO\$ = "IT" AND TRUE => "I"	6	6
Row8	\$TRABAJO\$ = "N" AND TRUE => "I"	2	1
Row9	\$TRABAJO\$ = "SOPORTE" AND TRUE => "I"	4	3
Row10	\$TRABAJO\$ = "INFRAESTRUCTURA" AND TRUE =...	3	3
Row11	\$TRABAJO\$ = "MESA DE AYUDA" AND TRUE => "D"	1	1
Row12	\$TRABAJO\$ = "ATENCION AL CLIENTE" AND TRUE...	1	1

Figura 21: Conjunto de reglas

En cuanto al conjunto de reglas se pudo comprobar que se generaron algunas, tanto para el perfil de desarrollador como para infraestructura, tal como se visualiza en las Figuras 22 y 23.

**Desarrollador:**

P = ">1.53" AND TRABAJO = "DESARROLLADOR" => "D"

H = "<=2.521" AND P = "<=1.53" AND TRABAJO = "DESARROLLADOR" => "D"

TRABAJO = "ANALISTA" AND TRABAJO = "MESA DE AYUDA" AND TRUE => "D"

Figura 22: Reglas para desarrollador

**Infraestructura**

H = ">2.521" AND P = "<=1.53" AND TRABAJO = "DESARROLLADOR" => "I"

TRABAJO = "ADMINISTRADOR" AND TRABAJO = "REDES" AND TRABAJO = "IT" AND TRABAJO = "N" AND TRABAJO = "SOPORTE" AND TRABAJO = "INFRAESTRUCTURA" AND TRABAJO = "ATENCION AL CLIENTE" AND TRUE => "I"

Figura 23: Reglas para infraestructura

Como se pudo apreciar en el desarrollo de este capítulo, el proceso de la aplicación de las técnicas de Minería de Datos permitió el análisis de la información de los estudiantes para determinar sus perfiles. Así, se generaron las reglas que sirvieron para establecer los patrones que permitieron identificar los estudiantes que se orientaron hacia desarrollo y los que optaron por infraestructura.

### ***Descripción del aplicativo***

El aplicativo es un ejecutable que se encuentra alojado en el ordenador del Coordinador de la Unidad de Titulación y es una pantalla con algunas opciones.

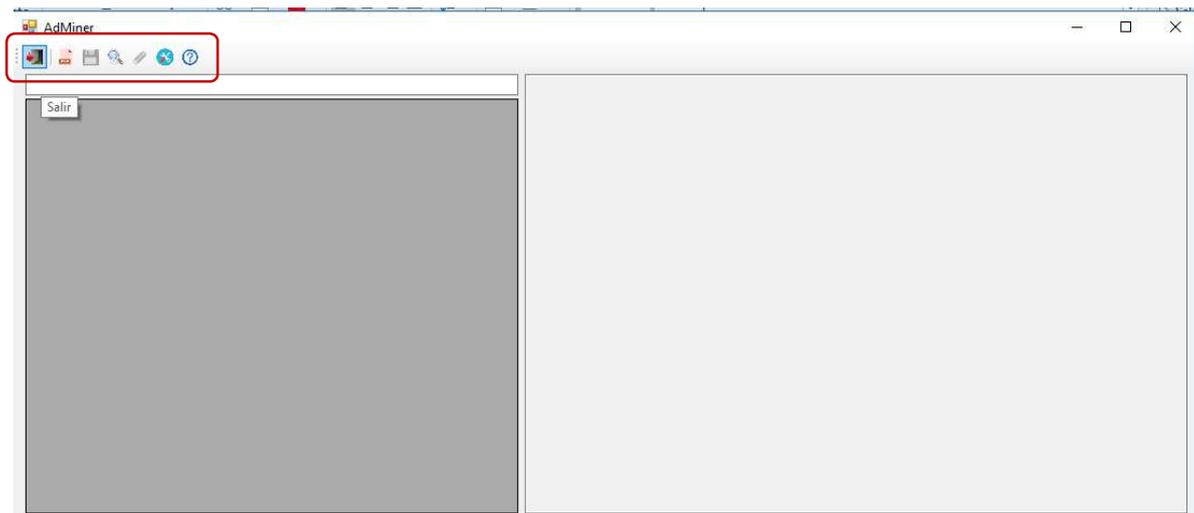


Figura 24: Opciones del aplicativo

En la opción Cargar archivo se debe subir la información de las materias y el número de veces que el estudiante se matricula en la materia. El archivo deberá tener formato CSV, con el orden definido previamente y se cargará de acuerdo a la ruta en la que se encuentre ubicado.

(Nombres, FUNDAMENTOS DE PROGRAMACIÓN, Matricula, LOGICA COMPUTACIONAL, Matricula, PROGRAMACION ORIENTADA A OBJETOS, Matricula, BASE DE DATOS, Matricula, PROGRAMACION EN CAPAS, Matricula, ADMINISTRACION DE BASE DE DATOS, Matricula, ANALISIS Y DISEÑO DE SISTEMAS, Matricula, PROGRAMACION DISTRIBUIDA, Matricula, ANALISIS Y DISEÑO DE SISTEMAS ORIENTADO A O, Matricula, FUNDAMENTO DE REDES DE COMPUTADORAS, Matricula, INGENIERIA DE SOFTWARE, Matricula, PROGRAMACION MOVIL, Matricula, ADMINISTRACION Y ORGANIZACION DE CENTROS, Matricula, ENRUTAMIENTO DE REDES, Matricula, TELECOMUNICACIONES, Matricula, PRACTICA PREPROFESIONAL, ASISTENCIA, TRABAJO).

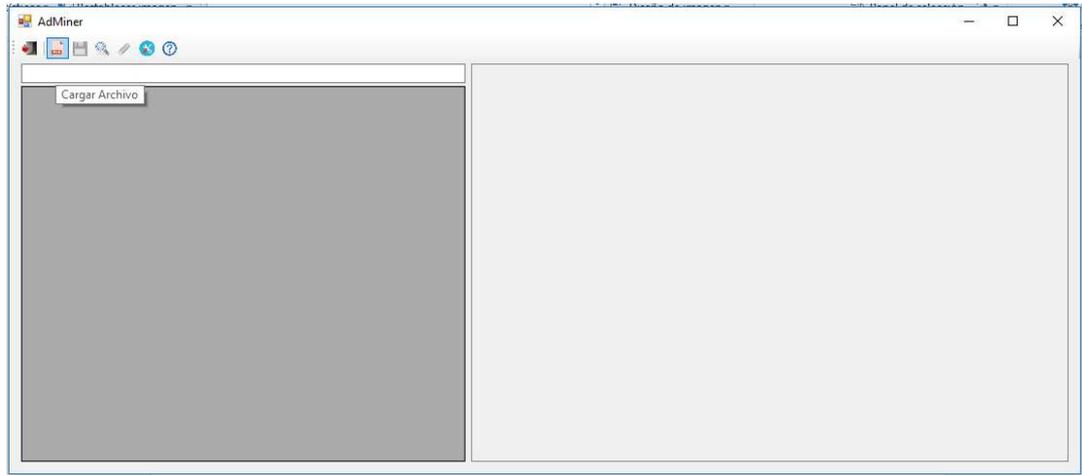


Figura 25: Opción Cargar archivo

Automáticamente al abrir el archivo se realizan los cálculos de cada estudiante para determinar su perfil.

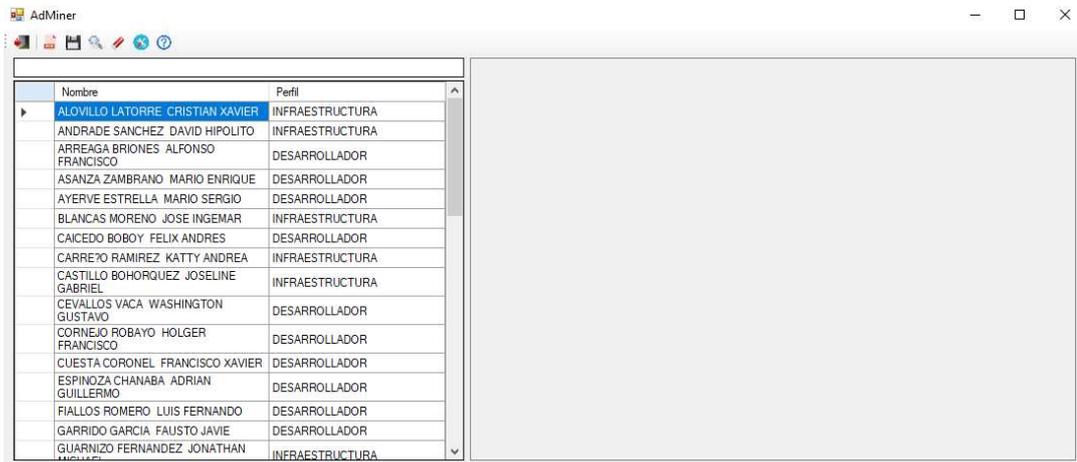


Figura 26: Cálculo del perfil

En el caso que se quiera ver el detalle de materias de cada estudiante se da doble clic en el mismo.

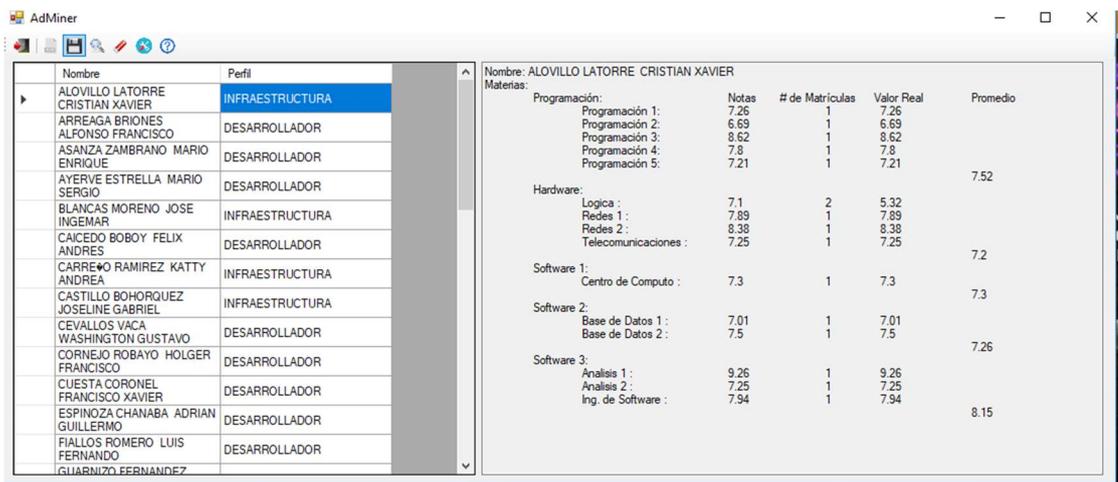
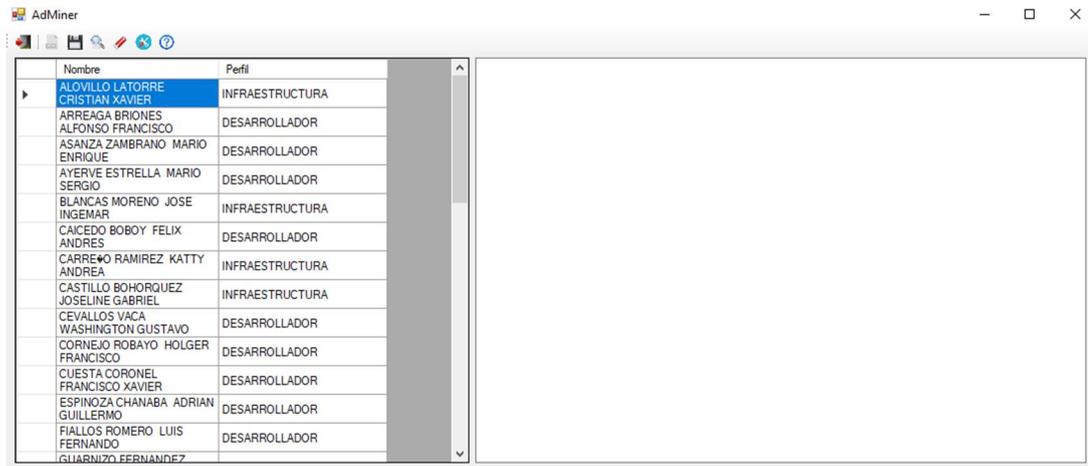


Figura 27: Detalle de materias

Para proceder a la búsqueda se da clic en el botón Buscar y por defecto se busca a todos los estudiantes y en caso de requerir un nombre en particular se escribe el texto y se realiza la búsqueda.



Nombre	Perfil
ALOVILLO LATORRE CRISTIAN XAVIER	INFRAESTRUCTURA
ARREAGA BRIONES ALFONSO FRANCISCO	DESARROLLADOR
ASANZA ZAMBRANO MARIO ENRIQUE	DESARROLLADOR
AYERVE ESTRELLA MARIO SERGIO	DESARROLLADOR
BLANCAS MORENO JOSE INGEMAR	INFRAESTRUCTURA
CAICEDO BOBOY FELIX ANDRES	DESARROLLADOR
CARREÑO RAMIREZ KATTY ANDREA	INFRAESTRUCTURA
CASTILLO BOHORQUEZ JOSELINE GABRIEL	INFRAESTRUCTURA
CEVALLOS VACA WASHINGTON GUSTAVO	DESARROLLADOR
CORNEJO ROBAYO HOLGER FRANCISCO	DESARROLLADOR
CUESTA CORONEL FRANCISCO XAVIER	DESARROLLADOR
ESPINOZA CHANABA ADRIAN GUILLERMO	DESARROLLADOR
FIALLOS ROMERO LUIS FERNANDO	DESARROLLADOR
GUARNIZO FERNANDEZ	

Figura 28: Opción Buscar

En caso de haber realizado algún cambio en los parámetros se lo puede realizar en Configuración, los mismos que por defecto son los del estudio. Si se desea cambiar, se aparecerá una pantalla y se deberán guardar los cambios.

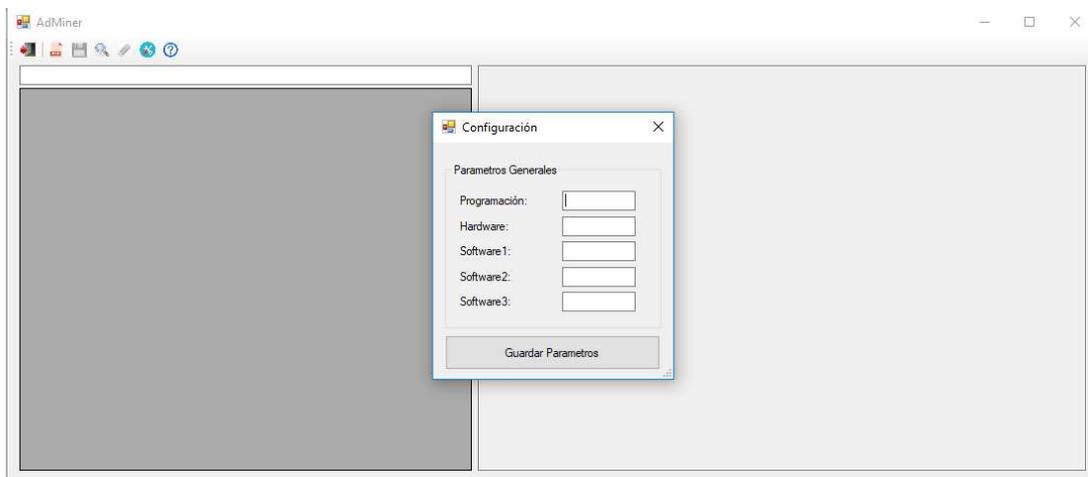


Figura 29: Opción Configuración

## CONCLUSIONES

Al término del proyecto, se determinó que la aplicación de la Minería de Datos en la determinación de los perfiles de estudiantes en la Unidad de Titulación permite considerar la orientación académica del estudiante para la asignación de los temas para Trabajo de Titulación.

La información que se consiguió para establecer las variables que sirvieron para determinar los perfiles de estudiantes, permitieron descubrir hacia dónde se dirigen las habilidades de los mismos, esto es, hacia desarrollo o infraestructura. De este modo el proceso de asignación de temas se realizará de forma rápida y óptima, de manera que se puedan aprovechar en su totalidad dichas habilidades.

Sobre el perfil del estudiante, éste se relaciona con el perfil profesional del graduado, en función de lo que ha conseguido durante los años de estudio de su carrera en las distintas ramas de investigación. Por lo tanto, este proyecto ha permitido conocer la orientación de conocimiento y habilidades de los graduados sobre su carrera, demostrados en las áreas de trabajo que en la actualidad desempeñan, especificando que el perfil está centrado hacia la programación e infraestructura.

La revisión bibliográfica sobre el tema de los perfiles de estudiantes, permitió establecer cuál es la información que se constituyó en las variables a analizar mediante las técnicas de Minería de Datos, la misma que refirió a los nombres del alumno, las materias, las notas y las pasantías realizadas.

Con las variables definidas para el modelo de entrenamiento se diseñó la vista minable, la misma que contiene las áreas de estudio de la carrera, las mismas que son programación, hardware, software y las pasantías de laboratorio y pre profesionales. Los cálculos del promedio de cada área corresponden a la cantidad de materias que tiene cada una.

Se generó el modelo de evaluación de perfil de estudiantes, aplicando las técnicas de Clustering y Árboles de decisión, las mismas que permitieron evaluar la información recolectada y establecer el perfil, de acuerdo a las áreas de estudio de la carrera, las prácticas de laboratorio y pre profesionales, y el trabajo actual.

## RECOMENDACIONES

El modelo de entrenamiento de generación del perfil del estudiante podría requerir de algunas consideraciones que serían de utilidad en dicho proceso, de modo que sea más específico en el momento de asignar los temas en la Unidad de Titulación.

Sería conveniente mayor recopilación de información de los estudiantes para para especificar más a fondo los perfiles. Por ejemplo, en el área de programación, especificar qué tipo de programación domina el alumno.

Además, la aplicación de instrumentos de seguimiento a graduados que permitan recoger datos socioeconómicos, de estudio, de trabajo, familiares y otra información de interés.

Al aplicar los instrumentos, se generarían nuevas variables, por lo que el modelo de entrenamiento debería actualizarse, ya que las nuevas variables generarían otras interpretaciones que aportarían más datos sobre el estudiante.

## REFERENCIAS BIBLIOGRÁFICAS

- Abet, J., Carrizo, B., Corso, C., & González, G. (2012). Estudio de fallos en el mantenimiento relacionando base de datos y fiabilidad. Fase I, 5. Recuperado de [http://sedici.unlp.edu.ar/bitstream/handle/10915/18604/Documento\\_completo.pdf?sequence=1&isAllowed=y](http://sedici.unlp.edu.ar/bitstream/handle/10915/18604/Documento_completo.pdf?sequence=1&isAllowed=y)
- Alcover, R., Benlloch, J., Blesa, P., Calduch, M. A., Celma, M., Ferri, C., ... Zúñica, L. R. (2007). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos, 8. Recuperado de <http://bioinfo.uib.es/~joemiro/aenui/procJenui/Jen2007/alanal.pdf>
- Asamblea Nacional Constituyente. Constitución del Ecuador, § Quinta (2008). Recuperado de <https://www.acnur.org/fileadmin/Documentos/BDL/2008/6716.pdf>
- Bazán, & Rojas. (2013). *PERFIL DEL ESTUDIANTE - UNIVERSITARIO*. Empleo. Recuperado de <https://es.slideshare.net/sirecita/perfil-del-estudiante-universitario>
- Bernal, C. A. (2010). *Metodología de la investigación, administración, economía humanidades y ciencias sociales* (Tercera). Bogotá: Pearson Educación de Colombia Ltda.
- Ceballos Sierra, F. (2011). *Microsoft C#. Curso de Programación*. (Segunda). Madrid: Grupo Editorial RA-MA. Recuperado de <https://books.google.es/books?hl=es&lr=&id=EK8-DwAAQBAJ&oi=fnd&pg=PT6&dq=qu%C3%A9+es+c%23&ots=aiOFOrwtHq&sig=nR8v9B70PLxmD9xHZW5cr-KKLC0#v=onepage&q=qu%C3%A9%20es%20c%23&f=false>
- CES. Reglamento de Régimen Académico, Pub. L. No. RPC-SE-13-No.051-Z013, CES-SG-2017-R-019 59 (2017).
- CES. Reglamento de Régimen Académico, Pub. L. No. 854 (2017). Recuperado de <http://www.ces.gob.ec/lotaip/2018/Enero/Anexos%20Procu/An-lit-a2-Reglamento%20de%20R%C3%A9gimen%20Acad%C3%A9mico.pdf>

- Chambi, R. (2016). IDE's para programadores para crear cualquier tipo de programas. Recuperado 6 de diciembre de 2018, de <http://www.gitmedio.com/gitmedio/ides-para-programadores/>
- CIT. (2011). Examen de Titulación o Trabajo de Tesis en las Ingenierías. *Formación universitaria*, 4(4), 1-1. <https://doi.org/10.4067/S0718-50062011000400001>
- Cobos, C. (2014). *Cuadro comparativo de herramientas de programacion eclipse, java*. Educación. Recuperado de <https://es.slideshare.net/CCCRiis/cuadro-comparativo-javi>
- Cornejo, J. (2012). *Cuadros comparativos*. Recuperado de <https://es.slideshare.net/jesuscornejosaravia/cuadros-comparativos-13337946>
- Dávila, F., & Sánchez, Y. (2012). Técnicas de minería de datos aplicadas al diagnóstico de entidades clínicas. *Revista Cubana de Informática Médica*, 4(2), 174-183. Recuperado de [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S1684-18592012000200007&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S1684-18592012000200007&lng=es&nrm=iso&tlng=es)
- Dimes, T. (2016). *PHP*. Babelcube Inc. Recuperado de [https://books.google.com.ec/books?id=m4AKDgAAQBAJ&printsec=frontcover&dq=php&hl=es-419&sa=X&ved=0ahUKEwjXjdiS7\\_DeAhWlzVkKHeJqCJAQ6AEIOTAD#v=onepage&q=php&f=false](https://books.google.com.ec/books?id=m4AKDgAAQBAJ&printsec=frontcover&dq=php&hl=es-419&sa=X&ved=0ahUKEwjXjdiS7_DeAhWlzVkKHeJqCJAQ6AEIOTAD#v=onepage&q=php&f=false)
- Duncan, O., & Cai, S. (2018). Conceptos de minería de datos [Tecnología]. Recuperado 1 de diciembre de 2018, de <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts>
- EcuRed. (s. f.). MySQL. Recuperado 7 de diciembre de 2018, de <https://www.ecured.cu/MySQL>
- Escobar, H., Alcivar, M., & Puris, A. (2016). Aplicaciones de Minería de Datos en Marketing. *Revista Publicando*, 3(8), 503-512. Recuperado de <https://www.rmlconsultores.com/revista/index.php/crv/article/view/169>
- Fingermann, H. (2011). Perfil del alumno. Recuperado 2 de diciembre de 2018, de <https://educacion.laguia2000.com/general/perfil-del-alumno>
- Hasperué, W. (2013). *Extracción de Conocimiento en Grandes Bases de Datos Utilizando Estrategias Adaptativas* (Primera). Buenos Aires: Editorial de la Universidad Nacional de La Plata (EduLP). Recuperado de

- [http://sedici.unlp.edu.ar/bitstream/handle/10915/35555/Documento\\_completo.pdf?sequence=1](http://sedici.unlp.edu.ar/bitstream/handle/10915/35555/Documento_completo.pdf?sequence=1)
- Hernández, J. (s. f.). Minería de Datos. El proceso de KDD. Recuperado de <http://polipapers.upv.es/index.php/IA/article/view/3293>
- Hernández, R., Fernández, C., & Baptista, P. (2014). *Metodología de la investigación* (Sexta). México: McGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.
- Kumar, A. (s. f.). Introducción Características Beneficios, 59. Recuperado de <http://www.gedlc.ulpgc.es/docencia/abd/Recursos/MySQL-Intro-features-benefits-SPANISH.pdf>
- Márquez, C., Romero, C., & Ventura, S. (2012). Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos, 7, 9. Recuperado de <http://rita.det.uvigo.es/201208/uploads/IEEE-RITA.2012.V7.N3.A1.pdf>
- Martín Pérez, A. A. (2016). Modelo del proceso de formación estadística investigativa del docente universitario en el posgrado. *Gaceta Médica Espirituana*, 18(1), 64-75.
- Martínez Ruiz, X. (2013). Juventud, desempleo y utopía: lecciones desde el pensamiento creativo y la ciudadanía. *Innovación educativa (México, DF)*, 13(61), 11-15. Recuperado de [http://www.scielo.org.mx/scielo.php?script=sci\\_abstract&pid=S1665-26732013000100002&lng=es&nrm=iso&tlng=es](http://www.scielo.org.mx/scielo.php?script=sci_abstract&pid=S1665-26732013000100002&lng=es&nrm=iso&tlng=es)
- Mauleon, O. (2013). *Tabla comparativa de herramientas case*. Recuperado de <https://es.slideshare.net/oswaldoyuneri/tabla-comparativa-de-herramientas-case-oswaldo-mauleon>
- Microsoft. (s. f.). SQL Server: características | Microsoft. Recuperado 7 de diciembre de 2018, de <https://www.microsoft.com/es-es/sql-server/sql-server-2017-features>
- Miranda, M. (2012). *Cuadro comparativo sgbd*. Recuperado de <https://es.slideshare.net/ManuelMirandaBuenabad/cuadro-comparativo-sgbd>
- Morales-Arencibia, A., Febles-Rodríguez, J., & Frómeta-Moreno, Y. (2017). Aplicación de reconocimiento de patrones para identificar la semejanza entre medicamentos teniendo en cuenta las reacciones adversas. *DYNA*, 84(201), 186-194. <https://doi.org/10.15446/dyna.v84n201.57060>

- 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>
- 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>
- Perry, S. (2012). Conceptos básicos del lenguaje Java. Recuperado 6 de diciembre de 2018, de <http://www.ibm.com/developerworks/ssa/java/tutorials/j-introtojava1/index.html>
- Planas, M., Rodríguez, T., & Lecha, M. (2004). La importancia de los datos. *Nutrición Hospitalaria*, 19(1), 11-13.
- Ramiro Vásquez, E., Caballero Núñez, A., & Herrera Villafranca, M. (2017). Transformación de variables binomiales para su análisis según un diseño de bloques al azar. *Cultivos Tropicales*, 38(1), 108-114.
- Ray, M., Milener, G., Roth, J., Steen, H., & Kumar, S. (2017). Editions and supported features of SQL Server 2017. Recuperado 7 de diciembre de 2018, de <https://docs.microsoft.com/en-us/sql/sql-server/editions-and-components-of-sql-server-2017>
- 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>
- Romero-Ortega, A., & Sanz-Cabrera, T. (2017). ¿Tesis o examen de grado? Un dilema para la formación investigativa. 2, 12, 238-247. <https://doi.org/10.17163/alt.v12n2.2017.09>
- Ruiz, Á. (2018). *Minería de Datos en Redes Sociales para PYMES* (Terer Nivel). Universidad de Jaén, España. Recuperado de <http://tauja.ujaen.es/bitstream/10953.1/7836/1/TFG%20-%20RUIZ%20JIMENEZ%2c%20ALVARO.pdf>
- San Martín H., F., & García P, M. (2006). La Tesis y su problemática en la Facultad de Medicina Veterinaria de la UNMSM. *Revista de Investigaciones Veterinarias del Perú*, 17(1), 81-88. Recuperado de

[http://www.scielo.org.pe/scielo.php?script=sci\\_abstract&pid=S1609-91172006000100015&lng=es&nrm=iso&tlng=es](http://www.scielo.org.pe/scielo.php?script=sci_abstract&pid=S1609-91172006000100015&lng=es&nrm=iso&tlng=es)

- Sánchez-Castañeda, A. (2014). Los jóvenes frente al empleo y el desempleo: la necesaria construcción de soluciones multidimensionales y multifactoriales. *Revista latinoamericana de derecho social*, (19), 133-162.
- Solar, G. (2010). *Aplicación de Minería de Datos Espacial en un estudio de cambio de uso del suelo* (Maestría). Universidad de Concepción, Chile. Recuperado de [http://repositorio.udec.cl/bitstream/handle/11594/788/tesis\\_Aplicacion\\_de\\_Mineria\\_de\\_datos\\_espacial.Image.Marked.pdf?sequence=1&isAllowed=y](http://repositorio.udec.cl/bitstream/handle/11594/788/tesis_Aplicacion_de_Mineria_de_datos_espacial.Image.Marked.pdf?sequence=1&isAllowed=y)
- Sommerville, I. (2011). *Ingeniería de software*. Madrid: Pearson Educación de México, S.A. de C.V.
- Suriá Martínez, R., Rosser Limiñana, A., & Villegas Castrillo, E. (2013). Análisis comparativo de estrategias de búsqueda de empleo de los estudiantes universitarios 2011/12 y 2012/13. En *XI Jornadas de Redes de Investigación en Docencia Universitaria: Retos de futuro en la enseñanza superior : Docencia e investigación para alcanzar la excelencia académica* (pp. 44-52). 2013: Instituto de Ciencias de la Educación. Recuperado de <https://dialnet.unirioja.es/servlet/libro?codigo=560420>
- Tecnologías Información. (2018). La minería de datos: Definición, usos y funciones. Recuperado 30 de noviembre de 2018, de <https://www.tecnologias-informacion.com/mineria-de-datos.html>
- 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>
- Troche, A. (2014). Aplicación de la Minería de datos sobre bases de datos transaccionales. *Fides et Ratio - Revista de Difusión cultural y científica de la Universidad La Salle en Bolivia*, 7(7), 58-66. Recuperado de [http://www.scielo.org.bo/scielo.php?script=sci\\_abstract&pid=S2071-081X2014000100005&lng=es&nrm=iso&tlng=es](http://www.scielo.org.bo/scielo.php?script=sci_abstract&pid=S2071-081X2014000100005&lng=es&nrm=iso&tlng=es)

- 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](http://www.ucsg.edu.ec/wp-content/uploads/transparencia/Reglamento_academico_pedagogicoUCSG.pdf)
- 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)
- Universidad de Monterrey. (2019). Perfil de ingreso | Ingeniero en Tecnologías Computacionales |. Recuperado 29 de enero de 2019, de <http://www.udem.edu.mx/Esp/Carreras/Ingenieria-y-Tecnologias/Ingeniero-en-Tecnologias-Computacionales/Pages/perfil-de-ingreso.aspx>
- Universidad del Norte. (s. f.). Perfil, Qué hace un Ingeniero de Sistemas y Computación -. Recuperado 29 de enero de 2019, de <https://www.uninorte.edu.co/web/ingenieria-de-sistemas-y-computacion/perfiles>
- Universidad Internacional de Valencia. (2018). Análisis de datos, entendiendo cada una de sus fases | VIU. Recuperado 2 de diciembre de 2018, de <https://www.universidadviu.com/analisis-datos-entendiendo-una-fases/>
- Universidad Latina de Costa Rica. (s. f.). Bachillerato en Ingenierías de Sistemas Informáticos – Sede San Pedro. Recuperado 29 de enero de 2019, de <https://ulatina.ac.cr/oferta-academica/grados/ingenierias-y-tics/bachillerato-en-ingenierias-sistemas-informaticos-sede-san-pedro/>
- Valenga, F., Perversi, I., Fernández, E., Merlino, H., Rodríguez, D., Britos, P., & García-Martínez. (2007). Aplicación de Minería de Datos para la exploración y detección de patrones delictivos en Argentina, 13. Recuperado de <http://sedici.unlp.edu.ar/bitstream/handle/10915/21783/05+Valenga.pdf?sequence=1>
- 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

Villar-Ledo, L., & Ledo-Ferrer, M. C. (2016). Aplicación de herramientas estadísticas para el análisis de indicadores. *Ingeniería Industrial*, 37(2), 138-150.



## DECLARACIÓN Y AUTORIZACIÓN

Yo, **Cordovilla Cordovilla, Juan Argenis** con C.C: # 0926021791 autor del trabajo de titulación: **Diseño de un Modelo Predictivo, mediante la técnica de Minería de Datos, para identificar el perfil de éxito del estudiante en la Unidad de Titulación en la Carrera de Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG** 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, 07 de marzo de 2019

f.

**Cordovilla Cordovilla, Juan Argenis**

C.C: 0926021791

<b>REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA</b>			
<b>FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN</b>			
<b>TEMA Y SUBTEMA:</b>	Diseño de un Modelo Predictivo, mediante la técnica de Minería de Datos, para identificar el perfil de éxito del estudiante en la Unidad de Titulación en la Carrera de Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG		
<b>AUTOR:</b>	Juan Argenis, Cordovilla Cordovilla		
<b>TUTOR:</b>	Ing. César Adriano, Salazar Tovar		
<b>INSTITUCIÓN:</b>	Universidad Católica de Santiago de Guayaquil		
<b>FACULTAD:</b>	Facultad de Ingeniería		
<b>CARRERA:</b>	Ingeniería en Sistemas Computacionales		
<b>TÍTULO OBTENIDO:</b>	Ingeniero en Sistemas Computacionales		
<b>FECHA DE PUBLICACIÓN:</b>		<b>No. DE PÁGINAS:</b>	72
<b>ÁREAS TEMÁTICAS:</b>	Sistemas de información, desarrollo de sistemas informáticos		
<b>PALABRAS CLAVES/ KEYWORDS:</b>	MODELO PREDICTIVO; MINERÍA DE DATOS; VISTA MINABLE, CLUSTERING; ÁRBOL DE DECISIÓN; PATRONES.		
<b>RESUMEN/ABSTRACT (150-250 palabras):</b>			
<p>La asignación de temas para Trabajo de Titulación a los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la Facultad de Ingeniería de la UCSG es un inconveniente que se presenta en la Unidad de Titulación ya que no se dispone del perfil del estudiante que ha terminado su malla curricular para la elaboración de su trabajo final y esto conlleva a que los temas sugeridos y analizados por la Comisión Académica de la Carrera no se ajusten al conocimiento de las aptitudes del mismo sobre el tema propuesto. Para tal efecto, se propuso el diseño de un modelo predictivo, mediante Minería de Datos, para identificar el perfil del estudiante, de manera que los temas de Trabajo de Titulación estén acordes a sus habilidades. Para realizar la investigación se utilizó el enfoque de investigación cuantitativo, de tipo analítico, descriptivo y exploratorio para estudiar el contexto en donde existe el problema de la identificación del perfil; como técnica de recolección de datos se usó la encuesta, que se realizó a los estudiantes de cohortes A y B 2017 y A 2018, técnicas de Minería de Datos no supervisadas como clustering y árboles de decisión, y la herramienta de Minería de Datos KNIME. Se diseñó la vista minable con las áreas de estudio de la carrera y se generó el modelo predictivo de evaluación para establecer el perfil de acuerdo a las áreas, prácticas y trabajo actual. Al final se plantearon recomendaciones a considerarse como mejoras del modelo predictivo.</p>			
<b>ADJUNTO PDF:</b>	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
<b>CONTACTO CON AUTOR:</b>	<b>Teléfono:</b> 593 93 968 5895	E-mail: <a href="mailto:juanargenis_cc@hotmail.com">juanargenis_cc@hotmail.com</a>	
<b>CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE):</b>	<b>Nombre:</b> Toala Quimí, Edison José		
	<b>Teléfono:</b> +593-4-2202763, ext. 1025		
	<b>E-mail:</b> <a href="mailto:edison.toala@cu.ucsg.edu.ec">edison.toala@cu.ucsg.edu.ec</a>		
<b>SECCIÓN PARA USO DE BIBLIOTECA</b>			
<b>Nº. DE REGISTRO (en base a datos):</b>			
<b>Nº. DE CLASIFICACIÓN:</b>			
<b>DIRECCIÓN URL (tesis en la web):</b>			