



UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

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

TEMA:

Minería de datos aplicada a la detección de patrones para el análisis de rendimiento académico de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la Universidad Católica Santiago de Guayaquil

AUTOR:

Solines Bernardino, Juan Jose

**Trabajo de titulación previo a la obtención del título de
Ingeniero en sistemas computacionales**

TUTOR:

Freire Cobo, Lenin Eduardo

Guayaquil, Ecuador

24 de Septiembre del 2018



UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

CERTIFICACIÓN


Certificamos que el presente trabajo de titulación fue realizado en su totalidad por Juan José Solines Bernardino, como requerimiento para la obtención del título de ingeniero en sistemas computacionales.

TUTOR



Ing. Freire Cobo, Lenin Eduardo, Mgs.

DIRECTORA DE LA CARRERA



Ing. Camacho Coronel, Ana Isabel, Mgs.

Guayaquil, a los 24 días del mes de septiembre del año 2018



UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

DECLARACIÓN DE RESPONSABILIDAD

Yo, Juan Jose Solines Bernardino

DECLARO QUE:

El Trabajo de Titulación, Minería de datos aplicada a la detección de patrones para el análisis de rendimiento académico de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la Universidad Católica Santiago de Guayaquil 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 24 días del mes de septiembre del año 2018

AUTOR

Solines Bernardino, Juan Jose



UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERÍA

**CARRERA DE INGENIERÍA EN SISTEMAS
COMPUTACIONALES**

AUTORIZACIÓN

Yo, Juan Jose Solines Bernardino

Autorizo a la Universidad Católica de Santiago de Guayaquil a la publicación en la biblioteca de la institución del Trabajo de Titulación, Minería de datos aplicada a la detección de patrones para el análisis de rendimiento académico de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la Universidad Católica Santiago de Guayaquil cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 24 días del mes de septiembre del año 2018

EL AUTOR:

Solines Bernardino, Juan Jose



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

CALIFICACION 20

Ing. Lenin Eduardo Freire Cobo

PROFESOR TUTOR




UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL


FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES


TRIBUNAL DE SUSTENTACIÓN



Ing. Lenin Eduardo Freire Cobo
PROFESOR TUTOR



Ing. Galo Enrique Cornejo Gomez
OPONENTE



Ing. Edison Jose Toala Quimi
COORDINADOR DEL AREA



Ing. Jose Miguel Erazo Ayon
DELEGADO DE LA CARRERA



UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL

FACULTAD DE INGENIERÍA

CARRERA DE INGENIERÍA EN SISTEMAS COMPUTACIONALES

INFORME URKUND

 URKUND

Urkund Analysis Result

Analysed Document:	Tesis Proyecto Solines.docx (D41022307)
Submitted:	8/29/2018 12:34:00 AM
Submitted By:	freirelenin@gmail.com
Significance:	2 %

AGRADECIMIENTO

Mi primer agradecimiento es a Dios, segundo lugar a mis padres Jose Luis y Monica por el apoyo siempre y tercero a mis abuelas Dora y Estrellita, en especial a mi abuela Estrellita que falleció el año pasado a quien quería muchísimo por apoyarme bastante.

DEDICATORIA

Esta tesis se la dedico a todo el que me apoyo y no me abandono en los momentos difíciles, y por último a todos los profesores, directora carrera y decana de la facultad de ingeniería en especial a los de sistemas.

INDICE GENERAL

RESUMEN	xi
ABSTRACT	xii
INTRODUCCIÓN	2
CAPITULO I Problemática de Investigación	3
1.1. Planteamiento del problema	3
1.2. Hipótesis de Investigación	4
1.3. Preguntas de Investigación.....	4
1.4. Variables de Investigación	4
1.5. Objetivos.....	4
1.5.1. General	4
1.5.2. Específicos	4
1.6. Justificación	5
1.7. Alcance	5
CAPITULO II Marco Teórico	6
2.1. Marco Conceptual.....	6
2.1.1. Metodología KDD (Knowledge Discovery in Databases)	6
2.1.2. Proceso de la Metodología KDD	6
2.2. Arboles de Decisión	8
2.3. Detección de Patrones	9
2.4. Rendimiento Académico	9
2.5. Bases de Datos	9
2.6. Tablas Dinámicas	10
2.7. Minería de Datos	10
2.7.1. Aplicaciones de la Minería de Datos	11
2.8. Universidad Católica Santiago de Guayaquil.....	12
2.9. Marco Legal	13
CAPITULO III Metodología de la Investigación y Análisis de Resultados... ..	14
3.1. Tipo y método de Investigación	14
3.2. Técnicas e Instrumentos para recolección de información	15
3.3. Herramienta para procesamiento de información	16
CAPITULO IV.....	17

4.1. Desarrollo de la Metodología	17
4.2. Análisis de los Resultados de la Investigación	26
4.3. Simulador Académico	31
CONCLUSIONES	35
RECOMENDACIONES.....	37
BIBLIOGRAFÍA.....	38

ÍNDICE DE TABLAS

Tabla 1. Vista Minable de las Asignaciones por Materia	18
Tabla2. <i>Reglas del modelo</i>	19
Tabla 3. Descripción de la variable de clasificación	19
Tabla4. <i>Justificación del Semáforo</i>	20
Tabla5. <i>Promedio de aprobación por Materias de la Malla 1 (A2018)</i>	21
Tabla6. <i>Promedio de aprobación por Materias de la Malla 10 (A2018)</i>	22
Tabla7. <i>Promedio de aprobación por Profesor (A2018)</i>	23
Tabla8. <i>Promedio de aprobación por Estudiante Malla 1 (A2018)</i>	24
Tabla9. <i>Promedio de aprobación por Estudiante Malla 10 (A2018)</i>	25
Tabla10. <i>Pruebas Malla 1 semestre (B2017)</i>	28
Tabla11. <i>Pruebas Malla 10 semestre (B2017)</i>	29
Tabla12. <i>Matriz falsos positivos/negativos Malla 1 y 10 (B2017)</i>	30

RESUMEN

Con el uso de la herramienta de Minería de Datos se analizaron patrones de rendimiento académico en los estudiantes, materias y profesores que permitieron elaborar un modelo predictivo en beneficio para los estudiantes, al momento que se inscriban en el semestre, para indicarle cuál es su probabilidad de éxito, alerta o fracaso basado en su rendimiento y en las materias que vaya a tomar. Para esta investigación se utilizó la metodología KDD que se basa en la extracción de conocimiento sobre una pequeña o gran cantidad de datos, se utilizó la herramienta de minería de datos Excel para recolección, procesamiento de datos, y la elaboración del simulador que aplico el modelo de predicción. En conclusión, el objetivo de este modelo lo que busca es mostrarle mediante un semáforo la probabilidad de éxito al estudiante basado en su rendimiento y las materias a escoger. Se realizaron 59 pruebas de los cuales mostro 16 falsos positivos (27,12%) y 6 falsos negativos (10,17%).

Palabras Claves: Minería de Datos, Detección de Patrones, Rendimiento Académico, Modelo Predictivo.

ABSTRACT

With the use of the Data Mining tool, patterns of academic performance were analyzed in the students, subjects and professors that allowed to elaborate a predictive model in benefit for the students, at the moment that they register in the semester, to indicate what is their probability of success, alert or failure based on their performance and the subjects they will take. For this research we used the KDD methodology that is based on the extraction of knowledge on a small or large amount of data, the Excel data mining tool was used for collection, data processing, and the development of the simulator that the model applied of prediction. In conclusion, the objective of this model is to show him through a traffic light the probability of success to the student based on his performance and the subjects to choose. There were 59 tests of which showed 16 false positives (27.12%) and 6 false negatives (10.17%).

Key Words: Data Mining, Pattern Detection, Academic Performance, Predictive Model.

INTRODUCCIÓN

Desde la creación de la Carrera de Sistemas Computacionales se ha cambiado en varias ocasiones su malla curricular en función de la evolución de las tecnologías de información que se ha venido dando aceleradamente en esta área. Durante todo ese tiempo, los estudiantes no han tenido una herramienta que los ayude en el proceso de registros conocer cuántas materias tomar para alcanzar una probabilidad de éxito en pasar todas las materias que hayan tomado. Tradicionalmente, los estudiantes han tenido que decidir hasta la actualidad que materias y con qué profesor tomar, preguntado a los estudiantes que ya pasaron o se quedaron en esas materias, la información que reciben es: esta materia es pesada, tal profesor es complicado, los proyectos son difíciles, etc.

Al no tener los estudiantes esta información en el proceso de registros, tienen la posibilidad de tomar demasiada carga académica, teniendo como consecuencia la posibilidad de que repitan una o más materias, ocasionando que se retrasen en la malla curricular, llegando hasta la posibilidad de cambiarse de universidad o llegar a repetir a un tercer crédito, significa que al no tener un aviso previo cuando vaya a tomar sus materias, puede ser que tenga un resultado que vaya en la dirección de bajar su desempeño académico tanto de la carrera como de la facultad.

Por esa razón, se desea realizar un semáforo de predicción que le vaya indicando a los estudiantes dependiendo de su rendimiento y a medida que vaya tomando su carga académica la posibilidad de éxito, alerta o posible fracaso en su semestre académico, este resultado va a ir en relación directa con su rendimiento académico en la carrera de Ingeniería en Sistemas Computacionales de la UCSG.

Sería importante, que se incluya esta información en el proceso de registros de la carrera de Ingeniería en Sistemas, para que los estudiantes tengan acceso a ello y puedan decidir con más información, con que carga académica están trabajando en el semestre.

CAPITULO I

Problemática de Investigación

No hay un sistema que avise al estudiante al momento de inscribirse, es decir que le vaya indicando al estudiante dependiendo de su rendimiento y a medida que vaya tomando su carga académica la posibilidad de éxito, alerta o posible fracaso en su semestre académico, para así mejorar el desempeño académico de los estudiantes de la Facultad de Ingeniería de la Carrera de Ingeniería en Ciencias de la Computación.

1.1. Planteamiento del problema

En las instituciones académicas públicas y privadas, siempre están preocupadas por mejorar el nivel académico de sus estudiantes, por eso están siempre invirtiendo en capacitación docente a los profesores, evaluaciones a los estudiantes, laboratorios de prácticas, y premiación a los mejores estudiantes, ya que ellos son los clientes de la institución, de lo contrario puede desencadenar en la pérdida o retiro de estudiantes o simplemente el cambio de carrera o universidad, causando un perjuicio académico y de imagen a la institución.

En la Facultad de Ingeniería en la carrera de ingeniería en sistemas computacionales e ingeniería en ciencias computacionales, no está ajena a la situación descrita, y se tienen estudiantes de bajo rendimiento académico, es decir que hay un gran grupo de estudiantes en segunda o tercera matrícula, no hay un sistema que avise al estudiante al momento de inscribirse cuál va a ser su probabilidad de éxito, alerta, o fracaso basado en su rendimiento académico y las materias que vayan a tomar.

Este modelo predictivo ayudaría a los estudiantes, antes de tomar una materia saber la probabilidad de éxito con esa materia y va a ir aumentando o disminuyendo mientras vaya tomando más materias dependiendo de su rendimiento académico. Esto garantizaría probabilísticamente un mejor rendimiento académico si los estudiantes previamente antes de registrarse conozcan su probabilidad de éxito.

1.2. Hipótesis de Investigación

Si al momento de escoger materias a los estudiantes se les muestra una probabilidad de aprobación (carga académica) pueden mejorar su rendimiento académico.

1.3. Preguntas de Investigación

¿Qué materias usted considera que son difíciles de aprobar?

¿Este Modelo Predictivo ayudaría a elevar el rendimiento académico de los estudiantes de Ingeniería en sistemas computacionales e ciencias de la computación de la UCSG?

1.4. Variables de Investigación

Dependientes: Semáforo de Aprobación.

Independientes: Malla 1, Malla 10, Dictada y Profesor Nuevo.

1.5. Objetivos

1.5.1. General

Analizar el historial académico de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la UCSG mediante el uso de la técnica de minería de datos, que contribuya a elaborar un modelo predictivo sobre rendimiento académico.

1.5.2. Específicos

- Realizar un diagnóstico de la información personal académica de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la UCSG con el fin de analizar el rendimiento académico.
- Recopilar los datos de los estudiantes de los diferentes niveles de la malla actual, de tal forma de agruparlos y establecer una correlación con ellos.

- Aplicar el modelo de árboles de decisión, para generar un modelo predictivo de rendimiento académico, por nivel.
- Elaborar una simulación que aplique el modelo predictivo.

1.6. Justificación

Proporcionar a través de la minería de datos un modelo predictivo para el análisis de rendimiento académico de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la UCSG. Los resultados pueden ser utilizados para analizar los factores que inciden directamente en el rendimiento académico de los estudiantes para así ejecutar las medidas correctivas necesarias.

Este trabajo de titulación pertenece al campo de investigación y desarrollo.

1.7. Alcance

Elaborar un análisis de rendimiento académico con la Información personal académica de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la UCSG que han cursado la malla curricular 10 (2012-B) entre semestre 2013-B hasta el semestre 2017-B y la carrera Ingeniería en Ciencias de la Computación de la presente malla curricular 1 (2017-A) entre semestre 2017-A hasta el semestre 2017-B.

CAPITULO II

Marco Teórico

La minería de datos nos ayudara a recoger información relevante para hacer un análisis de rendimiento académico y analizar los factores que inciden directamente para así poder ejecutar medidas correctivas necesarias.

2.1. Marco Conceptual

2.1.1. Metodología KDD (Knowledge Discovery in Databases)

El KDD es la extracción no trivial de conocimiento, implícito en los datos, previamente desconocido y potencialmente útil. El KDD es un campo multidisciplinario de investigación, donde contribuyen: Aprendizaje de máquina, estadística, tecnología de las bases de datos, sistemas expertos y visualización de los datos. (Matsudo, 2001, p. 2)

2.1.2. Proceso de la Metodología KDD

Según el artículo de Matsudo (2001) sobre árboles de decisión, una técnica de minería de datos desde una perspectiva informática y estadística menciono:

El proceso de KDD es un proceso de soporte de decisión en el cual se buscan patrones de información en los datos. Este proceso consta de seis pasos:

1. Selección de los datos: Una vez que se formularon los requerimientos de información, el paso siguiente es coleccionar y seleccionar los datos que se necesitan. No siempre es una tarea fácil reunir esta información en una base de datos centralizada, ya que esto puede llevar a conversiones de bajo nivel.
2. Limpieza: Una vez que se recolectaron los datos, el siguiente paso es la limpieza. Probablemente no se tiene conciencia de la cantidad de suciedad que existe en los datos. Es bueno invertir tiempo en examinar los datos para tener una idea de las posibilidades de extraer

información, quizá en la práctica, es difícil si se tienen grandes conjuntos de datos.

3. Enriquecimiento: Una vez efectuada la limpieza de datos, se puede enriquecer la base de datos. En algunos países obtener información adicional es un aspecto comercial, por ejemplo, se puede obtener información adicional sobre varios sujetos (como edad, ingresos, cantidad de créditos, etc.).

4. Codificado: Si para algunos individuos de la base de datos, falta algún tipo de información, por ejemplo, posesión de auto o casa, es necesario hacer un análisis de las posibles consecuencias antes de tomar la decisión de borrar dichos registros. Si esta información se distribuye aleatoriamente sobre la base de datos, remover todos los registros con falta de información no afectará el tipo de agrupamiento a encontrar.

5. Análisis Exploratorio – Minería de Datos: combina métodos y herramientas de al menos tres áreas: aprendizaje de máquina, estadística y bases de datos. La relación entre aprendizaje de máquina, estadística y minería de datos es un poco obvia, las tres áreas apuntan a localizar regularidades importantes, patrones o conceptos de datos empíricos. Los métodos de aprendizaje de máquina forman parte de la minería de datos. árboles de decisión o reglas de inducción son unos de los componentes de varios algoritmos de la minería de datos.

6. Información: El paso de información combina dos funciones diferentes:

- Análisis de los resultados de algoritmos de reconocimiento de patrón.
- Aplicaciones de los resultados de los algoritmos de reconocimiento de patrón a nuevos datos.

Se puede dar la información usando herramientas de consulta tradicional para base de datos (SQL). Aunque ahora, aparecieron nuevas técnicas de visualización de datos, desde simples diagramas de puntos mostrando diferentes agrupaciones en dos dimensiones

hasta complejos entornos interactivos conteniendo información del conjunto de datos. (pp. 2-4)

Esta metodología nos ayudara identificando los patrones críticos de rendimiento de los estudiantes a partir del historial académico de los estudiantes de una manera más fácil y ágil.

2.2. Árboles de Decisión

Según la Universidad Tecnológica de la Mixteca (2017) menciono que:

Árboles de decisión es una técnica que permite analizar decisiones secuenciales basadas en el uso de resultados y probabilidades asociadas. Los árboles de decisión se pueden usar para generar sistemas expertos, búsquedas binarias y árboles de juegos.

Las ventajas de un árbol de decisión son:

- Facilita la interpretación de la decisión adoptada.
- Proporciona un alto grado de comprensión del conocimiento utilizado en la toma de decisiones.
- Explica el comportamiento respecto a una determinada tarea de decisión.
- Reduce el número de variables independientes.
- Es una magnífica herramienta para el control de gestión empresarial.

Los árboles de decisión se utilizan en cualquier proceso que implique toma de decisiones como:

- Búsqueda Binaria.
- Sistemas expertos.
- Árboles de juegos. (p. 1)

La técnica de árboles de decisión nos permitirá analizar y tomar decisiones sobre los resultados obtenidos sobre el historial de los estudiantes de una forma más analítica y precisa.

2.3. Detección de Patrones

“El reconocimiento de patrones es un problema a nivel computacional ya que se puede ver afectado por el ruido y procesamiento de imágenes, sus ventajas son efectividad y velocidad de procesamiento” (Miranda & Encalada, 2009, p. 23).

En esta investigación sobre minería de datos se aplica un algoritmo de detección de patrones para identificar patrones de rendimiento de los estudiantes, patrones históricos de materias y profesores.

2.4. Rendimiento Académico

Según el artículo de Chungat (2001), sobre el Análisis multivariante para medir rendimiento académico de los estudiantes de una carrera universitaria: caso ingeniería en estadística informática menciona:

En síntesis, el rendimiento académico es la representación, en una escala, de las capacidades o puesta en práctica de los conocimientos alcanzados por el estudiante al final de un curso.

Es necesario establecer la existencia de toda una serie de factores diferenciales que pueden explicar el rendimiento académico. Entre ellos los factores psicosociales relativos a la percepción que el alumno posee de su ambiente familiar, escolar y social, sin olvidar factores de tipo personal tales como la inteligencia y el autoconcepto. (pp. 16-19)

En conclusión, el rendimiento académico se basa en las cualidades personales y psicosociales del estudiante que pueden afectar el rendimiento al final del curso.

2.5. Bases de Datos

Según el portal de Microsoft Office (2016) mencionaron:

Una base de datos es una herramienta para almacenar y organizar información sobre personas, artículos, productos o cualquier otro dato. Las bases al inicio eran hojas de cálculo y mientras su tamaño aumentaba comenzaban a haber problemas de redundancia y

búsqueda con los datos almacenados dentro, por ese motivo los datos se comenzaron a trasladar hacia una estructura de base datos actual para mejorar administración y manejo de los datos como por ejemplo OracleDB.

Las bases de datos son un elemento necesario para almacenar datos de estudiantes, materias, profesores o cualquier otro dato ya que así se nos facilita hacer análisis de cualquier tipo en determinada situación.

2.6. Tablas Dinámicas

Según el portal de Microsoft office (2016) menciona lo siguiente sobre las tablas dinámicas:

La posibilidad de analizar rápidamente los datos puede ayudarle a tomar mejores decisiones empresariales. Aunque a veces es difícil saber por dónde empezar, especialmente cuando se tiene gran cantidad de datos. Las tablas dinámicas son un excelente modo de resumir, analizar, explorar y presentar los datos que, además, se pueden crear con unos pocos clics. Las tablas dinámicas son muy flexibles y se pueden ajustar rápidamente en función de cómo se tengan que mostrar los resultados.

Esto nos ayudara a sintetizar nuestro trabajo con un gran volumen de datos, ya que al seleccionar los datos correspondientes Excel nos crea automáticamente una tabla dinámica con su análisis de datos correspondiente, es de gran ayuda e imprescindible cuando se quieren crear tablas con datos estadísticos.

2.7. Minería de Datos

Según el artículo de Troche (2014) sobre la Aplicación de la minería de datos sobre bases datos transaccionales, menciono:

La minería de datos es un campo de las ciencias orientada a la informática que intenta encontrar patrones sobre algún dato

específicos ya sean grandes o pequeños. Implementa métodos de estadísticas, inteligencias artificiales o aprendizaje automatizado para obtener información sobre algunos datos específicos y mostrarla en una matriz comprensible. La minería de datos comprende las etapas de: análisis y procesamiento de datos, modelo y métricas de interés, post procesamiento y visualización de los datos. (p. 58)

La minería de datos en si se basa en encontrar patrones sobre cualquier tipo de datos ya sean grandes o pequeños mediante estadísticas para así extraer los datos críticos para así obtener resultados más precisos.

2.7.1. Aplicaciones de la Minería de Datos

Según el artículo de Riquelme, Ruiz y Gilbert (2006) sobre conceptos de minería de datos, mencionaron:

Existen numerosas áreas donde la minería de datos se puede aplicar, prácticamente en todas las actividades humanas que generen datos:

- Comercio y banca: segmentación de clientes, previsión de ventas, análisis de riesgo.
- Medicina y Farmacia: diagnóstico de enfermedades y la efectividad de los tratamientos.
- Seguridad y detección de fraude: reconocimiento facial, identificaciones biométricas, accesos a redes no permitidos, etc.
- Recuperación de información no numérica: minería de texto, minería web, búsqueda e identificación de imagen, video, voz y texto de bases de datos multimedia.
- Astronomía: identificación de nuevas estrellas y galaxias.
- Geología, minería, agricultura y pesca: identificación de áreas de uso para distintos cultivos o de pesca o de explotación minera en bases de datos de imágenes de satélites.

- Ciencias Ambientales: identificación de modelos de funcionamiento de ecosistemas naturales y/o artificiales (plantas depuradoras de aguas residuales) para mejorar su observación, gestión y/o control.
- Ciencias Sociales: Estudio de los flujos de la opinión pública. (p. 14)

La minería de datos abarca varias áreas donde exista actividad humana y su objetivo es extraer datos mediante un análisis profundo para obtener resultados más precisos, como por ejemplo también sobre determinar los índices de tipos de crímenes en cualquier país.

2.8. Universidad Católica Santiago de Guayaquil

La Universidad Católica Santiago de Guayaquil fue fundada el 17 de mayo de 1962 por una junta pro-universidad católica presidida por Mons. Mosquera, Dr. Ortega y P. Flor, fue aprobada por el Acuerdo Ejecutivo #936 del Presidente de la República Dr. Carlos Arosemena. (Universidad Católica Santiago de Guayaquil, 2017a)

Misión

“Generar, promover, difundir y preservar la ciencia, tecnología, arte y cultura, formando personas competentes y profesionales socialmente responsables para el desarrollo sustentable del país, inspirados en la fe cristiana de la Iglesia Católica” (Universidad Católica Santiago de Guayaquil, 2017b).

Visión

“Ser una Universidad católica, emprendedora y líder en Latinoamérica que incida en la construcción de una sociedad nacional e internacional eficiente, justa y sustentable” (Universidad Católica Santiago de Guayaquil, 2017b).

2.9. Marco Legal

El art. 26 de la Constitución de la República del Ecuador menciona que la educación es un derecho fundamental para todas las personas que no puede negarse por algún tipo de discriminación ya que se promueve la igualdad para el buen vivir. (Asamblea Nacional Constituyente de Ecuador, 2008)

El art. 344 de la Constitución de la República del Ecuador menciona que el sistema nacional de educación abarca todos los profesores, autoridades, edificaciones del proceso educativo ya sea en nivel de educación básico, bachillerato o superior. (Asamblea Nacional Constituyente de Ecuador, 2008)

Art. 5 determina que los estudiantes tienen derecho de acceder, permanecer, egresar y titularse en una educación superior de calidad sin discriminación alguna. (Ley Orgánica de Educación Superior, 2010)

Art. 6 determina que profesores o investigadores tienen deben contar con las condiciones necesarias para ejercer su actividad y participar del sistema institucional. (Ley Orgánica de Educación Superior, 2010)

Art. 14 determina que todas las universidades públicas y particulares debidamente evaluadas y acreditadas forman parte de la presente ley. (Ley Orgánica de Educación Superior, 2010)

CAPITULO III

Metodología de la Investigación y Análisis de Resultados

En este capítulo se detalla la metodología utilizada para realizar la investigación, técnicas e instrumentos para recolección, procesamiento y análisis de la información. Esta investigación es analítica, estadística, cuantitativa e inductiva.

3.1. Tipo y método de Investigación

Esta investigación es analítica debido a que se tendrán que analizar e interpretar los datos obtenidos de los estudiantes, es cuantitativa debido a que se analizarán y recopilarán registros numéricos y se sacarán probabilidades porcentuales, y es inductiva debido a que nos mostrará una probabilidad general por estudiante, materia, y docente.

El tipo de investigación analítica nos permitirá analizar los datos obtenidos del SIU (Sistema Integrado Universitario) de la facultad de ingeniería de la carrera de ingeniería en sistemas computacionales e ingenieros de ciencias de la computación para determinar patrones de rendimiento de los estudiantes, esto ayudará a los estudiantes a tener un nivel de predicción de aprobación (éxito/fracaso) del semestre antes de matricularse al nuevo semestre.

“El método cuantitativo permite de manera eficaz, mediante números, obtener la información requerida, al efectuar el procesamiento de la información mediante las encuestas” (Andrade, 2017, p. 28).

Según el artículo de Cervera (1998) sobre el Método Inductivo, menciono:

El método inductivo consiste en conocer las características generales o comunes a una diversidad de realidades, tal y como se obtienen a partir del empleo del método comparativo, para articularlas mediante relaciones de causalidad y formular así proposiciones de validez general o leyes científicas. En la inducción el razonamiento es,

lógicamente, ascendente desde lo particular y concreto a lo general y abstracto. (p. 33)

“El método estadístico visualiza los resultados generados, mediante gráficos estadísticos que dan un mejor sustento al análisis e interpretación de los resultados” (Andrade, 2017, p.28).

3.2. Técnicas e Instrumentos para recolección de información

Las técnicas proponen normas que ayudan a ordenar etapas del proceso de investigación constituyendo el conjunto de mecanismos y recursos dirigidos a recolectar, conservar, clasificar y analizar los datos e información de la investigación para así saber que método se va a aplicar. (Godínez, 2013, citado por García & López 2016, p. 25)

Se solicitaron los registros del historial académico de los estudiantes a la directora de la carrera de ingeniería en ciencias de la computación la cual nos entregó unos PDF con dicha información, para luego poder realizar la extracción y análisis de los datos de los estudiantes posteriormente.

Según el último registro de estudiantes activos y no activos en el sistema de la facultad de ingeniería de la carrera de ingeniería en sistemas computacionales (malla 10) e ingeniería de las ciencias de la computación (malla 1) había un total de 186 estudiantes activos y 194 no activos hasta el semestre 2018-A.

El análisis de rendimiento se realizó a los 69 estudiantes activos y no activos de la malla 1 desde semestre 2017-A hasta 2018-A y 17 estudiantes activos de la malla 10 desde semestre 2013-B hasta el semestre 2018-A. Para tomar decisiones en el análisis de rendimiento académico se utilizó la técnica de observación y la herramienta de minería de datos que se utilizó fue Microsoft Excel.

3.3. Herramienta para procesamiento de información

La herramienta de algoritmos de minería de datos que se utilizó para la extracción de datos fue Microsoft Excel 2016, ya que nos permitió ordenar y dividir en cuadros, y establecer los índices porcentuales mediante fórmulas de una manera más sencilla.

Según el portal de Microsoft Office (2016) la herramienta Microsoft Excel permite:

- Crear hojas de cálculo con facilidad a partir de plantillas o desde cero, y realizar cálculos con fórmulas modernas.
- Los nuevos gráficos te ayudan a presentar tus datos de una forma convincente, con formato, minigráficos y tablas que permiten entender mejor la información. Crea pronósticos con un solo clic para predecir las tendencias.

CAPITULO IV

En este capítulo se detalla cómo se realizó este trabajo de investigación sobre el rendimiento académico de los estudiantes y se muestra el análisis de resultados de las pruebas realizadas.

4.1. Desarrollo de la Metodología

En esta sección explicaremos paso a paso el desarrollo de la metodología utilizada.

En la metodología KDD se aplican 6 pasos para el análisis de datos, que los vamos a detallar individualmente para explicar nuestro trabajo realizado:

Selección de Datos:

Nuestro objetivo era hacer una predicción para que cuando el Estudiantes se va a inscribir en dicho semestre obtener una predicción de cómo le va a ir el próximo semestre evaluando su carga académica y mostrándole su posible desempeño.

Para esto se solicitó el historial de rendimiento académico de los estudiantes que estuvieron desde el semestre B-2013 hasta el semestre B-2017 de la malla 1 y 10, para luego hacer la extracción de datos y obtener resultados más precisos (probabilidades).

Limpieza:

Para la limpieza de datos se descartaron algunas variables como los motivos personales o económicos. Tampoco se tomaron en cuenta el flujo de las materias, los horarios de cada una, pasantías, materias de otras facultades, ni perfil de ingreso en el caso de estudiante nuevo.

Codificación:

En esta sección explicaremos el modelo de árboles de decisión y las reglas que lo condicionaron.

Hay 3 probabilidades generales que utilizamos y son las siguientes:

- **PMateria:** primero se obtuvo la probabilidad de éxito de los n semestre, segundo se sumaron todas las probabilidades y se dividió para la cantidad de semestres dictada la materia, y luego si la materia era dada por dos docentes diferentes se sumaban esas probabilidades generales y se dividió para la cantidad de los profesores que la dicto.
- **PDocente:** se sumó todas las probabilidades generales de éxito de las materias dadas por dicho profesor y se dividió para las n materia dadas.
- **PEstudiante:** se sumó la probabilidad de éxito de cada semestre que curso y se la dividió para los n semestre que curso, si el estudiante es nuevo se le asigna una probabilidad del 50%.

Una vez establecido el modelo, se establecieron las reglas que lo condicionaron.

Tabla 1.
Vista Minable de las Asignaciones por Materia

MALLA 1	MALLA 10	DICTADA	PROFESOR NUEVO
SI	SI	SI	SI
SI	SI	SI	NO
SI	SI	NO	SI
SI	SI	NO	NO
SI	NO	SI	SI
SI	NO	SI	NO
SI	NO	NO	SI
SI	NO	NO	NO
NO	SI	SI	SI
NO	SI	SI	NO

Como se aprecia en la **Tabla 1** se establecen las reglas que condicionan los elementos que permiten asignar las probabilidades generales de éxito o fracaso de la materia.

Tabla2.*Reglas del modelo*

Reglas
Malla 1 (20) + Malla 10 (30)
Malla 1 (15) + Malla 10 (25) + Profesor (10)
Malla 10 (50)
Malla 10 (30) + Profesor (20)
Malla 1 (50)
Malla 1 (30) + Profesor (20)
50%
Profesor (50)
Malla 10 (50)
Malla 10 (30) + Profesor (20)

En la **Tabla 2** se especifican las reglas del modelo de árboles de decisión, lo que está entre paréntesis es el peso que se le ha dado a cada variable.

Tabla 3.*Descripción de la variable de clasificación*

ESTUDIANTE	SEMAFORO	EVALUACION
EXITO	VERDE	70-100%
ALERTA	AMARILLO	50-69,99%
FRACASO	ROJO	0-49,99%

Como se aprecia en la **Tabla 3** se establecen las reglas del semáforo para realizar la evaluación del semestre que se aplicaron en el simulador de carga académica.

Tabla4.*Justificación del Semáforo*

VERDE		AMARILLO		ROJO	
TOTAL	A%	TOTAL	A%	TOTAL	A%
7	5-7 (71,42-100%)	7	4 (57,14%)	7	0-3 (0-42,85%)
6	5-6 (83,33-100%)	6	3-4 (50-66,66%)	6	0-2 (0-33,33%)
5	4-5 (80-100%)	5	3 (60%)	5	0-2 (0-40%)
4	3-4 (75-100%)	4	2 (50%)	4	0-1 (0-25%)
3	3 (100%)	3	2 (66,66%)	3	0-1 (0-33,33%)
2	2 (100%)	2	1 (50%)	2	0 (0%)
1	1 (100%)			1	0 (0%)

En la **Tabla 4** se puede observar la justificación del semáforo, de cuantas materias tienen que aprobar según el color para confirmar la predicción.

Análisis Exploratorio:

Después de explicar el modelo y sus reglas se procedió a hacer el análisis y la extracción de los datos por materia, profesor y estudiantes. Los resultados de la extracción de datos se mostrarán en esta sección en las siguientes tablas: **Tabla 5, Tabla 6, Tabla 7, Tabla 8, Tabla 9.**

Tabla5.

Promedio de aprobación por Materias de la Malla 1 (A2018)

CICLO	LISTA DE MATERIAS	% A
1	FISICA	75,42
1	HUMANISMO INTEGRAL 1	91,12
1	MATEMATICAS	70,36
1	OFIMATICA EN LA EDUC SUP	96,00
1	QUIMICA	75,52
1	RESOLUCION PROBLEMAS ING	83,50
2	CALCULO DIFERENCIAL FUNDAMEN	86,67
2	PROGRAMACION	81,25
2	FISICA APLICADA A LA COMP	100,00
2	HUMANISMO INTEGRAL 2 INTROD A LAS CIENCIAS	100,00
2	COMP	85,00
2	MATEMATICAS DISCRETAS	73,33

Estas probabilidades se obtuvieron sobre el historial de cada materia de la malla 1 entre semestre A-2017 hasta B-2017.

Tabla6.*Promedio de aprobación por Materias de la Malla 10 (A2018)*

CICLO	LISTA DE MATERIAS	% A
1	Fundam De Investig	77,16
1	Matemat Discretas	50,87
1	Calc Diferencial e Integral	45,65
1	Fundamen Program	71,60
1	Introduc Comput	56,83
1	Estructura Datos	67,07
2	Metod De Investig	89,76
2	Probabili Y Estadist	74,37
2	Calc Multivariab	54,85
2	Program Orien Obj	55,99
2	Logica Computac	39,70
2	Sistem De Informac	81,64
2	Contab Basica	75,66
2	Humanismo Integral 1	91,51
3	Ecuac Diferenciales	60,17
3	Program En Capas	59,61
3	Base Datos	72,05
3	Contab Costos	86,57
3	Humanismo Integral 2	92,91
4	Algebra Lineal	49,69
4	Program Distribuid	76,24
4	Org Y Arquit Comput	83,94
4	Admin Base Datos	79,09
4	Analisis Y Disen Sist	83,81
4	Finanzas	71,45
5	Investig Operación	81,55
5	Fundam De Redes	96,70
5	Sistemas Operativos	84,49
5	Analisis Y Disen O.O.	87,22
5	Entorno Economico	93,27
5	Comunicación Oral	89,54
6	Simulación	81,08
6	Metodos Numericos	97,17
6	Programac Movil	73,74
6	Sistem Operat Distrib	90,99
6	Ing Software	99,33
6	Tributacion	89,78
6	Organiz Y Administ	98,46
7	Ingles Tecnico	95,45
7	Enrutam De Redes	96,17
7	Adm Y Org Centro Com	72,21
7	Evaluac Y Gest Proyec	70,70
7	Etica Y Aspect Legal	92,87
7	Optativa 1 Form Gen	91,03
8	Telecomunicaciones	95,37
8	Evaluacion Sistemas	95,45
8	Formac De Empresas	99,04
8	Admin Proyec Technol	100,00
9	Auditoria Informat	95,83
9	Optativa 2	89,49
9	Optativa 3	89,81

Estas probabilidades se obtuvieron sobre el historial de cada materia de la malla 10 entre semestre B-2013 hasta B-2017.

Tabla7.*Promedio de aprobación por Profesor (A2018)*

LISTA DE PROFESORES	% A
Ing. Castro	99,33
Ing. Viteri	99,23
Ing. Molina	96,45
Ing. Serrano	96,16
Ing. Yanza	95,54
Rvdo. Salcedo	93,89
Ing. Toala	93,47
Ing. Zambrano	91,51
Ab. Castillo	91,21
Ing. Almeida	91,00
Ing. Villacis	90,99
Ing. Garcia	89,81
Ing. Santacruz	87,18
Ing. Manrique	86,46
Ing. Freire	85,62
Ing. Yong	84,02
Ing. Guerrero	82,72
Ing. Salazar	82,34
Ing. Zurita	81,91
Ing. Sion	80,26
Ing. Miranda	79,09
Ing. Cornejo	76,62
Ing. Armijo	76,43
Ing. Villacres	73,33
Ing. Gonzalez	72,17
Ing. Gallardo	72,05
Ing. Pesantes	71,76
Ing. Sanchez	70,16
Ing. Celleri	67,72
Ing. Camacho	62,16
Ing. Torres	58,80
Ing. Morejon	58,33
Ing. Zambrano	57,34

Estas probabilidades se obtuvieron sobre el historial de materias dadas de cada Profesor (33 Profesores).

Tabla8.*Promedio de aprobación por Estudiante Malla 1 (A2018)*

ESTUDIANTES	A%
A14	100,00
A16	100,00
A17	100,00
B11	100,00
B13	100,00
A18	100,00
B14	100,00
B15	100,00
A19	100,00
A20	100,00
A22	100,00
A23	100,00
B17	100,00
A27	100,00
B18	100,00
B20	100,00
A29	100,00
A30	100,00
B21	100,00
A32	100,00
B22	100,00
B23	100,00
B24	100,00
B25	100,00
B27	100,00
A36	100,00
B28	100,00
A25	91,67
A11	90,91
A12	83,33
B19	83,33
A31	83,33
A33	80,00
A24	77,78
A26	75,00
B12	66,67
A21	66,67
B16	66,67
A13	63,64
A28	33,33
A34	33,33
B26	33,33
B29	33,33
A37	33,33
A35	30,00
A15	0,00

Esta probabilidad se obtuvo con el historial académico de cada uno de los 46 (34 activos y 12 inactivos) estudiantes de la malla 1 en transcurso de los semestres que han estado activo en la carrera de ciencias de la computación, si el estudiante es nuevo se le asigna la probabilidad del 50% ya que no existen registros (23 estudiantes nuevos).

Tabla9.

Promedio de aprobación por Estudiante Malla 10 (A2018)

ESTUDIANTES	A%
C22	95,65
C20	95,45
C23	94,59
C11	92,31
C14	91,89
C12	89,47
C26	89,47
C15	86,49
C27	86,36
C21	84,62
C18	84,21
C19	82,93
C17	80,85
C16	80,65
C13	77,50
C25	75,00
C24	60,00

Esta probabilidad se obtuvo con el historial académico de 17 (17 activos) estudiantes de la malla 10 en transcurso de los semestres que han estado activo en la carrera de ingeniería en sistemas computacionales.

Después de obtener todos los datos de los estudiantes de la Malla 1 y 10, profesores y estudiantes (8454 registros académicos analizados), se procedió a crear el simulador académico en Excel, se detalla paso a paso como se creo y como funciona en la sección 4.3.

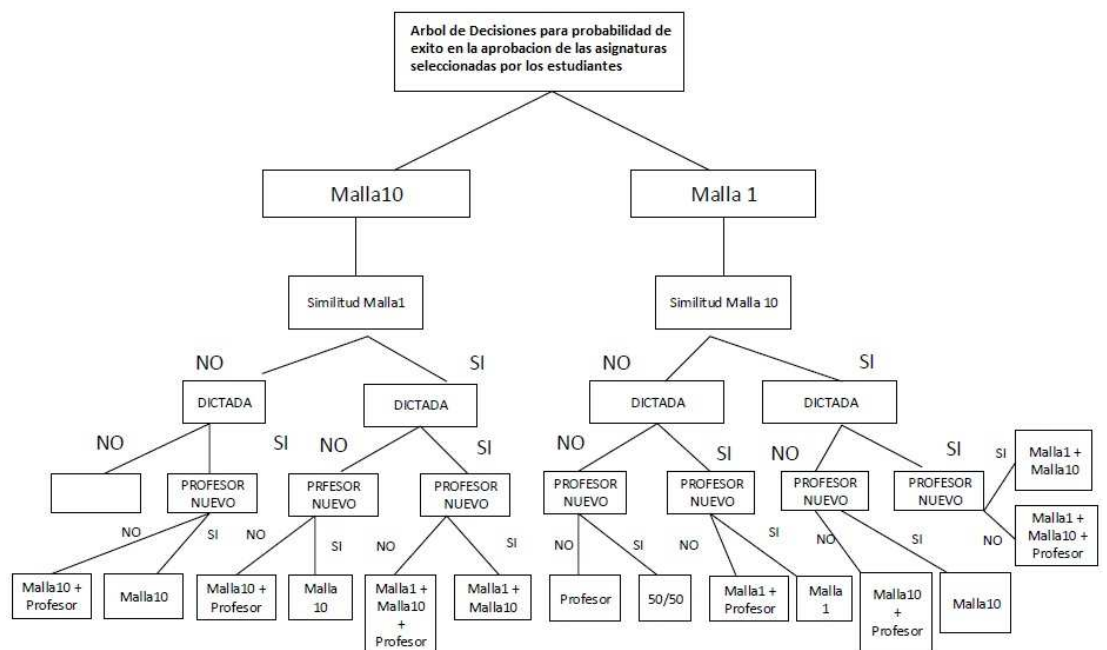
4.2. Análisis de los Resultados de la Investigación

Información:

En esta sección se muestra los resultados de las pruebas sobre el análisis de rendimiento académico de los estudiantes.

En esta simulación se usaron los datos del rendimiento académico de los estudiantes de la malla 1 del semestre A-2017 para simular inscribirse en el semestre B-2017 y los datos del rendimiento académico de la malla 10 desde semestre B-2013 hasta A-2017 para simular inscribirse en el semestre B-2017.

El árbol de decisiones para probabilidad de éxito en la aprobación de las asignaturas seleccionadas por los estudiantes quedo de la siguiente manera:



Las pruebas se realizaron de la siguiente manera se escogen el estudiante en la tabla dinámica de la izquierda luego se procede a coger las mismas materias que el estudiante escogió en el semestre B-2017 en la tabla dinámica de la derecha, luego inmediatamente muestra abajo del nombre del estudiante es decir debajo de la tabla dinámica de la izquierda te sale la probabilidad general dentro de un semáforo que te indica mediante colores el

posible rendimiento, luego se comparó con la realidad y se evaluó si estaba correcto.

Los resultados de las pruebas se muestran en la **Tabla 10**, **Tabla 11**, **Tabla 12**, se realizaron un total de 59 pruebas.

Tabla10.*Pruebas Malla 1 semestre (B2017)*

ESTUDIANTES	PROBAB	A-2017			PREDICCIÓN	B-2017		
		TOTAL	% A	% R		TOTAL	% A	% R
NUEVO*	50,00							
A11	83,33	6	5	1		5	5	0
A13	50,00	6	3	3		5	4	1
A14	100,00	6	6	0		6	6	0
A15	0,00	6	0	6		4	0	4
A16	100,00	4	4	0		2	2	0
A18	100,00	3	3	0		6	6	0
A19	100,00	6	6	0		6	6	0
A20	100,00	6	6	0		6	6	0
A22	100,00	6	6	0		6	6	0
A23	100,00	6	6	0		6	6	0
A24	80,00	5	4	1		4	3	1
A25	100,00	6	6	0		6	5	1
A26	100,00	6	6	0		6	3	3
A27	100,00	6	6	0		6	6	0
A28	50,00	6	3	3		4	0	4
A29	100,00	6	6	0		6	6	0
A30	100,00	6	6	0		6	6	0
A31	100,00	6	6	0		6	4	2
A32	100,00	6	6	0		6	6	0
A33	66,67	3	2	1		2	2	0
A35	0,00	6	0	6		4	3	1
A36	100,00	6	6	0		6	6	0
A37	50,00	6	3	3		3	0	3

En la **Tabla 10** se muestran los resultados de la prueba realizadas a los 42 estudiantes de la malla 1 en el semestre B2017, de los cuales 18 estudiantes obtuvieron semáforo verde, 22 estudiantes amarillo, y 2 estudiantes rojo.

Tabla11.*Pruebas Malla 10 semestre (B2017)*

ESTUDIANTES	B2013-A2017				B-2017			
	PROBAB	TOTAL	% A	% R	PREDICCION TOTAL	% A	% R	
C11	91,43	35	32	3	4	4	0	
C12	90,63	32	29	3	6	5	1	
C13	76,32	38	29	9	2	2	0	
C14	91,43	35	32	3	2	2	0	
C15	85,29	34	29	5	3	3	0	
C16	77,78	27	21	6	4	4	0	
C17	78,57	42	33	9	5	5	0	
C18	81,82	33	27	6	5	5	0	
C19	81,08	37	30	7	4	4	0	
C20	100,00	17	17	0	5	4	1	
C21	82,86	35	29	6	4	4	0	
C22	95,12	41	39	2	5	5	0	
C23	94,12	34	32	2	3	3	0	
C24	59,38	32	19	13	3	2	1	
C25	71,43	35	25	10	5	5	0	
C26	88,57	35	31	4	3	3	0	
C27	84,21	38	32	6	6	6	0	

En la **Tabla 11** se muestran los resultados de la prueba realizadas a 17 estudiantes de la malla 10 en el semestre B2017, de los cuales 16 estudiantes obtuvieron semáforo verde y 1 estudiante obtuvo amarillo.

Tabla12.

Matriz falsos positivos/negativos Malla 1 y 10 (B2017)

ESTUDIANTES	PROBAB	A-2017			PREDICCIÓN	B-2017		
		TOTAL	% A	% R		TOTAL	% A	% R
A13	50,00	6	3	3		5	4	1
A26	100,00	6	6	0		6	3	3
A28	50,00	6	3	3		4	0	4
A31	100,00	6	6	0		6	4	2
A35	0,00	6	0	6		4	3	1
A37	50,00	6	3	3		3	0	3
B11	50,00					5	5	0
B13	50,00					5	5	0
B14	50,00					4	4	0
B15	50,00					5	5	0
B17	50,00					5	5	0
B18	50,00					5	5	0
B20	50,00					3	3	0
B21	50,00					5	5	0
B22	50,00					5	5	0
B23	50,00					5	5	0
B24	50,00					5	5	0
B25	50,00					5	5	0
B26	50,00					5	1	4
B27	50,00					5	5	0
B28	50,00					4	4	0
B29	50,00					5	1	4

La **Tabla 12** muestra los falsos positivos y falsos negativos arrojados en las 59 pruebas que se realizaron en el semestre B2017 a los estudiantes de la Malla 1 y 10, de los cuales salieron 16 falsos positivos (27,12%) y 6 falsos negativos (10,17%) (**Negrita**).

4.3. Simulador Académico

En esta sección se va a explicar cómo se creó el simulador Académico, que predice el desempeño del estudiante al momento de inscribirse en el semestre.

La simulación se la realizo en Microsoft Excel con tablas dinámicas.

Lo que primero que se hizo fue la recopilación de datos que fue lo que se mostró en el capítulo 4.1. en el análisis exploratorio, ahí se sacó el promedio general de aprobación de cada materia y el promedio de rendimiento de cada estudiante de la malla 1, luego se procedió a crear dos tablas dinámicas, una con todos los datos de los estudiantes y otra con los datos de las materias.

Cuando se crearon las tablas dinámicas quedo de la siguiente manera:

The screenshot shows an Excel spreadsheet with two dynamic tables. The left table lists students and their probabilities, and the right table lists subjects and their probabilities. A formula bar shows a calculation for the overall probability.

LISTADO DE ESTUDIANTES	PROBABILIDAD%	MATERIAS A ESCOGER	PROBABILIDAD
CAYETANO AGUIRRE, GEANELLA GABRIELA	100,00	CALCULO DIFERENCIAL (ING. GUSTAVO ZAMBRANO)	58,83
ProbabilidadEstudiante%	100,00	FISICA (ING. MARIO CELERRI)	72,85
		FUNDAMENTOS DE PROGRAMACION (ING. TANYA ARMUJO)	76,43
		HUMANISMO INTEGRAL 2 (RVDO. SALACEDO)	95,60
		INTRODUCCION A LAS CIENCIAS COMP (ING. LUIS MANRIQUE)	85,73
		MATEMATICAS (ING. ANA CAMACHO)	67,63
Probabilidad General	76,18	ProbabilidadMaterias	76,18

Como podemos ver las dos tablas, la de la izquierda es para elegir el estudiante que se va a inscribir y la de la derecha es para elegir las materias que desea escoger.

En la tabla de la izquierda seleccionamos el estudiante que se va a inscribir en el semestre de la siguiente manera:

3	LISTADO DE ESTUDIANTES	PROBABILIDAD%
4	CAYETANO AGUIRRE , GEANELLA GABRIELA	100,00
5	ProbabilidadEstudiante%	100,00

Con dicho botón se selecciona al estudiante y luego se da en el botón aceptar para confirmar, luego me muestra en la columna de la derecha su probabilidad de aprobar materias.

En la tabla de la derecha es para seleccionar la cantidad de materias a escoger, como podemos ver existen dos columnas sobre la materia a escoger en el semestre y la probabilidad de cada materia, al final de la tabla dinámica muestra la probabilidad general del semestre.

El método para elegir las materias sería el siguiente:

MATERIAS A ESCOGER	PROBABILIDAD
CALCULO INTEGRAL (ING. GUSTAVO ZAMBRANO)	58,83
COMPUTACION Y SOCIEDAD (ING. GUSTAVO MOLINA)	92,83
ESTRUCTURA DE DATOS (ING. MARIO SION)	73,66
FUNDAMENTOS DE INVESTIGACION (ING. ADELA ZURITA)	79,54
INTRODUCCION A LOS SISTEMAS COMP (ING. MARCOS MIRANDA)	79,09
LOGICA DIGITAL (ING. LENIN MOREJON)	49,02
PROGRAMACION ORIENTADA OBJET (ING. GALO CORNEJO)	66,30
ProbabilidadMaterias	71,32

Ordenar de A a Z

Ordenar de Z a A

Más opciones de ordenación...

Borrar filtro de "MATERIAS MALLA NUEVA"

Filtros de etiqueta

Filtros de valor

Buscar

- (Seleccionar todo)
- ADMIN DE BASE DATOS
- ADMIN Y ORGANIZ DE CENTROS COM
- ANALISIS DE DATOS
- ANALISIS NUMERICO
- ANALISIS Y DISENO DE SISTEMAS
- ANALISIS Y OPTIMIZACION DE ALGORI
- AUDITORIA DE TIC
- BASE DATOS

ACEPTAR Cancelar

Con dicho botón se seleccionan las materias a través de la lista desplegable que contiene todas las materias y se da clic en el botón aceptar para confirmar las materias elegidas.

Después de seleccionar las materias deseadas la tabla dinámica nos muestra la probabilidad por materia:

MATERIAS A ESCOGER	PROBABILIDAD
CALCULO INTEGRAL (ING. GUSTAVO ZAMBRANO)	58,83
COMPUTACION Y SOCIEDAD (ING. GUSTAVO MOLINA)	92,83
ESTRUCTURA DE DATOS (ING. MARIO SION)	73,66
FUNDAMENTOS DE INVESTIGACION (ING. ADELA ZURITA)	79,54
INTRODUCCION A LOS SISTEMAS COMP (ING. MARCOS MIRANDA)	79,09
LOGICA DIGITAL (ING. LENIN MOREJON)	49,02
PROGRAMACION ORIENTADA OBJET (ING. GALO CORNEJO)	66,30
ProbabilidadMaterias	71,32

Y la probabilidad general de las materias escogidas al final de la tabla dinámica:

MATERIAS A ESCOGER	PROBABILIDAD
CALCULO INTEGRAL (ING. GUSTAVO ZAMBRANO)	58,83
COMPUTACION Y SOCIEDAD (ING. GUSTAVO MOLINA)	92,83
ESTRUCTURA DE DATOS (ING. MARIO SION)	73,66
FUNDAMENTOS DE INVESTIGACION (ING. ADELA ZURITA)	79,54
INTRODUCCION A LOS SISTEMAS COMP (ING. MARCOS MIRANDA)	79,09
LOGICA DIGITAL (ING. LENIN MOREJON)	49,02
PROGRAMACION ORIENTADA OBJET (ING. GALO CORNEJO)	66,30
ProbabilidadMaterias	71,32

Para implementación del semáforo se utilizó la herramienta formato condicional para aplicarle las condiciones, las reglas del semáforo condicional se definieron en la **Tabla 3**, entonces quedo de la siguiente manera:



Por último, se mostrará el respectivo color del semáforo sobre la probabilidad general depende de cuál sea su porcentaje.

LISTADO DE ESTUDIANTES	PROBABILIDAD%	MATERIAS A ESCOGER	PROBABILIDAD
PAREDES CALDERON , KAREN LISSETTE	100,00	CALCULO INTEGRAL (ING. GUSTAVO ZAMBRANO)	58,83
ProbabilidadEstudiante%	100,00	COMPUTACION Y SOCIEDAD (ING. GUSTAVO MOLINA)	92,83
		ESTRUCTURA DE DATOS (ING. MARIO SION)	73,66
		FUNDAMENTOS DE INVESTIGACION (ING. ADELA ZURITA)	79,54
		INTRODUCCION A LOS SISTEMAS COMP (ING. MARCOS MIRANDA)	79,09
Probabilidad General	85,66	LOGICA DIGITAL (ING. LENIN MOREJON)	49,02
		PROGRAMACION ORIENTADA OBJET (ING. GALO CORNEJO)	66,30
		ProbabilidadMaterias	71,32

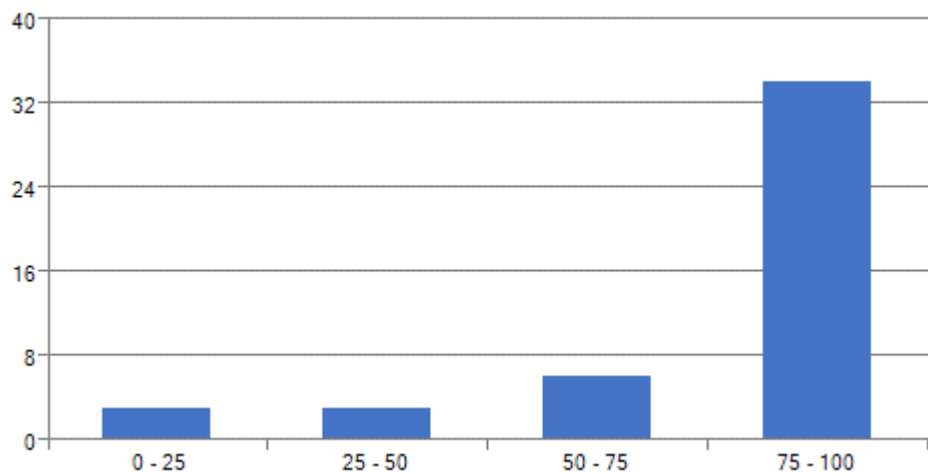
Mostrando la probabilidad general debajo de la tabla dinámica de la izquierda y mostrando cada probabilidad de la materia en la tabla dinámica de la izquierda, finaliza nuestro simulador de carga académica.

CONCLUSIONES

Al hacer una revisión de los resultados académicos de los estudiantes de las carreras de ingeniería en sistemas computacionales e ingeniería en ciencias de la computación, durante los periodos entre B-2013 al B-2017 se pudo observar el comportamiento de las probabilidades de aprobación de las asignaturas cursadas durante esos periodos, llegando a determinar que el nivel cuantitativo más bajo les corresponde a las asignaturas al área de matemáticas (Malla10) y fundamentos teóricos (Malla 1).

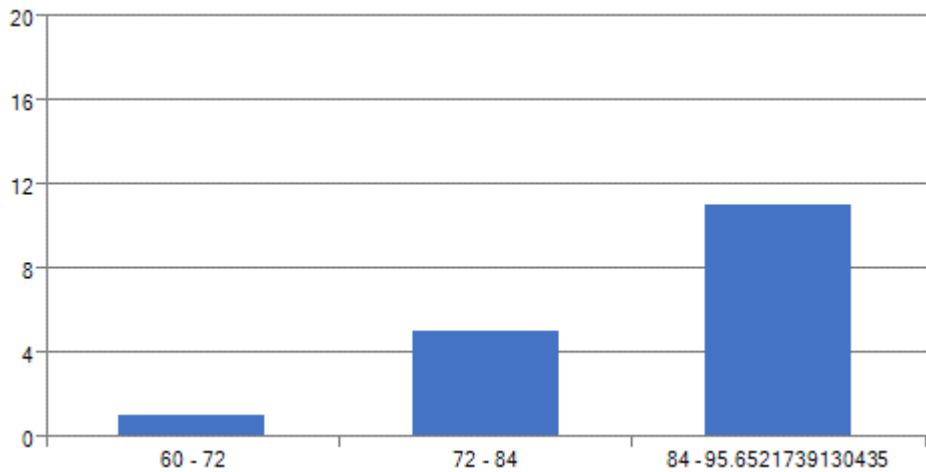
También se puede observar que la asignatura a nivel cuantitativo de probabilidad de aprobación más bajo corresponde a Lógica Computacional (Malla 10) y Matemáticas (Malla 1).

El rendimiento académico de los estudiantes de la malla 1 para el semestre A-2018



De los 46 estudiantes analizados 34 estudiantes tienen entre 75-100, 5 están entre 50-75, 4 entre 25-50 y 3 entre 0-25.

El rendimiento académico de los estudiantes de la malla 10 para el semestre A-2018



De los 17 estudiantes analizados 11 estudiantes tienen entre 84-96, 5 están entre 72-84, y 1 entre 60-72.

Después de hacer las pruebas del simulador académico, se procedió a analizar los resultados obtenidos determinando que el modelo si es confiable para predecir a los estudiantes sobre su carga académica antes de confirmar sus materias elegidas al inscribirse.

Nuestro objetivo es predecir cual va a hacer el rendimiento general del semestre, para esto se evalúa su carga académica general frente a su historial de rendimiento académico personal, para luego así advertir al estudiante mediante un semáforo como le ira en dicho semestre.

RECOMENDACIONES

- El modelo se debe validar utilizando otra herramienta de minería de datos por ejemplo redes neuronales, algoritmos genéticos entre otros.
- El modelo se debería utilizar en el proceso de registros para que los estudiantes tengan una predicción en el semestre inscrito.
- El responsable de la asesoría pedagógica de la carrera podría dar seguimiento a aquellos estudiantes que al momento de registrarse su semáforo le haya indicado rojo, para ofrecerles algún tipo de ayuda.
- El modelo funciona a través de unas reglas que se irán depurando en función de su utilización.

BIBLIOGRAFÍA

- Andrade. (2017). CERTIFICACIÓN DEL TRABAJO DE TITULACIÓN, 78.
- Asamblea Nacional Constituyente de Ecuador. (2008). La Educación Superior en la Constitución de la República del Ecuador. Recuperado 13 de junio de 2018, de <http://ecuadoruniversitario.com/directivos-y-docentes/legislacion/constitucion-de-la-republica-del-ecuador/la-educacion-superior-en-la-constitucion-de-la-republica/>
- Cervera. (1998). MÉTODOS Y TÉCNICAS DE INVESTIGACIÓN INTERNACIONAL, 180.
- Chungat, J. W. (2001). INGENIERO EN ESTADÍSTICA INFORMÁTICA, 462.
- Godinez. (2013). Tecnicas de Investigacion. Recuperado de <http://repositorio.ug.edu.ec/bitstream/redug/19855/1/TESIS%20LAS%20P%C3%8DLADORAS%20%28KAREN%20GARC%C3%8DA%20P.%20y%20D%C3%81MARIS%20L%C3%93PEZ%20L.%29.pdf>
- Ley Orgánica de Educación Superior. (2010). Ley Orgánica de Educación Superior, 39.
- Matsudo. (2001). *ÁRBOLES DE DECISIÓN, UNA TÉCNICA DE DATA MINING DESDE UNA PERSPECTIVA INFORMATICA y ESTADISTICA*. Universidad Nacional de Buenos Aires. Recuperado de <https://www.dc.uba.ar/academica/tesis-de-licenciatura/2001/matsudo.pdf>
- Microsoft. (2016a). Conceptos básicos sobre bases de datos. Recuperado 15 de junio de 2018, de <https://support.office.com/es-es/article/conceptos-b%C3%A1sicos-sobre-bases-de-datos-a849ac16-07c7-4a31-9948-3c8c94a7c204>

- Microsoft. (2016b). Crear una tabla dinámica para analizar datos de una hoja de cálculo. Recuperado 16 de julio de 2018, de <https://support.office.com/es-es/article/crear-una-tabla-din%C3%A1mica-para-analizar-datos-de-una-hoja-de-c%C3%A1lculo-a9a84538-bfe9-40a9-a8e9-f99134456576>
- Microsoft. (2016c). Microsoft Excel 2016, hojas de cálculo, prueba gratuita. Recuperado 26 de junio de 2018, de <https://products.office.com/es/excel>
- Miranda, A., & Encalada, L. (2009). INGENIERO EN ELECTRÓNICA Y TELECOMUNICACIONES, 77.
- Riquelme, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 10(29). Recuperado de <http://www.redalyc.org/resumen.oa?id=92502902>
- Troche. (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.
- Universidad Católica Santiago de Guayaquil. (2017a). La Universidad – Universidad Católica Santiago de Guayaquil. Recuperado 13 de junio de 2018, de <http://www.ucsg.edu.ec/la-universidad/>
- Universidad Católica Santiago de Guayaquil. (2017b). Misión, Visión y Objetivos – Universidad Católica Santiago de Guayaquil. Recuperado 13 de junio de 2018, de <http://www.ucsg.edu.ec/la-universidad/mision/>



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT
Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

DECLARACIÓN Y AUTORIZACIÓN

Yo, **Solines Bernardino Juan Jose**, con C.C: # **0924108301** autor/a del trabajo de titulación: **Minería de datos aplicada a la detección de patrones para el análisis de rendimiento académico de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la Universidad Católica Santiago de Guayaquil** 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, 24 de septiembre del 2018

f. _____

Nombre: Solines Bernardino Juan Jose

C.C: 0924108301



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT
Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA			
FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN			
TEMA Y SUBTEMA:	Minería de datos aplicada a la detección de patrones para el análisis de rendimiento académico de los estudiantes de la carrera de Ingeniería en Sistemas Computacionales de la Universidad Católica Santiago de Guayaquil.		
AUTOR(ES)	Solines Bernardino Juan Jose		
REVISOR(ES)/TUTOR(ES)	Freire Cobo Lenin Eduardo, Mgs.		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Ingeniería		
CARRERA:	Ingeniería en Sistemas Computacionales		
TÍTULO OBTENIDO:	Ingeniero en Sistemas Computacionales		
FECHA DE PUBLICACIÓN:	24 de septiembre del 2018	No. DE PÁGINAS:	51
ÁREAS TEMÁTICAS:	Modelo Predictivo, Base de Datos, Rendimiento Académico		
PALABRAS CLAVES/KEYWORDS:	Minería de Datos, Detección de Patrones, Rendimiento Académico, Modelo Predictivo		
RESUMEN/ABSTRACT (150-250 palabras):			
<p>Con el uso de la herramienta de Minería de Datos se analizaron patrones de rendimiento académico en los estudiantes, materias y profesores que permitieron elaborar un modelo predictivo en beneficio para los estudiantes, al momento que se inscriban en el semestre, para indicarle cuál es su probabilidad de éxito, alerta o fracaso basado en su rendimiento y en las materias que vaya a tomar. Para esta investigación se utilizó la metodología KDD que se basa en la extracción de conocimiento sobre una pequeña o gran cantidad de datos, se utilizó la herramienta de minería de datos Excel para recolección, procesamiento de datos, y la elaboración del simulador que aplico el modelo de predicción. En conclusión, el objetivo de este modelo lo que busca es mostrarle mediante un semáforo la probabilidad de éxito al estudiante basado en su rendimiento y las materias a escoger. Se realizaron 59 pruebas de los cuales mostro 16 falsos positivos (27,12%) y 6 falsos negativos (10,17%).</p> <p>Palabras Claves: Minería de Datos, Detección de Patrones, Rendimiento Académico, Modelo Predictivo.</p>			
ADJUNTO PDF:	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
CONTACTO CON AUTOR/ES:	CON	Teléfono: +593-4-0995735286	E-mail: juan.solines@cu.ucsg.edu.ec
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE):	LA	Nombre: Toala Quimi Edison Jose, Mgs.	
		Teléfono: +593-4-99 09 76 776	
		E-mail: edison.toala@cu.ucsg.edu.ec	
SECCIÓN PARA USO DE BIBLIOTECA			
Nº. DE REGISTRO (en base a datos):			
Nº. DE CLASIFICACIÓN:			
DIRECCIÓN URL (tesis en la web):			