



Universidad
Nacional
de Loja

Facultad de la Energía, las Industrias y los Recursos Naturales no Renovables

CARRERA DE INGENIERÍA EN SISTEMAS

Machine Learning para el análisis del rendimiento académico de estudiantes de Ingeniería

Línea de investigación: Sistemas Inteligentes

PROYECTO DE TITULACIÓN EN
INGENIERÍA EN SISTEMAS.

Autor:

- ◊ Bladimir Stanislao Minga Medina, bsmingam@unl.edu.ec

Tutor:

- Luis Antonio Chamba Eras

Cotutor:

- Edwin René Guamán Quinche



Carrera de Ingeniería en
Sistemas / Computación

LOJA - ECUADOR
2020

Certificación del Tutor

Luis Antonio Chamba Eras en calidad de tutor voluntario del proyecto de trabajo de titulación, certifico la tutela a Bladimir Stanislao Minga Medina, con el tema **Machine Learning para el análisis del rendimiento académico de estudiantes de Ingeniería**, quien ha cumplido con todas las observaciones requeridas. Es todo cuanto puedo decir en honor a la verdad, facultando al interesado hacer uso de la presente, así como el trámite de pertinencia del presente proyecto.

Loja, 18 de marzo de 2020

Atentamente,
Luis Antonio Chamba Eras
TUTOR

Certificación de Autoría del Proyecto

Yo Bladimir Stanislao Minga Medina, estudiante de la Universidad Nacional de Loja, declaro en forma libre y voluntaria que el presente proyecto de trabajo de titulación que versa sobre **Machine Learning para el análisis del rendimiento académico de estudiantes de Ingeniería**, así como la expresiones vertidas en la misma son autoría del compareciente, quien ha realizado en base a recopilación bibliográfica primaria y secundaria. En consecuencia asumo la responsabilidad de la originalidad de la misma y el cuidado al remitirse a las fuentes bibliográficas respectivas para fundamentar el contenido expuesto.

Atentamente,
Bladimir Stanislao Minga Medina

Índice general

1	Problemática	1
1.1	Situación Problemática	1
1.2	Problema de Investigación	2
2	Justificación	3
3	Objetivos	4
3.1	General	4
3.2	Específicos	4
4	Alcance	5
5	Marco Teórico	6
5.1	Rendimiento académico y la deserción universitaria en Ecuador	6
5.1.1	Rendimiento Académico	6
5.1.2	Factores que determinan el rendimiento académico	7
5.1.3	Deserción universitaria en Ecuador	7
5.2	Minería de datos y el Machine Learning	8
5.2.1	Minería de datos	8
5.2.2	Minería de datos Educativa	8
5.2.3	Analítica de aprendizaje (Learning analytics)	9
5.2.4	Machine Learning (Aprendizaje automático)	9
5.3	Trabajos Relacionados	10
6	Metodología	16
7	Cronograma	18
8	Presupuesto	20
8.1	Recursos Humanos	20
8.2	Recursos Físicos	20
8.3	Recursos Software	21
	Bibliografía	23
	Lista de Acrónimos y Abreviaturas	26

Índice de tablas

5.1	Trabajo relacionado 1.	10
5.2	Trabajo relacionado 2	11
5.3	Trabajo relacionado 3	12
5.4	Trabajo relacionado 4	12
5.5	Trabajo relacionado 5	13
5.6	Trabajo relacionado 6	14
5.7	Trabajo relacionado 7	15
7.1	Codificación de las actividades	18
8.1	Recursos humanos	20
8.2	Recursos físicos	21
8.3	Recursos de software	21
8.4	Presupuesto total	21

**Machine Learning para el análisis
del rendimiento académico de
estudiantes de Ingeniería**
Línea de investigación: Sistemas Inteligentes

1 Problemática

1.1 Situación Problemática

El Rendimiento Académico (RA) es fundamental en el marco de la educación superior por sus implicaciones en el cumplimiento de la función formativa de las instituciones educativas y el proyecto pedagógico de los estudiantes, debido a que contribuye a la instrucción profesional [1]. Por lo tanto, el RA es un indicador de progreso durante el proceso de enseñanza, y a su vez, es un pronosticador de la posibilidad de completar exitosamente un plan de estudios [2].

A nivel mundial, el fenómeno de la deserción universitaria conjuntamente con el bajo rendimiento académico afecta a las Instituciones de Educación Superior (IES). Es por ello, que actualmente estas IES muestran un gran interés para contrarrestar la deserción y el bajo rendimiento de los estudiantes [3]. En América Latina, a pesar de que durante los últimos 25 años se ha logrado avances significativos en lo que se refiere al acceso a la educación, [4] señala que toda la región presenta problemas de estudiantes que abandonan los estudios.

En Ecuador¹ las cifras de deserción universitaria bordean el 40%, un claro ejemplo: en 2015, en la Escuela Politécnica Nacional, la tasa de deserción fue el 50%. En la Universidad Nacional de Loja² (UNL) para disminuir la deserción escolar se está potenciando la infraestructura tecnológica con aulas virtuales e iconográficas para motivar a los estudiantes.

La eficiencia de un sistema educativo (SE) se da cuando los objetivos de enseñanza son alcanzados dentro de un tiempo oportuno y sin derroche de recursos. La repetición y la deserción son dos graves obstáculos que afectan a la eficiencia del SE, en lo cual, el estudiante que se inscribe a un centro universitario tiene las siguientes opciones: aprobar, repetir o desertar [5].

La deserción universitaria es el eslabón terminante en la serie del fracaso universitario y tiene una relación estrecha con la repetición, lo que genera un desaprovechamiento de recursos económicos y humanos [6].

¹El Telégrafo, “El Telégrafo - Noticias del Ecuador y del mundo - La deserción universitaria bordea el 40%,” El Telégrafo, 2016. [Online]. Available: <https://www.eltelegrafo.com.ec/noticias/sociedad/4/la-desercion-universitaria-bordea-el-40>. [Accessed: 21-Nov-2019].

²P. Gúixa, “Periódicos de alrededor del mundo.,” La Vanguardia, 2019. [Online]. Available: <https://www.pressreader.com/ecuador/la-hora-loja/20181023/282295321183993>. [Accessed: 27-Nov-2019].

La deserción universitaria se ha venido presentando desde hace varios años en las universidades ecuatorianas y ha ido intensificándose de forma evidente, con una proporción alta de estudiantes que han renunciado sus estudios, que por lo general sucede en los primeros semestres de la carrera [7]. Los motivos para que un estudiante abandone la universidad son muy diversos, aunque existen ciertos sucesos que se vuelven extremadamente comunes en la mayoría de los grupos. Esta situación afecta, no solo al estudiante, sino que también a la IES, en el no cumplimiento de su visión y a la sociedad ecuatoriana, ya que al desertar de la educación superior los avances socio-económicos y tecnológicos del país se retrasan [8].

El Consejo de Aseguramiento de la Calidad de la Educación Superior (CACES) bajo la Ley Orgánica de Educación Superior (LOES), señala que las IES deben realizar el seguimiento del desempeño estudiantil conforme a su normativa vigente, y los resultados deben ser utilizados en la toma de acciones para mejorar las tasas de permanencia estudiantil. Además, [9] señala que: “las universidades deben desarrollar actividades para promover la permanencia del estudiantado, el mejoramiento del desempeño estudiantil y la titulación”.

Existen diferentes factores que influyen en la decisión de los estudiantes de abandonar la universidad. En [8], [10], [11] señalan a la situación económica como la principal causa de la deserción. Mientras que, en [12] y [13] mencionan factores como: rendimiento académico, relación entre estudiantes y con sus docentes, el nivel de involucramiento con la vida universitaria, entre otros. Finalmente, [14] expone como otra causa a las insuficiencias en la integración del estudiante con el contexto social e intelectual de la comunidad universitaria.

Desde todo punto de vista, es conveniente analizar las diferentes causas (académicas, económicas, etc.) que permiten reconocer factores de carácter individual que se presentan cuando un alumno abandona los estudios. Lo que nos lleva a plantear la siguiente pregunta de investigación: ¿Mediante el uso de algoritmos de Machine Learning se podrá predecir el rendimiento académico de estudiantes de la Carrera de Ingeniería en Sistemas y la Carrera de Computación de la Universidad Nacional de Loja?

1.2 Problema de Investigación

¿Mediante el uso de algoritmos de Machine Learning se podrá predecir el rendimiento académico de estudiantes de la Carrera de Ingeniería en Sistemas y la Carrera de Computación de la Universidad Nacional de Loja?

2 Justificación

El presente Trabajo de Titulación (TT), busca identificar patrones en los datos académicos, y así, poder predecir el resultado (rendimiento académico) que tendrán los estudiantes al terminar un periodo escolar, con el fin que las autoridades de la UNL potencien la toma de decisiones para: mejorar el rendimiento académico, fomentar la permanencia estudiantil y garantizar calidad en la educación.

Con el desarrollo de la presente propuesta de TT se pretende generar conocimiento tecnológico para que en el futuro sirva de base para resolver problemas que se presentan en el contexto educativo y la sociedad en general, mediante la implementación de soluciones inteligentes y aprovechando al máximo las tecnologías emergentes. Además, permitirá a las autoridades mejorar la toma de decisiones y plantear estrategias con el fin de mantener la permanencia estudiantil, mejorando así la tasa de estudiantes que cumplen con éxito los programas académicos.

De igual manera, se busca reducir el despilfarro de recursos económicos que actualmente el estado ecuatoriano invierte en las Universidades públicas. Así como también, evitar gastos excusados en los estudiantes y sus representantes.

En lo que corresponde a los Objetivos de Desarrollo Sostenible (ODS), el desarrollo del TT aportará a conseguir el objetivo 4 “Educación de calidad”, con el fin de incrementar el número de personas con competencias técnicas y profesionales, que puedan acceder al trabajo decente y el aumentar el emprendimiento. En Ecuador la presente propuesta de TT se ampara sobre la línea de investigación “TIC Y Educación” y la línea “Big Data”. En la UNL pertenece al grupo “Energía, Industrias y Recursos Naturales no Renovables” y se asocia a la línea de investigación “Sistemas Inteligentes” y la línea “Informática Educativa Inteligente”, pertenecientes a la CIS. Finalmente, el presente se relaciona directamente con el proyecto LALA, porque busca mejorar la calidad y eficiencia de la Educación Superior mediante la gestión de herramientas de Learning Analytics para mejorar el proceso de toma de decisiones académicas.

3 Objetivos

3.1 General

Analizar el rendimiento académico de estudiantes de la Carrera de Ingeniería en Sistemas y la Carrera de Computación de la Universidad Nacional de Loja mediante algoritmos de Machine Learning.

3.2 Específicos

- Identificar una metodología que permita extraer conocimiento a partir de los datos académicos almacenados.
- Seleccionar el algoritmo de Machine Learning más eficiente para predecir el rendimiento académico.
- Simular el comportamiento del modelo generado.

4 Alcance

La presente propuesta de Trabajo de Titulación (TT) tiene la finalidad de identificar patrones que se encuentran presentes en la información académica y así, poder predecir el rendimiento académico de estudiantes de la Carrera de Ingeniería en Sistemas y la Carrera de Computación.

El tiempo estimado para el desarrollo del TT es de 6 meses, en la cual se pretende completar las siguientes etapas:

1. Identificar una metodología que permita extraer conocimiento a partir de los datos académicos almacenados:
 - Revisión bibliográfica para identificar trabajos relacionados y casos de éxito.
 - Realizar una comparación entre las diferentes metodologías encontradas.
 - Identificar una metodología.
2. Seleccionar el algoritmo de Machine Learning más eficiente para predecir el rendimiento académico:
 - Revisión bibliográfica para identificar los principales algoritmos de Machine Learning (aprendizaje automático).
 - Realizar una comparación entre los diversos algoritmos de Machine Learning identificados.
 - Seleccionar el algoritmo que brinde una mejor precisión.
3. Simular el comportamiento del modelo generado:
 - Evaluar la calidad del modelo, mediante la asignación de parámetros que pueden ser reales o ficticios de algún caso particular, en donde se conozca el resultado esperado (si aprobó o reprobó).

5 Marco Teórico

En las dos primeras secciones que se presentan a continuación se describe las bases teóricas referentes al caso de estudio y necesarias para comprender la presente propuesta de Trabajo de Titulación. En la sección tercera se presenta siete trabajos que de alguna u otra manera tienen relación con el presente proyecto.

5.1 Rendimiento académico y la deserción universitaria en Ecuador

5.1.1 Rendimiento Académico

El Rendimiento Académico (RA) es el resultado del aprendizaje producido por la correlación didáctica y pedagógica del docente con el estudiante. Lo cual significa que, el RA permite identificar si el estudiante cumple con los modelos de aprendizaje que determina el currículo de educación para ser promovido al siguiente nivel [15].

El estudio del RA es de suma importancia debido a la necesidad de aumentar la excelencia académica de las instituciones, por tal motivo, [16] menciona que el rendimiento académico en la educación superior hay que estudiarlo desde un contexto sociocultural-económico-político, a la vez familiar, personal y académico, ya que los factores que intervienen en el mismo son numerosos y se encuentran muy interrelacionados.

Como indica [16], los estudios realizados acerca del rendimiento académico son fundamentalmente sociológicos, ya existen investigaciones y tesis que no son puramente descriptivos sino que más bien y con el progreso de la tecnología y el auge de la Inteligencia artificial, minería de datos, algoritmos de predicción, entre otros, se empezaron a llevar a cabo investigaciones que relacionan los factores y las causas del fracaso o éxito de los estudiantes en su vida académica, y hasta se intenta predecir posibles resultados.

Sin embargo, [16] también relata algo muy importante, y es que hasta la fecha no se ha podido encontrar investigaciones que describan o relacionen el rendimiento académico de los estudiantes en la universidad con el puesto de trabajo que ocupará fuera de ella, es decir relacionar el rendimiento académico de un estudiante con el éxito o fracaso en la vida ocupacional del individuo.

5.1.2 Factores que determinan el rendimiento académico

En [17] se alude que los antecedentes familiares determinan en gran medida el desempeño académico de los estudiantes y se sostiene que el rendimiento escolar está fuertemente relacionado con los antecedentes familiares y fragilmente con las características de la escuela.

Durante las últimas décadas, los estudios acerca del desempeño académico han aumentado notablemente y aunque los resultados obtenidos son heterogéneos, existen investigaciones en las que se concluye que las características del estudiante o las de su entorno familiar se establecen como el factor más relevante [17].

En el caso de las características individuales como: el sexo, la repetición de curso y la raza, constituyen como los factores que más influyen en el desempeño escolar. De esta manera, se observa lo siguiente [17]: 1) en promedio, las mujeres obtienen mejores resultados en lectura y los hombres lo hacen en matemáticas y ciencias; 2) la condición de repetidor de curso afecta negativamente la puntuación media obtenida en cada una de las áreas evaluadas, y 3) los estudiantes de raza negra tienen mayores desventajas que los estudiantes de otras razas y, por tanto, obtienen peores resultados.

En [18] se expone que los factores más influyentes en la deserción estudiantil son los problemas económicos con un porcentaje de 25,29%, seguido por problemas personales (19,54%) y de trabajo (12,64%). Haciendo énfasis a [18] el factor más relevante es el problema económico y es por este motivo que el gobierno ecuatoriano ha tomado acciones para contrarrestar esta situación, a través de leyes que indican que la educación pública universitaria en Ecuador es universal y gratuita, para incrementar el número de estudiantes en el sistema educativo superior, beneficiando de esta manera a los sectores más vulnerables de la sociedad ecuatoriana. También se puede destacar que el Ecuador ha incrementado la inversión en educación superior de 1,6% del PIB a 2% en el 2015, siendo uno de los países de la región que mayor invierte en la Educación Superior, haciendo una inversión de \$ 12 mil millones de dólares entre los años 2007 al 2015 [8].

5.1.3 Deserción universitaria en Ecuador

El fracaso académico, es uno de los puntos más delicados que el sistema de educación superior tiene, por tal motivo es merecedor de un profundo análisis. La deserción es el fenómeno en el cual un individuo abandona su proceso de formación académica, debido a los diversos problemas acarreados durante el transcurso de su vida estudiantil [19].

La deserción estudiantil en la educación superior genera pérdidas económica, social y productiva del país. Según [8], la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (Unesco) indica que el abandono a nivel universitario en Ecuador llega a 40%.

La aptitud que tiene un estudiante para cumplir con los requerimientos implantados para los cursos dictados en una universidad y lograr superar dichos desafíos, se refleja en las calificaciones y se lo podrá considerar como un alumno con un buen rendimiento

académico. Por otro lado, aquellos que no muestran los conocimientos y habilidades para aprobar dichos requerimientos presentarán un bajo rendimiento académico, lo que podría favorecer a la deserción [20]. Por tanto, el bajo rendimiento académico es uno de los componentes que más influye en el fracaso universitario, afectando a una multitud de países [21].

En [22] se observa que, siempre la deserción universitaria ocurre mayormente en el primer año, en relación al segundo año de haber iniciado los estudios. Distintos estudios que han intentado investigar las causas y factores relacionados con la deserción estudiantil en las IES, concluyen en que la deserción estudiantil se puede explicar a partir de cinco dimensiones: psicológicas, sociológicas, económicas, institucionales y de interacción [23].

En [19] se indica que la deserción universitaria, sucede en los semestres inferiores a quinto, aumentando a medida que avanza la edad del estudiante, el número de materias y la complejidad de las mismas, llegando incluso a ubicarse en rangos del 30,8%. Es de esta manera, que se puede decir que la deserción es un punto crítico en Ecuador, el cual genera la duda de que ocurre y que medidas se puede implementar para reducir estos indicadores preocupantes, indicándonos también, que hay una pérdida en el sistema de educación.

5.2 Minería de datos y el Machine Learning

5.2.1 Minería de datos

La minería de datos en ciencias de la computación, es el proceso de descubrir patrones útiles e interesantes en grandes volúmenes de datos relacionados, en donde, el propósito de los métodos de minería de datos, es extraer conocimiento significativo [24].

5.2.2 Minería de datos Educativa

La minería de datos aplicada al sector educativo se la conoce como Educational Data Mining (EDM) y es una disciplina emergente que busca desarrollar métodos para examinar datos [20]. La EDM se considera como una de las mejores alternativas para examinar y comprender de mejor forma los problemas del RA de los estudiantes [25].

La EDM es un área de investigación muy importante que ayuda a predecir la información útil de las bases de datos educativas, para aumentar el rendimiento educativo y tener una mejor evaluación del proceso enseñanza-aprendizaje del alumnado [26].

LA EDM es un paradigma orientado a la generalización de modelos, métodos y algoritmos para la exploración de datos dentro de un entorno educativo, con el fin de encontrar y examinar patrones que describen el proceder de los estudiantes en base a sus metas alcanzadas, evaluaciones y el dominio de conocimiento obtenido con los diversos mecanismos de enseñanza-aprendizaje que actualmente se encuentran disponibles en las instituciones educativas, sobre todo de nivel superior [27].

La implementación de la EDM, está favoreciendo a todos los agentes educativos que trabajan en distintas líneas de educación, tratando de ayudar a los administradores a gestionar con efectividad los recursos y la capacidad para analizar el rendimiento de los estudiantes y el entorno académico. La implementación de la EDM se convierte en un aspecto significativo, ya que permite abordajes personales, sociales, psicológicos y de factores ambientales que influyen en el aprendizaje de los estudiantes y su capacidad para superar las limitaciones y triunfar, o en su defecto, caracterizar las determinantes del fracaso [28].

5.2.3 Analítica de aprendizaje (Learning analytics)

La Analítica del Aprendizaje es un campo de investigación reciente que analiza información relacionada a personas que están aprendiendo y sus contextos, con el propósito de comprender y optimizar el aprendizaje y los entornos en los que se da dicho aprendizaje. En los últimos años, las instituciones educativas han decidido explorar sus grandes agrupaciones de datos, con el fin de aumentar los índices de retención y proporcionar una experiencia personalizada y de mayor calidad para los estudiantes [29].

La analítica del aprendizaje se apoya en un conjunto más extenso de disciplinas académicas que la EDM, integrando conceptos y técnicas de la ciencia de la información y la sociología, además de la informática, la estadística, la psicología y las ciencias del aprendizaje [27].

5.2.4 Machine Learning (Aprendizaje automático)

El machine learning (ML) es una rama del conocimiento que permite a las máquinas desarrollar técnicas para poder aprender mediante procesos automatizados, extrayendo patrones de los datos para la construcción de modelos que faciliten la predicción empleando algoritmos [17]. ML es un campo de investigación que juega un rol fundamental en las instituciones educativas y es uno de los dominios más importantes de exploración con el objetivo de descubrir hechos significativos tomados de datos históricamente almacenados [27].

Un sistema que aprende de forma automatizada, se dice que es un instrumento (algoritmos) que para resolver problemas toma decisiones basadas en la experiencia acaparada. Estos sistemas debe tener la capacidad de trabajar con una variedad amplia de tipos de datos de entrada, que pueden incluir datos incompletos, inciertos, ruido, inconsistencias, etc. [27].

El propósito de una predicción es deducir el valor de un atributo de los datos desde una combinación de otra dimensión. La predicción en EDM usualmente se realiza a través de la elaboración de modelos mediante técnicas como clasificación, regresión y categorización [20]. La técnica más utilizada, en la tarea de predicción, es la clasificación y entre los algoritmos que se pueden utilizar se encuentran: árbol de decisiones, redes neuronales, bayesiano, k-nearest neighbor y Suppor Vector Machine, entre otros. [20].

5.3 Trabajos Relacionados

En las tablas 5.1-5.7 se presenta investigaciones preliminares relacionadas con el presente proyecto, de las cuales se menciona lo siguiente: el título, la metodología o técnicas aplicadas, las herramientas utilizadas, la cantidad de instancias (registros de datos) que manejaron para el estudio y finalmente, las conclusiones a las que llegaron cada uno de los diferentes autores.

Tabla 5.1: Trabajo relacionado 1.

Título:	Predicción del rendimiento académico de los Estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático [16]
Metodología / Técnicas:	CRISP-DM
Herramientas:	Weka
Instancias:	12609 instancias y 79 atributos
Conclusiones:	<p>El rendimiento académico de los estudiantes es un tema bastante complejo y que depende de muchas variables no solo socio demográficas y socio educativas, sino que dependen también de otros factores como el aspecto emocional de los estudiantes y también la familia, como se verificó en la revisión bibliográfica, sin embargo, el rendimiento académico se puede predecir a través de los datos de ingreso o de admisión a la UNSAAC, utilizando los algoritmos de aprendizaje automático hasta en un 69% de efectividad.</p> <p>Los factores claves de los datos de ingreso que determinan el rendimiento académico de los estudiantes a partir de los datos de ingreso son diversos, los principales son: la Nota de ingreso, la escuela profesional que se estudia, el semestre, el género, y la modalidad de ingreso, esta información fue generada por el análisis estadístico utilizando chi cuadrado y el coeficiente de correlación de Pearson, cumpliéndose con la hipótesis planteada.</p> <p>Según los resultados encontrados, el algoritmo de árboles de decisión “Random Forest”, fue el algoritmo que tuvo el mejor performance para la predicción del rendimiento académico de los ingresantes en los primeros semestres a la UNSAAC con un 69% de predicción, el segundo algoritmo con mejor performance fue algoritmo de Regresión Logística con un 68% para el presente caso de estudio.</p>

Tabla 5.2: Trabajo relacionado 2

Título:	Predicción del rendimiento académico en las nuevas titulaciones de grado de la EPS de la Universidad de Córdoba [30]
Metodología / Técnicas:	Técnicas de clasificación
Herramientas:	Weka
Instancias:	280 instancias y 24 atributos
Conclusiones:	<p>Con respecto a la utilidad de los modelos obtenidos, los dos modelos mostrados permiten detectar a los alumnos que pueden fracasar o abandonar. Estos nos indican las variables o factores que más afectan, y más concretamente qué valores y cómo están relacionados con otras variables o factores. También queremos indicar que algunos de los atributos discriminantes no han aparecido en ninguno de los dos modelos obtenidos, como por ejemplo el nivel de utilización de Moodle.</p> <p>Con respecto a la exactitud de los modelos, los porcentajes de clasificación obtenidos no han sido muy altos, lo que nos indica la dificultad del problema de la predicción del rendimiento académico al estar afectado por muchos factores. Creemos que algunas formas de mejorar la exactitud podrían ser:</p> <ul style="list-style-type: none"> • Utilizar sólo información sobre una única asignatura o una misma titulación, ya que hay factores que pueden variar entre una titulación y otra. Para ello se necesitaría tener más datos/alumnos de cada titulación. • Utilizar las notas obtenidas en las siguientes convocatorias de las asignaturas durante los siguientes 2 años. De esta forma se tendrá una mayor fiabilidad sobre los alumnos que siguen sin aprobar y cuales abandonan. • Recoger la información necesaria de otras fuentes más fiables que desde una encuesta realizada a los propios alumnos, ya que se basa en su sinceridad. Aunque variables como las notas anteriores se podrían obtener desde otras fuentes, hay variables como la situación económica familiar que son mucho más difícil de obtener.

Tabla 5.3: Trabajo relacionado 3

Título:	Prediction of Students Performance using Educational Data Mining [31]
Metodología / Técnicas:	Classification methodologies
Herramientas:	Weka
Instancias:	700 instancias y 19 atributos
Conclusiones:	<p>En este trabajo, la clasificación se emplea en la información de los estudiantes para predecir la división de los estudiantes sobre la base de la información previa. Como hay varios enfoques que esa unidad de área utiliza para la clasificación del conocimiento, se emplea aquí el teorema de la ingenuidad. Información como acción de grupo, prueba de clase, seminario y notas de la tarea se recogieron de la información previa de los estudiantes, para predecir el rendimiento en la parte superior del semestre.</p> <p>Este estudio puede facilitar a los estudiantes y a los profesores a impulsar a los estudiantes de todas las categorías a rendir bien. Este estudio ayuda a detectar a los estudiantes que requieren una atención especial, a minimizar la tasa de fracaso y a tomar medidas aceptables para los próximos exámenes semestrales.</p> <p>El trabajo futuro incluye la aplicación de técnicas de procesamiento de datos para el conjunto de conocimientos distendidos de enfermería con atributos típicos adicionales para instar a resultados correctos y económicos.</p>

Tabla 5.4: Trabajo relacionado 4

Título:	Analysis of Educational Data Mining using Classification [26]
Metodología / Técnicas:	Classification techniques
Herramientas:	Weka
Instancias:	163 instancias y 16 atributos
Conclusiones:	<p>La minería de datos tiene una importancia significativa en la educación. El conocimiento adquirido por el uso de técnicas de minería de datos se puede utilizar para tomar decisiones exitosas y efectivas que mejorarán y progresarán el desempeño del estudiante en la educación.</p>

Tabla 5.5: Trabajo relacionado 5

Título:	Performance Analysis of Feature Selection Algorithm for Educational Data Mining Maryam [25]
Metodología / Técnicas:	Classification algorithms
Herramientas:	Weka
Instancias:	500 instancias y 16 atributos
Conclusiones:	Los resultados del conjunto de datos de los estudiantes han demostrado que no hay ningún cambio significativo en el rendimiento de los algoritmos de selección de características disponibles en la herramienta weka. Sin embargo, entre todos estos métodos de selección de rasgos disponibles, los componentes principales han mostrado mejores resultados al utilizarlo con el clasificador de Random Forest. Este estudio también ha mostrado que el clasificador MLP se desempeñó ligeramente mejor que otros clasificadores en el conjunto de datos de los estudiantes. Los resultados muestran que es necesario un sutil ajuste de los parámetros de estos métodos de selección de características, en para conseguir mejores resultados rendimiento.

Tabla 5.6: Trabajo relacionado 6

Título:	Estudio del Rendimiento Académico Aplicando Técnicas de Minería de Datos [2]
Metodología / Técnicas:	CRISP-DM
Herramientas:	Weka, RAPIDMINER
Instancias:	2887 instancias y 13 atributos
Conclusiones:	<p>El empleo de técnicas de minería de datos ayuda a obtener modelos computacionales, ya que contiene diferentes algoritmos que ayudan a manipular los datos para obtener resultados que permita determinar el rendimiento académico. Sin embargo, mediante la técnica de Clasificación a través de los diferentes tipos de algoritmos evaluados, C4.5 permitió obtener los mejores resultados de acuerdo a los datos académicos, institucionales y personales de los estudiantes.</p> <p>Mediante el modelo de minería de datos obtenido y evaluado a través de datos reales se pudo comprobar que el rendimiento académico de los estudiantes está considerado bueno, con promedio de notas entre 7.00 - 9.00.</p>

Tabla 5.7: Trabajo relacionado 7

Título:	Identificación de Factores en la Reprobación y Deserción mediante técnicas de Minería de Datos en el Área de la Energía de la Universidad Nacional de Loja [32]
Metodología / Técnicas:	CRISP-DM
Herramientas:	Algoritmos basados en reglas
Instancias:	9545 instancias y 15 atributos
Conclusiones:	<p>En el trabajo realizado se llegó a la conclusión de que el factor que incide mayormente en la deserción de los estudiantes es el académico, seguido por el institucional dejando por último al factor individual, mientras que en el ambiente de reprobación los estudiantes que reprueban presentan la siguiente condición: Si el estudiante tiene un promedio de notas regular, un promedio de asistencia bajo y pertenecen a la carrera de Ingeniería en Sistemas entonces si reprueba.</p> <p>Se analizaron los datos que posee la institución y se pudo confirmar que en el área de energía los estudiantes reprueban en cada periodo académico un promedio de 23%.</p> <p>En el transcurso de la tarea de modelado se han obtenido excelentes resultados con los algoritmos basados en reglas como Ridor, Part y JRip.</p> <p>Se han examinado distintos algoritmos de clasificación y una de las conclusiones que se obtiene del análisis de los resultados obtenidos es que en este caso los algoritmos que generaban modelos más sencillos son los que presentan mejores resultados.</p> <p>Con el desarrollo del presente trabajo se ha aprendido a seguir la metodología CRISP-DM en un proyecto. Ya que permitió retomar y repetir fases anteriores, para generar nuevos modelos que permitan alcanzar los objetivos propuestos.</p>

6 Metodología

Para el desarrollo de la presente PTT se utilizará un conjunto de datos perteneciente a estudiantes de la Carrera de Ingeniería en Sistemas y la Carrera de Computación. Estos datos se encuentran almacenados en el Sistema de Gestión Académica (SGA) de la UNL y se administran desde la Unidad de Telecomunicaciones e Información (UTI).

Para alcanzar el objetivo general de la presente propuesta, se debe concluir con cada uno de los objetivos específicos. Por lo tanto, se propone utilizar los siguientes métodos:

1. Identificar una metodología que permita extraer conocimiento a partir de los datos académicos almacenados:
 - Revisión bibliográfica: Se utilizará este método para determinar el estado del arte sobre el área de minería de datos. En esta etapa se recopilará información sobre las metodologías más utilizadas y recomendadas para el desarrollo de proyectos enfocados a la extracción de conocimiento a partir de los datos.
2. Seleccionar el algoritmo de Machine Learning más eficiente que permita predecir el rendimiento académico:
 - Revisión bibliográfica: Se utilizará este método para recolectar información acerca de los algoritmos de aprendizaje automático más utilizados para estudiar el comportamiento académico.
 - Estudio de casos: A través de este método se pretende utilizar diferentes fuentes de evidencia que permitan conocer a más profundidad el fenómeno estudiado y con ello, identificar cómo se trabajó con problemas semejantes.
 - Método Inductivo: Se utilizará para comprender los datos académicos de los estudiantes y hacer un análisis de las variables presentes en la información de los alumnos que no completan exitosamente un plan de estudios.
3. Simular el comportamiento del modelo generado:
 - Experimentos: Se propone utilizar este método para analizar el comportamiento del modelo mediante la modificación de valores en las variables de entrada.

La técnica de la observación se utilizará para conocer la realidad académica de los estudiantes, docentes y administrativos de la UNL. Así, como también, para seguir obteniendo información adicional a lo largo del desarrollo del presente proyecto.

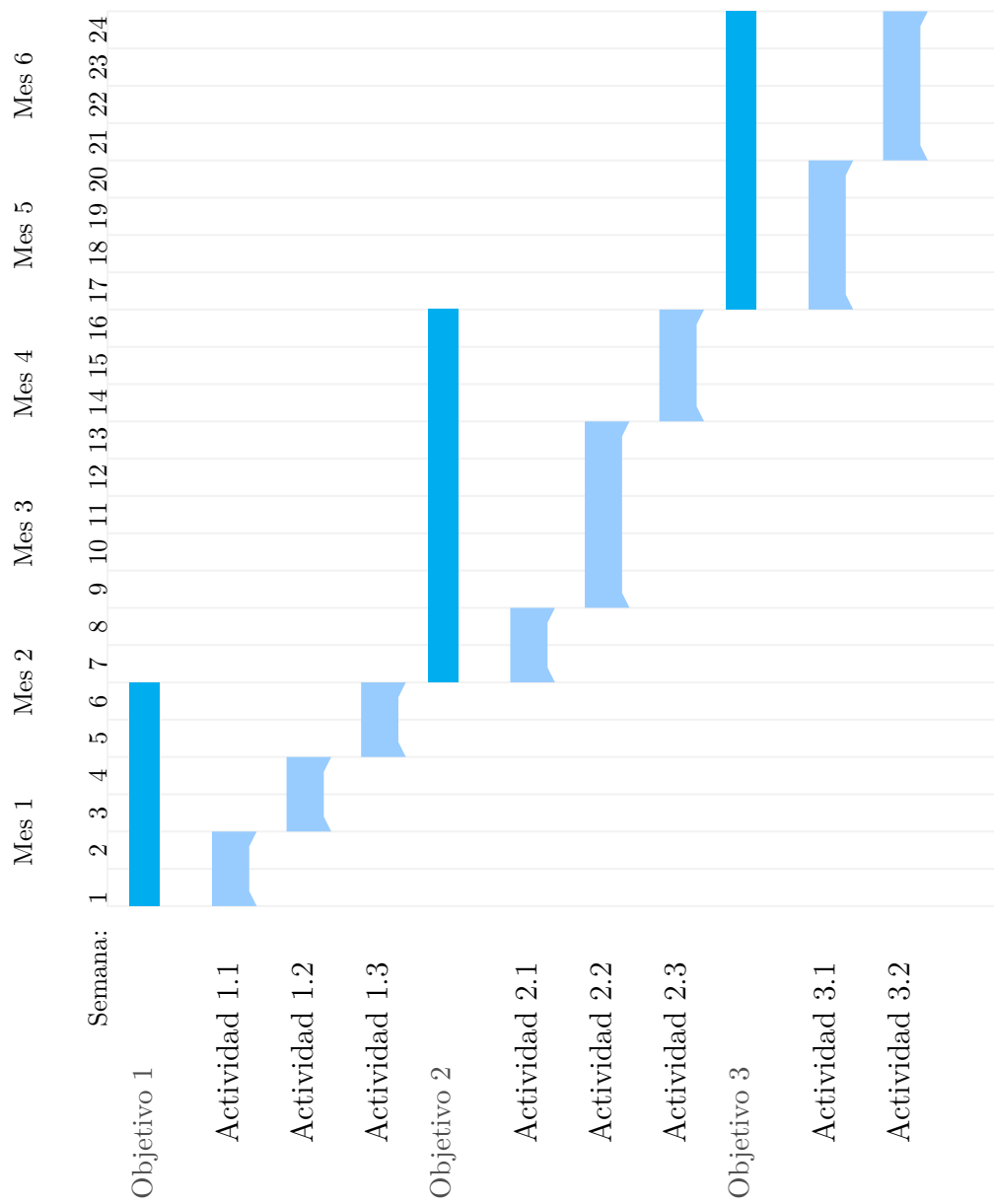
De igual manera, la revisión bibliográfica permitirá recopilar información complementaria y conducente para la elaboración del marco teórico.

7 Cronograma

Para un buen entendimiento y una mejor organización de las actividades que se deben realizar durante el desarrollo del presente proyecto, se procede a codificar las actividades según la tabla 7.1. Esta codificación se utilizará en la siguiente página, en donde se describe el cronograma mediante un diagrama de gantt.

Tabla 7.1: Codificación de las actividades

CÓDIGO	DESCRIPCIÓN
Objetivo 1	Identificar una metodología que permita extraer conocimiento a partir de los datos académicos almacenados.
Actividad 1.1	Revisión bibliográfica para identificar trabajos relacionados y casos de éxito.
Actividad 1.2	Realizar una comparación entre las diferentes metodologías encontradas.
Actividad 1.3	Identificar una metodología.
Objetivo 2	Seleccionar el algoritmo de Machine Learning más eficiente para predecir el rendimiento académico.
Actividad 2.1	Revisión bibliográfica para identificar los principales algoritmos de Machine Learning.
Actividad 2.2	Realizar una comparación entre los diversos algoritmos de Machine Learning identificados.
Actividad 2.3	Seleccionar el algoritmo que brinde una mejor precisión.
Objetivo 3	Simular el comportamiento del modelo generado.
Actividad 3.1	Simular el comportamiento del modelo generado.
Actividad 3.2	Terminar y presentar el Informe Final



8 Presupuesto

A continuación se detalla el presupuesto y financiamiento necesario para desarrollar el proyecto denominado: “Machine Learning para el análisis del rendimiento académico de estudiantes de Ingeniería”.

8.1 Recursos Humanos

Para el desarrollo de la presente propuesta de TT se necesita personal docente universitario y un estudiante investigador (autor de la propuesta). En base a lo descrito en [33] se procede a calcular los salarios.

Tabla 8.1: Recursos humanos

TALENTO HUMANO			
Recurso	Horas	Precio	Total
Director de TT	40	\$ 14.77	\$ 590.8
Tutor académico	120	\$ 14.77	\$ 1772.4
Estudiante	420	\$ 2.5	\$ 1050
Total:			\$ 3413.2

8.2 Recursos Físicos

Los recursos físicos necesarios para el desarrollo de la presente propuesta son los siguientes:

Tabla 8.2: Recursos físicos

FÍSICOS		
Recurso	Precio	Total
Dispositivos tecnológicos	\$ 2000	\$ 2000
Materiales de oficina	\$ 200	\$ 200
Efectivo para transporte	\$ 200	\$ 200
Total:		\$ 2400

8.3 Recursos Software

Para el desarrollo del presente proyecto se propone utilizar herramientas de código abierto (Open Source):

Tabla 8.3: Recursos de software

SOFTWARE		
Recurso	Precio	Total
Instrumentos para machine learning	\$ 0	\$ 0
Instrumentos para minería de datos	\$ 0	\$ 0
Aplicación para el manejo de hojas de cálculo	\$ 0	\$ 0
Total:		\$ 0

Tabla 8.4: Presupuesto total

PRESUPUESTO TOTAL	
Recursos humanos	\$ 3413.2
Recursos físicos	\$ 2400
Recursos software	\$ 0
Total:	\$ 5813.2
Imprevistos (15%)	\$ 871.98
Presupuesto total	\$ 6685.18

Por lo tanto, la ejecución del presente tendrá un costo aproximado de \$ 6685.18 dólares americanos. En lo que vale aclarar, que el financiamiento del docente y director de tesis corresponden a la UNL y los demás costos serán responsabilidad del investigador.

Bibliografía

- [1] J. Lerner Matíz, F. Cano Rodríguez, S. Moreno Osorio, A. M. Vargas Betancur, E. Flórez Zapata, G. E. Betancur Jiménez, C. D. S. Bolívar Gómez, and D. A. Osorio Suárez, “Rendimiento académico de los estudiantes de Pregrado de la Universidad EAFIT,” *Cuadernos de Investigación*, vol. 91, pp. 1–189, 2012.
- [2] D. A. Becerra, “Estudio del rendimiento académico aplicando técnicas de Minería de Datos,” *Dspace UNL*, 2015.
- [3] E. A. Carvajal, N. A. Cabezas, S. G. Soledispa, M. S. Chóez, and V. M. García, “La deserción estudiantil en los procesos de nivelación de la educación superior,” 2018.
- [4] K. Josephson, R. Francis, and S. Jayaram, “Políticas para promover la culminación de la educación media en América Latina y el Caribe. Lecciones desde México y Chile,” *Banco de Desarrollo de América Latina*, aug 2018.
- [5] M. M. Ferreyra, C. Avitabile, J. B. Álvarez, F. H. Paz, and S. Urzúa, “La educación superior en América Latina y el Caribe,” p. 59, 2017.
- [6] R. GARIJO-RODENAS, “Análisis de los factores determinantes del bajo rendimiento académico en educación secundaria obligatoria en la provincia de Albacete,” *UNED*, 2016.
- [7] Q. INGRITH, “Análisis de las causas de deserción Universitaria,” *IOSR Journal of Economics and Finance*, vol. 3, no. 1, p. 56, 2016.
- [8] N. S. Ortiz, X. F. Trujillo, and S. E. Tello, “Análisis de la deserción estudiantil en la carrera de ingeniería en telecomunicaciones para proponer nueva metodología de enseñanza,” *Conferencia Iberoamericana en Sistemas, Cibernética e Informática*, vol. 2, pp. 177–182, 2018.
- [9] LOES, “Ley Organica De Educacion Superior, Loes,” www.ces.gob.ec/, 2018.
- [10] E. R. Sinchi and G. P. Gómez Ceballos, “Acceso y deserción en las universidades. Alternativas de financiamiento,” *Alteridad*, vol. 13, no. 2, pp. 274–287, 2018.
- [11] E. V. García, R. C. Rodríguez, S. C. Tumbaco, Y. P. Santana, R. L. Fernández, and L. B. Benítez, “Factores que influyen en la deserción de los estudiantes en

- la Universidad de Guayaquil,” *Revista de la Facultad de Ciencias Médicas de la Universidad de Guayaquil*, vol. 0, no. 0, 2017.
- [12] X. F. O. Diego del Alcázar León, Edgar Silva, “Sistema inteligente para perfilar la deserción en estudiantes universitarios de carreras técnicas,” *Cuadernos del Contrato Social por la Educación*, vol. 10, pp. 50–65, 2014.
- [13] A. Peña Fernández, “Los factores pedagógicos influyen en la deserción universitaria,” *INNOVA Research Journal*, vol. 4, no. 3, pp. 108–115, 2019.
- [14] P. P. Motorik, H. Anak, U. Dini, M. Kegiatan, M. Di, T. Kanak-Kanak, P. Agam, and L. Indraswari, “Análisis de la deserción estudiantil en carreras de ciencias técnicas de la Universidad Regional Autónoma de los Andes (UNIANDES),” *Jurnal Pesona PAUD*, vol. 1, no. 1.
- [15] A. E. García, “Estilos de aprendizaje y rendimiento académico,” *Revista Boletín Redipe*, vol. 7, no. 7, pp. 218–228, 2018.
- [16] Dennis, “Predicción del rendimiento académico de los estudiantes de la unsaac a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático,” 2019.
- [17] G. Castro Aristizabal, G. Giménez, and D. Pérez Ximénez-De-Embún, “Estimación de los factores condicionantes de la adquisición de competencias académicas en América Latina en presencia de endogeneidad,” *Cepal Review*, vol. 2018, no. 124, pp. 36–59, 2018.
- [18] M. F. B. Preciado, “Deserción universitaria en la carrera de Ingeniería Comercial de la PUCESE,” 2015.
- [19] L. D. C. P. GUEVARA, “Estudio de las causas que generan la deserción académica de los estudiantes de la escuela de psicología: caso PUCESA,” 2014.
- [20] E. Yamao, “Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la escuela profesional de ingeniería de computación y sistemas, universidad de san martín de porres, Lima-Perú,” p. 158, 2014.
- [21] Á. P. Albarracín and D. A. Montoya, “Programas de intervención para estudiantes universitarios con bajo rendimiento académico,” *Informes psicológicos*, vol. 16, no. 1, pp. 13–34, 2015.
- [22] A. B. Rodriguez, J. Espinoza, L. J. Ramirez, and A. Ganga, “University dropout: New analysis methodology,” *Formacion Universitaria*, vol. 11, pp. 107–118, dec 2018.
-

-
- [23] S. Cortés Cáceres, P. Alvarez, M. Llanos Ortiz, and L. Castillo Collado, "CI Deserción universitaria: La epidemia que aqueja a los sistemas de educación superior," *Revista Perspectiva*, vol. 20, no. 1, pp. 13–25, 2019.
- [24] D. E. Gallardo-Corrales, "Análisis de patrones de deserción estudiantil de la Unidad Educativa Lenin School aplicando minería de datos," *ESPE*, 2017.
- [25] M. Zaffar, M. A. Hashmani, and K. S. Savita, "Performance analysis of feature selection algorithm for educational data mining," *IEEE Conference on Big Data and Analytics, ICBDA 2017*, vol. 2018-Janua, pp. 7–12, feb 2018.
- [26] C. Jalota and R. Agrawal, "Analysis of Educational Data Mining using Classification," *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India*, pp. 243–247, feb 2019.
- [27] G. E. Cordero, "Proceso para el descubrimiento de conocimiento en las bases de datos de la universidad de cuenca mediante técnicas de data mining," *UNIVERSIDAD DE CUENCA*, p. 89, 2017.
- [28] H. E. Escobar, M. A. Saltos, M. De la Plata, and C. E. Escobar, "Implementación de minería de datos en la gestión académica de las instituciones de educación superior," 2016.
- [29] J. Díaz, L. Lanzarini, M. E. Charnelli, G. Baldino, A. Schiavoni, and P. Amadeo, "Analítica del Aprendizaje y la personalización de la Educación," *Repositorio institucional de la UNLP*, 2015.
- [30] C. Romero, A. Zafra, E. Gibaja, M. Luque, and S. Ventura, "Predicción del rendimiento académico en las nuevas titulaciones de grado de la EPS de la Universidad de Córdoba," 2012.
- [31] T. Devasia, T. P. Vinushree, and V. Hegde, "Prediction of students performance using Educational Data Mining," *Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016*, pp. 91–95, oct 2016.
- [32] A. I. González-Pineda, "Identificación de factores en la reprobación y deserción mediante técnicas de minería de datos en el Área de la Energía de la Universidad Nacional de Loja," *Dspace UNL*, 2014.
- [33] M. del Trabajo, "Acuerdo ministerial No. MDT-2015-0226," <http://www.trabajo.gob.ec/>, 2015.
-

Lista de Acrónimos y Abreviaturas

CACES	Consejo de Aseguramiento de la Calidad de la Educación.
EDM	Minería de datos educativa.
IES	Institución de Educación superior.
LALA	Learning Analytics Latin America.
LOES	Ley Orgánica de Educación superior.
ML	Machine Learning.
ODS	Objetivos de Desarrollo Sostenible.
PIB	Producto interno bruto.
PTT	Propuesta de Trabajo de titulación.
RA	Rendimiento académico.
SE	Sistema educativo.
SGA	Sistema de gestión académica.
TIC	Tecnologías de Información y Comunicación.
TT	Trabajo de Titulación.
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura.
UNL	Universidad Nacional de Loja.
UTI	Unidad de Telecomunicaciones e Información.