UNIVERSIDAD NACIONAL DEL SANTA

FACULTAD DE INGENIERIA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

MODELO DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTUDIANTES DE ACUERDO A SU RENDIMIENTO ACADÉMICO EN EL CENTRO DE IDIOMAS DE LA UNIVERSIDAD NACIONAL DEL SANTA

Tesis para Optar el Título Profesional de Ingeniero de Sistemas e Informática

TESISTAS:

- **BACH. GIANIRA XIOMARA ESPINOZA AIRAC**
- **BACH. EDUAR FABIÁN LEON MUÑOZ**

ASESOR:

Ms. MIRKO MARTÍN MANRIQUE RONCEROS

Nvo. Chimbote - PERÚ

2020

UNIVERSIDAD NACIONAL DEL SANTA

FACULTAD DE INGENIERIA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

MODELO DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTUDIANTES DE ACUERDO A SU RENDIMIENTO ACADÉMICO EN EL CENTRO DE IDIOMAS DE LA UNIVERSIDAD NACIONAL DEL SANTA

Tesis para Optar el Título Profesional de Ingeniero de Sistemas e Informática

Revisado y Aprobado por Asesor:

Ms. Mirko Martín Manrique Ronceros

Asesor

Nvo. Chimbote - PERÚ

2020

UNIVERSIDAD NACIONAL DEL SANTA

FACULTAD DE INGENIERIA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

MODELO DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTUDIANTES DE ACUERDO A SU RENDIMIENTO ACADÉMICO EN EL CENTRO DE IDIOMAS DE LA UNIVERSIDAD NACIONAL DEL SANTA

Tesis para Optar el Título Profesional de Ingeniero de Sistemas e Informática

Revisado y Aprobado por el Jurado Evaluador:

Dr. Juan Pablo Sánchez Chávez
Presidente

Ms. Mirko Martín Manrique Ronceros
Secretario

Ms. Carlos Alfredo Gil Narváez

Integrante

Ms. Pedro Glicerio Manco Pulido

Accesitario

FACULTAD DE INGENIERIA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

ACTA DE EVALUACIÓN PARA SUSTENTACIÓN DE TESIS

En el Campus Universitario de la Universidad Nacional del Santa, siendo las 5:00 p.m. del día jueves 16 de diciembre de 2019, en el Aula S3 del Pabellón nuevo de la EPISI, en atención a la Resolución Decanal N° 780-2019-UNS-FI de Declaración de Expedito de fecha 18.12.19; se llevó a cabo la instalación del jurado Evaluador, designado mediante Resolución N° 601 – 2019 -UNS-CFI de fecha 10.12.2019, integrado por el **Dr. Juan Pablo Sánchez Chávez (Presidente), Ms. Mirko Martín Manrique Ronceros (Secretario), Ms. Carlos Alfredo Gil Narváez (Integrante),** para dar inicio a la sustentación del Informe Final de Tesis, cuyo título es: **MODELO DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTUDIANTES DE ACUERDO A SU RENDIMIENTO ACADÉMICO EN EL CENTRO DE IDIOMAS DE LA UNIVERSIDAD NACIONAL DEL SANTA,** perteneciente a la bachiller: **GIANIRA XIOMARA ESPINOZA AIRAC,** código de matrícula **Nº 0201214033**; tiene como **ASESOR** al **Ms. Mirko Martín Manrique Ronceros,** según **T/R. D. N° 107 – 2019 -UNS-FI** de fecha 04.04.2019.

Terminada la sustentación, la tesista respondió a las preguntas formuladas por los miembros del Jurado Evaluador y el público presente.

El Jurado después de deliberar sobre aspectos relacionados con el trabajo, contenido y sustentación del mismo y con las sugerencias pertinentes y en concordancia con el artículo 73º y 103º del Reglamento General de Grados y Títulos, vigente de la Universidad Nacional del Santa; considera la siguiente nota final de Evaluación:

BACHILLER	CALIFICACIÓN	CONDICIÓN
GIANIRA XIOMARA ESPINOZA AIRAC	17	MUY BUENO

Siendo la 6: 00 p.m. se dio por terminado el Acto de Sustentación y en señal de conformidad, firma el Jurado la presente Acta.

Nuevo Chimbote, 19 de diciembre de 2019

Dr. JUAN PABLO SÁNCHEZ CHÁVEZ
PRESIDENTE

MS. MIRKO MARTÍN MANRIQUE RONCEROS
SECRETARIO

MS. CARLOS ALFREDO GIL NARVÁEZ
INTEGRANTE

FACULTAD DE INGENIERIA

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

ACTA DE EVALUACIÓN PARA SUSTENTACIÓN DE TESIS

En el Campus Universitario de la Universidad Nacional del Santa, siendo las 5:00 p.m. del día jueves 19 de diciembre de 2019, en el Aula S3 del Pabellón nuevo de la EPISI, en atención a la Resolución Decanal N° 780-2019-UNS-FI de Declaración de Expedito de fecha 18.12.19; se llevó a cabo la instalación del jurado Evaluador, designado mediante Resolución N° 601 – 2019 -UNS-CFI de fecha 10.12.2019, integrado por el **Dr. Juan Pablo Sánchez Chávez (Presidente), Ms. Mirko Martín Manrique Ronceros (Secretario), Ms. Carlos Alfredo Gil Narváez (Integrante)**, para dar inicio a la sustentación del Informe Final de Tesis, cuyo título es: **MODELO DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTUDIANTES DE ACUERDO A SU RENDIMIENTO ACADÉMICO EN EL CENTRO DE IDIOMAS DE LA UNIVERSIDAD NACIONAL DEL SANTA**, perteneciente al bachiller: **EDUAR FABIAN LEÓN MUÑOZ,** código de matrícula **Nº 0201014047**; tiene como **ASESOR** al **Ms. Mirko Martín Manrique Ronceros,** según **T/R. D. N° 107 – 2019 -UNS-FI** de fecha 04.04.2019.

Terminada la sustentación, el tesista respondió a las preguntas formuladas por los miembros del Jurado Evaluador y el público presente.

El Jurado después de deliberar sobre aspectos relacionados con el trabajo, contenido y sustentación del mismo y con las sugerencias pertinentes y en concordancia con el artículo 73º y 103º del Reglamento General de Grados y Títulos, vigente de la Universidad Nacional del Santa; considera la siguiente nota final de Evaluación:

BACHILLER	CALIFICACIÓN	CONDICIÓN
EDUAR FABIAN LEÓN MUÑOZ	17	MUY BUENO

Siendo la 6: 00 p.m. se dio por terminado el Acto de Sustentación y en señal de conformidad, firma el Jurado la presente Acta.

Nuevo Chimbote, 19 de diciembre de 2019

Dr. JUAN PABLO SÁNCHEZ CHÁVEZ
PRESIDENTE

MS. MIRKO MARTÍN MANRIQUE RONCEROS
SECRETARIO

MS. CARLOS ALFREDO GIL NARVÁEZ
INTEGRANTE

DEDICATORIA

A Dios por la vida y todas las oportunidades que me ha regalado, gracias a las cuales he podido mejorar en lo personal y profesional. Por las personas que pone en mi camino y con quienes he comprendido que la familia se puede escoger.

A mi madre Maximina Airac Chauca, con el amor más grande que tengo en el corazón, le dedico este trabajo, por representar en mi vida el amor más grande, una entrega desinteresada y el anhelo constante de superación. Por tanto y todo, esta tesis es tuya mamá.

A mi hijo, por llegar en el momento más importante, por convertirse en mi luz al final del túnel y ser la razón más importante para buscar mi superación profesional y personal. Al compañero que llega para quedarse y acompañarme en las buenas, en las malas y en las peores etapas de mi vida.

A mi prometido, Noé por iniciar conmigo este trabajo y celebrar siempre mis triunfos, por demostrar siempre su interés en ayudarme a lograr mis metas.

Gianira

DEDICATORIA

Dedico esta tesis a Dios por regalarme la vida y darme sabiduría.

A mis padres Manuel y Haydee por ser mi soporte, por sus consejos y sobre todo por inculcarme buenos principios y valores.

A mi novia Judith por impulsarme a seguir siempre adelante, por su paciencia y comprensión.

A todas las personas que me apoyaron a realizar esta tesis

Eduar

AGRADECIMIENTO

En el desarrollo de ésta investigación hemos tenido mucho apoyo incondicional de personas, para las cuales queremos dejar constancia de nuestro agradecimiento:

Agradecemos a Dios por concedernos cada día vida y salud, sin él nada fuera posible.

Al Centro de Idiomas de la Universidad Nacional del Santa en especial; a la Alta Gerencia por tu valiosa colaboración y siempre la buena disposición de todo el personal en facilitarnos la información que solicitamos.

A nuestro asesor Ms. Mirko Manrique Ronceros, por su orientación, apoyo y supervisión permanente, por compartir con nosotros su experiencia profesional y sus conocimientos.

Al Dr. Juan Pablo Sánchez Chávez, quien preside el jurado evaluador de esta tesis y de quien hemos recibido las recomendaciones necesarias para optimizar el resultado de nuestro trabajo.

A la Escuela Profesional de Ingeniería de Sistemas e Informática de la Universidad Nacional del Santa, por la formación profesional que recibí durante el curso de mi carrera.

A nuestros buenos amigos, por los buenos consejos y sus buenos deseos. Por celebrar con nosotros nuestros logros y acompañarnos en los momentos difíciles. En el mundo aún existen personas que valen la pena conservar en la vida.

Bach. Gianira Espinoza Airac y Bach. Eduar León Muñoz

INDICE

INDICE DE FIGURAS

INDICE DE TABLAS

RESUMEN

La presente Tesis denominada Modelo de Machine Learning para la clasificación de estudiantes de acuerdo a su rendimiento académico en el Centro de Idiomas de la Universidad Nacional del Santa, tiene como objetivo central mejorar el proceso de clasificar a los estudiantes del centro de idiomas utilizando Aprendizaje Automático, centrándose en diferentes niveles, para así cumplir el objetivo principal del Centro de Idiomas que es el de brindar enseñanza de un idioma extranjero o nativo. Gracias a las nuevas tecnologías de Información, y en especial a los sistemas inteligentes, es posible obtener predicciones de clasificación que podrían pasar a futuro en base a los registros de datos históricos; y con ello desarrollar una solución que plasme esta información y sirva como herramienta de conocimiento en el proceso de clasificación de estudiantes según su rendimiento académico en el CEIDUNS.

Como resultado se obtuvo la reducción del Tiempo Promedio de clasificación de los estudiantes según su rendimiento académico en un 74.60% (de 218.19 segundos a 55.42 segundos); también el aumento del número de clasificaciones acertadas en un 82.08%), y por último en cuanto al nivel de satisfacción del personal docente, se aumentó en un 71.35% y del estudiante en un 66.30% utilizando el modelo predictivo.

El modelo de Machine Learning facilitó al personal del CEIDUNS en la identificación de estudiantes, asignación de aulas e incrementando el número docentes. El sistema de predicción de clasificación cumplió con su objetivo principal (de mejora del proceso).

Palabras claves: Machine Learning, Aprendizaje Automático, Modelo, Algoritmo, idiomas.

ABSTRACT

This thesis called Machine Learning Model for the classification of students according to their academic performance in the Language Center of the National University of Santa, has as its main objective to improve the process of classifying students of the language center using Machine Learning, focusing on different levels, in order to fulfill the main objective of the Language Center, which is to provide teaching of a foreign or native language. Thanks to new information technologies, and especially intelligent systems, it is possible to obtain classification predictions that could happen in the future based on historical data records; and thereby develop a solution that captures this information and serves as a knowledge tool in the process of classifying students according to their academic performance at CEIDUNS.

As a result, the reduction of the Average Classification Time of the students was obtained according to their academic performance by 74.60% (from 218.19 seconds to 55.42 seconds); also the increase in the number of successful classifications by 82.08%), and finally in terms of the level of satisfaction of the teaching staff, it was increased by 71.35% and the student by 66.30% using the predictive model.

The Machine Learning model facilitated CEIDUNS staff in identifying students, assigning classrooms and increasing the number of teachers. The classification prediction system met its main objective (process improvement).

Keywords: Machine Learning, prediction system, Model, Algorithms, languages

PRESENTACION

Señores miembros del Jurado Evaluador:

En cumplimiento a lo dispuesto por el Reglamento General de Grados y Títulos de la Universidad Nacional del Santa, ponemos a vuestra consideración el presente informe de Tesis titulado: **"MODELO DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTUDIANTES DE ACUERDO A SU RENDIMIENTO ACADÉMICO EN EL CENTRO DE IDIOMAS DE LA UNIVERSIDAD NACIONAL DEL SANTA"** como, requisito para optar el Título Profesional de Ingeniero de Sistemas e Informática

El objetivo de esta investigación es permitir al Centro de Idiomas de la Universidad Nacional del Santa, realizar la clasificación de los estudiantes según su rendimiento académico, además de contar con una plataforma tecnológica y tener una mejor toma de decisiones. Para esta investigación se utilizó la metodología CRISP, las herramientas y plataformas tecnológicas que permitieron resolver la problemática planteada.

Por lo expuesto, a ustedes señores miembros del jurado evaluador, presentamos nuestra tesis, para su revisión, esperando cumpla con los requisitos mínimos para su aprobación

Atentamente,

Los Autores

INTRODUCCIÓN

La minería de datos, como tecnología para analizar grandes volúmenes de datos orientado a la predicción de sucesos, ha sido muy utilizada en los últimos años (Caridad, 2014) en problemas de índole académico, como por ejemplo en predicción de deserción estudiantes, del rendimiento académico, etc.

En la presente tesis se ha realizado un análisis de datos en base a los registros históricos de las matrículas realizadas en el año 2018 para poder crear un algoritmo predictivo dle rendimiento académico que se pueda visualizar en una aplicación informática intuitiva para el personal de CEIDUNS y así ellos puedan mejorar sus tomas de decisiones en lo referido a la clasificación de estudiantes.

El presente informe está estructurado en seis capítulos, cada uno de los cuales se detallan a continuación:

EL CAPITULO I, presenta hace una descripción general del Centro de Idiomas de la Universidad Nacional del Santa.
EL CAPITULO II, describe el Plan de tesis especificando la realidad del problema, el enunciado del problema del proyecto, se plantea la hipótesis, se describe también los objetivos generales y específicos, la justificación, antecedentes e importancia del trabajo.
EL CAPITULO III, plasma el Marco Teórico necesario para el desarrollo de la tesis, describiendo los conceptos teóricos, Metodología y las Herramientas tecnológicas usados para el desarrollo de la aplicación.
El CAPITULO IV, trata del desarrollo de la metodología CRISP el cual contempla cada una de sus fases para el modelo de Machine Learning.
El CAPITULO V, trata de los Materiales y Métodos donde se realiza Contrastación de la hipótesis y se muestran los resultados obtenidos.
El CAPITULO VI, trata de la Discusión y Resultados de la Tesis.

Finalmente se hace mención a las Conclusiones y Recomendaciones finales del estudio realizado.

DATOS GENERALES DEL ESTUDIO

TITULO DEL PROYECTO

MODELO DE MACHINE LEARNING PARA LA CLASIFICACIÓN DE ESTUDIANTES DE ACUERDO A SU RENDIMIENTO ACADÉMICO EN EL CENTRO DE IDIOMAS DE LA UNIVERSIDAD NACIONAL DEL SANTA

TESISTAS

- ✓ BACH. EDUAR FABIAN LEON MUÑOZ
- ✓ BACH. GIANIRA XIOMARA ESPINOZA AIRAC

ASESOR

Ms. Mirko Martin Manrique Ronceros

TIPO DE INVESTIGACIÓN

a) Según su Naturaleza:

Pre - Experimental: Esta investigación es Pre - Experimental por motivos que se solo se analiza una sola variable o grupo, porque se levantará la información que necesitamos en un rango de tiempo específico, que luego será procesado y validado. No se manipulará la información de la variable o grupo.

Para esta investigación realizaremos un Pre – test y un Post – test, para poder asegurar la validez de la encuesta o cuestionario. Se medirán los resultados obtenidos en el proceso de clasificación de estudiantes (la población serán los estudiantes de idiomas), que se realizará en el mes de diciembre del 2018, con el fin de detectar valores imprevistos de las variables planteadas.

b) Según su fin o propósito:

Aplicada: Porque permite establecer la relación causal entre el sistema actual y el método de predicción.

Esta investigación es de tipo aplicada por el motivo que lograremos dar solución a un problema mediante un algoritmo de machine learning (variable independiente) y así lograr mejorar la clasificación de estudiantes (variable dependiente).

METODO DE INVESTIGACION

Es **Inductivo – Deductivo** porque luego de definir la realidad problemática se planteó una hipótesis y se realizó las observaciones respectivas, con las cuales se planteó el desarrollo de algoritmo de machine learning.

CAPÍTULO I
LA EMPRESA

1.1. RESEÑA HISTÓRICA

Según OCEI (2007) "El CEIDUNS nace por decisión de las tres docentes de la especialidad de Idiomas de la Facultad de Educación y Humanidades quienes mostraron un plan de funcionamiento y un estudio de mercado indicando la demanda insatisfecha de la enseñanza del idioma inglés en forma libre"

En el año 1991 en la ciudad de Chimbote únicamente poseía cinco academias de inglés como John F Kennedy, Daniel Alcides Carrión, Abraham Valdelomar, IDICOMP y SIDERPERU los cuales en total únicamente tenían capacidad para una media de 936 alumnos que equivalía al 0,4% de la población de Chimbote."

"La Resolución Nº 197-91-UNS del 10 de septiembre del año 1991 autoriza del funcionamiento del Centro de Idiomas a partir del primero de octubre de 1991."

Por un año y sin remuneración alguna la Dirección del CEIDUNS estuvo a cargo de la Prof. Lila Maguiña Alvarado (octubre 1991-octubre 1992). Luego le sucedió la Prof. Susana Gutiérrez Saldaña (noviembre 1992 - 1994), luego continuaron la Prof. Janna Tolentino (período 1995-1996), la Prof. Susana Gutiérrez Saldaña (período 1997-1998), Prof. Betty Risco Rodríguez (período 1999-2000), Prof. Lila Maguiña Alvarado (período 2001-2003), Prof. Susana Gutiérrez Saldaña (período 2004-2005) Prof. Lila Maguiña Alvarado (año 2006), Prof. Betty Risco Rodríguez, (2007), Prof. Juan Martínez Guillén (2008-mitad del 2010, Prof. Susana Gutiérrez Saldaña (mitad del 2010), Prof. Hermes Lozano Alvarado (2011), Prof James Solis Godoy (2012-abril 2013), Prof. Esmila Calderón Reyes (mayo 2013- diciembre 2013), Prof Rosendo Daniel Ramos (mayo 2014- 13 abril 2014), Prof Lila Maguiña Alvarado (desde 16 abril 2014).

El CEIDUNS cuenta con un local propio de dos plantas con 11 aulas, hay un ambiente para la biblioteca y otro para la sala de profesores. Tiene una capacidad por turno de 330 alumnos y una proyección de atención máxima de 2310 alumnos si funcionaran los siete turnos diarios. Durante el año 2014 se han registrado 13 407 matrículas, con un promedio de 1117 alumnos al mes.

1.2. PERFIL DE LA EMPRESA

1.2.1. Datos Generales de la Empresa

1.2.1.1. Razón Social

Centro de Idiomas (CEIDUNS)

1.2.1.2. Domicilio Legal

Urb. Bellamar del Distrito de Nuevo Chimbote dentro de la Universidad Nacional del Santa

Figura N° 01: Ubicación Geográfica

Fuente: Google Maps

1.2.2. Actividad de la Empresa

Educación

1.2.3. RUC

20148309109

1.2.4. Logotipo

Figura N° 02: Logotipo de la Organización

1.3. VALORES

Los valores que siempre se han promovido desde su creación 1991 son:

- ✓ Puntualidad
- ✓ Cumplimiento
- ✓ Perseverancia.
- ✓ Honestidad
- ✓ Respeto

1.4. FUNCIONES DEL CENTRO DE IDIOMAS

Según Ceidum las funciones de Centro de idiomas son las siguientes:

- ✓ "Proyectar, constituir, destinar e inspeccionar las acciones del Centro de Idiomas"
- ✓ "Vigilar la asistencia de servicios consignados a la acción de recursos del Centro de Idiomas."

- ✓ “Gestionar los recursos humanos, materiales y económicos del Centro de Idiomas.”
- ✓ “Plantear y proveer oportuna inversión.”

1.5. BASE LEGAL

Según Ceidum Su base legal es:

- ✓ “Ley General de Educación N° 28044.”
- ✓ “Ley Universitaria N° 30222.”
- ✓ “Estatuto de la Universidad Nacional del Santa, aprobado mediante Resolución Presidencial Nº 518-98-UNS”
- ✓ “Reglamento de Organización y Funciones de la Universidad Nacional del Santa, aprobado mediante Resolución Nº 071-2001-CU-R-UNS y sus modificatorias Nº 175-2001 y 109-2002-CU-R-UNS”
- ✓ “Proyecto de Funcionamiento del Centro de Idiomas de la Universidad Nacional del Santa (CEIDUNS), aprobado mediante Resolución Rectoral Nº 197-91-UNS-R.”

1.6. LA ESTRUCTURA ORGÁNICA DEL CENTRO DE IDIOMAS

- **Órgano de Dirección:** Dirección del Centro de Idiomas
- **Órgano de Coordinación:** Coordinación General
- **Órgano de Apoyo:** Centro de Logística, Oficina de Secretaria y de Servicios y Unidad de Información y Documentación.
- **Órgano de Línea:** Plana Docente

1.7. ORGANIGRAMA

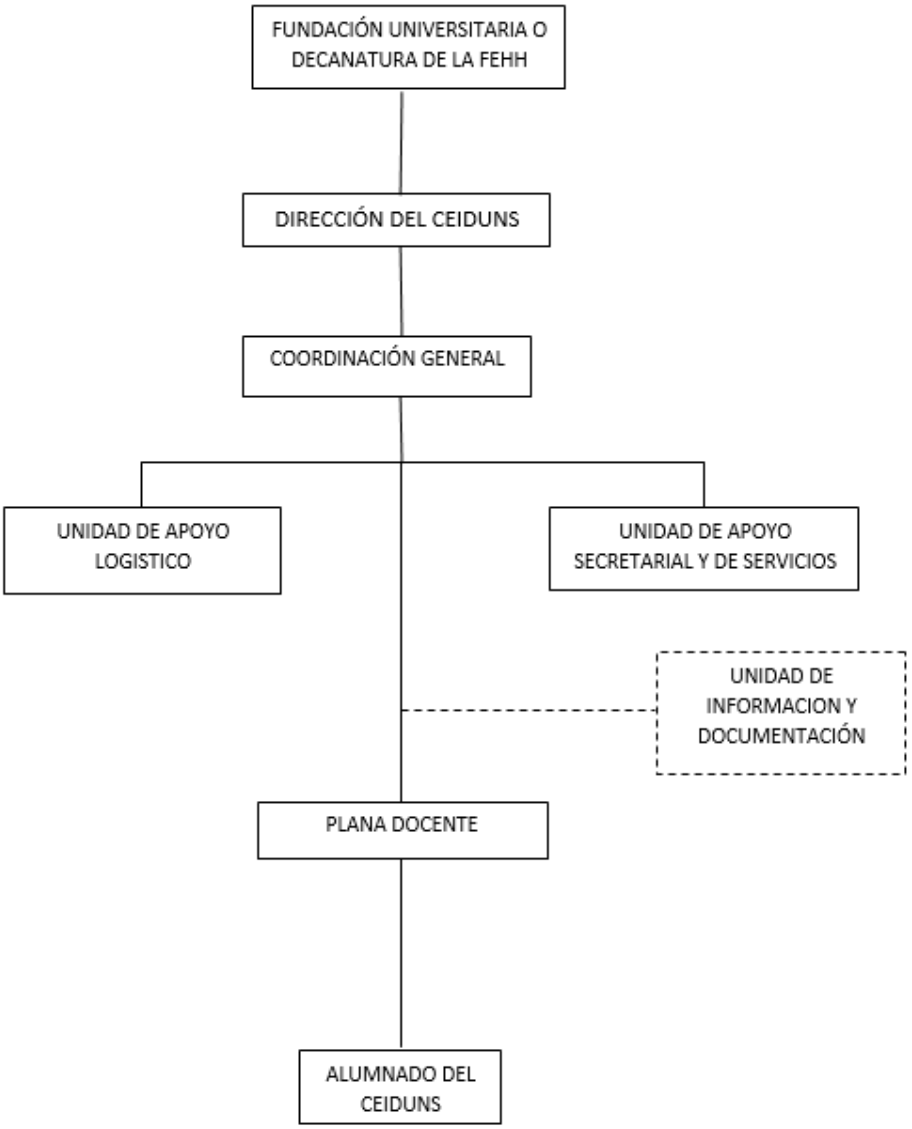


Figura N° 03: Organigrama de Ceiduns

1.7.1. Directivos

El Centro de Idiomas es conducido por un Director que ha sido favorecidos por el Consejo de Facultad a consejo del Decano y cuyos requisitos son: profesor principal o coligado a asignación exclusiva o período completo. Su mandamiento dura dos años, pudiendo ampliarse a un periodo más. Desde su funcionamiento 1991 hasta el año 2010 el Director fue un profesor de la especialidad de Idiomas. Desde el 2011 hasta el 15 de abril del 2014 la Dirección estuvo a cargo de dos profesores

de lengua y dos profesores de Comunicación Social. A partir del 16 de abril del 2014 la Dirección del CEIDUNS está bajo la conducción de un profesor de la especialidad de Idiomas

1.7.2. Profesores

El CEIDUNS ha cuenta con 25 profesores de inglés, 01 profesor de portugués, 01 profesora de italiano, 02 profesores de francés y 02 profesoras de quechua, totalizándose 31 profesores.

1.7.3. Alumnos

- ✓ Son estudiantes del CEIDUNS, las personas a partir de los tres años de edad y que desean aprender un segundo idioma. Es decir que los alumnos son aquellos que asisten al grado inicial, grado primario, grado secundario, grado superior y toda la comunidad en general.

CAPÍTULO II

PLANTEAMIENTO DEL PROBLEMA

2.1. REALIDAD DEL PROBLEMA

Según Paredes Dulce (2015) afirma: “Hoy en día el lenguaje inglés es el idioma más hablado al nivel mundial. El motivo principal del por qué es importante, es que cada día se emplea más en casi todas las áreas de conocimiento y desarrollo humano, esto tiene varias implicaciones sobretodo en el mundo laboral, de los negocios, la informática u otros fines, haciendo uso así de esta herramienta tan útil como lo es el idioma inglés, por lo expuesto, tener dominio del idioma inglés es muy positivo para un currículum laboral o cualquiera que sea la necesidad, en nuestro país y también en el extranjero, porque se trata de un lenguaje al nivel mundial más hablado. Como se ha podido observar la validación internacional del aprendizaje de este idioma es esencial para ascender en desarrollo profesional o laboral es decir necesitamos una certificación de valor internacional donde nos permita demostrar el dominio de la competencia comunicativa en el idioma inglés y que además respalde que nuestro aprendizaje ha sido el óptimo y que el estudiante está en condiciones de usar el idioma en cualquier situación que se presente.”

Según Camborda (2014) afirma que: “El rendimiento académico es un claro indicador del avance exitoso en la carrera de estudios de un estudiante en un momento particular, y a su vez también es un pronosticador de la posibilidad de completar exitosamente una carrera de estudios.” “El término "rendimiento" tiene muchas implicancias, principalmente si se considera a las notas obtenidas por los alumnos como el referente casi exclusivo. Esta pesquisa puede crear, incluso una lección inexperta, que se concentra sólo en el trabajo académica en el alumno. Sin embargo, la responsabilidad institucional es clave para evaluar lo que se entiende por rendimiento.”

En el centro de Idiomas de la Universidad Nacional del Santa (CEPUNS) existe un alto porcentaje de alumnos que desaprueban los cursos de idiomas, tal es el caso de los estudiantes nuevos que proceden de otras universidades, institutos y colegios; también de personas adultas o alumnos que retoman sus estudios y llegan a repetir uno, dos e incluso tres veces el mismo ciclo o estudiantes que desean convalidar sus estudios de algún idioma extranjero para matricularse en un nivel superior al que ya tienen cursado en otro centro de idiomas. Es preocupante saber que cada ciclo este porcentaje se va incrementando. Diversos indicadores expresan de manera obvia, preocupante y permanente que no existe un adecuado proceso de clasificación para los estudiantes, de tal manera que el bajo nivel académico que se suscita por este problema conlleva a que en el siguiente ciclo que tienen que cursar, haya grupos sobresaturados y grupos totalmente vacíos, porque no existe una herramienta de clasificación de estudiantes con similar nivel académico que permita prever este tipo de escenarios y proyectar planes de contingencia. Considerando que el idioma inglés (que es el idioma donde se refleja más la problemática) en cualquier sea el nivel, tiene 12 ciclos en modalidad regular y 06 ciclos en modalidad intensivo, además que cada ciclo se desarrolla en un mes y que todos los meses se inician nuevos grupos, es decir inicia un primer ciclo de cada nivel.

El no contar con un sistema de clasificación de estudiantes que permita prever el índice de estudiantes aprobados y desaprobados cada ciclo, conlleva a enfrentar todos los meses una realidad incierta donde se tiene: exceso de demanda para algunos ciclos, déficit de inscripciones para otros, estudiantes que no pueden acceder al servicio educativo y deben esperar para ello, además de estudiantes que optan por abandonar sus estudios.

De continuarse presentando esta problemática en el Centro de Idiomas de la UNS, se verán afectados los estudiantes que eligen estudiar en dicha institución, además de verse comprometida la rentabilidad económica que todo centro de producción debe tener para cubrir sus gastos y continuar sus actividades.

Por esta razón esta investigación tiene como propósito un "Modelo de Machine Learning para la clasificación de estudiantes de acuerdo a su rendimiento académico en el Centro de Idiomas de la Universidad Nacional del Santa".

2.2. ANALISIS DEL PROBLEMA

A. Pocas posibilidades de obtener una vacante

- Causa

  Al no contar con un mecanismo de clasificación de estudiantes, mensualmente se obtiene una cantidad incierta de estudiantes desaprobados que tienen que repetir el ciclo y buscarán una vacante para dicho ciclo. Sin embargo, por ciclo hay un límite de vacantes y no todos los estudiantes que repitieron de ciclo alcanzan a matricularse.

- Consecuencia

  La administración del Centro de Idiomas no puede prever cuantos alumnos podrían quedarse sin vacantes, por lo que no dispone la apertura de nuevos grupos, pues desconoce cuál serán los ciclos de mayor demanda el siguiente mes. Además, se ve afectada la continuidad de los estudios de quienes no alcanzaron a matricularse por falta de vacantes.

B. Ciclos que no se apertura

- Causa

  Debido a que en algunos ciclos más de la mitad de los estudiantes desaprobaron y no se logra cubrir el número mínimo de inscripciones, dichos ciclos no se apertura y los estudiantes que se hayan matriculado deberán esperar a que los ciclos menores avancen para poder incluirse.

- Consecuencia

  Se genera descontento en los estudiantes aprobados que se ven perjudicados por el cierre de sus ciclos.

C. Alumnos que retoman estudios

- Causa

  Debido a los puntos anteriores, los estudiantes se ven en la necesidad de suspender sus estudios hasta obtener una vacante para el ciclo que desaprobaron, lo cual puede tardar uno, dos, tres y en ocasiones más tiempo, generándose de esta manera un grupo mayor de estudiantes que cada mes acudirá en busca de una vacante para diferentes ciclos.

- Consecuencia

  El hecho de tener que esperar para poder matricularse genera un grave problema en el rendimiento académico de los estudiantes que dejan de estudiar por mucho tiempo.

D. Bajo nivel académico

- Causa

  El desorden que se genera a raíz del incremento de estudiantes que retoman estudios y la problemática de las vacantes, ocasiona que el nivel académico

de los estudiantes al finalizar sus estudios sea menor en comparación de aquellos que no tuvieron que suspender sus clases.

- Consecuencia

  Tanto para el estudiante como para el Centro de Idiomas, el nivel académico de los egresados es la carta de presentación de la institución y un factor importante en la vida de todo profesional, que, de verse perjudicado, los hace menos competitivos.

E. Estudiante que desertan

- Causa

  Al verse afectada la continuidad de los estudios de quienes no alcanzaron una vacante, ocasiona que muchos de ellos que opten por desertar y dejar su preparación académica trunca, pues para cuando encuentran vacantes disponibles ya tienen planeado cumplir con otras actividades.

- Consecuencias

  El prestigio del Centro de Idiomas de la UNS se rige por el nivel académico de sus egresados y no es conveniente tener una población de estudiantes que desertan por una mala administración del servicio educativo, pues esto ocasiona que los competidores (otros centros de idiomas locales) tomen este aspecto como una oportunidad para posicionarse en el mercado.

También podemos mencionar los siguientes problemas generados:

- Imposibilidad de captar con exactitud las condiciones académicas primarias para una predicción.
- Demora en la búsqueda y/o consulta de la clasificación de un estudiante por motivos que se realiza de forma manual.
- Desconocimiento del avance del desempeño académico del estudiante.

- Tiempo de procesamiento largo.
- No existe mecanismos correctivos por parte de la dirección que aporten a mejorar los indicadores de bajo rendimiento académico y abandono.
- No se puede establecer los elementos demográficos, socioculturales, estudiosos, pedagógicos que admitan predecir el provecho universitario de los alumnos.
- Prexiste un grado de porcentaje de estudiantes que desaprueban materias de idiomas principalmente en los primeros ciclos del idioma inglés.
- Reportes no funcionales innecesario a la restringida y pausada indagación adquirida que no admite poseer la sistematización de las medias aritméticas, calificaciones y valoración, de los estudiantes quienes tienen los principales puestos.
- Insatisfacción en el personal administrativo porque tiene demora en los procesos académicos, debido se realizan manualmente.

2.3. ANTECEDENTES DEL PROBLEMA

2.3.1. Antecedentes Internacionales

Tesis 01

Autor : Carlos J. Villagrá Arnedo

Título : Sistema predictivo progresivo de clasificación probabilística como guía para el aprendizaje

Institución : Universidad de Alicante

Grado : Doctor en Ingeniería Informática y Computación

Año : 2015

Resumen u Objetivo

Según el trabajo realizado Villagrá (2015) afirma que “En esta tesis está basado en el desarrollo de un modelo de predicción progresiva que mejora el

proceso de enseñanza-aprendizaje a través del uso de las tecnologías de la información y, en particular, de las técnicas de inteligencia artificial."

"Este modelo tiene como base un sistema interactivo gamificado que gestiona las prácticas de la asignatura Matemáticas I, en las que se aprende razonamiento lógico a través de un videojuego llamado PLMan, muy similar al comecocos (PacMan). Los estudiantes acceden durante el curso a este sistema y van progresando y acumulando nota en las prácticas de la asignatura mediante la resolución de mapas del videojuego PLMan."

Relación:

La tesis de Villagrá obtenida como antecedente de referencia tiene mucho en común con ésta investigación, teniendo como relación que ambas tesis tienen el objetivo de realizar un rendimiento de clasificaciones para poder mostrar información actualizada de aprendizaje disponibles logrando fidelizar a los alumnos.

Tesis 02

Autor : Gisela Isabel García Gazabón

Título : Modelo de Machine Learning para la Clasificación de pacientes en términos del nivel asistencial requerido en una urgencia pediátrica con Área de Cuidados Mínimos

Institución: Universidad Tecnológica de Bolívar

Grado : Maestría en Ingeniería

Año : 2014

Resumen u Objetivo

Según García (2014) afirma que: "La clasificación de las urgencias en las clínicas de la ciudad es un problema que va en aumento, pero como se ha

demostrado durante el desarrollo de esta investigación desde diferentes perspectivas, podemos aportar soluciones tendientes a la mejora continua del proceso de clasificación para el ingreso de los pacientes en un Departamento de Urgencias."

"Un aspecto para destacar, fue la utilización del análisis de componentes principales, que permitió disminuir de 145 variables a 74 componentes principales, esto permitió disminuir la dimensionalidad del conjunto de datos sin poner en riesgo los resultados esperados, además utilizando todos los criterios identificados para evaluar el nivel asistencia de un paciente en la Urgencia."

Relación:

La tesis como referencia de García Gazabon, ayuda a esta investigación a comparar los diferentes tipos de Clasificación de una entidad, como asimismo información respectiva a los interesados inscriptos. Adquiriendo una gran acogida de los alumnos a la hora de ver su clasificación a base de su rendimiento, logrando optimar el tiempo del proceso de gestión de los mismos.

Tesis 03

Autor : Andrés Rico Páez y Daniel Sánchez Guzmán

Título : Diseño de un modelo para automatizar la predicción del rendimiento académico en estudiantes del IPN

Institución: Instituto Politécnico Nacional, México

Grado : Maestría en Ingeniería

Año : 2018

Resumen u Objetivo

Según Rico & Sánchez (2018) afirma: “La minería de datos educativa permite extraer conocimiento útil y comprensible a partir de datos académicos para la solución de problemas acerca de diversos procesos de enseñanza y de aprendizaje. Una de las aplicaciones más populares de la minería de datos educativa es la predicción del rendimiento académico. El principal objetivo de este trabajo fue diseñar y automatizar un modelo predictivo del rendimiento académico de estudiantes del Instituto Politécnico Nacional (IPN).”

“Para la construcción del modelo, se analizaron las calificaciones de actividades académicas y la calificación final de 94 estudiantes inscritos en una carrera de ingeniería perteneciente al IPN. Este modelo se aplicó a 86 estudiantes para predecir su rendimiento académico. Posteriormente, se compararon estas predicciones con los resultados reales obtenidos por los estudiantes al final del curso. Se obtuvieron exactitudes de las predicciones de la aprobación del curso de hasta 73%, únicamente con cinco atributos correspondientes a las calificaciones de las actividades académicas iniciales del mismo.”

Relación:

Esta tesis ayudará a detallar los requerimientos, el diseño, sobre la predicción del rendimiento académico en estudiantes, analizando las notas de las tareas académicas y la nota final.

2.3.2. Antecedentes Nacionales

Tesis 04

Autor : Eiriku Yamao

Título : 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.

Institución: Universidad San Martín de Porres

Grado : Maestro en Ingeniería de Computación y Sistemas

Año : 2018

Resumen u Objetivo

Segun Yomao (2018) afirma: "El rendimiento académico es un tema estudiado desde hace mucho tiempo. Los alumnos ingresantes de las universidades son los más vulnerables a enfrentar problemas de rendimiento, resultando en posible deserción. La minería de datos en educación aplica técnicas de minería de datos en la información generada en el sector educación. El presente trabajo consiste en realizar la predicción del rendimiento académico de los alumnos que ingresaron a la Escuela Profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martin de Porres en el primer ciclo utilizando minería de datos. Se extrajeron datos de 1304 ingresantes que fueron clasificados en tres factores: sociales, económicos y académicos."

Relación:

Esta tesis permitirá tener una visión más clara sobre el proceso de Rendimiento Académico.

Tesis 05

Autor : Jorge Rodríguez Castillo y Milagros Miñano Ochoa

Título : Desarrollo de una aplicación informática basada en un modelo de Mahine Learning para mejorar la evaluación de préstamos crediticios.

Institución : Universidad Privada del Norte

Grado : Ingeniero en Sistemas Computacionales.

Año : 2017

Resumen u Objetivo

Según Rodríguez & Miñano (2017) afirma que: "El presente trabajo de investigación está enfocado en el estudio de un modelo de machine learning para desarrollar una aplicación informática, que permita mejorar la evaluación de préstamos crediticios brindando un mejor análisis de la rentabilidad y el riesgo crediticio; la cual sea usada por la empresa financiera que por motivos de privacidad de sus datos llamaremos Financiera Nuestro Crédito."

"El problema radica en determinar como una aplicación informática basada en un modelo de machine learning contribuye a mejorar la evaluación de préstamos crediticios."

"El problema radica en determinar como una aplicación informática basada en un modelo de machine learning contribuye a mejorar la evaluación de préstamos crediticios."

Relación:

Esta tesis nos permitirá tener una mejorar el modelo de Machine Learning, y tener en cuenta los parámetros para poder realizar una Clasificación de rendimiento Académico más confiable y seguro.

Tesis 06

Autor : Sandra Zoraida Medrano Parado

Título : Modelo de minería de datos usando machine learning con reconocimiento de patrones de síntomas y enfermedades respiratorias en las historias clínicas para mejorar el diagnóstico de pacientes en la ciudad de Trujillo 2016

Institución: Universidad César Vallejo

Grado : Ingeniero en Sistemas

Año : 2016

Resumen u Objetivo

Según Medrano (2016) Dice que: "Esta investigación se desarrolló en la ciudad de Trujillo desde abril hasta diciembre del 2016, centrándose en el problema de diagnósticos médicos y el uso de machine learning para contribuir en la mejora. La metodología que se usó para desarrollar la investigación fue CRISP. El tipo de investigación de acuerdo al fin es aplicado, de acuerdo a la técnica de contrastación es Pre Experimental, la población es la ciudad de Trujillo, los instrumentos usados en la investigación son el cuestionario y el análisis documental."

Relación:

Esta tesis nos permitirá a aumentar el modelo e implementación de Machine learning para contribuir en la mejora de clasificación de rendimiento académico.

2.3.3. Antecedentes Locales

Tesis 07

Autor : Sixto Díaz Tello

Título : Sistema integral bajo el enfoque de minería de datos y redes neuronales para la predicción y control de la contaminación atmosférica por pm10 en la ciudad de Chimbote

Institución: Universidad Nacional del Santa

Grado : Doctor en Ingeniería de Sistemas e Informática

Año : 2014

Resumen u Objetivo

Según Díaz (2014) dice que: "En la presente tesis doctoral se propone la implementación de un sistema integral, que permita realizar la predicción de la contaminación atmosférica por concentración del contaminante crítico PM10 para Chimbote, haciendo uso de redes neuronales y minería de datos. El modelo implementado plantea el uso de las variables meteorológicas y variables estacionales; como factores que influyen en la concentración de los contaminantes."

2.4. FORMULACIÓN DEL PROBLEMA

¿De qué manera un Modelo de Machine Learning logrará la Clasificación de Estudiantes de acuerdo a su Rendimiento Académico en el Centro de Idiomas de la Universidad Nacional del Santa?

2.5. HIPÓTESIS

El desarrollo de un modelo de Machine Learning logra la clasificación de estudiantes de acuerdo a su rendimiento académico en el centro de idiomas de la Universidad Nacional del Santa.

2.6. OPERACIONALIZACIÓN DE LAS VARIABLES

Tabla N° 01: Operacionalización de las Variables

Variables	Indicadores
V.I: Machine Learning	1. Pruebas Unitarias
	2. Tiempo de respuesta
	3. Categoría de Defensa de los datos.
	4. Grado de Protección de los datos.
V.D: Clasificación de Estudiantes	1. Tiempo medio en realizar la clasificación académica.
	2. Número de clasificaciones académicas acertadas.
	3. Grado de satisfacción del Estudiante.
	4. Grado de satisfacción de la plana docente

2.7. OBJETIVOS DEL PROYECTO

2.7.1. Objetivo General

Desarrollar un modelo de machine learning para lograr la clasificación de estudiantes de acuerdo a su rendimiento académico en el centro de idiomas de la Universidad Nacional del Santa.

2.7.2. Objetivos Específicos

- Identificar las características y patrones que pueden ser usados como predictores de clasificación de un alumno de acuerdo con el grado académico solicitado.
- Determinar cuáles son los modelos de Machine Learning que se logran utilizar para efectuar las notas de un estudiante de acuerdo a su rendimiento académico.
- Utilizar la metodología CRISP para cada una de sus fases para el modelo de Machine Learning.
- Disminuir el tiempo de clasificación académica de los estudiantes.
- Disminuir el número de errores en la clasificación de estudiantes.
- Extender el grado de satisfacción de la plana docente y estudiantes.

2.8. JUSTIFICACIÓN

2.8.1. Justificación Social

- Perfeccionar la imagen institucional.
- Mejora el flujo de estudiantes disminuyendo el tiempo de estancia.

2.8.2. Justificación Tecnológica

- Fomentar la apropiación tecnológica y la usabilidad de las tecnologías de información y las comunicaciones.
- Contribuye con la determinación de modelos que generen patrones de datos.
- Compatibilidad con todos los sistemas operativos.

2.8.3. Justificación Operativa

- Procesar la información de manera más segura y acertada.
- Reducción de errores al ejecutar las tareas de aprendizaje.

- Resultados más fidedignos en las tareas de aprendizaje.

2.8.4. Justificación Técnica

- Disponibilidad inmediata de los datos informativos para los alumnos en tiempo real.
- Reducir los tiempos de búsqueda referente a su rendimiento académico.
- Optimizar la ejecución de registro y gestión de la información de los alumnos y docentes del centro de idiomas.
- Evitar la pérdida o alteración de las notas de los alumnos del centro de idiomas.
- Reducir la duplicidad de datos en el centro de idiomas.
- Mayor conocimiento de los componentes que contribuyen en el rendimiento académico.

2.8.5. Justificación Económica

- La realización del presente proyecto accederá a disminuir el consumo de recursos en el centro de idiomas CEIDUNS de la universidad Nacional del Santa.

2.8.6. Justificación Personal

- Permitirá que los indagadores ahonden en los argumentos pertinentes a ciencia de datos.

2.9. Importancia de la Investigación

La gestión que se realiza en el centro de idiomas de la universidad Nacional del Santa es muy ineficiente ya que causa malestar en los padres de familia y alumnos cuando realizan el trámite modificación del Registro de datos del alumno porque no saben en donde se encuentra archivado ese registro, también no saben cómo se van

desempeñando cada alumno y cuál es el grado de nivel de conocimientos que tienen y poder clasificarlos de una manera rápida y eficaz, por lo que es de gran importancia saber cómo Aplicar técnicas de machine learning para generar modelos que permitan predecir la clasificación de estudiantes, estimando guías sociales, económicos y académicos para el rendimiento académico en estudiantes del centro de idiomas de la Universidad Nacional del Santa.

También es importante considerar que un alumno satisfecho mejora la imagen del centro de Idiomas. El desarrollo de un modelo Machine Learning logra clasificar de acuerdo a su rendimiento académico a los estudiantes del Centro de Idiomas de la Universidad Nacional del Santa.

2.10. Limitaciones

Limitado tiempo del personal del centro de idiomas de la Universidad Nacional del Santa para la realización de entrevistas y cuestionarios que implica en la ejecución de la investigación.

CAPÍTULO III
MARCO TEÓRICO

3.1. CIENCIA DE DATOS

3.1.1. Definición

Según Dhar (2013) “La ciencia de datos es un método procedente multidisciplinaria que comprende desde el desarrollo de software, hasta la inteligencia artificial, el manejo de datos y la estadística”. “Los proyectos con ciencia de datos se basan en asemejar similitudes, recomendaciones impensadas, especificar y pronosticar eventos, equiparar patrones y anomalías e inferir posibilidades, Estos proyectos se enfocan en la obtención de conocimiento de los datos y están estrechamente relacionados con datos masivos (big data en inglés)” (Schoenherr y Speier-Pero, 2015). “La ciencia de datos cubre el proceso de la recopilación de los datos, su análisis y los resultados de los experimentos. Con el aumento de la capacidad de recoger, acumular y analizar una diversidad es aún mayor de datos, la ciencia de datos se está incrementando rápidamente. 18 Promueve el uso de algoritmos que pueden generar resultados útiles. Por desgracia, en el campo de la ciencia de datos no se conoce el "mejor" proceso para realizar un proyecto científico”. (Saltz, Shamshurin y Crowston, 2017).

3.1.2. Metodología de ciencia de datos

Según Rollins (2015) “La IBM (International business machines) propone una metodología para la aplicación de ciencia de datos a nivel industrial y científico. Esta metodología se organiza en diez etapas que representan un proceso interactivo, las cuales se presentan a continuación.”

1. **“Compresión del proyecto**. Es la etapa donde se definen tanto los problemas y los propósitos que proyecta como también los requerimientos

del resultado desde un punto de vista comercial. Por eso, esta etapa se considera como la más difícil."

2. **"Enfoque analítico.** En esta etapa, el científico de datos define el enfoque analítico que utilizará para resolver el problema. Esto implica saber expresar la problemática con el argumento de métodos estadísticos y de instrucción automático para que puedan asemejar los mecanismos adecuados para lograr una conclusión exitosa."
3. **"Requerimiento de datos.** La elección de la aproximación analítica se determina a partir de los requisitos de los datos para poder conocer la técnica analítica que se usará. Las especialidades de los datos son específicos en diversos formatos y representaciones; pueden ser constituir, semi-constituido y no constituido."
4. "**Recolección de datos**. En este periodo, el científico de datos identifica y reúne datos constituidos, semi-constituido o no constituido, relevantes en el dominio del problema."
5. "**Comprensión de los datos.** En este período, la estadística descriptiva y las técnicas de visualización pueden ayudar a entender al científico de datos el contenido de los datos, evaluar su calidad y descubrir una idea inicial de estos."
6. "**Preparación de los datos**. Esta etapa comprende actividades enfocadas en la preparación del conjunto de datos que serán usados en la etapa de modelado. Esto incluye la limpieza de la información, la combinación de la información de múltiples fuentes y la transformación de la información."

7. “**Modelamiento.** En esta etapa, se generan modelos predictivos mediante algoritmos de aprendizaje automático.”
8. “**Evaluación.** En esta etapa, el científico de datos evalúa la calidad de los modelos generados.”
9. “**Despliegue**. Después de haber elegido un modelo satisfactorio que es aprobado por los patrocinadores del proyecto, el modelo se implementa en el ambiente de producción o de prueba para su utilización.”
10. “**Retroalimentación**. Mediante la recopilación de los resultados del modelo implementado, la organización recibe información sobre el rendimiento del modelo y se observa la forma en que este afecta su entorno. El análisis de esta información le muestra al científico de datos si se debe refinar el modelo. Esto aumenta su precisión y, por tanto, su utilidad.” (Rollins, 2015).

3.2. MACHINE LEARNING (ML)

Según Rivero (2017) “Machine Learning se originó en el campo de Inteligencia Artificial. Conforma un conjunto de modelos matemáticos y estadísticos cuyas tareas involucran reconocimiento, diagnostico, control de robots, predicción, etc.” “Machine Learning es utilizada en la ciencia de los datos como una herramienta, debido a que produce resultados autónomos sin necesidad de ser explícitamente programados, habiendo pasado por una etapa de entrenamiento. Un ejemplo tradicional de Machine Learning es un sistema de predicción de mails spam, donde la entrada es un conjunto de mails que han sido etiquetados como spam por el humano. El proceso de aprendizaje consiste en reconocer un conjunto de palabras y características que aparecen en un spam mail. Luego, cuando un nuevo mail es recibido, se revisarán sus características y si encajan en el conjunto de contenido

sospechoso, será etiquetado como mail spam. Este sistema, con una base adecuada de datos de entrenamiento, está capacitado para correctamente predecir la etiqueta de todo nuevo mail entrante. El campo de Machine Learning se ha dividido en algunos sub campos para tratar con los diferentes tipos de tareas de aprendizaje. Por objetivos de la investigación, se hará énfasis en el par: Supervisados y No supervisados." (Rivero, 2017)

3.2.1. Clasificación de Machine Learning

Según Stephen (2014) Dice que "Machine Learning se clasifica de acuerdo al tipo de aprendizaje; de los cuales tenemos:" (Stephen, 2014)

- **"Aprendizaje supervisado:** Dado un conjunto de datos de entrada X, compuesto por el vector $\vec{x}$ = {x (1), x (2), x (3) … x (m) } (también llamado vector de características); donde m indica el número total de muestras; y un vector de salida (output o labels) $\vec{y}$ = {y (1) , y (2) , y (3) …y (m) } ; la tarea del aprendizaje supervisado, consiste en encontrar una función h; tal que h: $X \rightarrow y$; dicha función asume el nombre de hipótesis"
- "**Aprendizaje no supervisado:** Dado un conjunto de datos de entrada X, compuesto por el vector $\vec{x}$ = {x (1) , x (2) , x (3) … x (m) }; la tarea del aprendizaje no supervisado, consiste en encontrar una función h; tal que h: $X \rightarrow y$; cabe destacar que, a diferencia del aprendizaje supervisado; la función h tiene que realizar el mapeo sin el vector de salida $\vec{y}$" (Geras, 2011)
- **"Aprendizaje por refuerzo**: Este tipo de aprendizaje se centra en la interacción, y el control de entornos estocásticos con conocimiento parcial; para lo cual se usa el modelo de proceso de decisión de

Markov; el cual define cuatro funciones (T,R, π, V) sobre el conjunto finito de estados S y acciones A; donde la función de transición T establece la probabilidad de distribución en el estado sucesor st+1 de la siguiente manera $P(st+1 = s'|st = s, at = a)$, al tiempo que, la recompensa rt es establecida por la función R de la forma $P(rt = r|st = s, at = a)$, asimismo la función de política π mapea los estados de las acciones, y finalmente la función V provee el valor de la recompensa esperada r cuando el proceso empezó en el estado s y el agente siguió la política π; finalmente a través de la interacción de las funciones y los estados, se logra encontrar un conjunto óptimo de políticas π; con lo cual el agente resuelve el problema" (Daw, 2003).

- **"Aprendizaje evolutivo:** Este enfoque de aprendizaje se centra en tomar modelos de aprendizaje biológicos; los cuales adapta a los problemas de aprendizaje; algunos de estos modelos son: inteligencia de enjambre, algoritmos genéticos, sistemas artificiales inmunes (Mo, 2009), etc.; cabe destacar que en la mayoría de los modelos se suele usar una función fitness f; la cual se encarga de medir la performance de la solución en cada generación de los algoritmos." (Stephen, 2014).

3.2.2. Aplicaciones Machine Learning

Según Rivero (2017) "El conjunto de algoritmos de Machine Learning es aplicado a numerosos problemas. Por ejemplo, al momento de publicar una fotografía de personas en las redes sociales, se ejecuta el reconocimiento de las características de los rostros para luego sugerir el nombre de cada persona. En caso de la detección de malware por los sistemas antivirus, se realiza el entrenamiento utilizando la colección de firmas o patrones de

comportamiento de los virus. Similarmente, la mitigación de fraude en el comercio electrónico, atraviesa por el entrenamiento de las actividades transaccionales del usuario, localidad, monto, periodicidad y otras variables de compra, las cuales son verificadas por el modelo para alertar una posible transacción fraudulenta. Otra aplicación es la predicción de embarazo de una mujer en base a los hábitos de su compra, si los pagos los hace con cupones de revistas, si compra por poca cantidad o si deja de comprar algunos productos, etc. En algunas aplicaciones la cantidad de datos supera los miles, por ejemplo: el análisis de sentimiento sobre blogs de campeonatos de fútbol. El algoritmo requerirá cada mensaje con su respectiva etiqueta: {alegría, decepción, antipatía, neutral}. Esta tarea demandará leer y etiquetar cada mensaje. Para esto son utilizados los servicios de Crowdsourcing, que consiste en obtener el trabajo de personas generalmente a través de Internet para desarrollar una tarea. Es necesario aclarar que tanto la obtención de datos, selección de características y etiquetado de datos de entrenamiento, conforman las tareas más finas y delicadas de realizar para obtener un modelo de predicción bien calificado." (Rivero, 2017)

3.3. ALGORITMOS DE APRENDIZAJE

Según Stephen (2014) describe a los algoritmos de aprendizaje "Como hemos visto existen diversos tipos de algoritmos de aprendizaje; los cuales están clasificados principalmente por el tipo de aprendizaje con el que se desea lidiar: supervisado, no supervisado, por refuerzo y finalmente evolutivo. Básicamente la idea de un algoritmo de aprendizaje (modelo) es encontrar una hipótesis *h*; la cual es una función que realiza la tarea 2 del modelo".

3.4. RENDIMIENTO ACADEMICO

Según (Pizarro & Clark, 1998) "Es una medida de la capacidad de respuesta del individuo, que expresa, en forma estimativa, lo que una persona ha aprendido como resultado de un proceso de instrucción o formación. Los mismos autores (1998, p.3), ahora desde la perspectiva del alumno, definen el rendimiento como la capacidad de respuesta que tiene un individuo, a estímulos, objetivos y propósitos educativos previamente establecidos."

Según Camborda (2014) "Además, el rendimiento académico es entendido por Pizarro como una medida de las capacidades respondientes o indicativas que manifiestan, en forma estimativa, lo que una persona ha aprendido como 24 consecuencia de un proceso de instrucción o formación. El mismo autor, ahora desde una perspectiva propia del alumno, define el rendimiento como una capacidad respondiente de éste frente a estímulos educativos, susceptible de ser interpretado según objetivos o propósitos educativos pre-establecidos. Este tipo de rendimiento académico puede ser entendido en relación con un grupo social que fija los niveles mínimos de aprobación ante un determinado cúmulo de conocimientos o aptitudes."

Según Herrán y Villarroel (1987), el rendimiento académico se define en forma operativa y tácita afirmando que se puede comprender el rendimiento escolar previo como el número de veces que el alumno ha repetido uno o más cursos."

Según Mined (2002), "El rendimiento académico determina el nivel de conocimiento alcanzado, y es tomado como único criterio para medir el éxito o fracaso escolar a través de un sistema de calificaciones de 0 a 10 en la mayoría de los centro educativos públicos y privados, en otras instituciones se utilizan el sistema de porcentaje de 0 a 100%,y los casos de las instituciones bilingües, se utiliza el sistema de letras que va desde la "A" a la "F", para evaluar al estudiante como deficiente,

Bueno, Muy bueno o Excelente en la comprobación y la evaluación tienen que ser una medida objetiva sobre el estado de los rendimientos de los alumnos".

3.4.1. Características del rendimiento académico

Según Camborda (2014) en su tesis "Después de realizar un análisis comparativo de diversas definiciones del rendimiento escolar, concluyen que hay un doble punto de vista, estático y dinámico, que atañen al sujeto de la educación como ser social. En general, el rendimiento escolar es caracterizado del siguiente modo: a) el rendimiento en su aspecto dinámico responde al proceso de aprendizaje, como tal está ligado a la capacidad y esfuerzo del alumno; b) en su aspecto estático comprende al producto del aprendizaje generado por el alumno y expresa una conducta de aprovechamiento; c) el rendimiento está ligado a medidas de calidad y a juicios de valoración; d) el rendimiento es un medio y no un fin en sí mismo; e) el rendimiento está relacionado a propósitos de carácter ético que incluye expectativas económicas, lo cual hace necesario un tipo de rendimiento en función al modelo social vigente." (Camborda, 2014)

3.4.2. Rendimiento Académico en el Perú

Según Camborda (2014) "En consonancia con esa caracterización y en directa relación con los propósitos de la investigación, es necesario conceptuar el rendimiento académico. Para ello se requiere previamente considerar dos aspectos básicos del rendimiento: el proceso de aprendizaje y la evaluación de dicho aprendizaje. El proceso de aprendizaje no será abordado en este estudio. Sobre la evaluación académica hay una variedad de postulados que pueden agruparse en dos categorías: aquellos dirigidos a la consecución de un valor numérico (u otro) y aquellos encaminados a

propiciar la comprensión (insight) en términos de utilizar también la evaluación como parte del aprendizaje. En el presente trabajo interesa la primera categoría, que se expresa en los calificativos escolares. Las calificaciones son las notas o expresiones cuantitativas o cualitativas con las que se valora o mide el nivel del rendimiento académico en los alumnos. Las calificaciones escolares son el resultado de los exámenes o de la evaluación continua a que se ven sometidos los estudiantes. Medir o evaluar los rendimientos escolares es una tarea compleja que exige del docente obrar con la máxima objetividad y precisión (Fernández Huerta, 1983; cit. por Aliaga, 1998b). En el sistema educativo peruano, en especial en las universidades -y en este caso específico, en la UNMSM-, la mayor parte de las calificaciones se basan en el sistema vigesimal, es decir de 0 a 20. Sistema en el cual el puntaje obtenido se traduce a la categorización del logro de aprendizaje, el cual puede variar desde aprendizaje bien logrado hasta aprendizaje deficiente."

3.4.3. Funcionamiento del Área de Gestión Académica

Según el Repositorio (2016). "El Área de Gestión Académica es una unidad administrativa con características muy específicas en la Escuela o Universidad, pues su misión no es otra que la de servir de hilo conductor de la gestión administrativa del alumno a lo largo de su vida académica."

Funciones del Área:

- "Gestión administrativa de las pruebas de acceso a una institución, Planes de Estudio y Oferta de Enseñanzas."

- “Gestión de los procesos de preinscripción, matrícula y normativa general académica de títulos de Grado, 1º y 2º Ciclo, Másteres Oficiales y Doctorado.”
- “Coordinación de las Secretarías de los Centros y de la información académica general del alumno.”
- “Gestión de las solicitudes y reclamaciones en materia de gestión académica.”
- “Gestión de las solicitudes de Evaluación Curricular por Compensación.”

CAPÍTULO IV

DESARROLLO DE LA METODOLOGIA

4.1. COMPRENSIÓN DEL NEGOCIO

4.1.1. Plan de Desarrollo De Software

4.1.1.1. Determinar los objetivos del negocio

Los objetivos del negocio considerados son aquellos que están directamente asociados con el propósito del desarrollo del modelo de machine learning.

Tabla N° 02: Objetivos del Negocio

CÓDIGO	OBJETIVOS DEL NEGOCIO
OBNG-01	Brindar un servicio de calidad con tecnologías modernas para los procesos que involucran a los Alumnos de CEIDUNS
OBNG-02	Mejorar el proceso de Clasificación académica en un 30% para minimizar el conflicto de calificaciones de los alumnos de CEIDUNS
OBNG-03	Mejorar el proceso de Clasificación para disminuir en un 50% el tiempo de registro de alumnos dependiendo de la clasificación obtenida.

4.1.1.2. Criterios de éxito del negocio

Para cada objetivo se han definido los criterios de aceptación

Tabla N° 03: Criterios de Aceptación

OBJETIVO	CRITERIO DE ACEPTACIÓN
OBNG-01	Implantar puesta en entorno de desarrollo para validar y verificar el funcionamiento de los procesos.
OBNG-02	Tener implementado un sistema informático automático que brinde un resultado de Clasificación mayor al 70%

OBNG-03	Tener una plataforma o sistema válido, que brinde un nivel de eficiencia mayor al 80%

4.1.2. Evaluación de la situación

En este punto se describen los recursos, requisitos, supuestos, restricciones, riesgos, terminologías y costes del proyecto

4.1.2.1. Inventario de recursos

El inventario contiene los requerimientos humanos también los orígenes de datos

Tabla N° 04: Recursos Humanos

OBJETIVO	RECURSOS HUMANOS
RH-01	Gianira Espinoza Airac
RH-02	Dante Escalante Roldán
OBNG-03	Estudiantes Matriculados en el Año 2018

Tabla N° 05: Fuente de Datos

OBJETIVO	FUENTE DE DATOS
FD-01	Base de datos en Excel con los datos de los estudiantes del centro de idiomas para su clasificación.

4.1.2.2. Requisitos, supuestos y restricciones

Tabla N° 06: Requisitos del Proyecto

OBJETIVO	REQUISITOS DEL PROYECTO
RQ-01	Contar con información de los alumnos del centro de idiomas periodo mínimo de 2 años.

RQ-02	Contar con información sobre el proceso de calificaciones del centro de idiomas
RQ-03	Tener acceso de lectura de las fuentes de datos
RQ-04	Utilizar tecnología que permite obtener resultados confiables

Tabla N° 07: Supuestos del Proyecto

OBJETIVO	**SUPUESTOS DEL PROYECTO**
ST-01	Se cuenta con datos relevantes para el desarrollo del proyecto.
ST-02	La aplicación informática sólo cumple con uno de los indicadores
ST-03	El costo del desarrollo de la aplicación informática es muy elevado para el negocio
ST-04	El centro de idiomas no brinda información suficiente del proceso de Clasificación de alumnos.

Tabla N° 08: Restricciones del Proyecto

OBJETIVO	**RESTRICCIONES DEL PROYECTO**
RT-01	No contar con acceso a datos actualizados.

4.1.2.3. Riesgos y contingencias

El desarrollo del proyecto se puede ver afectado por algunos riesgos, para los cuales es necesario determinar una solución

Tabla N° 09: Riesgos y Contingencia

OBJETIVO	RIESGOS	CONTINGENCIA
RG-01	La información con la que se cuenta no esté completa	Establecer un intervalo de tiempo

RG-02	Datos pocos relevantes.	con la información recogida para realizar un modelo más confiable

4.1.2.4. Terminología

Tabla N° 10: Términos del Proyectos

CÓDIGO	TÉRMINO
ML	Machine Learning
CRISP	Cross-Industry Standard Process for Data Mining
CA	Clasificación Académica
FEATURES	Características

4.1.2.5. Costes

Tabla N° 11: Costos de Hardware

RECURSOS DE **HARDWARE**			
Equipo	Cantidad	Precio	Total (depreciación 3 años)
Laptop	2	2700.00	1800.00
Impresora/Scanner	1	300.00	100.00
TOTAL			**1900.00**

Tabla N° 12: Costos de Software

RECURSOS DE SOFTWARE			
Descripción	**Cantidad**	**Precio**	**Total**
Windows 10	1	215,82	215,82
Microsoft Office 2016	1	120,00	120.00
TOTAL			**335.82**

Tabla N° 13: Costos de Internet

INTERNET			
Descripción	**Precio (mes)**	**Meses**	**Total**
Internet Movistar de 15 mbps	176.00	4	704.00
TOTAL			704.00

Tabla N° 14: Costos de Recursos Humanos

RECURSOS HUMANOS			
Personal	Meses	Precio	Total
Analista/Programador	8	750.00	6000.00
Tester	8	750.00	6000.00
TOTAL			**3600.00**

Tabla N° 15: Costos de Materiales

Lapiceros	**Cantidad**	**Unidad**	**Total**
Folder Manila	8	750.00	6000.00
Tester	8	750.00	6000.00
TOTAL			**3600.00**

Tabla N° 16: Ahorro en personal

Recursos de Materiales				
Personal	Sueldo	Tiempo Ahorrado	Monto	Total
	Hora (s/)	Mensual (horas)		
Analista de crédito	12,50	40,00	500,00	6000,00
Total			**500.00**	**6000.00**

4.1.3. Realizar el plan del proyecto

4.1.3.1. Conformación del equipo

Tabla N° 17: Miembros del Equipo

Nº	MIEMBRO	ROL
01	Eduar Leon Muñoz	• Analista / Programador • Tester
02	Gianira Espinoza	• Analista / Programador • Tester
03	Juan Martinez	Director Ceiduns

4.1.3.2. Requerimientos

4.1.3.2.1. Requerimientos Funcionales

- La Aplicación informática debe permitir registrar a los alumnos.
- La aplicación debe permitir ingresar las notas del centro de idiomas
- La aplicación debe Clasificar a los alumnos de acuerdo a sus notas

4.1.3.2.2. Requerimientos no funcionales

- El diseño del sistema informático tiene que ser amigables.
- El sistema informático debe ser de alto rendimiento
- La Clasificación de los alumnos tiene que realizarse de forma rápida

- La clasificación de los alumnos debe ser eficiente.

4.1.3.3. Fases de desarrollo

Tabla N° 18: Comprensión del negocio

FASE 01
Nombre de la fase: Comprensión del negocio
Encargados: Eduar León Gianira Espinoza
Descripción: Se establece los requerimientos del establecimiento y la estimación del negocio. Tareas de la fase de desarrollo: • Establecer los requerimientos del negocio • Apreciación del ambiente del establecimiento • Establecer los Requerimientos • Realizar el plan del proyecto

Tabla N° 19: Comprensión de la información

FASE 02
Nombre de la fase: Comprensión de la información
Encargados: Eduar León Gianira Espinoza
Descripción: Acceder y explorar los datos. Tareas de la fase de desarrollo: • Recopilación de datos iniciales • Representación de la información • Investigación de la información • Verificar la característica de la información

Tabla N° 20: Elaboración de la Información

FASE 03
Nombre de la fase: Elaboración de la información
Encargados: Eduar León Gianira Espinoza
Descripción: Seleccionar, limpiar e integrar los datos. Tareas de la fase de desarrollo: • Elegir la Información • Purificar la Información • Exaltar la Información • Completar la Información • Formateo de la información

Tabla N° 21: Modelado

FASE 04
Nombre de la fase: Modelado
Encargados: Eduar León Gianira Espinoza
Descripción: Determinar y aplicar un modelo. Tareas de la fase de desarrollo: • Elegir la habilidad del modelamiento. • Crear una interfaz de comprobación. • Arquitectura del modelo. • Generar los modelos.

Tabla N° 22: Evaluación del modelo

FASE 05
Nombre de la fase: Evaluación del modelo
Encargados: Eduar León

Gianira Espinoza
Descripción: Evaluar los resultados obtenidos, según lo planteado en la primera iteración

Tabla N° 23: Implantación

FASE 06
Nombre de la fase: Implantación
Encargados: Eduar León Gianira Espinoza
Descripción: Implantación en entorno de desarrollo y/o producción. Tareas de la fase de desarrollo: Planear la implantación • Planear la monitorización y supervisión • Producir el informe final • Analizar el proyecto

4.1.3.4. Planificación inicial

Se define la preferencia (Bajo, Media o Alta dependiendo de la jerarquía que posea), inseguridad (Bajo, Medio o Alto es la posibilidad de algún fallo en cada fase), esfuerzo (Se califica Bajo, Medio o Alto según el tiempo y trabajo que nos demandará en desarrollar la fase) e iteración (Es el orden de la implementación de cada fase, se califica del 1 al 6) de cada historia de usuario.

Tabla N° 24: Planificación inicial

Nº	**Fase de desarrollo**	**Prioridad**	**Riesgo**	**Esfuerzo**	**Iteración**
01	Compresión del Negocio	Alta	Bajo	Bajo	1
02	Compresión de los Datos	Alta	Medio	Alto	2
03	Preparación de los datos	Alta	Alto	Alto	3
04	Modelado	Alta	Alto	Alto	4
05	Evaluación del Modelado	Alta	Medio	Medio	5
06	Implementación	Alta	Medio	Medio	6

4.1.3.5. Velocidad del proyecto

Con las ponderaciones obtenidas de los datos de prioridad, esfuerzo y riesgo se ha logrado estimar el tiempo de desarrollo de cada historia.

Tabla N° 25: Tiempo estimado en el desarrollo

Nº	Fase de desarrollo	Tiempo estimado
01	Compresión del Negocio	04 Días
02	Compresión de los Datos	15 Días
03	Preparación de los Datos	20 Días
04	Modelado	12 Días
05	Evaluación del modelado	08 Días
06	Implantación	04 Días

4.2. COMPRENSIÓN DE LOS DATOS

4.2.1. Recopilación de datos iniciales

La recolección de la información se hizo mediante las sucesivas fuentes:

- Base de datos – BDCLASIFICACIÓN: La base de datos tiene datos con la solicitud de Calificaciones, los datos de los Alumnos. En algunos casos, los registros no están completos. Se cuenta con registros desde el 2016 hasta el 2019.

4.2.2. Descripción de los datos

- La Ejecución del proyecto, se manejan los datos dentro del tiempo de enero hasta diciembre de los años 2017 y 2019, contando con una gran cantidad de alumnos.
- Registro de datos (solo se considera la información precisa para la ejecución del proyecto).

4.2.3. Verificar la calidad de los datos

Siguiendo con la lista se menciona los puntos apreciados para confirmar la eficacia de la información:

Tabla N° 26: Calidad de los Datos

CÓDIGO	TÉRMINO
CLD-01	Los campos de las tablas no tienen que contener información o valor nulo
CLD-02	Los campos de las tablas deben cuidar el criterio solicitado o aprobado, no deben tener valor cero
CLD-03	La información referente a la solicitud de datos no debe estar duplicada

4.3. PREPARACIÓN DE LOS DATOS

4.3.1. Seleccionar los datos

Una vez ejecutada la demostración de los datos, en el período anterior, se completó el comunicado en una solo generalidad de los datos utilizando el código del estudiante. Se elimina la información inconclusa, carente o impropio para esquivar que se examinen la información que logren difundir errores al momento de tramitar los algoritmos de Machine Learning.

Se creó un ID único para cada uno de los estudiantes, en la tabla, a fin de diferenciar la información y además eliminar la filiación de otros identificadores que lograran concurrir en la data, en esta cuestión, el código del alumno.

- **Obtener datos desde la base de Datos**

```
Select ALUMNO. NOMBRECOMPLETO , ALUMNO.CATEGORIA,
ALUMNO. UBIGEONACIMIENTO, ALUMNO.SEXO,
ALUMNO. ID_ESTADO, ALUMNO. ESTADO_ALUMNO,
ALUMNO. MODALIDAD_INGRESO, ALUMNO.MODALIDAD,
ALUMNO. CODALUMNO,
(select first 1 mov_alumno.promediop
From mov_alumno where alumno. codigo = mov_alumno. Alumno)
ponderado,
(select first 1 mov_alumno.estado
From mov_alumno where alumno.codigo = mov_alumno.alumno) condición
From ALUMNO
Where (select first 1 mov_alumno. promediop
From mov_alumno where alumno.codigo = mov_alumno.alumno) is not null
And ALUMNO.MODALIDAD_INGRESO is not null
```

Order by alumno.nombrecompleto

Tabla 27: Descripción de la Tabla Alumno

DATOS	DESCRIPCIÓN
:IDCodigo_Alumno	Código o DNI del alumno como llave principal
:Dni	Documento de Identidad
:Sem_Ingreso	Semestre de ingreso al CEIDUNS
:Ape_Partenos	Apellido Paterno
:Ape_Materno	Apellido Materno
:Nombres	Nombres del Alumno
:Fec_Naci	Fecha de Nacimiento
:Genero	Genero del Alumno
:Mod_ingreso	Modalidad de Ingreso
:Estado_Matricula	Estado de Matricula (Activo/Inactivo)
:Curso	Curso Cursando
:Nota	Nota del estudiante

4.3.2. Limpiar los datos

En este periodo, se emplearon las funciones de EXCEL: TRIM(), CLEAN(), SUBSTITUTE() para descartar las áreas en blanco efectuadas en los datos de tipo textual y se emplearon las funciones de termino para los valores numerales con decimal.

Se eliminaron los datos personales que consiguieran identificar, personalmente, a cada estudiante como código del escolar, apellido paterno, patronímico materno, popularidad y dirección.

En una primera revisión de la información, mediante este periodo, se localizó que algunos estudiantes que sí estaban matriculados, pero aparentemente no concurrieron a clases, ya que en todos los cursos plasmaba la calificación de cero. La información de estos estudiantes fue eliminada de la totalidad de la información a fin de obviar posibles errores.

4.3.2.1. Verificación de duplicidad

Este proceso se ha realizado manualmente, para identificar la duplicidad en los registros se ha verificado si el nombre, monto solicitado, fecha y hora de proceso son iguales. En la base datos BD_CLASIFICACIÓN se encontró con este tipo de inconsistencia y se dejó el último registro, como se puede observar en las imágenes.

Para eliminar los registros duplicados se utilizó la siguiente consulta:

4.3.2.2. Exclusión de características

Al analizar la base datos BD_CLASIFICACIÓN se determinó que existen características que no son necesarias para el desarrollo del proyecto, por tal motivo, se realizó la exclusión de características. Las características que se utilizará finalmente son las siguientes.

4.3.2.3. Asignar datos

En esta fase se formó con la asignación de datos nuevos con unas características para engrandecer los datos con la finalidad de acomodar los datos a que sea adaptable para la aplicación de Machine Learning.

En estas características incorporadas se encuentran:

- **EDAD:** Es la edad estimada al período de haber empezado las clases, al cual se deduce restando de la fecha del primer día de clases y la fecha de nacimiento.

  Por ejemplo, un estudiante ingresante del ciclo 2016-I con fecha de nacimiento 01 de enero de 1992 se calcula:

  Primer día de clases – fecha de nacimiento = tiempo en años (redondeado hacia abajo)

  01/03/2016 – 01/01/1992 = 18.1853425 -> 24 años

- **SEXO:** Para proporcionar la práctica de algoritmos se solicita que todas las informaciones sean numéricas se convirtió a binario establecido 1 para masculino y 0 para femenino.
- **APROBADO:** El estado de aprobado para los cursos indicados en binario, 1 para aprobado (nota >= 14) y 0 para desaprobado (nota <14). Se ejecutaron cambios para cada curso que produjeron durante el primer ciclo

4.3.3. Construir los datos

El método utilizado es derivación de datos, pues es necesario contar con la edad de los solicitantes, por eso se construye una nueva columna en la base datos, Edad, tomando como referencia la fecha de nacimiento y la fecha en el que se empieza el proceso de Registro de Alumnos.

4.4. MODELADO

4.4.1. Técnica del modelo

En el presente proyecto se ha considera el uso del algoritmo de machine learning denominado Clasificación Académica. Este algoritmo se ajusta a los requerimientos planeados en el proyecto, debido a que es un algoritmo eficaz de clasificación, es rápido y sencillo de implementar

4.4.2. Plan de pruebas del modelo

En el plan de pruebas se ha dividido la información en dos conjuntos, uno de ejercicio y otro de ensayo de predicción, para posteriormente edificar la interfaz basada en el conjunto de preparación y calcular la calidad de la interfaz generada. Este plan es utilizado en el punto Evaluación del modelado.

4.4.3. Arquitectura de la aplicación

4.4.3.1. Componentes de la aplicación

Para el proyecto se han realizado tres componentes fundamente**s**

a) Componente Web

Este componente contiene la página web para ingresar los datos que serán usados para clasificar el nuevo préstamo crediticio e indicar si ha sido rechazado o aprobado. Para lo cual se ha usado Python en la parte de la página web.

b) Componente API

Este componente contiene los métodos que serán consumidos por el sistema web. Los métodos implementados permiten calcular el nivel de exactitud de la clasificación de nuevos Notas, indica si el alumno será aprobado o desaprobado

clasificando académicamente a los alumnos según los datos recibidos. Se ha usado el microframework Flask para la implementació**n.**

c) Componente ML

Este componente es el más importante y trabaja en el nivel más bajo de la aplicación, ya que por medio de este componente se realiza toda la lógica del algoritmo de machine learning que se implementará, además de consumir la información con la cual se entrenará el algoritmo. Se ha desarrollado con Python y numpy.

4.4.4. Construcción del modelo

El modelo de machine learning desarrollado es una regresión logística la cual ha sido construida con los siguientes algoritmos:

4.4.4.1. Algoritmo de escalamiento de datos

El algoritmo de escalamiento permite reducir el rango de dispersión de los datos y de esa manera el algoritmo de regresión logística puede hacer los cálculos de manera más rápida, además que ayuda a reducir el sobre ajuste del conjunto de datos de entrenamiento.

4.4.4.2. Función Sigmoidal

La función sigmoidal servirá para obtener los valores predictivos de nuevas Calificaciones de los alumnos y también para calcular el valor óptimo de clasificación de cada coeficiente de entrenamiento, entiéndase por coeficientes a los valores θ (theta) que se tendrán que calcular por medio de las demás funciones.

Tener en cuenta que la función sigmoidal usada para la regresión logística es uniporlar, es decir, los valores estarán en el rango [0,1].

4.4.4.3. Función de costo para el modelo de Regresión Logística

La función de costo sirve para verificar que el modelo converja, es decir, minimizar en cada iteración para que se obtengan los valores de los coeficientes θ. Dado que se cuenta con dos clases $y \in \{0,1\}$ para clasificar, la función de costo se reduce a dos casos de evaluación de donde se deduce la siguiente fórmula para el modelo propuesto.

4.5. INTERFAZ DE USUARIO DE LA APLICACIÓN

En la siguiente imagen se puede apreciar el diseño de la aplicación, la cual tiene un formulario con características necesarias para que el algoritmo de machine learning realice la clasificación. Además, es de fácil uso para el usuario.

4.6. PRUEBAS

Las pruebas realizadas al modelo de regresión logística que se han desarrollado en el presente proyecto fueron hechas en la herramienta Weka, la cual brinda los datos de las calificaciones de los estudiantes de centro de idiomas para el conjunto de datos pruebas (25%). Los pasos para realizar las pruebas del modelo fueron las siguientes:

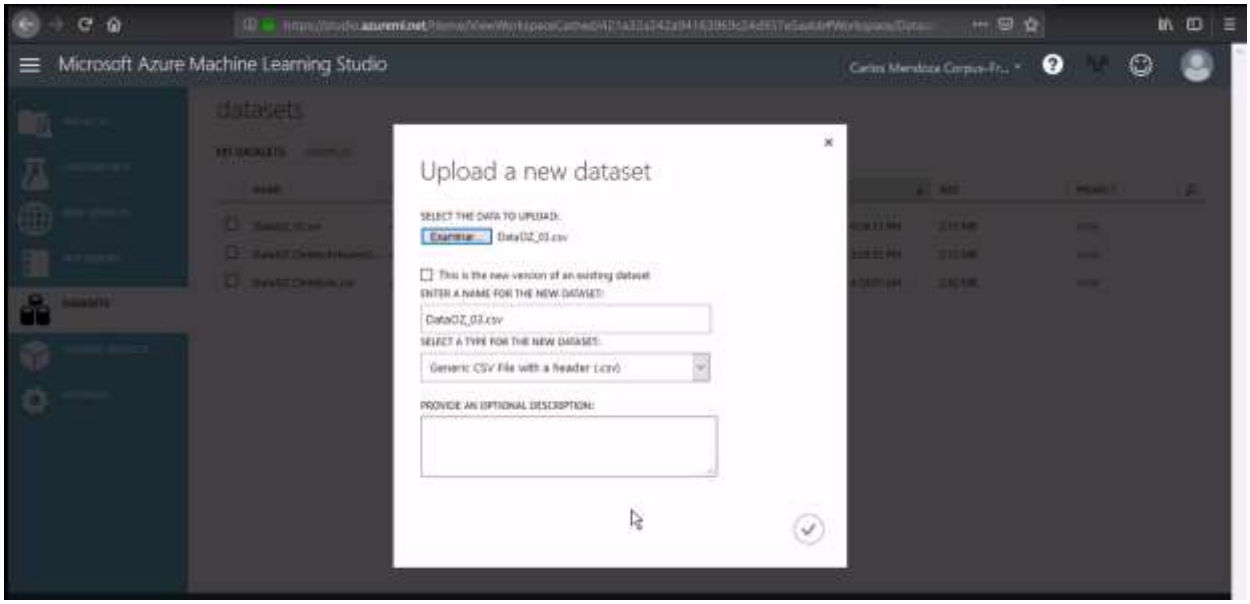


Figura N° 04: Carga del DataSet

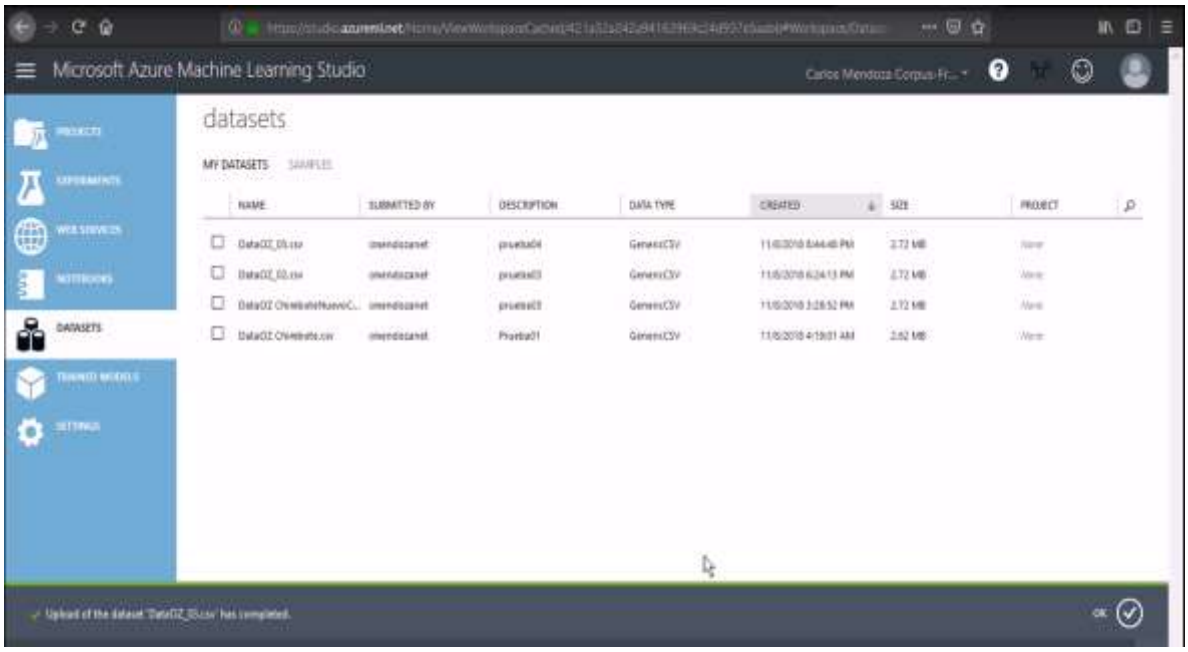


Figura N° 05: Repositorio del Machine Learning

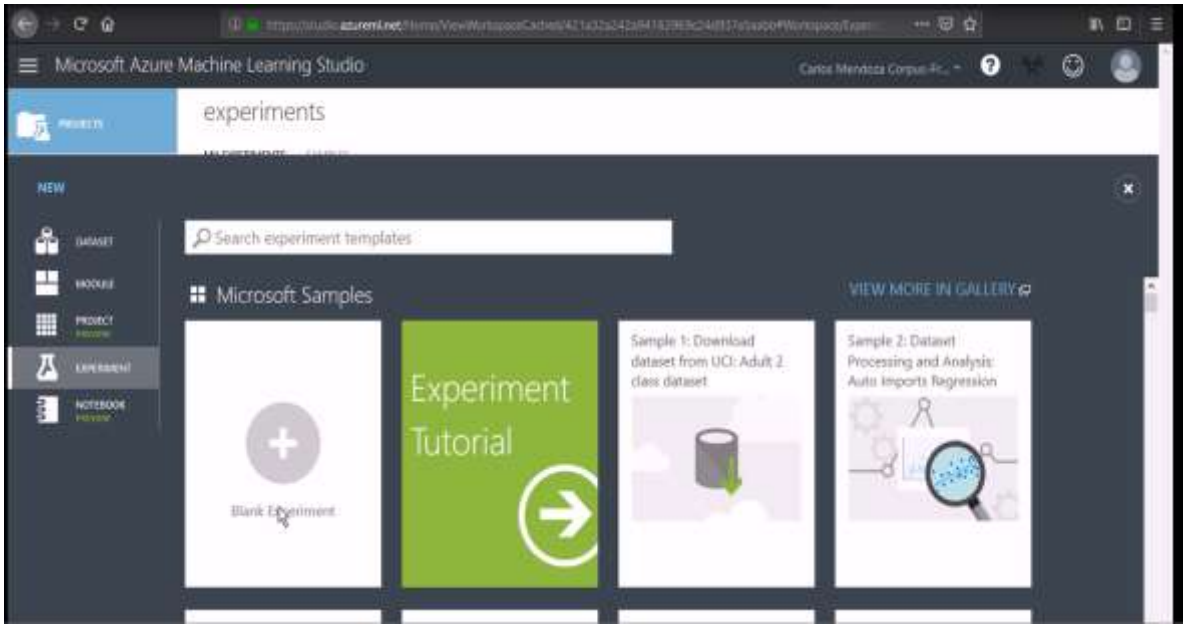


Figura N° 06: Creación de Nuevo Experimento

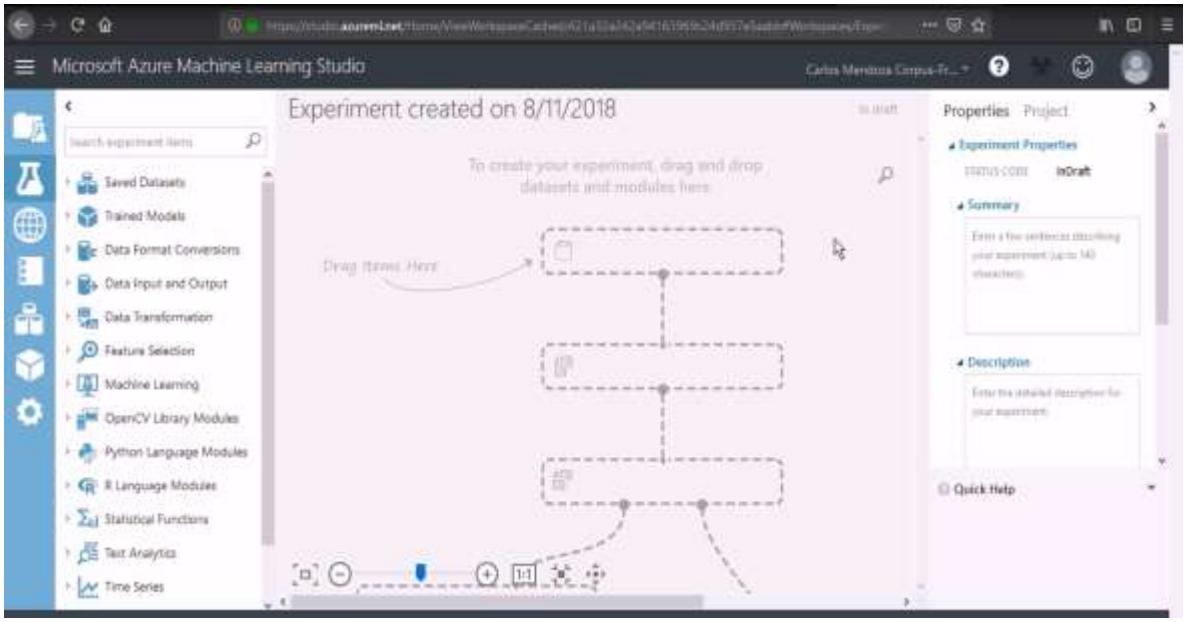


Figura N° 07: Entorno de Desarrollo del Experimento

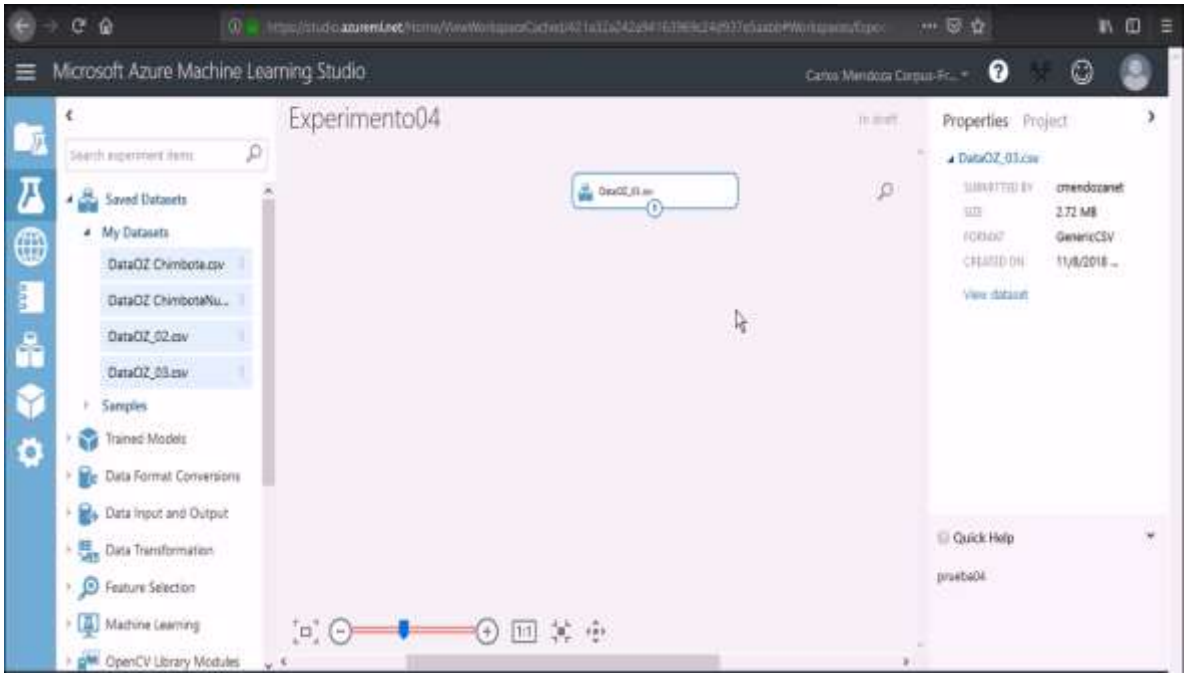


Figura N° 08: Incorporación del DataSet

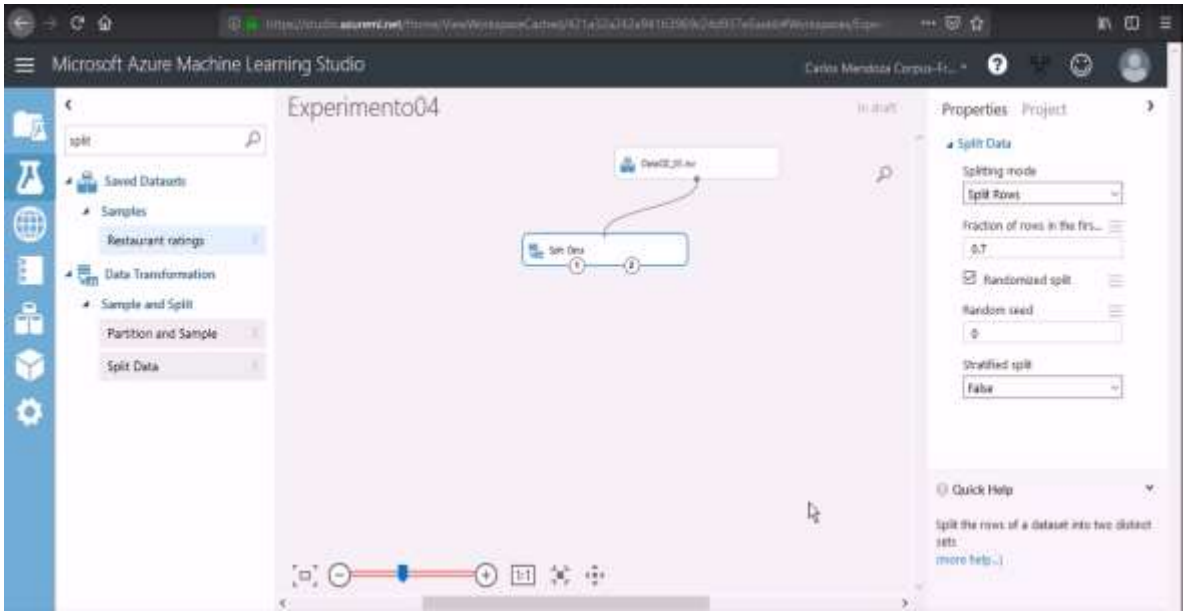


Figura N° 09: Incorporación de Split Data al experimento

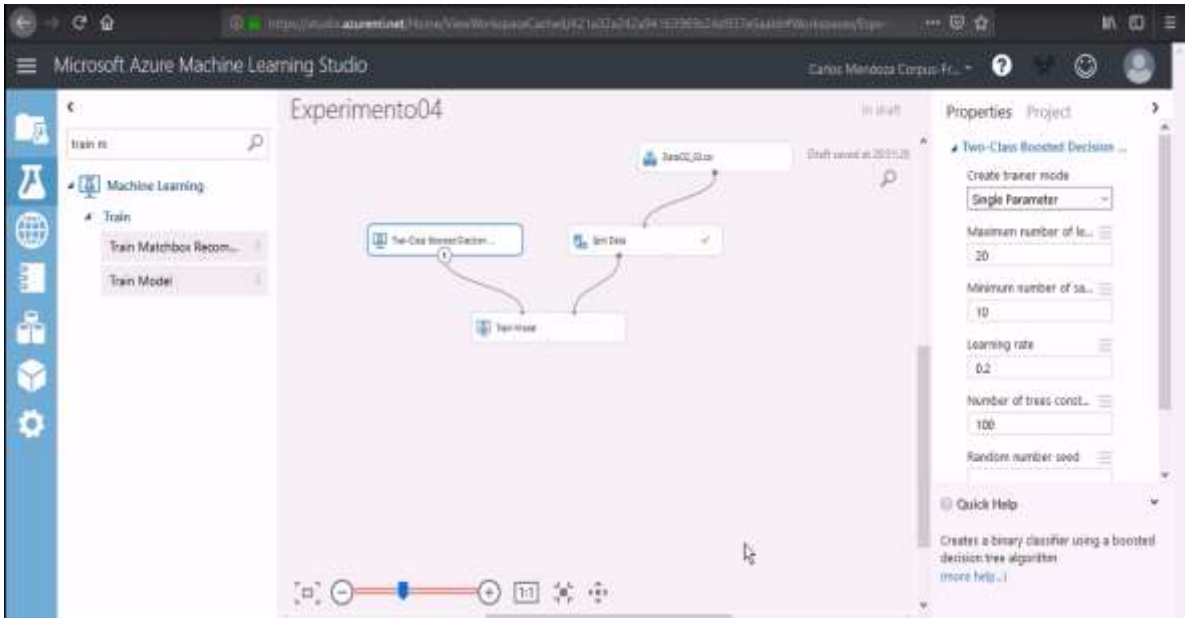


Figura N° 10: Agregando el modelo y el entrenador del modelo

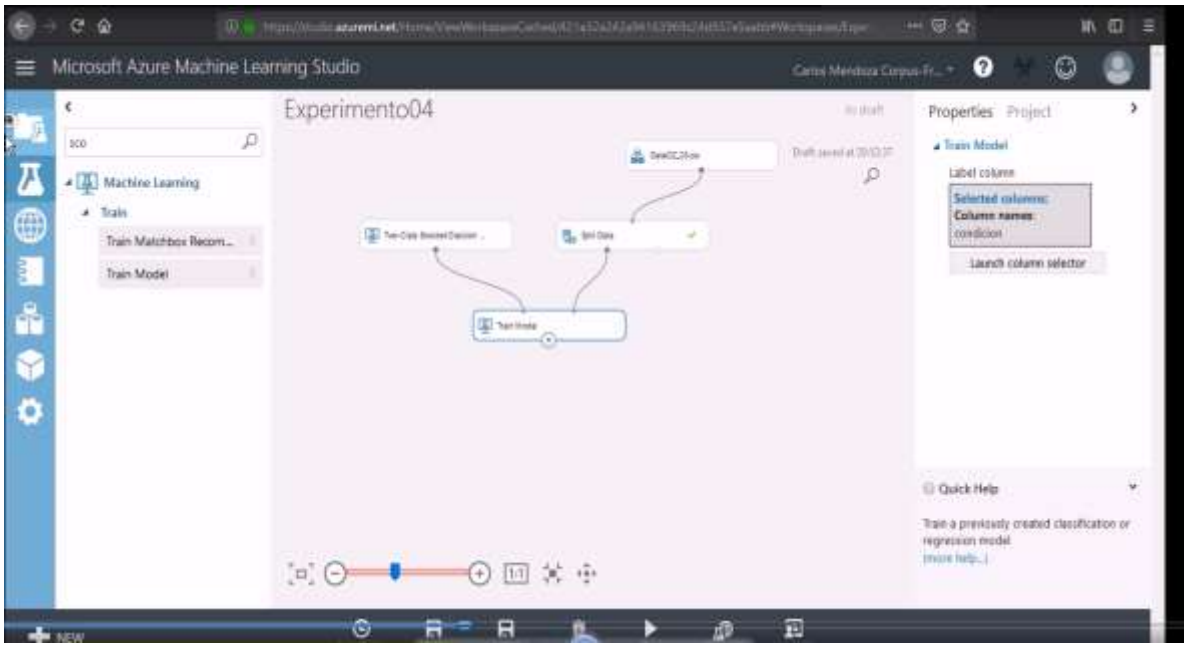


Figura N° 11: Entrenando el Modelo

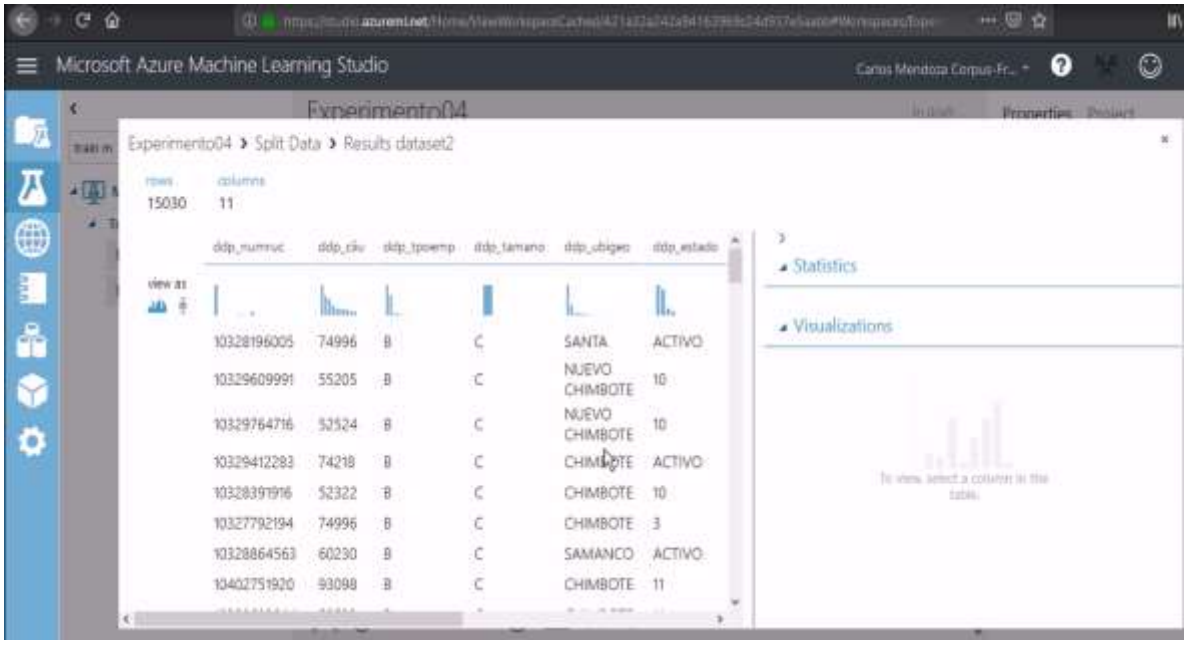


Figura N° 12: Resultados de la prueba

4.7. Implantación

4.7.1. Planificar la implantación

4.7.1.1. Descripción de resultados, modelo y descubrimientos

Los rendimientos logrados después de efectuar las evidencias sobre el modelo de machine learning – Regresión Logística – se logró con exactitud de 88.1594% al 25% del vinculado de datos, de esta manera se valida que el modelo escogido para el presente propósito tiene un grado de predicción conforme con lo que se solicita según los objetivos del negocio.

El modelo de Regresión Logística no es un modelo complejo, pero ayuda al cumplimiento de los objetivos del proyecto, por lo cual fue seleccionado. Este modelo está basado en clasificación de clases (sí, no), además se ha incluido el monto del riesgo de cada Clasificación de alumnos.

Los descubrimientos encontrados están en los patrones entre los idiomas requeridos, la edad de los alumnos y sus calificaciones. En las siguientes gráficas se muestran los patrones de los datos.

4.7.2. Despliegue de la aplicación

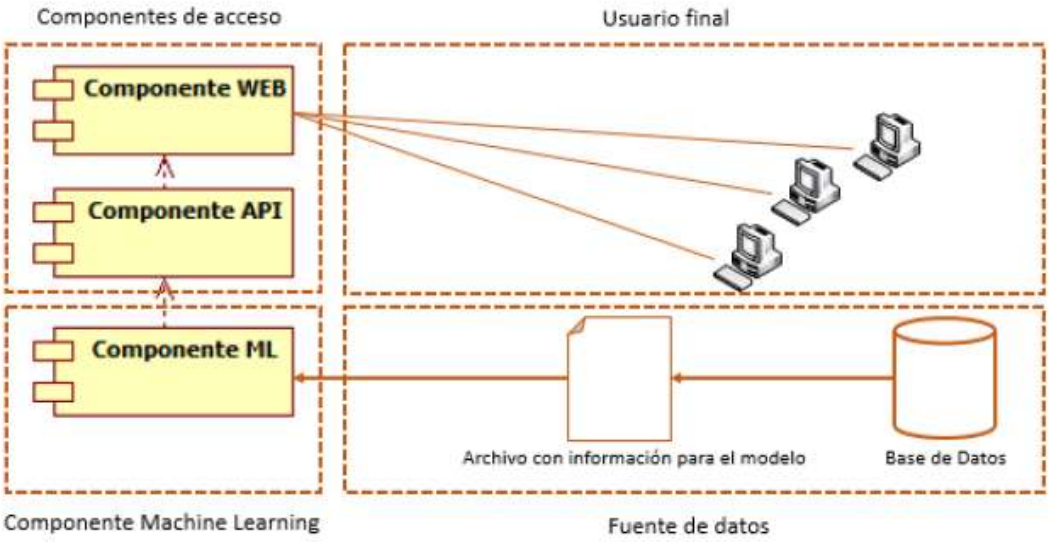


Figura N° 13: Despliegue de la Aplicación

CAPITULO V

PRUEBA DE HIPOTESIS

La Contrastación de Hipótesis se elaboró de conforme al Método Propuesto Pre Test - Pos Test, para poder admitir o rechazar la hipótesis. También, para la ejecución de este diseño se asemejaron indicadores cuantitativos, los cuales se refieren a continuación:

Tabla N° 28: Tipo de Indicadores

INDICADOR	TIPO
Tiempo Promedio de realizar la clasificación Académica	Cuantitativo
Número de Clasificaciones académicas Acertadas	Cuantitativo

5.1. Prueba de Hipótesis para el indicador de tiempo de Realizar la Clasificación académica.

5.1.1. Definición de Variables

T_a = Tiempo promedio antes de emplear el modelo de Machine Learning

T_d = Tiempo promedio después de emplear el modelo de Machine Learning

5.1.2. Hipótesis Estadística

- **Hipótesis H_o=** El tiempo promedio de Clasificación antes de emplear el modelo de Machine Learning es inferior al tiempo de Clasificación posteriormente de aplicar el modelo de Machine Learning (Minutos)

$$H_o = T_a - T_d \leq 0$$

- **Hipótesis H_a =** El tiempo de Clasificación antes de emplear el modelo de Machine Learning es superior que el tiempo de Clasificación después de emplear el modelo de Machine Learning (Minutos)

$$H_a = T_a - T_d > 0$$

5.1.3. Nivel de Significancia

Se especifica un nivel de confiabilidad de 95%.

Utilizando el nivel de significancia (∞=0.05) del 5 % será el margen de error.

Por lo ello, el nivel de confianza (1 - ∞ = 0.95) será del 95%.

5.1.4. Estadígrafo de contraste

Tabla N° 29: Tipo de registro de estudiantes (Datos, notas, ponderado)

N°	Pre Test	Post Test	$(T_a - \overline{CA})^2$	$(T_d - \overline{CP})^2$
	T_a	T_d		
01	180	40	1458.75	237.76
02	260	70	1747.78	212.60
03	184	56	1169.20	0.34
04	224	66	33.71	111.95
05	188	62	911.65	43.30
06	232	58	190.62	6.66
07	176	54	1780.30	2.01
08	216	52	4.81	11.69
09	256	42	1429.33	180.08
10	184	56	1169.20	0.34
11	228	58	96.17	6.66
12	260	60	1747.78	20.98
13	180	66	1458.75	111.95
14	204	68	201.46	158.27
15	224	60	33.71	20.98
16	248	54	888.42	2.01

17	244	52	665.97	11.69
18	208	40	103.91	237.76
19	232	44	190.62	130.40
20	184	54	1169.20	2.01
21	192	60	686.10	20.98
22	200	70	331.01	212.60
23	208	62	103.91	43.30
24	220	56	3.26	0.34
25	228	50	96.17	29.37
26	240	46	475.52	88.72
27	252	42	1142.88	180.08
28	244	40	665.97	237.76
29	236	54	317.07	2.01
30	224	60	33.71	20.98
31	208	66	103.91	111.95

- **Cálculo del Promedio**

$$\bar{T} = \frac{\sum_{i=1}^{n} Ti}{n}$$

$$\bar{T}_a = \frac{6764}{31} = 218.19$$

$$\overline{T_d} = \frac{1718}{31} = 55.42$$

- **Varianza**

$$\boldsymbol{\sigma^2} = \frac{\sum_{i=1}^{n}(\boldsymbol{T1} - \bar{\boldsymbol{T}})}{\boldsymbol{n}}$$

$$\sigma Ta^2 = \frac{20410.84}{31} = 658.41$$

$$\sigma Td^2 = \frac{2457.55}{31} = 79.28$$

- **Desviación Estándar**

$$\sigma TA^2 = \sqrt{658.41} = 25.66$$

$$\sigma TP^2 = \sqrt{79.28} = 8.90$$

- **Cálculo de Z**

$$z = \frac{x_2 - x_1}{\sqrt{\frac{S_1^{\,2}}{N_1} + \frac{S_2^{\,2}}{N_2}}}$$

$$z = \frac{218.19 - 55.42}{\sqrt{\frac{658.41}{31} + \frac{79.28}{31}}}$$

$$z = 33.37$$

5.1.5. Región Critica

Para α = 0.05, en la tabla Z, (Anexo N°3) hallamos Zα = 1.645. Entonces la región critica de la prueba es Zα <1.645, ∞>

5.1.6. Conclusión

Punto que Zc = 33.37 calculado, es superior que Zα = 1.645 y existiendo este valor adentro de la región de rechazo > 1.645 <=, Por lo tanto, se rechaza el Ho y por consecuente se acepta Ha. Se finaliza entonces que el tiempo de Clasificación académica de los alumnos posteriormente de aplicar el modelo Machine Learning es inferior al tiempo de Clasificación antes de emplear el modelo con un nivel de error del 5% y un nivel de confianza del 95%.

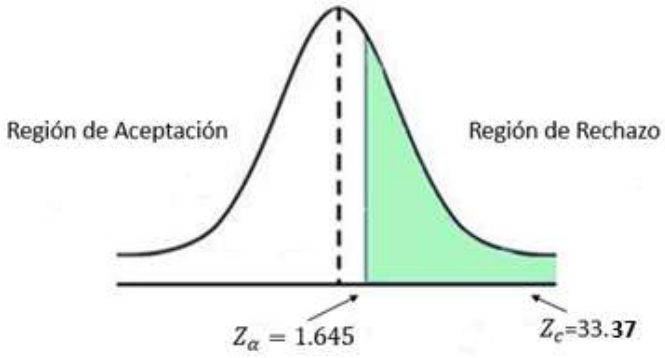


Figura N° 14: Área de Aceptación y Rechazo II (Indicador Cuantitativo)

Fuente: Elaboración Propia

5.2. Prueba de Hipótesis para el indicador de Número de Clasificaciones Acertadas

5.2.1. Definición de Variables

C_A = Número de clasificaciones acertadas

5.2.2. Hipótesis Estadística

Hipótesis H_o = el número de clasificaciones acertadas es superior al 80% de lo evaluado

$$\mathbf{H_o = C_A > 80\%}$$

Hipótesis H_a = el número de clasificaciones acertadas es inferior o igual al 80% de lo evaluado.

$$\mathbf{H_a = C_A \leq 80\%}$$

5.2.3. Nivel de Significancia

Utilizando un nivel de significancia del 5% ($\propto = 0.05$). Por consiguiente, el nivel de confianza ($1-\propto = 0.95$) será del 95%

5.2.4. Estadígrafo de contraste

n = 2000 y sitúo la formula

Para conjeturar el rendimiento de alzo un algoritmo y se sacara un promedio de ello con cada árbol de decisión

Para Registro_notas = 600

Para Registro_alumno = 600

Para Clasificación = 80

Para Ponderado = 20

Tabla N° 30: Tabulación de Clasificación académica

Registro Nota		**Registro Alumnos**		**Clasificación**		**Ponderado**	
Aciertos	**Promedio %**	**Aciertos**	**Promedio %**	**Acierto**	**Promedio %**	**Acierto**	**Promedio %**
495	82.5%	486	81.00%	69	86.25%	17	85.00%
Total	100%	1300					

Calculo C_a:

$$C_a = \frac{\sum_{i=1}^{n\ aciertos}}{N} * 100$$

$$C_a = \frac{1067}{13} * 100$$

$$C_a = 82.08\%$$

5.2.5. Conclusión

Puesto que C_a = 82.08 es superior que 80% por consecuente se acepta la H_o y se rechaza la H_a.

5.3. Grado de Satisfacción

5.3.1. Grado de Satisfacción del Personal Docente

1. **¿Cree Ud. que el desarrollo de un modelo predictivo para el proceso de matrícula de los estudiantes en CEIDUNS será beneficioso?**

Tabla N° 31. Pregunta 01

¿Cree Ud. que el desarrollo de un modelo predictivo para el proceso de matrícula de los estudiantes en CEIDUNS será beneficioso?				
Pregunta	**VALORACION**			**%**
01	1	Siempre	22	92
	2	Casi Siempre	2	8
	3	Rara Vez	0	0
	4	Nunca	0	0
	TOTAL		**24**	**100%**

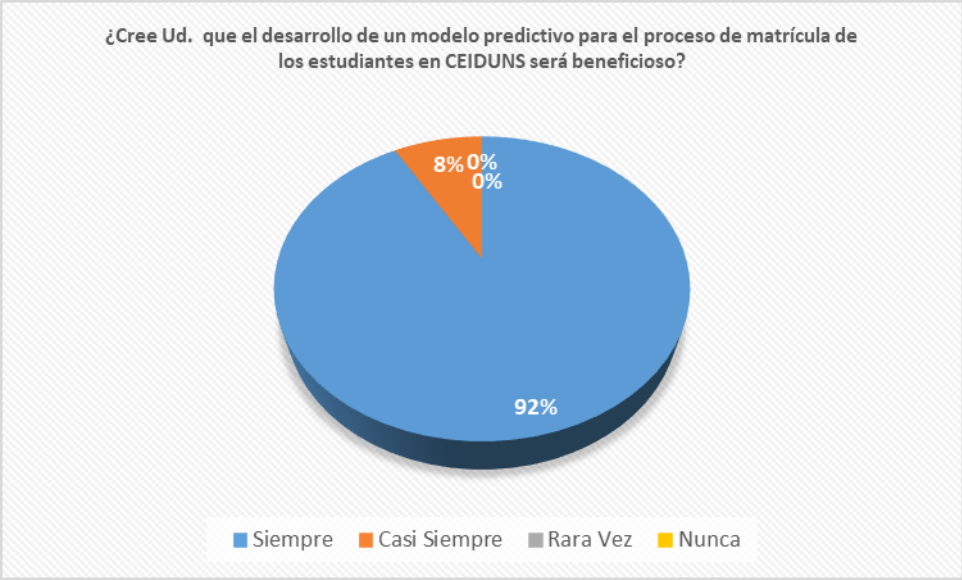


Figura N° 15. Grafico 01

Análisis: En el gráfico se establece que el 92% del personal docente muestran que será beneficioso en el CEIDUNS si prexiste un modelo predictivo hacia la matrícula de sus personificados, el 8% muestran que casi siempre será beneficioso el desarrollo de un modelo predictivo de clasificación de estudiantes y el 0% siempre y casi siempre.

Interpretación: El personal administrativo encuestados dan a conocer que el CEIDUNS no tiene un aplicativo informático que les acceda a inscribir a sus estudiantes de manera veloz, eficiente y sin pérdida de tiempo.

2. **¿Está de acuerdo con la elaboración de un modelo predictivo para la clasificación de estudiantes según su rendimiento académico?**

Tabla N° 32. Pregunta 02

¿Está de acuerdo con la elaboración de un modelo predictivo para la clasificación de estudiantes según su rendimiento académico?				
Pregunta	**VALORACION**			**%**
02	1	Siempre	18	75%
	2	Casi Siempre	4	17%
	3	Rara Vez	1	4%
	4	Nunca	1	4%
	TOTAL		**24**	**100%**

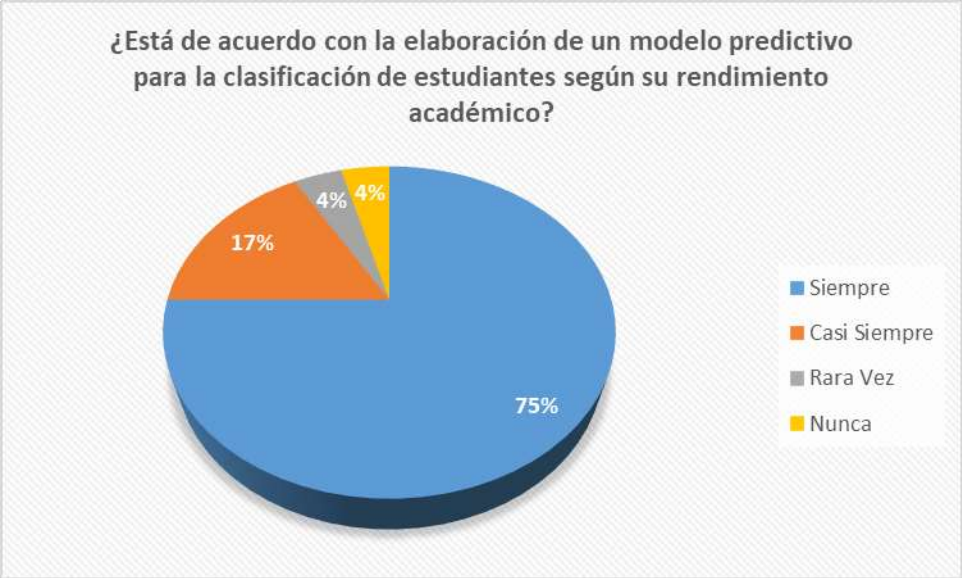


Figura N° 16: Grafico 02

Análisis: En el gráfico de los encuestados muestra que el 75% consta que se encuentran de acuerdo que se procese un modelo predictivo para la clasificación de estudiantes según su rendimiento académico, el 17% que casi siempre, el 4% que rara vez y nunca.

Interpretación: Es vital que se brinde una asistencia de calidad a los estudiantes del CEIDUNS por este motivo existe el apoyo, ya que se dan cuenta de las ventajas que va a obtener el CEIDUNS.

3. **¿Considera necesario la elaboración de un modelo predictivo para la clasificación de estudiantes según su rendimiento académico para mejorar el proceso de matrícula en el CEIDUNS?**

Tabla N° 33. Pregunta 03

¿Considera necesario la elaboración de un modelo predictivo para la clasificación de estudiantes según su rendimiento académico para mejorar el proceso de matrícula en el CEIDUNS?				
Pregunta	**VALORACION**			**%**
03	1	Siempre	19	79%
	2	Casi Siempre	3	13%
	3	Rara Vez	1	4%
	4	Nunca	1	4%
	TOTAL		**24**	**100%**

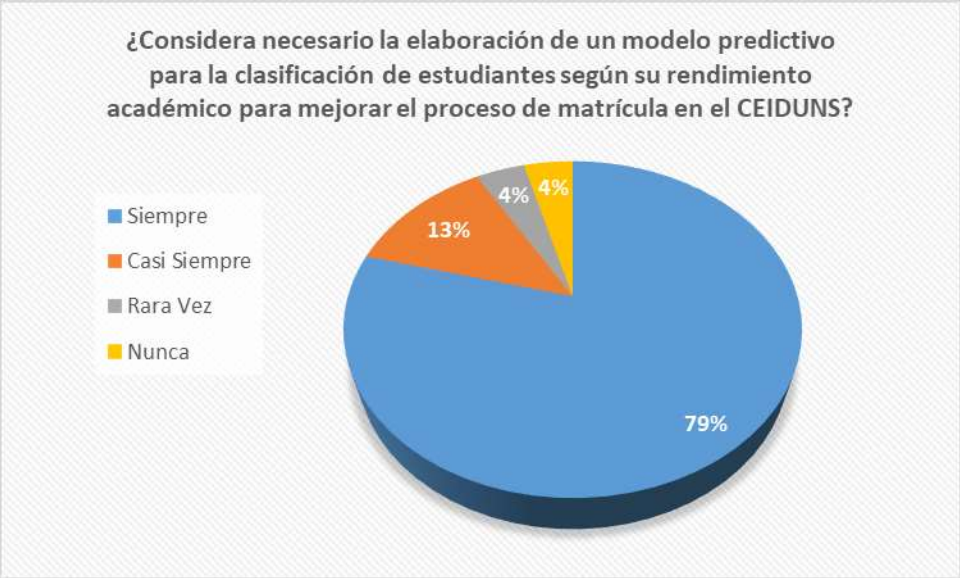


Figura N° 17. Grafico 03

Análisis: En el gráfico de los encuestados indica que el 79% estiman que es sustancial la elaboración de un modelo predictivo para la clasificación de estudiantes según su rendimiento académico para optimizar el proceso de

matrícula en el CEIDUNS, el 13% que casi siempre, el 4% que rara vez y nunca.

Interpretación: Al tener un modelo predictivo se aligerarán los procesos de matrícula de los alumnos y de esta forma crear que el personal administrativo realice sus procesos sin pérdida de tiempo.

4. **¿Cree usted que existe el personal idóneo para manipular el aplicativo informático?**

Tabla N° 34. Pregunta 04

¿ Cree usted que existe el personal idóneo para manipular el aplicativo informático?				
Pregunta	**VALORACION**			**%**
04	1	Siempre	8	33%
	2	Casi Siempre	7	29%
	3	Rara Vez	6	25%
	4	Nunca	3	13%
	TOTAL		**24**	**100%**

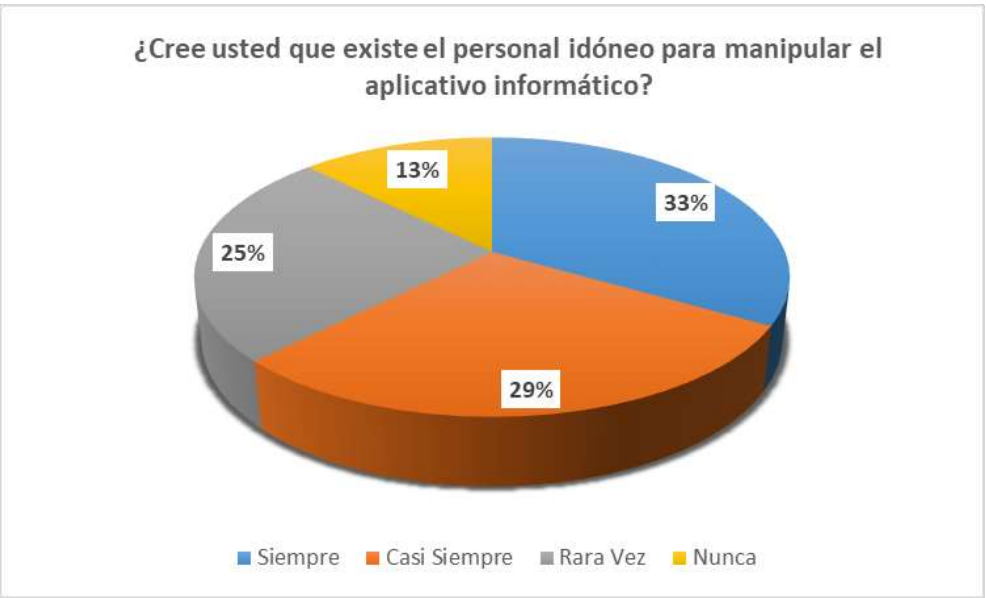


Figura N° 18. Grafico 04

Análisis: En el grafico se muestran como resultado que el 33% de los encuestados piensan que, si prexiste la persona adecuada para manejar el aplicativo informático, 29% que casi siempre, el 25% que rara vez y 13% que nunca.

Interpretación: Toda institución tiene que tener con una persona encargada de las Tics, preexistiendo en la realización de los diferentes trabajos informáticos.

5. **¿Se llevará un mejor control de los estudiantes matriculados con la elaboración de un modelo predictivo?**

Tabla N° 35. Pregunta 05

¿ Se llevará un mejor control de los estudiantes matriculados con la elaboración de un modelo predictivo?				
Pregunta	**VALORACION**			**%**
05	1	Siempre	19	79%
	2	Casi Siempre	3	13%
	3	Rara Vez	2	8%
	4	Nunca	0	0%
		TOTAL	**24**	**100%**

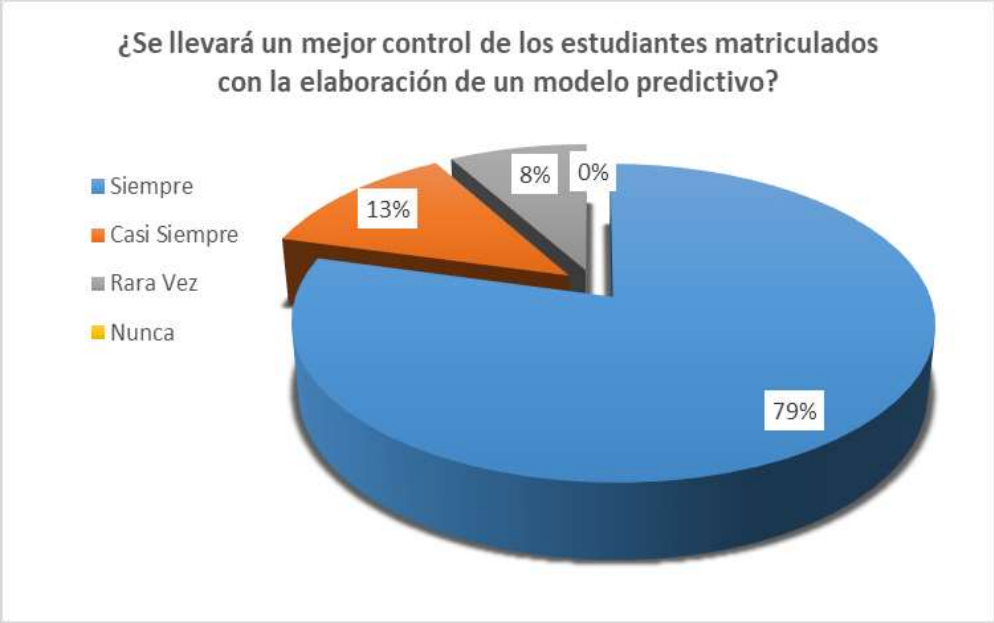


Figura N° 19. Grafico 05

Análisis: En el gráfico estadístico nos muestra que el 79% del personal docente piensan que el modelo predictivo permitirá verificar los archivos del CEIDUNS, el 13% muestran que casi siempre; el 8% que rara vez y el 0% que nunca.

Interpretación: Es primordial un modelo predictivo de matrícula de los que se almacenarán la información de los alumnos en un medio digital, para ejecutar consultas apropiadas de los datos inscritos en la computadora.

6. **¿Cree usted que con el modelo predictivo de matrícula mejorará la atención a los usuarios?**

Tabla N° 36. Pregunta 06

¿Cree usted que con el modelo predictivo de matrícula mejorará la atención a los usuarios?				
Pregunta	**VALORACION**			**%**
06	1	Siempre	20	83%
	2	Casi Siempre	3	13%
	3	Rara Vez	1	4%
	4	Nunca	0	0%
	TOTAL		**24**	**100%**

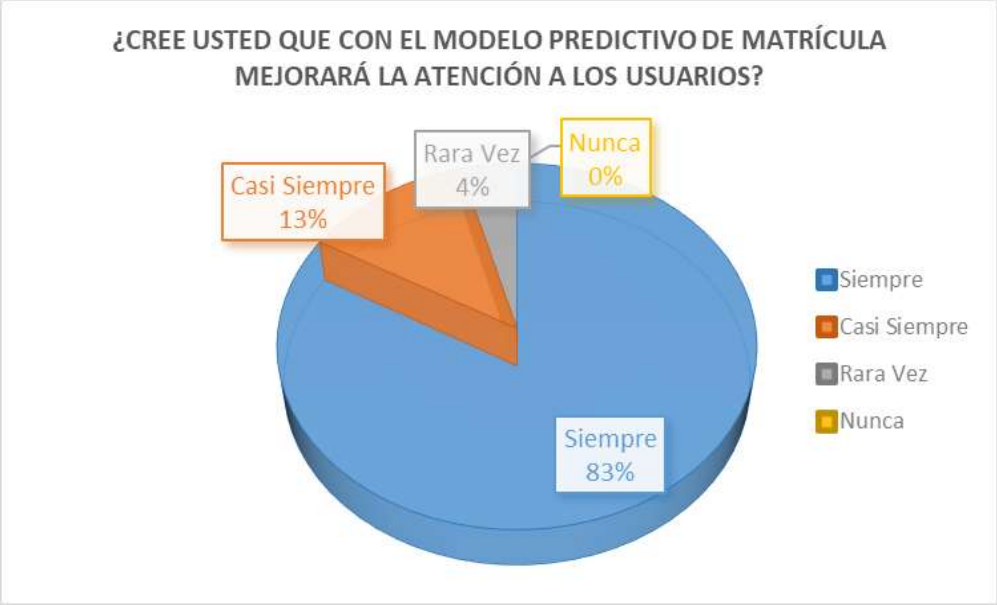


Figura N° 20. Grafico 06

Análisis: El gráfico estadístico muestra que el 83% de encuestados se encuentran de acuerdo que este modelo predictivo optimizará la atención al usuario, el 13% que casi siempre, el 4% que rara vez y el 0% que nunca.

Interpretación: Un modelo predictivo ofrecerá una mejora en el servicio a la Sociedad estudiantil, recopilando la información en un medio digital y agilizando el trabajo de la persona delegada al momento de ejecutar una consulta.

7. **¿Será necesaria la capacitación de la secretaria para el uso del modelo predictivo en el proceso de clasificación de los estudiantes?**

Tabla N° 37. Pregunta 07

¿ Será necesaria la capacitación de la secretaria para el uso del modelo predictivo en el proceso de clasificación de los estudiantes?				
Pregunta	**VALORACION**			**%**
7	1	Siempre	12	50%
	2	Casi Siempre	7	29%
	3	Rara Vez	4	17%
	4	Nunca	1	4%
	TOTAL		**24**	**100%**

Figura N° 21. Grafico 07

Análisis: En el gráfico porcentual se observa que el 50% de encuestados piensan que se debe instruir a la secretaria, el 29% casi siempre, el 17% que rara vez y el 4% nunca.

Interpretación: Al efectuarse un modelo predictivo, se tiene que enseñar a la secretaria ya que va a ser la delegada para de registrar todos los datos de los alumnos.

8. **¿Cambiará el control administrativo con el desarrollo del modelo predictivo de clasificación de estudiantes del CEIDUNS?**

Tabla N° 38. Pregunta 08

¿ Cambiará el control administrativo con el desarrollo del modelo predictivo de clasificación de estudiantes del CEIDUNS?				
Pregunta	**VALORACION**			**%**
8	1	Siempre	19	79%
	2	Casi Siempre	4	17%
	3	Rara Vez	1	4%

	4	Nunca	0	0%
	TOTAL		**24**	**100%**

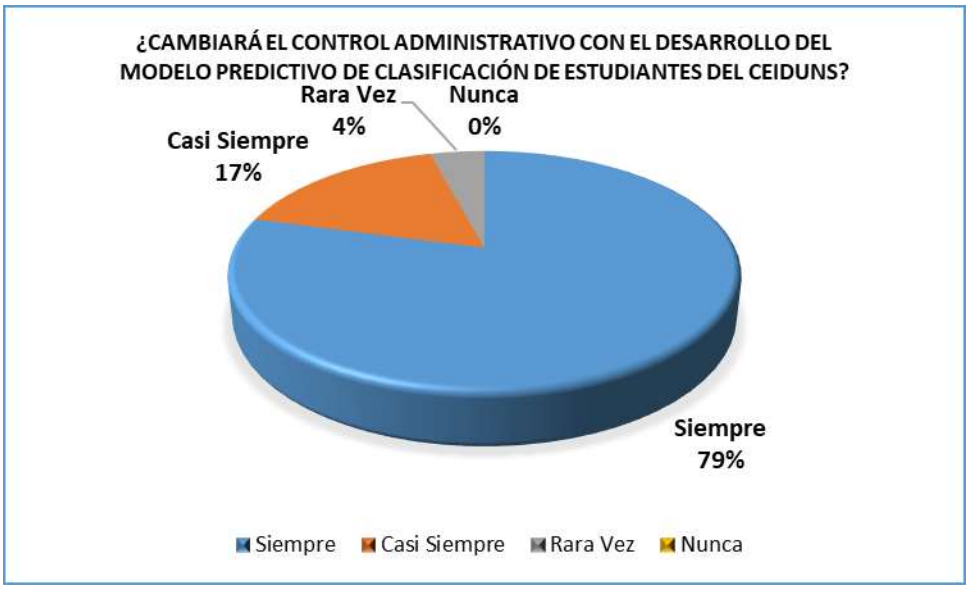


Figura N° 22. Grafico 08

Análisis: En el gráfico se plasma que el 79% del personal docente piensan que existirá un cambio en el proceso, el 17% que casi siempre, el 4% que rara vez y el 0% que nunca.

Interpretación: Al localizar la información ingresadas en medios digitales será más sencillo de controlar y encontrar información en cualquier instante, renovando el control administrativo.

RESULTADOS DE LA ENTREVISTA DIRIGIDAS A ESTUDIANTES

1. **¿Consideras necesario que se implemente una clasificación de estudiantes según rendimiento académico en el CEIDUNS?**

Tabla N° 39. Pregunta 09

¿ Consideras necesario que se implemente una clasificación de estudiantes según rendimiento académico en el CEIDUNS?				
Pregunta	**VALORACION**			**%**
1	1	Siempre	29	66%
	2	Casi Siempre	11	25%
	3	Rara Vez	4	9%
	4	Nunca	0	0%
	TOTAL		**44**	**100%**

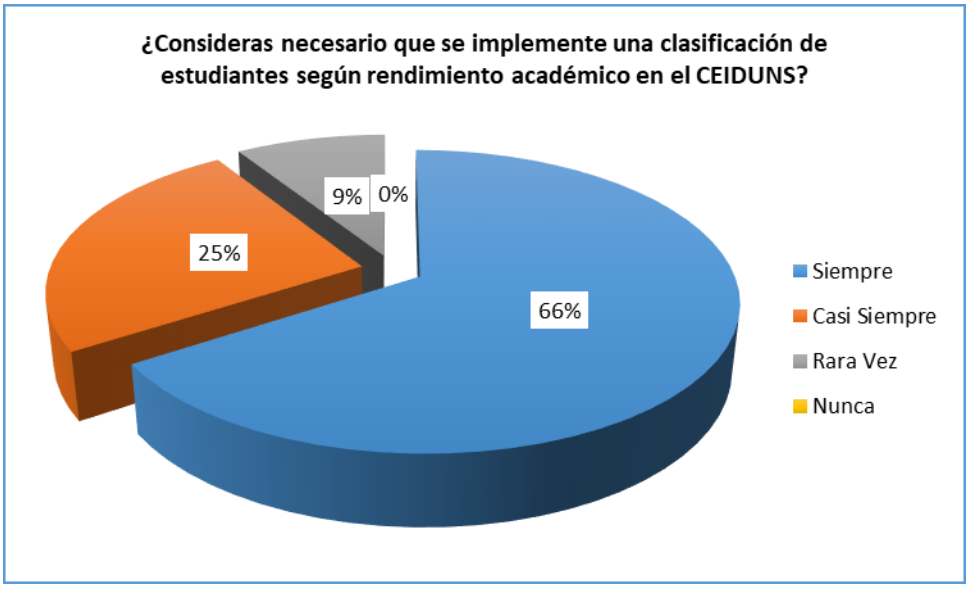


Figura N° 23. Gráfico 09

Análisis: El gráfico estadístico da como resultado que el 66% de los alumnos está de acuerdo que se realice una clasificación de estudiantes según rendimiento académico en el CEIDUNS, el 25% casi siempre, el 9% rara vez y el 0% nunca.

Interpretación: Para poder tener los datos protegidos en un ambiente seguro, existe la necesidad de realizar una clasificación de estudiantes dependiendo del rendimiento académico en el CEIDUNS.

2. **¿Crees qué se está utilizando la tecnología al implementar un modelo predictivo de clasificación de estudiantes de acuerdo a su rendimiento académico?**

Tabla N° 40. Pregunta 10

¿ Crees qué se está utilizando la tecnología al implementar un modelo predictivo de clasificación de estudiantes de acuerdo a su rendimiento académico?				
Pregunta	**VALORACION**			**%**
2	1	Siempre	28	64%
	2	Casi Siempre	13	29%
	3	Rara Vez	3	7%
	4	Nunca	0	0%
	TOTAL		**44**	**100%**

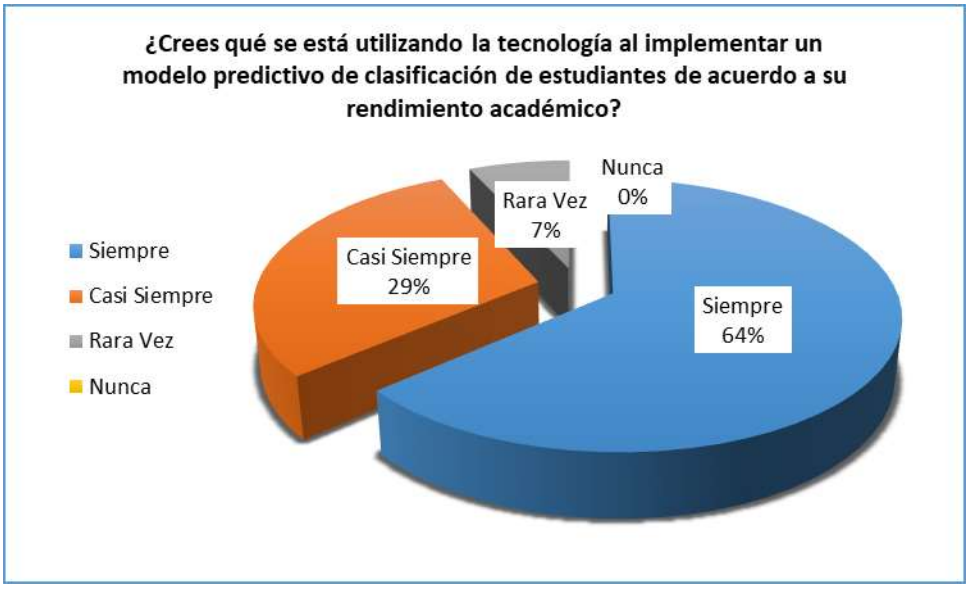


Figura N° 24. Grafico 10

Análisis: El gráfico estadístico muestra que de los alumnos encuestados para el 64% se está manejando la tecnología al implementar un modelo predictivo de clasificación de los alumnos de acuerdo a su rendimiento académico, el 29% casi siempre, el 7% rara vez y el 0% nunca.

Interpretación: Actualmente es fundamental la práctica de la tecnología en un Centro de estudios principalmente en el campo administrativo para compensar las insuficiencias de los usuarios que acuden a la Institución

3. **¿El modelo predictivo de clasificación de estudiantes ayudará a prestar una mejor atención a los estudiantes?**

Tabla N° 41. Pregunta 11

¿ El modelo predictivo de clasificación de estudiantes ayudará a prestar una mejor atención a los estudiantes?				
Pregunta	**VALORACION**			**%**
3	1	Siempre	29	66%
	2	Casi Siempre	11	25%
	3	Rara Vez	3	7%
	4	Nunca	1	2%
		TOTAL	**44**	**100%**

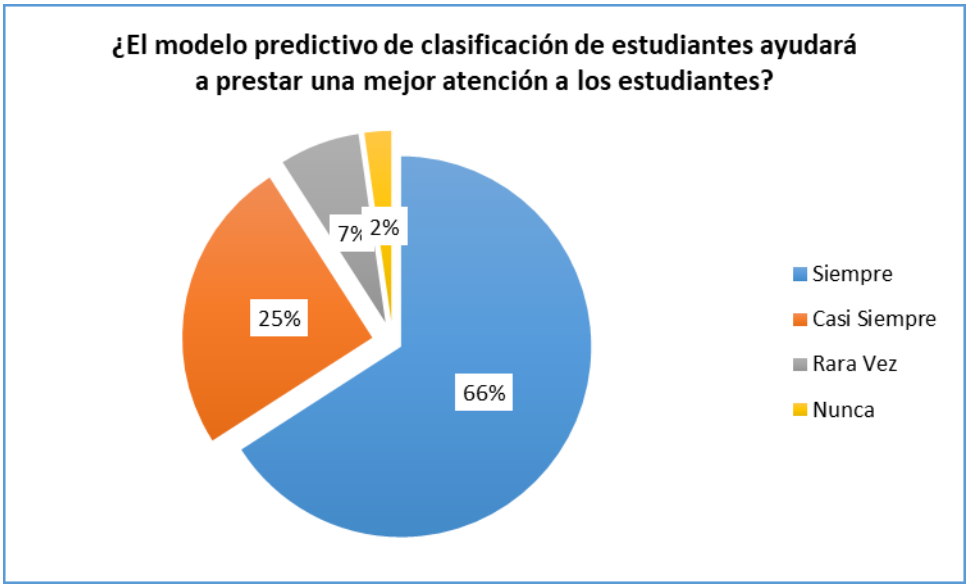


Figura N° 25. Grafico 11

Análisis: El grafico estadístico muestra que los alumnos encuestados dicen que el 66% está de acuerdo que El modelo predictivo de clasificación de estudiantes favorecerá el proceso de matrícula a sus representantes, el 25% que casi siempre, el 7% que rara vez y el 2% nunca.

Interpretación: Actualmente es necesario que los institutos cuenten con un ambiente computarizado y ofrezca bienestar a los estudiantes y administrativos.

4. **¿Es beneficioso el uso de la tecnología en el proceso de clasificación de estudiantes según se rendimiento académico en el CEIDUNS?**

Tabla N° 42. Pregunta 12

¿ Es beneficioso el uso de la tecnología en el proceso de clasificación de estudiantes según se rendimiento académico en el CEIDUNS?				
Pregunta	**VALORACION**			**%**
4	1	Siempre	35	80%
	2	Casi Siempre	7	16%
	3	Rara Vez	2	4%
	4	Nunca	0	0%
	TOTAL		**44**	**100%**

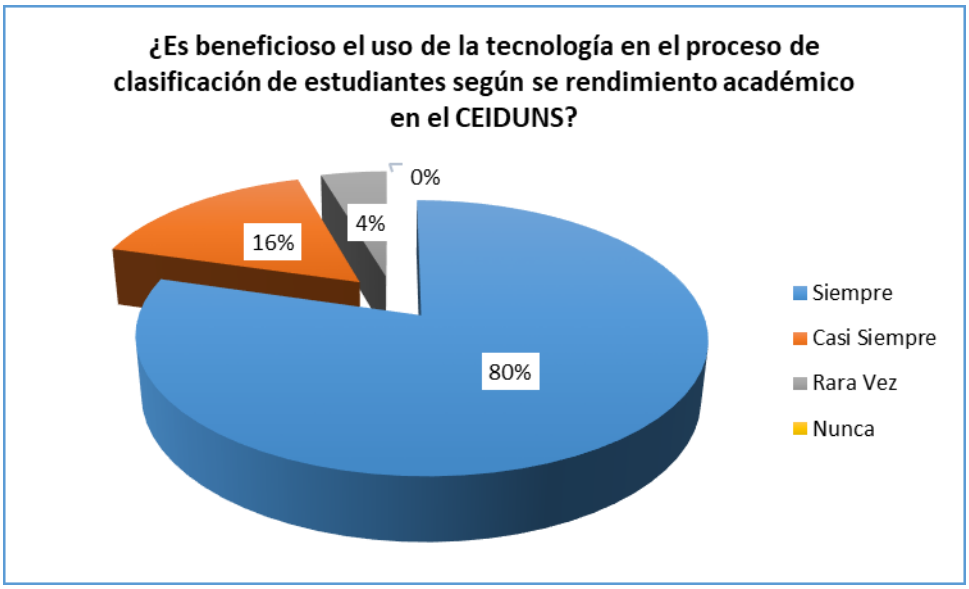


Figura N° 26. Grafico 12

Análisis: El gráfico de estadística nos muetsra que el 80% de los alumnos encuestados se encuentran de acuerdo que la tecnología es favorable para el proceso de clasificación de estudiantes según su rendimiento académico, el 16% casi siempre, el 4% que rara vez y el 0% que nunca.

Interpretación: Al implementar un modelo predictivo de clasificación de estudiantes según su rendimiento académico en el CEIDUNS se está realizando la utilización de las tecnologías para mejorar el proceso de matrícula y brindar un servicio superior a la comunidad.

5. ¿El desarrollo de un modelo predictivo de clasificación de estudiantes según rendimiento académico favorece al CEIDUNS como a los estudiantes?

Tabla N° 43. Pregunta 13

¿ El desarrollo de un modelo predictivo de clasificación de estudiantes según rendimiento académico favorece al CEIDUNS como a los estudiantes?				
Pregunta	**VALORACION**			**%**
6	1	Siempre	20	45%
	2	Casi Siempre	17	39%
	3	Rara Vez	7	16%
	4	Nunca	0	0%
	TOTAL		**44**	**100%**

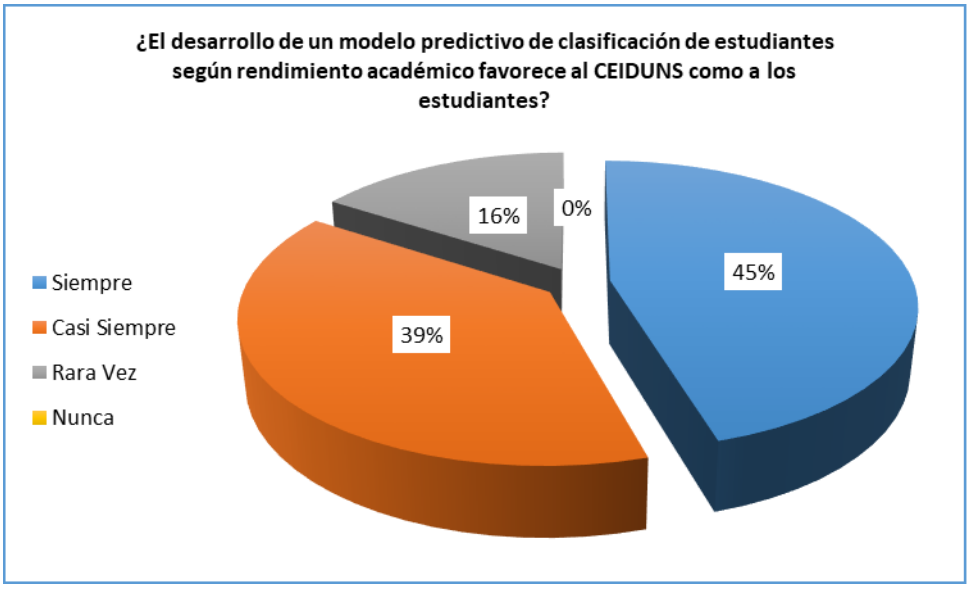


Figura N° 27. Grafico 13

Análisis: En el gráfico estadístico nos muestra que el 45% de alumnos encuestados considerando que en el desarrollo de un modelo predictivo de clasificación de estudiantes según rendimiento académico favorece a la corporación estudiantil, el 39% casi siempre, el 16% rara vez y el 0% nunca.

Interpretación: En el momento de desarrollar un modelo predictivo logra acceder a la agilización de los trámites internos del CEIDUNS, se consigue un progreso en el sistema educativo.

6. **¿La automatización de la clasificación de los estudiantes será un avance en el sistema institucional?**

Tabla N° 44. Pregunta 14

¿ La automatización de la clasificación de los estudiantes será un avance en el sistema institucional?				
Pregunta	**VALORACION**			**%**
7	1	Siempre	34	77%
	2	Casi Siempre	8	18%
	3	Rara Vez	2	5%
	4	Nunca	0	0%
	TOTAL		**44**	**100%**

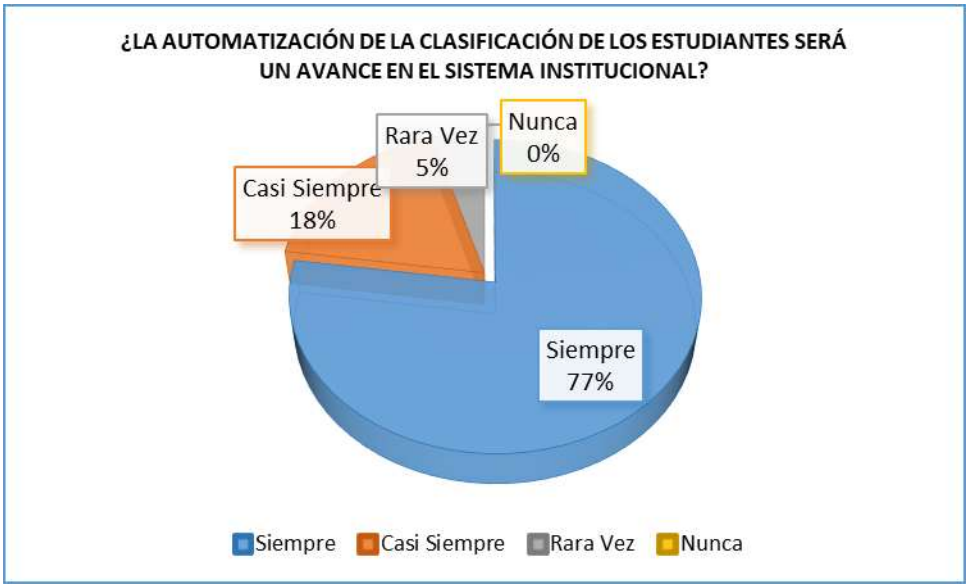


Figura N° 28. Grafico 14

Análisis: En el gráfico estadístico nos indica que el 77% de encuestados están conforme con la sistematización es un progreso en el sistema educativo, el 18% que casi siempre, el 5% que rara vez y 0% nunca.

Interpretación: Al ejecutarse un método de clasificación de estudiantes según su rendimiento académico se está avanzando y se da un paso más adelante en la educación; fundamentalmente en los ambientes administrativos.

CAPITULO VI

RESULTADOS Y DISCUSIÓN

6.1. Indicadores Cuantitativos

6.1.1. Tiempo Promedio de realizar la clasificación Académica

Tabla N° 45. Indicador de Tiempo Promedio de realizar la clasificación Académica

Indicador	Tiempo Promedio del Sistema Actual	Tiempo Promedio del Sistema Propuesto	Tiempo Ganado (Segundos)	Porcentaje Ganado
Tiempo promedio de realizar la clasificación Académica	218.19	55.42	162.77	74.60

Fuente: Elaboración Propia

Tabla N° 46. Análisis Estadístico del Indicador de Tiempo Promedio de realizar la clasificación Académica

Indicador	Antes			Después			Zα	Zc	Conclusión
	n_A	$\overline{T}_A$	σTA^2	n_P	$\overline{T_P}$	σTP^2			
Tiempo promedio de realizar la clasificación Académica	31	218.19	658.41	31	55.42	79.28	1.645	33.37	Se Acepta

Fuente: Elaboración Propia

Discusión

La gráfica consecutiva nos indica que los tiempos y los porcentajes conseguidos del análisis del indicador Tiempo Promedio de realizar la clasificación Académica, de donde podemos observar que el Tiempo Promedio de realizar la clasificación Académica con el sistema actualmente es de 218.19 segundos y con el sistema propuesto es de 55.42 segundos logrando un tiempo ganado de 162.77 segundos.

Podemos concluir que se ha logrado obtener un 74.60 % de disminución en el Tiempo Promedio de realizar la clasificación Académica; lo que logrará tener la información actualizada, en tiempo real para su respectivo análisis que influya en la toma de juicios en CEIDUNS.

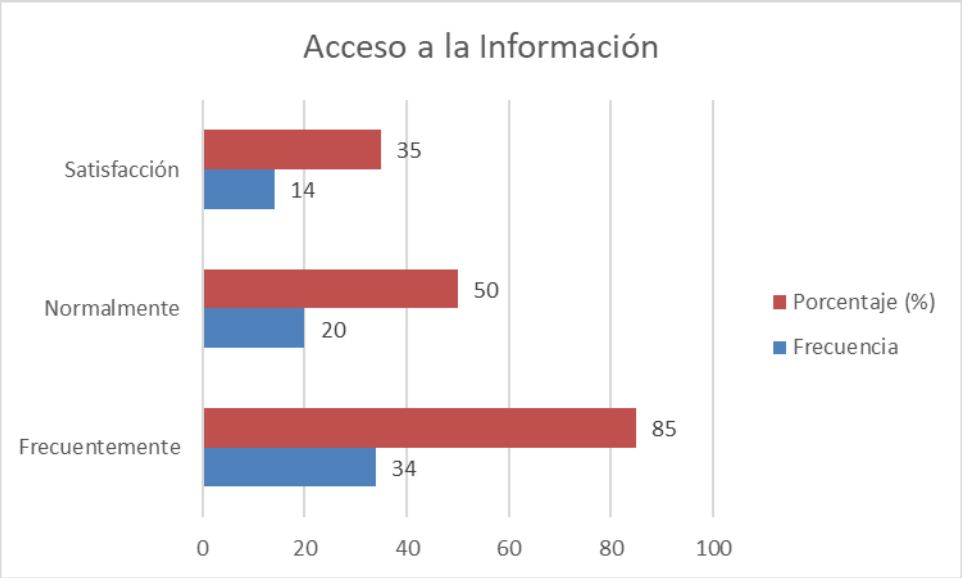


Figura N° 29. Diagrama de Barra Indicador Tiempo Promedio de realizar la clasificación Académica

Fuente: Elaboración Propia

6.1.2. Diagnósticos Acertados

Referente a este indicador se observa que el dígito de aciertos por cada uno de los item sobrepasa el 80% dando como proporción final cociente de 82.08%, por lo consiguiente, el modelo predictivo de clasificación de estudiantes aplicado es admisible de acuerdo al cuadro de Operacionalización de variable.

6.2. Indicadores Cualitativos

6.2.1. Grado de Satisfacción del Personal Docente

Tabla N° 47. Indicador de Grado de Satisfacción del Personal Docente

VALOR CUALITATIVO	P1	P2	P3	P4	P5	P6	P7	P8	VALOR	Frecuencia
Muy Satisfecho	22	18	19	08	19	20	12	19	17.125	71.35 %
Satisfecho	02	04	03	07	03	03	07	04	4.125	17.19 %
Poco Satisfecho	00	01	01	06	02	01	04	01	2.00	8.33 %
Nada Satisfecho	00	01	01	03	00	00	01	00	0.75	3.13 %
Total	**24**	**24**	**24**	**24**	**24**	**24**	**24**	**24**	**24**	**100 %**

Discusión

Al Analizar este indicador, podemos concluir que el desarrollo de un modelo predictivo de clasificación de estudiantes propuesto es la mejor solución para la satisfacción del personal docente del Centro de Idiomas de la Universidad Nacional del Santa, donde el índice de satisfacción fluctúa entre 71.35% al 88.54%.

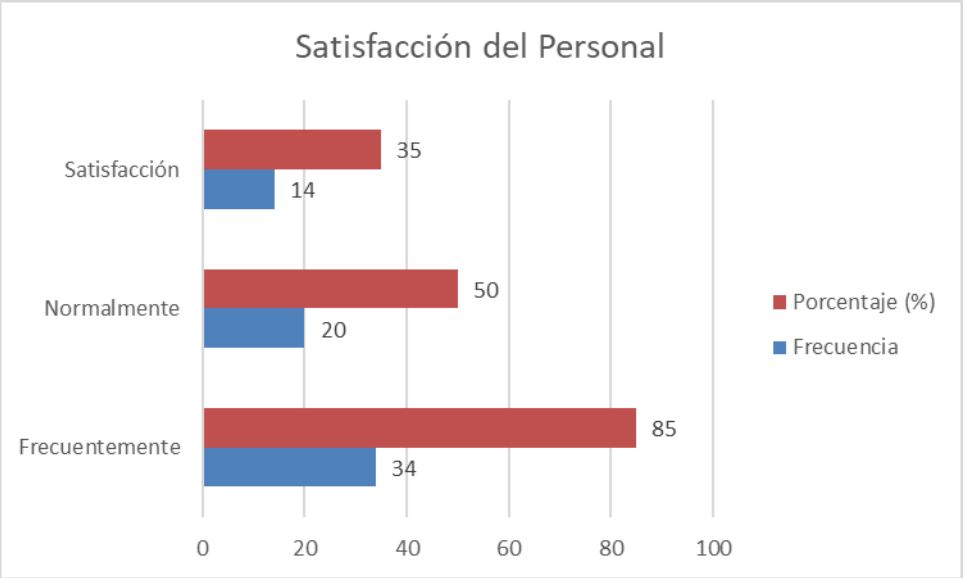


Figura 30. Diagrama de Barra Indicador Satisfacción del Personal Docente

6.2.2. Grado de Satisfacción del Estudiante

Tabla N° 48. Indicador de Grado de Satisfacción del Estudiante

VALOR CUALITATIVO	P1	P2	P3	P4	P5	P6	VALOR	Frecuencia
Muy Satisfecho	29	28	29	35	20	34	29.17	66.30 %
Satisfecho	11	13	11	07	17	08	11.17	25.39 %
Poco Satisfecho	04	03	03	02	07	02	3.50	7.95 %
Nada Satisfecho	00	00	01	00	00	00	0.17	0.39 %
Total	**44**	**44**	**44**	**44**	**44**	**44**	**44**	**100 %**

Fuente: Elaboración Propia

Discusión

Al Analizar este indicador, podemos concluir que el desarrollo de un modelo predictivo de clasificación de estudiantes según su rendimiento académico es la mejor solución para la satisfacción del estudiante del

Centro de Idiomas de la Universidad Nacional del Santa, donde el índice de satisfacción de satisfacción fluctúa entre 66.30% al 91.69%.

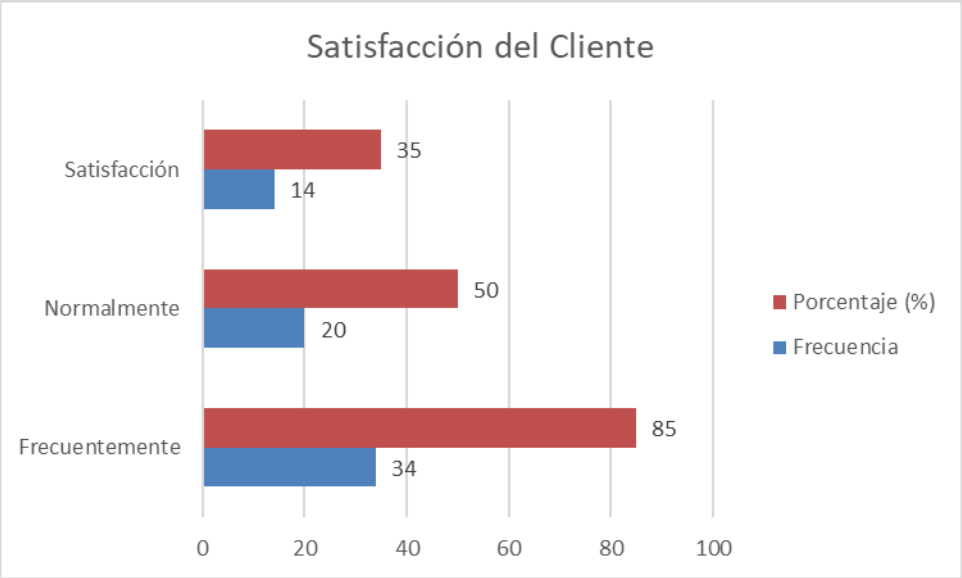


Figura N° 31. Diagrama de Barra Indicador Satisfacción del Estudiante

Fuente: Elaboración Propia

CONCLUSIONES

- Se logró mejorar la clasificación de estudiantes de acuerdo a su rendimiento académico mediante el desarrollo de un modelo de Machine Learning en el Centro de Idiomas de la Universidad Nacional del Santa; en el cual se logró una exactitud de 88.1594% de la vinculación de datos requeridas.
- El modelo utilizado fue de regresión logística que permitió el cumplimiento de los objetivos de la investigación. Este modelo está basado en la clasificación de clases. Además, tiene un alto grado de predicción.
- Se utilizó la metodología CRISP para cada una de las fases en el Modelo de Machine Learning de Clasificación de Estudiantes.
- El tiempo promedio de clasificación académica de estudiantes disminuyó en un 74.60%. Con el sistema actual se utilizaban en promedio 218.19 segundos y con el modelo predictivo 55.42 segundos para obtener la proyección de la clasificación de estudiantes. Por lo que se obtuvo un tiempo ganado de 162.77 segundos en razón de reemplazar el proceso manual por un modelo predictivo que facilita en realizar este proceso que será muy beneficioso tanto para el estudiante al matricularse como para el personal docente y administrativo del Centro de Idiomas de la Universidad Nacional del Santa.
- El número de clasificaciones académicas acertadas fue de 82.08%, por lo que se demuestra la consistencia de datos, confiabilidad y usabilidad de Machine Learning por contar con información confiable. En tal sentido, la eficacia del personal del CEIDUNS ha aumentado debido a que el modelo predictivo permite prever la asignación de salones y docentes para cada ciclo académico.
- Se logró aumentar la satisfacción del personal docente en un 71.35% y los estudiantes en un 66.30% quienes a través de las encuestas manifiestan que la

utilización del modelo predictivo influye de manera positiva en la toma de decisiones.

- Se logró realizar la factibilidad económica con una recuperación de la inversión de unos 10 meses aproximadamente.

RECOMENDACIONES

- La clasificación de las estudiantes según su rendimiento académico es un problema que va en extendiendo, pero como se ha verificado durante el desarrollo de esta investigación desde distintas perspectivas, logramos contribuir soluciones propensas a la mejora continua del proceso de clasificación para el ingreso de los estudiantes en el centro de Idiomas.
- Con respecto a la seguridad de los datos, se debe de tomar en cuenta las normas para definir contraseñas de los usuarios, hacer Backus periódicos de la información.
- Se debe capacitar el personal a fin de garantizar la eficiencia en su trabajo. De igual manera en lo que respecta a la calidad de atención académica
- Se recomienda agregar más reportes mensuales y anuales de los alumnos al módulo de reportes, con el fin de llevar un mejor el registro de alumnos de CEIDUNS.
- Estimar la posibilidad de liberar la aplicación y sea utilizada por otras empresas utilizando la tecnología Cloud Computing.
- El proceso de clasificación de estudiantes por rendimiento académico podría tener controles y métricas relacionadas a la mejora que se obtiene según nuevas medidas adoptadas, en este caso, el sistema de predicción.

BIBLIOGRAFÍA

Alvarado, S. (2016). *"El Desarrollo - Progreso"*. Obtenido de http://alvaradosandra0908.blogspot.pe/2016/03/desarrollo-desarrollo-progreso.html

Benavides Gaibor, L. (2011). *"Gestión, liderazgo y valores en la administración de la unidad educativa "san juan de bucay" del Canton general Antonio Elizalde (bucay). Durante periodo 2010 -2011"*. Obtenido de Universidad Técnica Particular de Loja – Guayaquil: http://dspace.utpl.edu.ec/bitstream/123456789/2039/3/Benavides_Gaibor_Luis_Hernan.pdf

Camborda Zamudio, M. (2014). *"Aplicación de Árboles de decisión para la predicción del Rendimiento Académico de los Estudiantes de los primeros ciclos de la carrera de Ingenieria Civil de la Universidad Continental"*. Obtenido de Universidad Nacional del Centro del Perú: http://repositorio.uncp.edu.pe/bitstream/handle/UNCP/1477/Tesis%20Maria%20Gabriela%20Camborda%20Zamudio.pdf?cv=1&isAllowed=y&sequence=1

Daw, N. (2013). *"Reinforcement learning models of the dopamine system and their behavioral implications"*. Obtenido de Carnegie Mellon University: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.94.6376&rep=rep1&ty

Dhar, V. (2013). *"Data Science and Prediction"*. 56(12), 64-73: Communications of the ACM.

Díaz Tello, S. (2014). *"Sistema Integral Bajo el enfoque de minería de datos y redes neuronales para predicción y control de la contaminación atmosférica por pm10 en la Ciudad de Chimbote"*. Obtenido de Universidad Nacional del Santa: http://repositorio.uns.edu.pe/bitstream/handle/UNS/3036/42884.pdf?sequence=1&isAllowed=y

García Gazabón, G. (2014). *"Modelo de Machine Learning para la Clasificación de pacientes en términos del nivel asistencial requerido en una urgencia pediátrica con Área de Cuidados Mínimos"*. Obtenido de Universidad Tecnologica de Bolivar: https://biblioteca.utb.edu.co/notas/tesis/0068210.pdf

Geras, K. J. (2001). *"Prediction Markets for Machine Learning - Artificial Intelligence"*. University of Warsaw.

Hernández Sampieri, R. (2006). *"Metodología de la Investigación"*. ISBN 9789701057337: Editorial Mc Graw Hill. Interamericana de México. .

Hernández, R., Fernández, C., & Baptista, P. (2006). *"Metodología de la Investigación"*. 5ta.ed. Mexico: D.F.: McGRAW-HILL, 2010. ISBN: 978-607-15-0291-9.

Medrano Parado, S. (2016). *"Modelo de minería de datos usando machine learning con reconocimiento de patrones de síntomas y enfermedades respiratorias en las historias clínicas para mejorar el diagnóstico de pacientes en la ciudad de Trujillo 2016"*. Obtenido de Universidad César Vallejo: http://repositorio.ucv.edu.pe/handle/UCV/9852

Mined. (2002). *"Lineamientos para la Evaluación del Aprendizaje en Educación Media, San salvador".* Ministerio de Educación de el Salvador: Primera Edición, Editorial Algier.

Paredes Dulce, E. (2015). *licación del examen internacional ket (key english test) A estudiantes, ciclo básico del centro de idiomas de la Universidad nacional del santa y el nivel de competencia Comunicativa, 2014"*. Obtenido de Universidad Nacional del Santa: http://repositorio.uns.edu.pe/bitstream/handle/UNS/2884/42767.pdf?cv=1&sequence=

Pizarro, R., & Clark, S. (1998). *"Currículo del hogar y aprendizajes educativos. Interacción versus estatus".* 7, 25-33: Revista de Psicología de la Universidad de Chile.

Rico, A., & D., S. (2018). *"Diseño de un modelo para automatizar la predicción del rendimiento académico en estudiantes del IPN"*. Obtenido de Instituto Politécnico Nacional – Mexico: http://www.scielo.org.mx/pdf/ride/v8n16/2007-7467-ride-8-16-00246.pdf

Rienzo, d., & Otros. (2009). *"Estadística para las Ciencias Agropecuarias"*. Impreso en Argentina: Séptima Edición. Editorial Brujas. ISBN 978-987-591-112-3.

Rivero, E. (2017). *"Detección de contenido malicioso mediante técnicas de Machine Learning en las redes sociales"*. Obtenido de Universidad de Buenos Aires. (p.11): http://bibliotecadigital.econ.uba.ar/download/tpos/1502-0560_RiveroE.pdf?fbclid=IwAR2jvyBevYwJP27QgVZuY1Ab66KsmCwq-BfBaJpNLLfYRdOlHx78euHQ2cs

Rodríguez, J., & Miñano, M. (2017). *“Desarrollo de una aplicación informática basada en un modelo de Machine Learning para mejorar la evaluación de préstamos crediticios”*. Obtenido de Universidad Privada del Norte: http://repositorio.upn.edu.pe/handle/11537/12294

Rollins, L. S. (2015). *"Foundational methodology for data science"*. Obtenido de IBM: http:// www01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IMW14824USEN

Saltz, J., Shamshurin, I., & Crowston, K. (2017). *"Comparing data science project management methodologies via a controlled experiment"*. Manoa, Hawaii: Ponencia presentada en la 50th Hawaii International Conference on System Sciences.

Schoenherr, T., & Speier Pero, C. (2015). *"Data science, predictive analytics, and big data iin supply chain management: Current state and future potential"*. Journal of Business Logistics. 36(1), 120-132.

Stephen, M. (2014). *"Chapman & Hall/CRC Machine Learning & Pattern Recognition Series Chapman & Hall/CRC Machine Learning & Pattern Recognition Series Machine Learning An Algoritgmic Persperctive"*. CRC Press. (2nd ed).

Villagrá Arnedo, C. (2015). *“Sistema predictivo progresivo de clasificación probabilística como guía para el aprendizaje”*. Obtenido de Universidad de Alicante: https://rua.ua.es/dspace/bitstream/10045/54256/1/tesis_carlos_j_villagra_arnedo.pdf

Yamao, E. (2018). *"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"*. Obtenido de Universidad San Martin de Porres: http://www.repositorioacademico.usmp.edu.pe/bitstream/handle/usmp/3555/yamao_e.pdf;jsessionid=DEC318E98AF168D43E809385E4A0DD31?sequence=3

ANEXOS

ANEXO 1: ANÁLISIS DE LA FACTIBILIDAD

1. FACTIBILIDAD TÉCNICA

- **Análisis de Software**

  En el desarrollo de este proyecto se tuvo en cuenta, herramientas de software altamente calificados y actualizados. Estas herramientas brindan resultados de buen rendimiento y confianza al momento de realizar sus funciones. Estas herramientas son totalmente administrables por los usuarios de la aplicación y la correcta visualización en cualquier tamaño de pantalla permite su administración más sencilla

  El uso de este tipo de software libre, permite un desarrollo fácil, rápido y eficiente de las aplicaciones. Tienen muy en cuenta las necesidades crecientes de los desarrolladores y al ser libre, no demanda un costo

- **Análisis de Hardware**

  El centro de Idiomas de la Universidad Nacional del Santa cuenta con todos los equipos necesarios para el desarrollo de proyecto, lo cual nos permite obviar la inversión para el aspecto de inversión en hardware.

  De acuerdo al análisis anterior se concluye que: este proyecto es ***técnicamente factible.***

2. FACTIBILIDAD OPERATIVA

El software para realizar el modelo predictivo cuenta con interfaces muy flexibles, adaptables e intuitivas para su correcta comprensión, teniendo en cuenta los requerimientos funcionales y no funcionales para su correcto desarrollo, brindando información detallada y organizada

De acuerdo a lo anteriormente mencionado se concluye que: este proyecto es operacionalmente factible.

3. FACTIBILIDAD ECONÓMICA

3.1. Inversión

A. Hardware

El centro de Idiomas de la Universidad Nacional del Santa ya cuenta con todo el equipo necesario para cumplir con los requerimientos del proyecto. El CEIDUNS cuenta con servicio de internet y electricidad, determinamos que es aspecto no demanda ninguna inversión

Tabla N° 49. Inversión en Hardware

Cantidad	Descripción	Costo (S/.)
2 Unid.	Desktop PC. Intel Core i7 (Sétima Generación), 4.10 GHz, 16 GB de RAM, 1 TB de disco duro. Monitor Samsung de 22" Full HD 1080 x 1920.	0.00
1 Unid.	Modem – Swtich 6 puertos	0.00
	Subtotal:	**0.00**

B. Software

Tabla N° 50. Inversión de Software

N°	Descripción	Costo (S/.)
1	WORDPRESS	0.00
2	PHP	0.00
3	PLUGIN ELEMENTOR	0.00
4	Wireframe.cc	0.00
5	SublimeText (versión de prueba)	0.00
	Subtotal	**0.00**

C. Recursos Humanos

Tabla N° 51. Inversión en Recursos Humanos

N°	Descripción	Costo (S/.)
1	Sitio Web	450.00
	Subtotal	**450.00**

D. Servicios (Costo Anual)

En cuanto a la inversión en servicios, se considera el pago por el dominio y el hosting, los cuales tienen un costo, por ambos servicios, de 700.00 nuevos soles anuales

Tabla N° 52. Inversión en Servicios

N°	Descripción	Costo (S/.)
1	Dominio y Hosting	700.00
	Subtotal	**700.00**

E. Resumen de Costos

Tabla N° 53. Estimación del Proyecto – Costo Total

N°	Concepto	Costo (S/.)
1	Hardware	0.00
2	Software	0.00
3	Recursos Humanos	450.00
4	Servicios	700.00
	Total	**1150.00**

3.2. Costo Anual Operativo

A. Mantenimiento

Teniendo en cuenta posibles cambios en el modelo de negocio o en caso la aplicación requiera algún cambio o actualización, el CEIDUNS debe programar el mantenimiento del sitio web, un promedio de 2 veces al año, lo que representa un costo total de 200.00 nuevos soles anuales

Tabla N° 54. Costo en Mantenimiento

N° Mantenimientos	Costo (S/.)	Total (S/.)
2	100.00	**200.00**

3.3. Beneficios del Proyecto

A. Beneficios Intangibles

- Posicionamiento del CEIDUNS en beneficio de la Universidad Nacional del Santa.
- Desempeño eficiente del personal administrativo y docente.
- Acceso a la información de manera rápida y eficaz

- Mejor atención a los estudiantes, que realizan su matrícula.

B. Beneficios Tangibles

- Brindar información inmediata a los estudiantes
- Mejora en el proceso de búsqueda de empleo de los postulantes

Tabla N° 55. Optimización de los Tiempos

N°	Descripción	Sin el Modelo Predictivo	Con el modelo predictivo	Duración óptima (sin tiempo muerto)
1	Búsqueda de información de estudiantes	4	3	1
2	Clasificación de estudiantes	6	3	3
3	Tiempo de espera por matrícula	7	2	5
Total		**17**	**8**	**9**

En base a los datos obtenidos se nota una mejora significativa en los procesos anteriormente expuestos, y esto se pudo conseguir gracias a la rapidez y facilidad para conseguir dicha información.

- Me: Porcentaje de mejora

$$Me = \left(1 - \frac{8}{17}\right) * 100$$

$$Me = 52.94\%$$

En la ejecución de los procesos se lograría una mejora de 52.94% en relación al tiempo

Do: Porcentaje de mejora en relación a la duración óptima

$$Do = \left(\frac{9}{17}\right) * 100$$

$$Do = 52.94\% \mid$$

La ejecución de los procesos sin la aplicación web y en relación al tiempo óptimo de ejecución es de 52.94%

$$Do = \left(\frac{9}{8}\right) * 100$$

$$Do = 99.12\%$$

La ejecución de los procesos con la aplicación web y en relación al tiempo óptimo de ejecución es de 99.12%.

3.4. Evaluación Económica

Se determina los siguientes flujos económicos de costo y beneficio del proyecto, los cuales se forman por los saldos netos anuales, estos flujos son utilizados para realizar el cálculo de los siguientes indicadores:

- Valor Actual Neto Económico (VANE)
- Tasa Interna de Retorno Económico (TIRE)
- Relación Costo – Beneficio (B/C)
- Periodo de Recuperación de Inversión

Para llevar a cabo este análisis se cuenta con los siguientes datos:

Inversión:	S/. 1150.00
Costo Operativo:	S/. 200.00
Beneficios anuales:	S/. 1920.00

Interés por defecto del Sistema Económico Peruano (i): 15%

Tiempo promedio de vida del Sistema (n): 3 años

Diagrama de Flujo Convencional

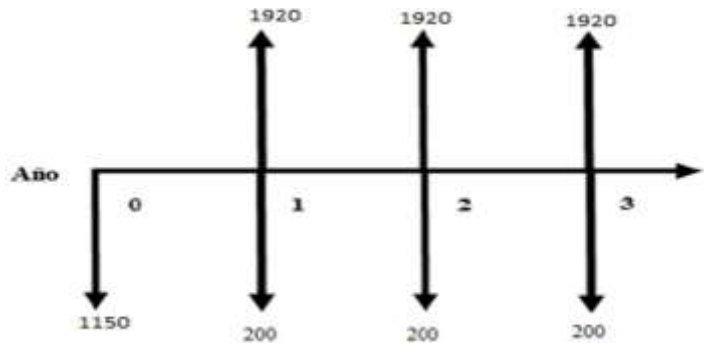


Figura N° 32. Diagrama de Flujo Convencional

Diagrama de Flujo Simplificado

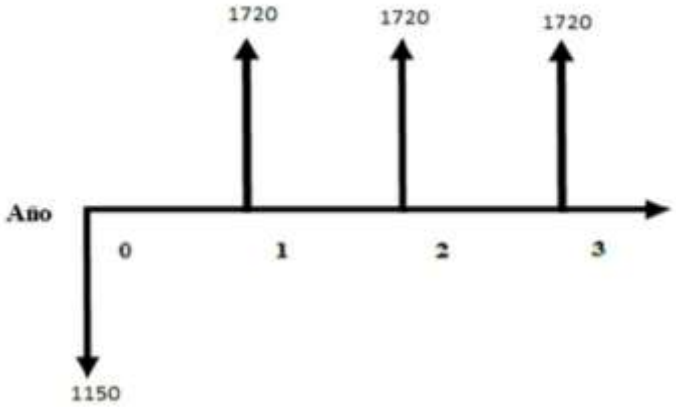


Figura N° 33. Diagrama de Flujo Simplificado

A. Valor Actual Neto Económico

El VAN representa la suma de los valores actualizados de los costos y beneficios que son generados por el proyecto durante el horizonte de planeamiento, sin tomar en cuenta los gastos financieros

$$VAN = -1 + \frac{(B - C)}{(1+i)^1} + \frac{(B - C)}{(1+i)^2} + \cdots + \frac{(B - C)}{(1+i)^n}$$

Vt= flujos de caja en cada periodo t

I0= valor del desembolso inicial de la inversión

n = número de periodos considerado

k= tipo de interés

$$VAN = \left(\frac{1720}{(1+0.15)^1} + \frac{1720}{(1+0.15^2)} + \frac{1720}{(1+0.15)^3}\right) - 1150$$

VAN = 3927.147201 – 1150

VAN = 2777.147201

Este resultado indica que el proyecto renta económicamente S/. 2777.15, debido a que este resultado es mayor a cero y el valor es alto, se determina que el proyecto es factible

B. Tasa Interna de Retorno Económico

TIR es la tasa de descuento que iguala las inversiones actualizadas con los beneficios actualizados

V_p de Ganancia - V_p de Inversión = 0

V_p de Ganancia = V_p de Inversión

$$\left(\frac{1720}{(1+0.15)^1} + \frac{1720}{(1+0.15^2)} + \frac{1720}{(1+0.15)^3}\right) = 1150$$

Usamos las funciones de Excel para calcular el TIR:

	-1150
1° Año	1495,65
2° Año	1300,57
3° Año	1130,93
TIR	107%

Figura N° 34. Cálculo del TIR

Por lo tanto: **TIR** = 107%

Este resultado TIR, a nivel económico nos indica la tasa de interés que el inversionista, puede pagar sin perder su dinero

C. Relación Costo Beneficio

El análisis de costo-beneficio es una técnica analítica que enumera y compara el costo neto de una intervención en salud con los beneficios que surgen como consecuencia de aplicar dicha intervención. (Christia Borja , 2015)

La relación Costo beneficio representa la razón entre el beneficio económico que otorga el proyecto y los costos de inversión, en base al cociente de las utilidades actualizadas y el monto de inversión.

$$\frac{VpB}{VpC} = 3.414$$

El resultado es 3.414, este valor es mayor a 1, lo cual indica que las utilidades económicas del proyecto están a razón de 3.414 veces mayor a los costos de inversión

D. Periodo de Recuperación

$$\boldsymbol{Periodo = \frac{(1 + TIR)^n - 1}{TIR\ x\ (1 + TIR)^n}}$$

Tenemos: TIR = 107% y N = 3, tenemos

Teniendo en cuenta en que el Año 1, se gana 1495.65 lo dividimos entre 12 y nos da 124.58, al multiplicarlo por 10 nos da 1245.8 dándonos como resultados que se recuperó la inversión

La inversión se recupera en 10 meses aproximadamente

De acuerdo a lo anteriormente mencionado se concluye que: este proyecto es ***económicamente factible***

Luego de terminar con el análisis, el proyecto ha pasado correctamente las tres evaluaciones de factibilidad, en consecuencia, se puede concluir que ***es viable realizar el Modelo de Machine Learning para la Clasificación de estudiantes según su rendimiento académico en el Centro de Idiomas de la Universidad Nacional del Santa.***

ANEXO 2: TABLA Z

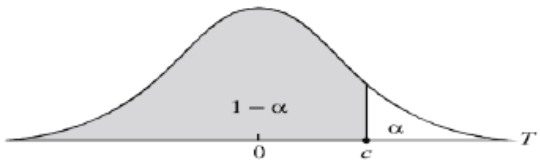
TABLA Probabilidades de una Normal Estándar

z	,00	,01	,02	,03	,04	,05	,06	,07	,08	,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990

Figura N° 35. Tabla Z

Fuente: https://jrvargas.files.wordpress.com/2010/07/tabla-z.pdf

ANEXO 3: TABLA DISTRIBUCIÓN T-STUDENT



	1 − α							
r	0.75	0.80	0.85	0.90	0.95	0.975	0.99	0.995
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576

Figura N° 36. Tabla de Distribución de T-Student

Fuente: www-eio.upc.edu/teaching/estad/MC/taules/TStud.pdf

ANEXO 4: GLOSARIO DE TÉRMINOS

- **APACHE:** Es un servidor de páginas web de distribución libre.
- **APLICACIÓN:** Son aquellos programas que permiten la interacción entre el usuario y la computadora, que están preparados para una utilización específica
- **APLICACIÓN WEB:** Es un sitio web que contiene páginas con contenido sin determinar parcialmente o en su totalidad. El contenido final de estas páginas se determina sólo cuando un visitante solicita una página del servidor Web.
- **BASE DE DATOS:** Es un "almacén" de datos que permite guardar grandes cantidades de información de forma organizada y así podamos encontrarlo y utilizarlo fácilmente.
- **CAJA BLANCA:** Se denomina cajas blancas a un tipo de pruebas de software que se realiza sobre las funciones internas de un módulo.
- **CAJA NEGRA:** Se denomina caja negra a aquel elemento que es estudiado desde el punto de vista de las entradas que recibe y las salidas o respuestas que produce, sin tener en cuenta su funcionamiento interno.
- **GUI:** del inglés graphical user interface, conocida en español como interfaz gráfica de usuario
- **I.D.E.:** Es un entorno de programación que ha sido empaquetado como un programa de aplicación; es decir, consiste en un editor de código, un compilador, un depurador y un constructor de interfaz gráfica.
- **INDICADORES:** Datos que nos permiten medir de forma objetiva los sucesos del mercado para poder respaldar acciones.
- **INFORMACIÓN**: La información es un conjunto organizado de datos procesados, que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema.

- **INTERFAZ**: Es la presentación gráfica que le permite al usuario interactuar con el hardware de la computadora.
- **INTERNET:** Conjunto de ordenadores o servidores, conectados en una red de redes mundiales que comparten un mismo protocolo de comunicaciones y que prestan servicios a los ordenadores que se conectan a esa red
- **INSATISFACCIÓN**: Es un sentimiento interior que experimenta una persona cuando siente que una realidad determinada no cumple sus expectativas.
- **METODOLOGÍA ÁGIL**: Método que permite incorporar cambios con rapidez en el desarrollo de software.
- **NAVEGADOR:** Un navegador web o web browser es una aplicación software que permite al usuario recuperar y visualizar documentos de hipertexto, comúnmente descritos en HTML, desde servidores web de todo el mundo a través de Internet.
- **PERIODO DE RECUPERACION**: Es un instrumento que permite medir el plazo de tiempo que se requiere para que los flujos netos de efectivo de una inversión recuperen su costo o inversión inicial
- **PROCESOS**: Una secuencia de pasos dispuesta con algún tipo de lógica que se enfoca en lograr algún resultado específico.
- **PRUEBAS DE ACEPTACIÓN:** Tiene como propósito demostrar al cliente el cumplimiento de un requisito del software.
- **PRUEBAS UNITARIAS**: Es una forma de comprobar el correcto funcionamiento de un módulo.
- **SERVICIO**: Los servicios son funciones ejercidas por las personas hacia otras personas con la finalidad de que estas cumplan con la satisfacción de recibirlos

- **SERVIDOR WEB:** Es un software que suministra páginas web en respuesta a las peticiones de los navegadores Web. La petición de una página se genera cuando un visitante hace clic en un vínculo de una página Web, en el navegador, elige un marcador en el navegador o introduce un URL ~en "el cuadro de texto Dirección del navegador. Entre los servidores Web más -utilizados se encuentran Microsoft Internet Information Server, Microsoft Personal Web Server, Apache HTTP Server, Netscape Enterprise Server y Sun ONE Web Server.
- **SISTEMA:** Conjunto de elementos interrelacionados entre sí para obtener un objetivo común
- **SISTEMA DE ADMINISTRACIÓN DE BASE DE DATOS:** (DBMS o sistema de base de datos) es un software que se utiliza. para crear y· manipular bases de datos. Entre los sistemas de bases de datos más habituales figuran Microsoft Access, Oracle 9i y MySQL.
- **SOFTWARE LIBRE:** Es una expresión que pertenece al ámbito de la informática. Aunque puede traducirse como “fuente abierta”, suele emplearse en nuestro idioma directamente en su versión original, sin su traducción correspondiente.