



UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA MECÁNICA ELÉCTRICA,
ELECTRÓNICA Y SISTEMAS
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



APLICACIÓN DE TÉCNICAS DE APRENDIZAJE PROFUNDO
PARA EL ANÁLISIS DE DATOS EDUCATIVOS EN EL
INSTITUTO DE EDUCACIÓN SUPERIOR PEDAGÓGICO
PÚBLICO PUNO

TESIS

PRESENTADA POR:

Bach. WALKER ALEXANDER HERRERA VILLANUEVA

Bach. LUIS JENS ARGAMA PAYE

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO DE SISTEMAS

PUNO – PERÚ

2024



LUIS JENS ARGAMA PAYE WALKER ALEXANDER HE... APLICACIÓN DE TÉCNICAS DE APRENDIZAJE PROFUNDO PARA EL ANÁLISIS DE DATOS EDUCATIVOS EN EL INSTITUT

- tesis 20204
- tesis 20204
- Universidad Nacional del Altiplano

Detalles del documento

Identificador de la entrega
trn:oid:::8254:417304731

Fecha de entrega
18 dic 2024, 2:45 p.m. GMT-5

Fecha de descarga
18 dic 2024, 2:48 p.m. GMT-5

Nombre de archivo
TESIS - FINAL. 14.44.pdf

Tamaño de archivo
2.3 MB

123 Páginas

21,704 Palabras

130,965 Caracteres








10% Similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para ca...

Filtrado desde el informe

- Bibliografía
- Texto citado
- Texto mencionado
- Coincidencias menores (menos de 10 palabras)

Fuentes principales

- 7%  Fuentes de Internet
- 1%  Publicaciones
- 7%  Trabajos entregados (trabajos del estudiante)

Marcas de integridad

N.º de alertas de integridad para revisión

No se han detectado manipulaciones de texto sospechosas.

Los algoritmos de nuestro sistema analizan un documento en profundidad para buscar inconsistencias que permitirían distinguirlo de una entrega normal. Si advertimos algo extraño, lo marcamos como una alerta para que pueda revisarlo.

Una marca de alerta no es necesariamente un indicador de problemas. Sin embargo, recomendamos que preste atención y la revise.



Firmado digitalmente por GOMEZ
QUISPE Hugo Yosel FAU
20145496170 soft
Motivo: Soy el autor del documento
Fecha: 18.12.2024 14:56:26 -05:00

V°B°

Firmado digitalmente por
SOTOMAYOR ALZAMORA Guina
Guadalupe FAU 20145496170 hard
Motivo: Doy V° B°
Fecha: 18.12.2024 23:39:31 -05:00





DEDICATORIA

Por la gracia y guía de Dios, quien ha sido mi fortaleza y fuente de entendimiento para superar cada reto, hoy celebro este importante hito en mi vida. Quiero dedicar este logro a mi familia, que siempre ha sido mi pilar fundamental. A mis padres, por su amor incondicional, sus sacrificios y su ejemplo constante, que me han inspirado a dar lo mejor de mí. A mis hermanos, por su aliento, su compañía y por ser una fuente de fuerza y motivación en mi camino. Cada palabra de ánimo y cada gesto de apoyo de ustedes han sido claves para que llegue hasta aquí. Este momento es tanto mío como de ustedes, porque han sido parte esencial de mi crecimiento y mis logros. Con gratitud en mi corazón, miro al futuro confiado en que, con la guía de Dios y el respaldo de quienes amo, puedo enfrentar nuevos desafíos y alcanzar nuevas metas.

Walker Alexander Herrera Villanueva



DEDICATORIA

A mis abuelos y tíos, quienes con su sabiduría, experiencia y amor han dejado una huella imborrable en mi vida. Sus consejos, siempre llenos de verdad y cariño, han sido una brújula que me ha orientado hacia el camino correcto. Gracias por transmitirme los principios y valores que hoy definen quién soy, y por enseñarme, con su ejemplo, la importancia del esfuerzo, la humildad y la perseverancia. Han estado presentes en cada paso de mi vida, no solo brindándome apoyo incondicional, sino también recordándome con sus palabras y acciones que los logros más significativos se construyen con dedicación y fe. Este logro no es solo mío, sino también de ustedes, porque su influencia ha sido clave en cada etapa de mi camino.

Luis Jens Argama Paye



AGRADECIMIENTOS

A la Universidad Nacional del Altiplano de Puno y a la Escuela Profesional de Ingeniería de Sistemas, por brindarnos el espacio para nuestra formación académica.

A nuestros docentes, quienes nos compartieron sus conocimientos y experiencias, contribuyendo significativamente a nuestra formación profesional.

A los miembros del jurado de este trabajo de investigación, por su tiempo, dedicación y esfuerzo en la revisión y evaluación de esta tesis.

A nuestros familiares, amigos y a todas aquellas personas que, de diversas maneras, nos brindaron su apoyo a lo largo de nuestra carrera profesional.

Y a todas aquellas personas que, de diversas maneras, nos brindaron su apoyo, ya sea con un consejo, una sonrisa o un gesto de confianza, queremos expresar nuestra más sincera gratitud. Cada uno de ustedes ha sido parte de este logro, y este momento es también una celebración de todo el esfuerzo compartido.

Walker Alexander Herrera Villanueva

Luis Jens Argama Paye



ÍNDICE GENERAL

	Pág.
DEDICATORIA	
AGRADECIMIENTOS	
ÍNDICE GENERAL	
ÍNDICE DE TABLAS	
ÍNDICE DE FIGURAS	
ÍNDICE DE ANEXOS	
ACRÓNIMOS	
RESUMEN	17
ABSTRACT.....	18
CAPÍTULO I	
INTRODUCCIÓN	
1.1. DESCRIPCIÓN DEL PROBLEMA DE INVESTIGACIÓN.....	21
1.2. FORMULACIÓN DEL PROBLEMA	23
1.2.1. Problema general.....	23
1.3. OBJETIVOS DE LA INVESTIGACIÓN.....	23
1.3.1. Objetivo general	23
1.3.2. Objetivos específicos	23
1.4. JUSTIFICACIÓN DE LA INVESTIGACIÓN	24
1.5. ALCANCES Y LIMITACIONES DE LA INVESTIGACIÓN	25
1.5.1. Alcances	25
1.5.2. Limitaciones	26
1.6. HIPÓTESIS DE LA INVESTIGACIÓN	26
1.6.1. Hipótesis general.....	26



CAPÍTULO II

REVISIÓN DE LITERATURA

2.1.	ANTECEDENTES DE LA INVESTIGACIÓN	27
2.1.1.	Antecedentes Locales	27
2.1.2.	Antecedentes Nacionales	29
2.1.3.	Antecedentes Internacionales	30
2.2.	MARCO TEÓRICO	33
2.2.1.	Definiciones generales	33
2.2.1.1.	Redes Neuronales Profundas	33
2.2.1.2.	Aprendizaje Supervisado	34
2.2.1.3.	Redes Neuronales Convolucionales.....	35
2.2.1.4.	Transformers en Procesamiento del Lenguaje Natural	35
2.2.1.5.	Redes Generativas Antagónicas.....	36
2.2.1.6.	Deep Learning en Dispositivos Móviles y en la Nube.....	37
2.2.1.7.	Reconocimiento de Voz y Procesamiento del Lenguaje.....	37
2.2.1.8.	Aprendizaje Automático Aplicado para Ingenieros.....	38
2.2.1.9.	Visión por Computadora Aplicada	39
2.2.1.10.	Inteligencia Artificial en la Educación.....	40
2.2.1.11.	Enseñanza de la Inteligencia Artificial	40
2.2.1.12.	Programación de Redes Neuronales en Python	41
2.2.1.13.	Desarrollo de Aplicaciones Basadas en IA	42
2.2.1.14.	Aprendizaje por Refuerzo Profundo	43
2.2.1.15.	Métodos de Redes Neuronales en Procesamiento del Lenguaje Natural.....	44
2.2.1.16.	Dominio del Aprendizaje por Refuerzo con Python.....	44



2.2.1.17. Introducción al Aprendizaje Automático con Python.....	45
2.2.1.18. Aprendizaje Profundo con JavaScript.....	46
2.2.1.19. Teoría del Aprendizaje Profundo	47
2.2.1.20. Inteligencia Artificial y sus Implicaciones Éticas en la Educación.....	47
2.2.1.21. Compatibilidad de la Inteligencia Artificial con la Autonomía Estudiantil	48
2.2.1.22. Liderazgo Global en el Desarrollo de IA para la Educación	49
2.2.1.23. Ética y Control de la Superinteligencia en la Educación	49
2.2.1.24. Aplicación de Redes Neuronales Convolucionales en la Educación.....	50
2.2.1.25. Aprendizaje por Refuerzo en Sistemas Educativos	51

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. TIPO Y DISEÑO DE INVESTIGACIÓN	52
3.2. POBLACIÓN Y MUESTRA.....	53
3.2.1. Población.....	53
3.2.2. Muestras	54
3.2.3. Ubicación y Descripción de la Población	54
3.3. TÉCNICAS E INSTRUMENTOS.....	55
3.3.1. Técnicas e instrumentos para la recolección de datos.....	55
3.3.2. Técnicas para el análisis de la información.....	56
3.3.3. Técnicas para el procesamiento de datos	57
3.4. PLAN DE TRATAMIENTO DE DATOS	57
3.4.1. Análisis de datos	58



3.5.	APLICACIÓN DE LA METODOLOGÍA	59
3.5.1.	Metodología en Prototipo.....	59
3.5.2.	Material Experimental.....	60
3.5.2.1.	Infraestructura Tecnológica:	60
3.5.2.2.	Recursos Digitales y Servicios:.....	61
3.5.2.3.	Materiales de Apoyo:	61
3.6.	ASPECTOS ÉTICOS	61

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1.	DISEÑAR LA ARQUITECTURA DE SOFTWARE DEL SISTEMA DE ANÁLISIS DE DATOS EDUCATIVOS.....	63
4.1.1.	Identificación de los componentes del sistema:	63
4.1.2.	Selección de las tecnologías:.....	66
4.1.3.	Diseño modular y flexible:.....	69
4.1.4.	Modelado de los flujos de datos:.....	73
4.1.5.	Dataset Inicial y Estructura de los datos fuentes.....	78
4.1.6.	Base de Datos	84
4.1.7.	Diseño del interfaz	86
4.2.	DESARROLLAR Y AJUSTAR MODELOS DE APRENDIZAJE PROFUNDO PARA LA PREDICCIÓN Y ANÁLISIS DE COMPORTAMIENTOS Y RENDIMIENTOS EDUCATIVOS.....	86
4.2.1.	Preprocesamiento de datos:.....	86
4.2.2.	Entrenamiento de modelos:.....	89
4.2.3.	Validación de los modelos:	90
4.2.4.	Evaluación y optimización:.....	91



4.3. IMPLEMENTAR PRUEBAS DEL SOFTWARE PARA GARANTIZAR SU FUNCIONALIDAD Y FIABILIDAD	95
4.3.1. Definición de los casos de prueba:	95
4.3.2. Pruebas unitarias y de integración:	96
4.3.3. Pruebas de rendimiento y estrés:	97
4.3.4. Depuración y corrección de errores:	101
4.4. DISCUSIÓN	102
V. CONCLUSIONES.....	106
VI. RECOMENDACIONES	107
VII. REFERENCIAS BIBLIOGRÁFICAS.....	108
ANEXOS.....	113

Área: Inteligencia Artificial Y Sistemas Bio-Inspirados

Tema: Aplicación De Técnicas De Aprendizaje Profundo Para El Análisis De Datos Educativos En El Instituto De Educación Superior Pedagógico Público Puno

Fecha de sustentación: 27/12/24



ÍNDICE DE TABLAS

	Pág.
Tabla 1 Población.....	53
Tabla 2 Técnicas e Instrumentos.....	56



ÍNDICE DE FIGURAS

	Pág.
Figura 1 Diagrama de casos de uso.....	65
Figura 2 Diagrama de Clases	68
Figura 3 Diagrama de Componentes.....	70
Figura 4 Diagrama de Despliegue.....	72
Figura 5 Diagrama de Actividades	75
Figura 6 Diagrama de Secuencia	77
Figura 7 Volumen de Datos	83
Figura 8 Base de Datos	84
Figura 9 Diseño del interfaz.....	86
Figura 10 Diagrama de estados	88
Figura 11 Comparación de Precisión	93
Figura 12 Matriz de confusión	94
Figura 13 Rendimiento del Sistema.....	99
Figura 14 Uso de recursos del sistema.....	100



ÍNDICE DE ANEXOS

	Pág.
ANEXO 1 Código Fuente	113
ANEXO 2 Declaración jurada de autenticidad de tesis.....	120
ANEXO 3 Autorización para el depósito de tesis en el Repositorio Institucional....	122



ACRÓNIMOS

DNN:	Redes Neuronales Profundas (Deep Neural Networks).
RNN:	Redes Neuronales Recurrentes (Recurrent Neural Networks).
CNN:	Redes Neuronales Convolucionales (Convolutional Neural Networks).
LSTM:	Memoria a Largo Corto Plazo (Long Short-Term Memory).
PLN:	Procesamiento de Lenguaje Natural.
F1-score:	Métrica que combina la precisión y el recall en un único valor para evaluar la precisión de un modelo de clasificación.
K-NN:	K-Nearest Neighbors, un algoritmo de aprendizaje automático supervisado.
MySQL:	Sistema de gestión de bases de datos relacional.
Redis:	Sistema de almacenamiento en memoria, utilizado para gestionar datos en tiempo real.
JSON:	JavaScript Object Notation, un formato ligero de intercambio de datos.
SQL:	Structured Query Language, lenguaje de consulta utilizado para gestionar bases de datos.
Pandas:	Biblioteca de Python utilizada para la manipulación y análisis de datos.
API:	Interfaz de Programación de Aplicaciones (Application Programming Interface), un conjunto de reglas que permite la interacción entre sistemas de software.



- ROC:** Curva Característica Operativa del Receptor (Receiver Operating Characteristic), utilizada para evaluar la capacidad predictiva de los modelos de clasificación.
- AUC:** Área Bajo la Curva (Area Under the Curve), métrica utilizada en la evaluación de modelos de clasificación.



RESUMEN

Esta investigación examina los impactos de la tecnología en la capacidad de aprendizaje de los estudiantes y la eficacia del profesor en la adquisición de conocimientos por parte del estudiante. Los estudiantes tienden a adoptar la tecnología como parte de lo que son y de lo que hacen debido a su conveniencia. Por lo tanto, el desarrollo de tales sistemas educativos se vuelve invaluable. El sistema muestra utilidad a través de las tasas de retención de estudiantes, donde un aumento en la tasa de retención también sirvió para reducir las tasas de deserción. La toma de decisiones empleada en la gestión del sistema desempeñó un papel clave en el mantenimiento del éxito de los servicios de apoyo a los estudiantes, así como del sistema educativo en su conjunto. La validación del sistema se realizó midiendo parámetros de rendimiento como la precisión y el puntaje F1, superando el 89%. Además, se garantizó el procesamiento eficiente de los tiempos, lo que optimizó la capacidad de análisis en tiempo real para mejorar la gestión educativa de la institución. Los resultados obtenidos confirman que el sistema implementado de análisis de datos educativos es muy efectivo para predecir el rendimiento académico y mejorar las intervenciones educativas dirigidas en el Instituto Pedagógico Público de Puno.

Palabras clave: Análisis de Datos, Aprendizaje Profundo, Gestión Educativa, Predicción Educativa, Redes Neuronales.



ABSTRACT

This research examines the impacts of technology on students' learning ability and teachers' effectiveness in facilitating knowledge acquisition. Students tend to adopt technology as an integral part of who they are and what they do due to its convenience. Therefore, the development of such educational systems becomes invaluable. The system demonstrates its usefulness through student retention rates, where an increase in retention also helped reduce dropout rates. Decision-making employed in the system's management played a key role in maintaining the success of student support services as well as the educational system as a whole. System validation was conducted by measuring performance parameters such as accuracy and F1 score, exceeding 89%. Additionally, efficient processing times were ensured, optimizing real-time analysis capabilities to enhance the institution's educational management. The results confirm that the implemented educational data analysis system is highly effective in predicting academic performance and improving targeted educational interventions at the Public Pedagogical Institute of Puno.

Keywords: Deep Learning, Neural Networks, Educational Prediction, Data Analysis, Educational Management.



CAPÍTULO I

INTRODUCCIÓN

El presente trabajo de investigación titulado: "Implementación de un Sistema de Análisis de Datos Educativos utilizando Técnicas de Aprendizaje Profundo para el Instituto de Educación Superior Pedagógico Público Puno" aborda la creciente necesidad de optimizar el análisis de grandes volúmenes de datos en el entorno educativo. Hay un problema al tratar de analizar el rendimiento académico de los alumnos del Instituto Pedagógico Superior de Puno, que no se puede evaluar con precisión y predecir continuamente el rendimiento académico. Esto aumenta la dificultad para tomar decisiones apropiadas y la evaluación de posibilidades para la intervención y la planificación de actividades para fortalecer el rendimiento académico.

En los últimos años, y debido a tecnologías novedosas desarrolladas como el aprendizaje profundo y la inteligencia artificial que han sido adoptadas, las instituciones educativas han podido cambiar y mejorar sus resultados de análisis de datos. Estos métodos son fundamentales en la predicción e identificación de conjuntos de datos que generan ciertos resultados, pronosticando el rendimiento académico y ayudando a planificar recursos y estrategias. En cualquier caso, todavía existen desafíos considerables para el análisis educativo que involucran grandes datos en instituciones con escasos recursos.

El propósito de esta investigación fue diseñar y desarrollar un sistema de análisis de datos educativos utilizando redes neuronales profundas. El sistema desarrollado hace posible procesar grandes volúmenes de datos educativos, predecir el rendimiento estudiantil con un alto nivel de precisión y crear alertas tempranas para los estudiantes que necesitan apoyo. Se espera que el establecimiento de este sistema cambie la gestión



educativa del instituto, ya que ofrece una herramienta impulsada por datos que mejora la toma de decisiones estratégicas y los logros de aprendizaje.

Este enfoque hizo posible optimizar los procesos de análisis de datos, acortar el tiempo requerido para realizar diagnósticos educativos y mejorar la precisión en la identificación de riesgos de los estudiantes. Además, la individualización de las estrategias de intervención educativa, habilitada por predicciones más precisas, ayuda a mejorar la calidad general de la educación.

La investigación se ha dividido en seis capítulos que se enfocan en la realización y puesta en marcha del sistema de análisis:

Capítulo I: Planteamiento del Problema

Se explica cuál es el problema al enfrentar el análisis tradicional en el instituto y a los problemas derivados de hacer proyecciones de rendimiento de los alumnos. Dentro de esta parte se explican los objetivos, la justificación y los límites de la investigación.

Capítulo II: Revisión de Literatura

Se presenta el marco teórico necesario para el desarrollo del sistema de análisis de datos educativos, basado en una revisión exhaustiva de estudios previos sobre aprendizaje profundo, técnicas de predicción en el ámbito educativo y su impacto en la mejora de la calidad educativa. También se analizan los modelos más avanzados utilizados en el procesamiento de datos académicos.



Capítulo III: Materiales y Métodos

Este capítulo lista los métodos utilizados para la recolección de datos académicos, los instrumentos y las herramientas tecnológicas usadas en el desarrollo del sistema, así como la población y la muestra de datos utilizados en la investigación.

Capítulo IV: Resultados y Discusión

Se presentan los resultados obtenidos tras la implementación del sistema, comparando la predicción del rendimiento académico antes y después de la implantación de los modelos. Los resultados son discutidos en relación con la hipótesis planteada y los estudios previos revisados.

Capítulo V: Conclusiones

En este capítulo se resume las conclusiones del presente trabajo, destacando las mejoras en la precisión del análisis de datos y su impacto en la gestión educativa.

Capítulo VI: Recomendaciones

Finalmente, se proponen recomendaciones para la mejora continua del sistema de análisis de datos y su posible adaptación a otros ámbitos educativos que busquen aplicar técnicas de aprendizaje profundo para la optimización del rendimiento estudiantil.

También se incluyeron bibliografía, apéndices y glosario que complementan el contenido de la redacción del documento y proporcionan una mejor comprensión de los aspectos técnicos y conceptuales tratados durante la investigación.

1.1. DESCRIPCIÓN DEL PROBLEMA DE INVESTIGACIÓN

En el Instituto de Educación Superior Pedagógico Público Puno, la gestión y análisis de los datos educativos se realizan principalmente mediante métodos



tradicionales, los mismos que presentan variadas limitaciones en su capacidad para procesar grandes volúmenes de información. Esto dificulta la extracción de patrones complejos y las relaciones significativas entre las diversas variables, lo que afecta a la toma de decisiones basadas en datos y a la optimización de los recursos educativos. La falta de una infraestructura tecnológica avanzada y de técnicas de análisis automatizadas impide una identificación precisa de los factores que inciden en el rendimiento académico y en las dinámicas de aprendizaje.

Un área de la inteligencia artificial conocido como aprendizaje profundo ha tenido, históricamente, un buen desempeño en el manejo de datos, especialmente con relaciones no lineales y en la obtención de variables latentes a partir de datos altamente diversos. El uso de enfoques de aprendizaje profundo en el análisis de datos educativos en varias instituciones, como el Instituto de Educación Superior Pedagógico Público Puno, parece estar severamente limitado, factores como la disponibilidad de suficientes recursos computacionales y el conocimiento técnico parecen ser un problema.

Por lo que es importante desarrollar un sistema que utilice técnicas de aprendizaje profundo en los procesos de análisis de datos educativos en el Instituto de Educación Superior Pedagógico Público Puno. Este sistema permitirá el procesamiento automático de grandes cantidades de datos de diversas fuentes, como registros académicos, evaluaciones y encuestas. La aplicación del modelo avanzado de redes neuronales nos permitirá clasificar a los estudiantes de manera más precisa, detectar tendencias en el rendimiento académico, evaluar los logros futuros de los estudiantes bajo diferentes condiciones remotas, cambiando así las formas de enseñanza o incluso las características de los estudiantes.



La mejora permite que los procesos de análisis de datos educativos se realicen de tal manera que se puedan producir informes más precisos y específicos para los docentes y los administradores, lo que puede influir significativamente en la elección de estrategias pedagógicas y en la asignación de recursos para el seguimiento de los estudiantes. Además, la aplicación de técnicas de aprendizaje profundo garantizará una mejor precisión y capacidad del sistema para acomodar diferentes tipos de datos en varios entornos, lo que mejorará la calidad de la educación del Instituto.

1.2. FORMULACIÓN DEL PROBLEMA

1.2.1. Problema general

¿Cómo la aplicación de técnicas de aprendizaje profundo para el análisis de datos educativos en el Instituto de Educación Superior Pedagógico Público Puno pueden optimizar la identificación de patrones complejos y la toma de decisiones en la gestión educativa?

1.3. OBJETIVOS DE LA INVESTIGACIÓN

1.3.1. Objetivo general

Implementar un sistema de análisis de datos educativos empleando técnicas de aprendizaje profundo para el Instituto de Educación Superior Pedagógico Público Puno.

1.3.2. Objetivos específicos

- Diseñar la arquitectura de software del sistema de análisis de datos educativos.



- Desarrollar y ajustar modelos de aprendizaje profundo para la predicción y análisis de comportamientos y rendimientos educativos.
- Implementar pruebas de software para garantizar funcionalidad y fiabilidad.

1.4. JUSTIFICACIÓN DE LA INVESTIGACIÓN

Esta investigación se basa en la necesidad de mejorar el procesamiento de datos de información educativa en el Instituto de Educación Superior Pedagógico Público Puno a través de técnicas de aprendizaje profundo. En la actualidad, las técnicas tradicionales de análisis de datos en el ámbito educativo no son suficientes para procesar eficientemente grandes datos. Tal deficiencia impide que las instituciones reconozcan patrones complejos que serían útiles para mejorar el rendimiento académico de los estudiantes, la adaptación de los enfoques pedagógicos y la gestión de los recursos educativos.

El aprendizaje profundo ha demostrado ser una tecnología apropiada que se puede utilizar para procesar y analizar datos multidimensionales a través de sus arquitecturas, como Redes Neuronales Profundas (DNN) y Redes Neuronales Recursivas (RNN). Estas técnicas tienen la capacidad de aprender representaciones latentes de grandes conjuntos de datos, que implican correlaciones complejas entre los conjuntos de datos donde tales patrones serían invisibles con métodos convencionales. Esto es bastante importante en el contexto del aprendizaje, ya que existen relaciones complejas entre el rendimiento académico de los estudiantes, su comportamiento y el entorno educativo.

La implementación de los modelos de aprendizaje profundo mejorará la segmentación de los estudiantes, de modo que estudiantes específicos serán identificados y se propondrán intervenciones educativas predictivas adicionales. Además, estos modelos son capaces de predecir el rendimiento de un estudiante en clase de manera



relativamente más precisa a largo plazo, lo que facilita establecer y hacer cumplir acciones preventivas y de apoyo que resulten en una gestión académica más adecuada.

A nivel técnico, las redes neuronales profundas ayudarán en la comprensión, integración y procesamiento de diversos datos no homogéneos, como calificaciones, evaluaciones de rendimiento, asistencia y datos demográficos, lo que resultará en un análisis multidimensional que va más allá de las técnicas estadísticas convencionales. Agilizar y automatizar el proceso de análisis de datos reducirá significativamente el tiempo requerido para analizarlos manualmente, mientras se crean oportunidades para avanzar en los informes educativos en formatos en tiempo real y proporcionar mejores y más detallados conocimientos a los que toman las decisiones.

Este enfoque mejora el análisis y la gestión de datos, al tiempo que tiene un efecto positivo en la planificación pedagógica y en la toma de decisiones estratégicas dentro del Instituto. Además, la introducción de nuevas tecnologías en la educación permitirá contar con un ejemplo para otras instituciones, ayudando a actualizar los procesos educativos y permitiendo que se adopten soluciones tecnológicas avanzadas para mejorar los resultados de aprendizaje.

1.5. ALCANCES Y LIMITACIONES DE LA INVESTIGACIÓN

1.5.1. Alcances

El desarrollo de la presente investigación se basó en la implementación de un sistema de análisis para datos educativos, a través del uso de herramientas de deep learning en el Instituto de Educación Superior Pedagógico Público Puno. El sistema incluyó el tratamiento de datos académicos y administrativos, con un particular interés en la identificación de variables del rendimiento académico de los estudiantes y esenciales para el proceso educativo. Los registros académicos,



las evaluaciones y la asistencia brindaron información actual e histórica sobre el estudiante. El sistema fue un desarrollo para el Instituto, contando con opciones de actualizaciones y mejoras en función de la demanda pedagógica y tecnológica que pudieran surgir en un futuro. También se brindó formación al personal administrativo y docente en el uso y mantenimiento del sistema.

1.5.2. Limitaciones

Las limitaciones de la investigación incluyeron la restricción del sistema al Instituto de Educación Superior Pedagógico Público Puno, sin tener en cuenta su implementación en otras instituciones educativas. No se anticipó la integración con plataformas externas en la primera etapa, lo que puede inhibir su crecimiento posterior en otras situaciones educativas. Se pretendía una expansión continua del sistema; sin embargo, el estudio se basó únicamente en datos del Instituto, lo que limita la posibilidad de extrapolar los hallazgos a otras situaciones. Además, la otra limitación del enfoque del estudio fue la disponibilidad de datos del Instituto; así, la relevancia y cantidad de datos podrían haber influido en la precisión de los resultados.

1.6. HIPÓTESIS DE LA INVESTIGACIÓN

1.6.1. Hipótesis general

Es posible implementar un sistema de análisis de datos educativos utilizando técnicas de aprendizaje profundo en el Instituto de Educación Superior Pedagógico Público Puno.



CAPÍTULO II

REVISIÓN DE LITERATURA

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

2.1.1. Antecedentes Locales

Bermejo Escobar & Vizcarra Aguilar, (2020) desarrollaron un modelo basado en aprendizaje profundo aplicando modelos de deep learning sobre tweets que estaban en español como una de las formas de mejorar el rango de resultados adquiridos utilizando los otros procesos que tratan con el lenguaje natural. Los enfoques que se utilizaron incluyeron las redes neuronales recurrentes (RNN), las Redes Neuronales Convolucionales (CNN) así como un modelo híbrido que involucraba el uso de las dos arquitecturas. Su estudio estuvo orientado a abordar los desafíos existentes con el análisis de sentimientos en español para usar el corpus InterTASS. Las métricas de análisis de sentimientos en español son bastante bajas en comparación con las contrapartes en inglés, lo que dificulta el análisis utilizando tweets en español. Las CNNs fueron entrenadas utilizando los datos del corpus InterTASS. Se clasificó en cuatro clases diferentes de los tweets que se recibieron. Se demostró que los modelos de deep learning eran precisos y predictivos mucho más que los métodos tradicionales de clasificación, demostrando la efectividad de las técnicas de deep learning para texto en español, estableciendo así benchmarks importantes para proyectos futuros en este dominio.

Mestas Yucra, (2024) a partir del rendimiento académico de los estudiantes, explica cómo ha desarrollado un modelo de aprendizaje profundo para pronosticar la matrícula estudiantil de la Universidad Nacional del Altiplano



Puno. Este estudio fue por iniciativa de la Unidad de Gestión Educativa Local de Melgar para evaluar el nivel de preparación de los solicitantes de las escuelas locales. Se llevó a cabo un análisis cuantitativo de la base de datos existente de 257 estudiantes con la aplicación de técnicas de agrupamiento no supervisado para la elaboración de perfiles de rendimiento académico. Además, se identificó un grupo significativo de estudiantes de alto rendimiento empleando la medida de distancia métrica de Russellrao. Sin embargo, se reconoció un desajuste notable entre las tasas de matrícula pronosticadas y reales, es decir, 58.36% y 3.5% respectivamente. Esta discrepancia motivó al autor a razonar que existían elementos externos, como la naturaleza competitiva del proceso de admisión, que afectaban la efectividad del modelo. Tales modelos educativos, como el presentado, subrayan la necesidad de adaptar modelos educativos al entorno apropiado, así como sugerir formas de mejorar los futuros enfoques a las técnicas de aprendizaje profundo en la predicción de la admisión universitaria.

Flores Dueñas & Pari Salazar, (2022) propusieron un modelo de aprendizaje automático con la finalidad de identificar operaciones extrañas de lavado de activos en una entidad financiera. El modelo usaba técnicas avanzadas de preprocesamiento de datos y aprendizaje automático para analizar grandes volúmenes de datos de transacciones y así detectar patrones sospechosos que podrían estar relacionados con actividades de ilícitas de lavado de activos. La investigación se centró en la aplicación de algoritmos de clasificación y de detección de anomalías, utilizando redes neuronales y árboles de decisión como herramientas de base. El proceso incluyó una validación cruzada de los modelos y la métrica de precisión alcanzada fue del 78.37%, lo que demostró la efectividad del enfoque en la identificación de operaciones inusuales. Este estudio resaltó la



importancia de actualizar constantemente la información de riesgo en las entidades financieras para hacer frente a la evolución de las amenazas relacionadas con el lavado de activos y el financiamiento del terrorismo, y destacó el papel de las técnicas de aprendizaje automático como herramientas clave en la gestión de riesgos financieros.

2.1.2. Antecedentes Nacionales

Borjas Ramos & Saqui Marin, (2023) desarrollaron un modelo para reducir la tasa de deserción estudiantil en instituciones de educación superior en Perú. Emplearon aprendizaje automático y este estudio se concentró en mejorar el proceso de predicción de deserción estudiantil a través de la comparación de varios algoritmos de aprendizaje automático. Con este propósito, se creó una tabla maestra que contenía el rendimiento académico actual, el rendimiento pasado, la información de recibos, las tasas de asistencia y los datos personales. RUSBoost, GausNB, Perceptron y LightGBM fueron algunos de los diferentes algoritmos examinados, siendo este último el que arrojó los mejores resultados con un promedio de Recall de 86.37% y un promedio de Precisión de 86.83%. Después de implementar el modelo, se pudo reducir la tasa de deserción estudiantil del 13.45 % al 9.45 % en un año. Este trabajo muestra las grandes posibilidades de las técnicas de aprendizaje automatizado para la mejora de la retención estudiantil en el contexto de las instituciones educativas, porque es fácil y eficiente encontrar estudiantes que probablemente abandonen.

Vargas Quispe, (2020) llevó a cabo un estudio de predicción del rendimiento académico utilizando algoritmos supervisados en estudiantes de ingeniería de la Universidad Nacional Amazónica de Madre de Dios. El análisis



incluyó el análisis de 861 registros reales de estudiantes que ingresaron entre los semestres 2010-1 y 2020-2, y se utilizaron algoritmos de árbol de decisión, K-NN y bayesiano ingenuo como herramientas para examinar relaciones entre múltiples variables y el rendimiento académico de los estudiantes. Se encontró que K-NN alcanzó la mayor precisión con un 81.97 por ciento. Los autores identificaron las dimensiones sociales, económicas y académicas como las variables más importantes relacionadas con el rendimiento de los estudiantes y señalaron los grandes beneficios de utilizar técnicas de aprendizaje automático para identificar patrones que podrían fundamentar estrategias de intervención temprana para estudiantes con bajo rendimiento.

Ruiz Campos, (2024) evaluó el impacto de la inteligencia artificial en la gestión educativa de calidad en I.E.P. Alipio Ponce ubicado en el distrito de San Juan de Lurigancho. La presente investigación, de tipo correlacional y no experimental, analizó las inversiones en tecnologías de inteligencia artificial y el impacto que esta tiene sobre los procesos de gestión educativa, centrandose su atención en la mejora de la calidad institucional. Utilizando herramientas de observación, se evaluaron ambas variables, encontrando una correlación moderada entre el uso de la inteligencia artificial y la mejora de la gestión educativa. Este trabajo hace hincapié en el papel que juega la inteligencia artificial como un instrumento para la mejora de la calidad de gestión en las instituciones de enseñanza primaria.

2.1.3. Antecedentes Internacionales

González Muñoz, (2023) en su investigación, presenta un estudio enfatizando el uso de Deep Learning en sistemas prácticos resaltando como estas



estrategias han contribuido en la optimización de operaciones y resolución de problemas grandes. Los mayores esfuerzos dentro de este campo consisten en análisis de grandes volúmenes de datos como por ejemplo imágenes generadas por sensores así como también señales y datos en tiempo real integrando redes neuronales convolucionales y recurrentes en sistemas de ingeniería. Uno de sus estudios más promisorios fue el uso del aprendizaje profundo en la evaluación de patrones en los industriales, estos informes complementan cuidadosamente el uso del aprendizaje profundo en la optimización industrial, reducción de tiempos de inactividad y en la toma de decisiones. Se enfatiza la importancia del marco eléctrico para la integración de las tecnologías de delay de tipo controlador, dispositivos de almacenamiento en la nube y sensores inteligentes, a los sistemas de ingeniería, argumentos que aseguran ese enfoque automatizado y de alta precisión.

Bustamante Domínguez, (2022) implementó estrategias basadas en aprendizaje profundo para la toma de decisiones en aplicaciones relacionadas con la pandemia de COVID-19. En su investigación se enfocó en el uso de deep learning para analizar los datos médicos y públicos, lo que permitió identificar patrones y tendencias que ayudaron en la toma de decisiones acertadas durante la crisis sanitaria. Bustamante exploró diversas aplicaciones del aprendizaje profundo en el contexto de la pandemia, como el análisis de imágenes médicas para la detección de la enfermedad del COVID-19, el monitoreo de la propagación del virus a través del análisis de datos públicos y la detección de información equivocada relacionada con el COVID-19 en redes sociales. Además, su estudio profundizó en el uso de redes neuronales convolucionales para procesar las imágenes de rayos X y las tomografías computarizadas, mejorando



significativamente la precisión en el diagnóstico de la enfermedad. Finalmente concluyó que el uso de aprendizaje profundo fue fundamental para agilizar la respuesta médica y pública ante la pandemia, y también sugirió que estas técnicas seguirán desempeñando un papel clave en la medicina preventiva y para el análisis de datos de salud en el futuro.

Ruiz Manosalva, (2019) desarrolló un modelo de análisis de datos en el que se analizaron patrones en datos generados por pacientes durante el uso de videojuegos como herramienta terapéutica. Su enfoque gira en torno a la aplicación práctica de algoritmos de agrupamiento y clasificación a grandes cantidades de información producidas a través de terapias de entretenimiento adaptadas para la rehabilitación de la atención y la cognición en pacientes con deterioro cognitivo leve. Para el proyecto sabático de Ruiz Manosalva, utilizó la metodología CRISP-DM como un marco dentro del cual se implementaron varios métodos de minería de datos, incluyendo la creación de variables y el procesamiento de datos en bruto para la detección de patrones de comportamiento con el objetivo de mejorar la técnica terapéutica. Los hallazgos del estudio indicaron que los juegos desarrollados para los pacientes durante el tratamiento tenían muchas similitudes que podían ser utilizados plenamente para adaptar las terapias según los requisitos de cada individuo. Los resultados también indicaron que los pacientes pudieron aumentar su eficiencia con cada sesión de juego subsiguiente, reforzando así la visión de que la utilidad del aprendizaje automático en el contexto de mejorar la terapia cognitiva es innegable. Este estudio enfatizó el poder del aprendizaje profundo y técnicas automáticas no solo como medios para examinar enormes volúmenes de datos, sino también como herramientas que



generan información útil para mejorar la calidad del tratamiento en entornos médicos.

2.2. MARCO TEÓRICO

2.2.1. Definiciones generales

2.2.1.1. Redes Neuronales Profundas

Se entiende como métodos aplicados para resolver tareas complejas con la ayuda de un gran número de neuronas agrupadas en varias capas. Este tipo de red neuronal tiene, como característica diferenciadora, la capacidad de aprender representaciones multilaterales de los datos que reciben; las capas inferiores aprenden características simples como bordes para imágenes, y las capas superiores pueden 'entender' objetos. El entrenamiento de la red neuronal, o retroalimentación recursiva, se prepara ajustando millones de parámetros llamados pesos que definen cómo reaccionan las neuronas respectivas a los estímulos. La reconstrucción algorítmica hacia atrás y la concentración de pesos para predecir etiquetas difíciles y desconocidas de muestras reales se utilizan para entrenar la red. Las tareas de consolidación y recuperación de memoria utilizan retropropagación para fines no supervisados con el objetivo de mejorar las habilidades predictivas complejas reales en el futuro, y convertir las redes de aprendizaje profundo en redes neuronales. Reconocer patrones como imágenes, sonidos o voces es ideal para redes neuronales artificiales. También en educación, tales sistemas pueden modelar de manera integral el rendimiento de los estudiantes basado en



datos históricos y cambiar el contenido según el tiempo y los requerimientos. (Howard & Gugger, 2020)

2.2.1.2. Aprendizaje Supervisado

El aprendizaje supervisado puede definirse como uno de los tipos importantes de técnicas de aprendizaje automático en las que se utilizan conjuntos de datos que están marcados o etiquetados para entrenar un modelo. Para cada conjunto de características de entrada de un modelo, existe una única característica de salida (la etiqueta), el objetivo del modelo es conocer cuál es la función que relaciona cada polinomio con su etiqueta. El modelo a ser implementado supera el modelo por la utilización de datos de aprendizaje, el modelo se entrena con el etiquetado de un conjunto de entrada y etiqueta de salida; las máquinas de soporte vectorial, por ejemplo, son generadoras de SVMs. Narain Shah (2022) explica que el aprendizaje supervisado se utiliza eficazmente en las funciones de clasificación y regresión. Esta técnica ayuda a poder pronosticar qué rendimiento académico tendrá el niño, así como su nivel de dominio en cada materia, y también permite que el maestro pueda obtener de forma más rápida alguna tendencia que pueda llevar al alumno a dejar de asistir a la escuela. Cuando se trabaja con modelos respaldados donde se dispone de una gran base de datos etiquetados de estudiantes, como una asistencia a clases, entonces se logra mejorar la pedagogía personalizada.



2.2.1.3. Redes Neuronales Convolucionales

Las imágenes y los videos son, en términos generales, datos que estructuran una red de tipo grilla; esto se produce a través de la ingeniería especializada en sistemas, a lo que se le conoce como la arquitectura de la red neuronal. Este proceso se logra a través de la instalación de capas que se conocen como capas convolucionales, donde se les coloca filtros a la información que se les brinda. Con esto, las redes neuronales convolucionales son capaces de aprender patrones espaciales, tales como bordes, formas y texturas, de una manera más efectiva que las arquitecturas de redes neuronales tradicionales. Las CNNs se han usado con éxito en aplicaciones de visión, como el reconocimiento de un ambiente tridimensional, reconocimiento facial, detección de objetos y clasificación de imágenes. Concerniente al ámbito pedagógico, las redes neuronales pueden contribuir en exámenes donde se requiere interpretar una imagen, videos y gráficos para la corrección de tales trabajos. En este contexto, las CNNs también pueden ser utilizadas en tareas educativas de visión por computador, como la realidad aumentada o el reconocimiento visual para una mejor interacción del estudiante con el contenido educativo. (Gulli et al., 2021)

2.2.1.4. Transformers en Procesamiento del Lenguaje Natural

Transformers es una arquitectura novedosa que desarrolla nuevas maneras de trabajar el lenguaje natural y ha mejorado este campo por el hecho de que puede procesar más cantidad de texto en cuestión de texto que los modelos secuenciales tradicionales, tales como RNNs y LSTMs.



Transformers emplean mecanismos de atención donde el modelo puede centrarse en partes diferentes de una secuencia de texto al mismo tiempo que optimiza el modelado de relaciones de gran alcance. Las traducciones automáticas, la evaluación de sentimientos o el análisis de textos-entre otros- han sido automatizados satisfactoriamente gracias a los modelos Transformer con BERT o GPT, mientras que el sector educativo se dirige en utilizar estos modelos para la evaluación automática de exámenes abiertos, feedback automatizado, análisis de grandes volúmenes de textos, preguntas de evaluación sobre un texto educativo y recomendación de contenido basado en las interacciones previas de un estudiante (Rothman, 2021)

2.2.1.5. Redes Generativas Antagónicas

Las Redes Generativas Antagónicas se basan en un método de aprendizaje profundo en el que se emplean dos redes neuronales en competencia entre sí, una generadora y otra discriminadora. A grandes rasgos, la primera red- la generadora- se encarga de obtener nuevos datos provenientes de un ruido aleatorio mientras que la segunda- la discriminadora- debe evaluar si esos datos son falsos o fueron actualizados genuinamente. Una vez que las dos redes son entrenadas en conjunto, la red generadora puede crear datos mucho más reales. Las GAN son muy importantes a la hora de crear imágenes, textos, raciones y videos. Dentro del ámbito educativo, las GAN pueden generar datos sintéticos de problemas matemáticos, exámenes personalizables o incluso simulaciones personalizadas adaptadas a los requerimientos específicos de un estudiante. También pueden ser utilizadas para crear una simulación



educativa interactiva o un entorno de realidad virtual. Además, se confía en las capacidades de creación de datos ultra realistas de las GAN cuando se desarrollan entornos de aprendizaje basados en simulaciones donde un estudiante puede perfeccionar una habilidad particular dentro de un entorno controlado. (Foster, 2022)

2.2.1.6. Deep Learning en Dispositivos Móviles y en la Nube

El aprendizaje profundo en dispositivos móviles y la nube ha aumentado debido a la creciente necesidad de aplicaciones con poder de IA que requieren inteligencia en tiempo real. El aprendizaje profundo en la nube permite el entrenamiento y despliegue de modelos complejos en servidores remotos mientras un dispositivo móvil puede utilizar una versión optimizada del modelo para predecir lo que es más probable que ocurra. Esta combinación de poder de IA y portabilidad ha permitido la presencia de aplicaciones de IA que pueden funcionar en una tableta, teléfono inteligente y/o otros dispositivos conectados. En el sector educativo el uso de smartphones permite que los estudiantes cuenten con tutorías en línea, además de contar con aplicaciones especializadas y análisis en tiempo real. Por el lado de los profesores, también pueden ser beneficiados por el uso de aplicaciones que generan evaluaciones a partir de datos de sus estudiantes que son recolectados en dispositivos móviles y se envían a la nube para un análisis adicional. (Moroney, 2020)

2.2.1.7. Reconocimiento de Voz y Procesamiento del Lenguaje

Es un proceso complementario en el procesamiento del lenguaje que se encarga de hacer que las máquinas entiendan y traduzcan el texto



hablado en audio. Este proceso requiere el uso de redes neuronales profundas, modelos secuenciales como los LSTMs o GRUs, y más recientemente transformadores, que son capaces de aprender de las relaciones temporales y contextuales en las secuencias de audio. Este sistema puede ser utilizado en educación para el subtitulado en directo de clases y conferencias, ofreciendo un mejor acceso a personas con faltas auditivas o mayores dificultades de aprendizaje. También puede ser utilizado en plataformas de educación a distancia para hacer que los alumnos que utilizan un sistema de tutoría automática, que responden oralmente, puedan obtener respuestas de manera más rápida. La combinación de técnicas de reconocimiento de voz y PLN permite a los estudiantes utilizar el lenguaje natural mientras interactúan con software educativo, mejorando así la experiencia educativa. (Kamath et al., 2020)

2.2.1.8. Aprendizaje Automático Aplicado para Ingenieros

El aprendizaje automático para ingenieros está orientado hacia la adopción práctica de soluciones de IA en el entorno de la ingeniería y los negocios. Este patrón incluye el uso de métodos de aprendizaje supervisados, no supervisados y de refuerzo para abordar situaciones prácticas. Es necesario que los ingenieros estén familiarizados con los conceptos básicos de IA como clasificación, regresión y agrupamiento para poder tratar problemas claramente definidos, como la predicción de fallos en sistemas industriales o la optimización del control de procesos de fabricación. Dentro de la educación, las tecnologías de aprendizaje automatizado pueden mejorar las actividades administrativas y docentes al mejorar la calidad de la toma de decisiones en el aula utilizando modelos



predictivos basados en los datos de los estudiantes, como el rendimiento académico y las actividades de interacción en el aula. Además, este enfoque ayuda a los ingenieros a crear herramientas de software que proporcionen evaluación automatizada de la adquisición de los estudiantes, adaptación de los materiales educativos y adecuación en la distribución de materiales según las necesidades del estudiante. (Prosise, 2022)

2.2.1.9. Visión por Computadora Aplicada

La visión por computadora es una subárea disciplinaria de la computación orientada a interpretar imágenes y videos, usando redes neuronales. Usando CNNs, los modelos de visión por computadora son capaces de aprender a localizar objetos, realizar tareas de reconocimiento de patrones en imágenes y realizar análisis de escenas. Estos modelos son de gran utilidad para la investigación educativa en el proceso de análisis de imágenes, donde su uso puede enfocarse en la búsqueda de patrones en los trabajos, automatizar la evaluación de documentos escritos o explorar elementos visuales que producen los estudiantes como escenas gráficas o gráficos. El uso de la visión por computadora en educación también se complementa con la realidad aumentada donde los alumnos pueden manipular objetos virtuales en tiempo real. Esta tecnología en el entorno educativo también puede contribuir a la mejora de la inclusión, mediante el uso de inteligencia artificial que proporcione explicaciones breves automáticas de las imágenes a personas con discapacidad visual que cubren todo ese material. (Shanmugamani, 2021)



2.2.1.10. Inteligencia Artificial en la Educación

En la era de educación, se observa que la IA está cambiando la forma en que se utiliza la educación ya que aparece un marco que facilita un aprendizaje más personalizado, adaptativo y más efectivo. Los sistemas basados en IA pueden procesar datos de estudiantes en masa y personalizar de forma independiente el contenido, el nivel de dificultad y el flujo de aprendizaje para ajustarse a las necesidades de cada estudiante. Esto es posible utilizando algoritmos de aprendizaje automático que ayudan a fomentar patrones en el rendimiento de los estudiantes y a predecir sus posibilidades de aprobar o fracasar. Los sistemas de tutoría inteligente también utilizan IA para brindar retroalimentación inmediata sobre las preguntas junto con sugerencias personalizadas basadas en el historial de aprendizaje del estudiante. Además, la IA también permite a los maestros automatizar tareas monótonas como calificar exámenes y marcar trabajos escritos para que puedan tener más tiempo para concentrarse en la enseñanza. La IA tiene la capacidad de ayudar a mejorar la diversidad y la inclusión en la educación al ofrecer materiales ajustables a estudiantes con diferentes necesidades de aprendizaje y velocidad de aprendizaje. (Holmes et al., 2020)

2.2.1.11. Enseñanza de la Inteligencia Artificial

La integración internacional de la educación en inteligencia artificial en las mejores clases de IA y cursos de IA está tomando progresivamente raíz porque se considera una de las habilidades fundamentales del siglo XXI para el futuro empleo e innovación. Integrar



la inteligencia artificial en el currículo implica más que simplemente aprender las habilidades técnicas como el aprendizaje automático y las redes neuronales, también incorpora los aspectos éticos y sociales relacionados con el uso de esta tecnología. Los estudiantes deben comprender la forma en que la IA afecta a diversos sectores y cómo se puede utilizar para abordar problemas globales como el cambio climático, la salud pública y la educación. Además, dentro del dominio educativo, también es posible aprovechar la IA con el propósito de adaptar el entorno de aprendizaje a las habilidades de cada aprendiz en particular. Desde esta perspectiva, a medida que los estudiantes desarrollan competencias en IA, pueden utilizar sus habilidades para crear soluciones novedosas en sus respectivas disciplinas, ya sean ciencias o humanidades, y ayudar a construir una mejor apreciación del concepto de IA y cómo puede servir para mejorar el mundo (Zimmerman, 2020).

2.2.1.12. Programación de Redes Neuronales en Python

Python ha surgido como un lenguaje de programación importante para desarrolladores de inteligencia artificial y científicos de datos, especialmente en el área de programación de redes neuronales. Gracias a un conjunto diverso de bibliotecas como Tensorflow y Keras, ajustar modelos de aprendizaje profundo se ha convertido en un proceso rápido. El paradigma de programación de aprendizaje profundo en Python brinda a los desarrolladores la capacidad de integrar fácilmente modelos como redes neuronales convolucionales, redes neuronales recurrentes e incluso modelos secuenciales para una plétora de funcionalidades, incluyendo visión por computadora y procesamiento de lenguaje natural. También



cuenta con herramientas de visualización y análisis que ayudan a los investigadores a estudiar el comportamiento de los modelos y optimizar los hiperparámetros para una mejor efectividad de los modelos. Con propósitos educativos, la programación de una red neuronal utilizando el lenguaje Python desarrolla un entorno práctico para maestros y estudiantes, quienes pueden trabajar con los modelos de IA en tiempo real para aprender o probarlos. (Chollet, 2021)

2.2.1.13. Desarrollo de Aplicaciones Basadas en IA

La aplicación de aprendizaje automático en desarrollo de aplicaciones ha cambiado el enfoque que adoptan las empresas y organizaciones para resolver problemas complejos. Las aplicaciones actuales brindan la posibilidad de que determinados sistemas mejoren y permanezcan en constante adaptación a la información disponible, sin la necesidad de contar con algún tipo de intervención por parte de una persona. Para concretar un desarrollo, se debe destacar un problema de interés, reunir la información relevante relacionada con este, elegir el modelo a utilizar, entrenarlo y ajustarlo si es necesario, y finalmente exponerlo al público. En el ámbito educativo, el desarrollo de aplicaciones que cuentan con el uso de IA tiene la posibilidad de realizar evaluaciones automáticas de los estudiantes, proporcionarles inteligencia encargada de darles tutorías, y adaptar el contenido imprimible a las necesidades de cada uno. Estas aplicaciones también realizan la función de recopilar y analizar grandes volúmenes de información acerca de los estudiantes para crear patrones y llegar a tener resultados predecibles. Pasar a un modelo en el que se priorice el desarrollo de productos por medio de una inteligencia



artificial tiene el potencial de transformar la educación a la que estamos acostumbrados. (Ameisen, 2020)

2.2.1.14. Aprendizaje por Refuerzo Profundo

Si se compara el aprendizaje reforzado tradicional y el profundo, el segundo consiste en crear un modelo, un agente, que domine las redes neuronales y que logre tomar decisiones en situaciones complicadas. El aprendizaje reforzado, al igual que el que posee supervisión, está encaminado a lograr un objetivo, aunque este tipo de aprendizaje no usa anotaciones. Su funcionamiento está basado en que el agente que se utiliza va a ir interactuando con su entorno, en el cual recibirá beneficios o quedará en opuestas por las acciones que realice. Esta visión ha demostrado eficacia en numerosas aplicaciones como el videojuego, control de robots y tareas de optimización de redes. En educación, el aprendizaje reforzado podría utilizarse para desarrollar simulaciones interactivas que permitan a los estudiantes navegar en entornos de aprendizaje y obtener retroalimentación sobre las elecciones que hacen casi instantáneamente. Además, los agentes de aprendizaje reforzado pueden responder a las acciones de un estudiante en tiempo real, cambiando el nivel de desafío de las tareas en función del rendimiento previo del estudiante. Esta posibilidad de adaptar el proceso educativo y proporcionar retroalimentación continua lo convierte en una herramienta efectiva o poderosa en la educación actual (Lapan, 2020).



2.2.1.15. Métodos de Redes Neuronales en Procesamiento del Lenguaje Natural

El uso de redes neuronales en el procesamiento del lenguaje natural (NLP) ha revolucionado el modo en que las máquinas entienden e interactúan con el lenguaje humano. Antes, el NLP se realizó mediante palabras clave así como el uso de estadísticas, pero con la llegada de las redes neuronales sobre todo de las redes neuronales recurrentes (RNN), LSTMs, y en los últimos tiempos los Transformers, los modelos de NLP han sido capaces de mejorar su comprensión y el contexto sobre la semántica del lenguaje. Estas redes son capaces de aprender representaciones vectoriales de palabras (embeddings), que capturan relaciones semánticas y sintácticas entre palabras. Las redes neuronales facilitan que los modelos puedan llevar a cabo tareas como la traducción automática, análisis de la componente de sentimientos, generación o redacción de textos así como la clasificación de documentos. En el campo educativo, se emplean métodos y técnicas de redes neuronales para el análisis de las respuestas escritas de los alumnos, corregir automáticamente ensayos escritos y ofrecer luces o ejercicios lectivos a partir del rendimiento logrado por los estudiantes, entre otros. (Goldberg, 2021)

2.2.1.16. Dominio del Aprendizaje por Refuerzo con Python

El aprendizaje por refuerzo es esa área del aprendizaje de máquinas que se ocupa con secuenciales, decisiones, es decir, una temporal interactiva en donde el agente aprende a interactuar en un ambiente dadas



unas determinadas condiciones para maximizar el premio que en dicho ambiente se acumula. El Deep Reinforcement Learning tiene colocado de modo mixto las técnicas con redes neuronales hacia el control de problemas que requieren un gran espacio en estados y acciones. En este sentido, Python proporciona gran versatilidad y un amplio rango de bibliotecas como Tensorflow y Pytorch para su desarrollo y entrenamiento de agentes de aprendizaje por refuerzo. En educación, dichos agentes pueden ser usados para crear trayectorias personalizadas de aprendizaje donde el contenido o la dificultad de las tareas se ajustan a la conducta del estudiante. Igualmente, los módulos de aprendizaje profundo pueden ayudar en el aprendizaje por refuerzo, donde los estudiantes son expuestos y/o utilizan simulaciones educativas en un ambiente controlado donde las acciones realizadas por el estudiante son fácilmente respondidas, permitiendo un aprendizaje efectivo a través de la práctica y la exploración. (Bilgin, 2021)

2.2.1.17. Introducción al Aprendizaje Automático con Python

El aprendizaje automático es una subdisciplina de la inteligencia artificial cuyo objetivo es la elaboración de modelos que a partir de datos son capaces de aprender. Dentro de este rubro, uno de los lenguajes mayormente utilizados es Python, lo que es atribuible a su facilidad de uso y la disponibilidad de librerías como scikit-learn que también simplifica la aplicación de algoritmos más comunes de clasificación, regresión y agrupamiento. Un desarrollo de modelos de aprendizaje automático sigue el proceso de recopilación de datos pertinentes, selección del modelo adecuado, entrenamiento del modelo con datos históricos, evaluación del



desempeño seguido de una calibración de hiperparámetros para lograr mejores resultados. En el ámbito educativo, el aprendizaje automático se ocupa de la gran cantidad de información de los estudiantes, busca patrones y produce estimaciones sobre su desempeño académico o la conducta de los estudiantes. Por otro lado, los sistemas de aprendizaje automático pueden enseñar a los estudiantes sobre algún contenido en específico e incluso ayudan a los maestros a diseñar sus clases analizando tendencias que ocurren dentro de su salón de clases. (Müller & Guido, 2020)

2.2.1.18. Aprendizaje Profundo con JavaScript

A pesar de que el aprendizaje profundo se asocia más con el uso de Python, el entrenamiento y despliegue de modelos de aprendizaje profundo utilizando JavaScript puro está ganando popularidad con la introducción de TensorFlow.js. Esta biblioteca hace posible realizar el entrenamiento y despliegue de modelos de aprendizaje profundo en los navegadores web de los clientes, lo que allana el camino para un nuevo cambio hacia la construcción de aplicaciones de IA disponibles en una amplia gama de dispositivos. Usando el navegador web y Node, los desarrolladores de TensorFlow.js pueden aprovechar el aprendizaje profundo al desarrollar aplicaciones web sin necesidad de instalar un software especial en los dispositivos de los usuarios. Permitir aplicaciones de aprendizaje profundo construidas con JavaScript permite a estudiantes y profesores probar diferentes modelos de IA a través de la web en un navegador compatible con JavaScript sin preocuparse por las herramientas



o la integración compleja para programar. Esto es ideal para simulaciones educativas interactivas o experimentos en la web. (Cai et al., 2020)

2.2.1.19. Teoría del Aprendizaje Profundo

La teoría de la imitación de la enseñanza del aprendizaje profundo se ocupa del análisis de los fundamentos matemáticos y conceptuales de las redes neuronales artificiales profundas. A pesar de la extensa investigación sobre las aplicaciones prácticas de estas redes, la teoría de cómo funcionan sigue siendo un área activa de investigación. Esta rama investiga las capacidades de generalización de las redes, los comportamientos de las funciones de pérdida, las optimizaciones en espacios de alta dimensión y la estabilidad de las soluciones. El aprendizaje profundo ha sido muy efectivo en la práctica, mostrando capacidades asombrosas en disciplinas con complejidades sin precedentes, y los investigadores tienen curiosidad por descubrir por qué estos modelos están funcionando tan bien en una amplia variedad de tareas. En este contexto, una comprensión exhaustiva de los principios del aprendizaje profundo permite a los investigadores construir modelos efectivos y precisos para procesar datos de estudiantes y mejorar los sistemas de tutoría inteligente al ajustar los parámetros del modelo para un rendimiento óptimo en los entornos de aprendizaje dados. (Roberts et al., 2022)

2.2.1.20. Inteligencia Artificial y sus Implicaciones Éticas en la Educación

A medida que la inteligencia artificial (IA) se integra en los sistemas educativos, surgen cuestiones éticas fundamentales. En la



educación, la IA está involucrada en la toma de decisiones automatizadas que van desde la evaluación del rendimiento estudiantil hasta la provisión de recomendaciones de contenido de aprendizaje. Sin embargo, pueden surgir sesgos en la IA si los datos relevantes para el entrenamiento del modelo no son representativos. Además, el uso extensivo de los datos personales de los estudiantes debe abordarse con cuidado en lo que respecta a la privacidad. La IA debe desarrollarse de tal manera que sea transparente y explicativa para educadores y administradores, y también equitativa para que los resultados sean razonables para ellos. En relación con la educación, esto es ideal, especialmente para asegurarse de que los sistemas de toma de decisiones algorítmicas no sean perjudiciales para grupos sensibles ni se utilicen para mantener las desigualdades del status quo. (Mitchell, 2020)

2.2.1.21. Compatibilidad de la Inteligencia Artificial con la Autonomía Estudiantil

La compatibilidad de los factores humanos de los sistemas de IA es importante en las aulas porque tales sistemas pueden alterar los caminos de aprendizaje de los estudiantes positivamente o negativamente. Al igual que con otras tecnologías educativas, los sistemas de IA deben estructurarse de tal manera que respeten la autogobernanza de los estudiantes al enfatizar elecciones informadas sobre su aprendizaje. Es necesario que los algoritmos de evaluación y recomendación educativa no sean sesgados y no limiten las opciones de los estudiantes. Además, los sistemas de IA deben ser amigables para el usuario y proporcionar



descripciones adecuadas para que los educadores puedan tomar decisiones sobre si corregir a los estudiantes o no. (Russell, 2020)

2.2.1.22.Liderazgo Global en el Desarrollo de IA para la Educación

El liderazgo global en inteligencia artificial (IA) tiene implicaciones de manera directa en el ámbito educativo. Los desarrollos en IA que se están llevando a cabo especialmente en naciones como China y Estados Unidos, están revolucionando los sistemas formacionales clásicos. La IA en la educación hace posible el diseño y expansión de los módulos. De este modo, cuando una plataforma de aprendizaje virtual se desarrolla de acuerdo a las capacidades individuales de cada estudiante, el contenido y el ritmo se ajustan individualmente según el rendimiento del estudiante. Además de eso, las tecnologías impulsadas por IA podrían asistir en la mejora de la gestión educativa al permitir la previsión del rendimiento, el análisis de patrones de comportamiento y la elaboración de perfiles de riesgo de estudiantes propensos a fracasos académicos. La posesión de tales tecnologías por parte del sector educativo global exige cambios dentro de los sistemas educativos para permitir que los estudiantes estén listos para trabajos en la era de la automatización y la IA. (Lee, 2021)

2.2.1.23.Ética y Control de la Superinteligencia en la Educación

La superinteligencia como un concepto plantea consideraciones éticas y prácticas significativas, especialmente para los educadores. Parece que el desarrollo de IA con capacidades superiores a las humanas refuerza la idea de que los sistemas educativos tienen que cambiar para que los



estudiantes puedan interactuar y colaborar con estas tecnologías. Por otra parte, sucede que no podemos pedir a la gente del futuro que se involucre en políticas en torno a la superinteligencia porque sea mejor que un ser humano a menos que exista un desarrollo con implicaciones éticas en el futuro; el uso de IA debe estar integrado en lo educativo de modo que siempre existan valores éticos del uso de estas tecnologías avanzadas así como el control y la gestión de las mismas para lograr un equilibrio y desarrollo responsable. (Bostrom, 2020)

2.2.1.24. Aplicación de Redes Neuronales Convolucionales en la Educación

Se estima que la educación tuvo una gran evolución en los 30 años precedentes a nuestra década. Ahora en el siglo XXI la integración del uso de software esencial en la enseñanza es de suma importancia a nivel profesional. El subcampo de la visión por computadora que estudia el comportamiento visual de un sistema informático se ha vuelto de gran importancia en el ámbito educativo. Una de las aplicaciones de este subcampo que trata sobre la interpretación automatizada de imágenes digitales es el uso de redes neuronales en la educación. Nuevamente esta tecnología de IA puede automatizar la corrección de pruebas, reconocer el estilo de los alumnos y otorgar comentarios a cada alumno en tiempo real. Asimismo, la red neuronal convolucional puede aplicarse en el aula y ser utilizada en sistemas de realidad mixta, donde el alumno interactúa con contenido visual que es en vivo. El análisis educativo mediante tecnologías de visión artificial permite personalizar el proceso educativo, así como democratizar el proceso educativo. (Rosebrock, 2021)



2.2.1.25. Aprendizaje por Refuerzo en Sistemas Educativos

Aprender haciendo a la vez es un dispositivo que tiene como objetivo la interacción con su medio y el aprendizaje por el logro de recompensas que se conceptualizan como correctas. En el contexto educativo, esta tecnología nos habilita a construir sistemas de tutoría, que son muy eficaces en adaptarse a las lecciones sobre el desempeño de los alumnos. Con el uso de aprendizaje por refuerzo, se puede volver autoajustable el sistema educativo de acuerdo con la manera como los alumnos respondan, de tal manera que el contenido y la dificultad del mecanismo de tareas será determinístico y de hecho para cada uno. Esta destreza de adaptación da lugar a un rango más ancho de modificaciones en las enseñanzas obtenidas, que modernicen las enseñanzas interactivas y modeladas, a lo que los niños necesitan para motivarlos cada vez más para un mejor rendimiento. (Sutton & Barto, 2020)



CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. TIPO Y DISEÑO DE INVESTIGACIÓN

Este trabajo de investigación fue de tipo aplicada, centrado en la implementación de técnicas de aprendizaje profundo para el análisis de datos educativos en el Instituto de Educación Superior Pedagógico Público Puno. La meta fue resolver la propuesta, haciendo uso de conocimientos modernos en inteligencia artificial para así mejorar los procesos de tratamiento y análisis de grandes masas de datos con el objetivo de lograr un entendimiento de los patrones de desempeño y factores de éxito del estudiante.

El diseño de la investigación fue de tipo cuasi experimental ya que se realizaron control y manipulación de determinadas variables en un escenario educativo real, con la finalidad de probarlas contra otras variables. Dentro de los factores dependientes se incluyeron, el uso de los algoritmos que mejoran la velocidad a la vez que la precisión que los modelos se convierten en predicciones del rendimiento académico y la localización de variables de gran interés, y los factores independientes fueron los algoritmos de aprendizaje profundo que fueron el foco de estudio. La muestra fue no probabilística, siguiendo las huellas de los datos históricos pertenecientes a los estudiantes del instituto de manera que fuera posible evaluar el efecto de la mejora del análisis computacional sobre el aspecto de la toma de decisiones educativas.

Método: Implementación de técnicas de aprendizaje profundo.

Tipo de investigación: Aplicada.

Nivel de investigación: Explicativo.

Diseño de investigación: Cuasi-experimental.

3.2. POBLACIÓN Y MUESTRA

3.2.1. Población

La población de esta investigación estuvo compuesta por los estudiantes matriculados en el Instituto de Educación Superior Pedagógico Público Puno en el año 2023, distribuidos en los distintos ciclos de estudios. Se incluyeron tanto estudiantes del ciclo inicial (I) hasta el ciclo final (X), abarcando un total de 497 estudiantes, de los cuales 174 son hombres y 323 son mujeres. Estos estudiantes fueron seleccionados como base de datos para realizar el análisis de patrones educativos mediante técnicas de aprendizaje profundo. A continuación, se presenta un desglose detallado de la población:

Tabla 1

Población

Grupo	Descripción de la población	Cantidad
Ciclo I	Estudiantes del primer ciclo	158
Ciclo III	Estudiantes del tercer ciclo	114
Ciclo V	Estudiantes del quinto ciclo	90
Ciclo VII	Estudiantes del séptimo ciclo	83
Ciclo IX	Estudiantes del noveno ciclo	135
Total	Total de estudiantes matriculados 2023	497

Nota: La tabla muestra una mayor concentración de estudiantes en el Ciclo I, una disminución progresiva en los ciclos superiores y un incremento en el Ciclo IX.



3.2.2. Muestras

Para esta investigación se estableció por medio de un enfoque no probabilístico de muestreo de manera intencional se seleccionó a los estudiantes en base a su grado de antigüedad, es decir de los ciclos contabilizados en que se encuentren en V, VII y IX, ya que en el camino han alcanzado un mayor progreso y se dispone de un menor error al analizar la predicción del rendimiento académico y sus factores. Estos estudiantes fueron elegidos para que pueden hacer uso de técnicas de aprendizaje profundo en los datos proporcionados. En total, la muestra estuvo compuesta por 308 estudiantes correspondientes al 62% de toda la población objetivo.

3.2.3. Ubicación y Descripción de la Población

El alcance de la investigación se limitó a los estudiantes del Instituto de Educación Superior Pedagógico Público Puno situado en la Ciudad Pedagógica Salcedo en el Departamento de Puno. Esta es una institución educativa de gestión pública y actúa bajo la autoridad del Departamento de Educación. Lo siguiente son las ubicaciones y otros detalles pertinentes sobre la población que se está considerando:

Departamento: Puno

Provincia: Puno

Distrito: Puno

Dirección: Ciudad Pedagógica Salcedo, Av. Pedagógica S/N

Código modular: 0474320

Esta institución ofrece cursos certificados en Educación Inicial, Docencia Primaria, Educación Física, Computación y Tecnología de la Información, y



varias disciplinas en Educación Secundaria. El entorno urbano de Ciudad Pedagógica Salcedo ofrece un buen ambiente para la búsqueda de la educación relacionada entre los estudiantes matriculados.

3.3. TÉCNICAS E INSTRUMENTOS

3.3.1. Técnicas e instrumentos para la recolección de datos

En la presente investigación, se emplearon diversas técnicas e instrumentos con el objetivo de recolectar datos relevantes para el análisis de patrones educativos mediante técnicas de aprendizaje profundo en el Instituto de Educación Superior Pedagógico Público Puno.

Planificación técnica del proyecto.

Recolección de datos académicos y sociodemográficos que ya han sido almacenados.

Procesamiento y análisis de la información obtenida.

A continuación, se describen las técnicas e instrumentos específicos utilizados en la recolección de datos:

Tabla 2

Técnicas e Instrumentos

TÉCNICAS	JUSTIFICACIÓN	INSTRUMENTO
Revisión de registros académicos	Recopilación de información sobre el historial académico y rendimiento estudiantil.	Base de datos de registros académicos
Análisis de datos educativos	Identificación de patrones en el rendimiento académico y factores de riesgo educativo.	Dataset estructurado de datos educativos
Análisis sociodemográfico	Análisis de datos previos sobre características demográficas y socioeconómicas.	Base de datos sociodemográfica

Nota: Se presenta técnicas, justificaciones e instrumentos utilizados para recopilar y analizar datos académicos, educativos y sociodemográficos.

3.3.2. Técnicas para el análisis de la información

Se realizó un análisis de los datos recolectados para poder validar los resultados obtenidos mediante las técnicas de aprendizaje profundo aplicadas en este trabajo. El análisis se centró en identificar diversos patrones de rendimiento académico y correlacionarlos con factores sociodemográficos. Se utilizaron métricas clave como la precisión de las predicciones, la identificación de factores de riesgo y la correlación entre variables académicas y sociodemográficas. Estos resultados fueron contrastados con el marco teórico y los objetivos propuestos.



3.3.3. Técnicas para el procesamiento de datos

- Base de Datos de Registros Académicos: Se utiliza para mantener y organizar la asistencia de los estudiantes y otros datos relacionados con la historia del rendimiento académico.
- Conjunto de datos estructurado del carácter educativo: Se trata de un conjunto de datos que ha sido obtenido con la ayuda del análisis de los registros académicos y la información semi demográfica que rastrea el conjunto de datos por medio de técnicas de aprendizaje avanzado.
- Base de Datos sociodemográfica: Teniendo los datos basados en el ambiente económico y social en el cual se desenvuelven los alumnos, dentro de este conjunto de datos se tiene como información adicional para el análisis y modelación de estos.

3.4. PLAN DE TRATAMIENTO DE DATOS

El tratamiento de los datos se realizó a través de procedimientos estructurados para asegurar la correcta interpretación y el análisis de la información educativa disponible en el Instituto de Educación Superior Pedagógico Público Puno. Las fases de dicho plan de tratamiento de datos comprendidas en el mismo fueron:

Codificación y tabulación: La información académica y sociodemográfica que contaba con datos útiles fue convertida a un formato que garantizaba la integridad de la información para la posterior utilización de esta en técnicas de aprendizaje profundo.

Análisis descriptivo: Este tipo de análisis fue necesario en el intento de contextualizar los datos en torno a ciertos indicadores como el promedio de las calificaciones y su dispersión entre los alumnos que luego sirvió para su posterior análisis



con el objetivo de detectar patrones elementales de la conducta del alumno en cuanto a su rendimiento.

Análisis inferencial: Un análisis inferencial fue realizado con el fin de verificar si el comportamiento observado en los datos académicos era reproducible en el total de estudiantes; para ello se analizaron las interrelaciones que existieron entre las variables dependientes e independientes.

Presentación de resultados: Los resultados fueron representados visualmente en forma de gráficos y tablas para entender rápidamente las tendencias y relaciones importantes en los resultados académicos.

Interpretación de métricas: Resultados promediados como la desviación estándar y correlaciones entre parámetros académicos y sociodemográficos fueron procesados en el contexto de los objetivos del estudio con los reclamos sobre la eficiencia de las técnicas de aprendizaje profundo en la detección de patrones de riesgo y éxito para los estudiantes.

3.4.1. Análisis de datos

El análisis de los datos se enfocó en evaluar la efectividad de las técnicas de aprendizaje profundo aplicadas en la identificación de patrones académicos. Se realizó una comparación de los resultados con los datos básicos y se exploraron las relaciones entre variables dependientes (rendimiento académico) y variables independientes (características demográficas y educativas). Este análisis permitió la obtención de conclusiones sólidas sobre los factores que afectan el rendimiento académico de los estudiantes y confirmó o desacreditó las hipótesis propuestas.



3.5. APLICACIÓN DE LA METODOLOGÍA

3.5.1. Metodología en Prototipo

Para este estudio en particular, se seleccionaron la metodología de prototipos que permitió el desarrollo gradual de un modelo para la evaluación de datos educativos mediante el uso de técnicas de aprendizaje profundo. Debido a su adaptabilidad y optimización incremental del modelo de predicción, se adoptó esta metodología. Las principales características que justificaron tal adscripción fueron las siguientes:

- Ciclo constante de retroalimentación: Se integraron elementos de evaluación y modificación en cada etapa del proceso de desarrollo. Tal enfoque proporcionó un ajuste continuo del modelo, basado en los resultados de etapas anteriores, lo que a su vez mejoró la precisión de la predicción del rendimiento académico de los estudiantes.
- Validación temprana: Se realizaron pruebas regulares sobre datos históricos, estimando la cercanía del modelo a la realidad y se realizaron ajustes temporales para mover sus parámetros. Esto hizo posible asegurar una validación efectiva dentro de un entorno controlado sin interacción directa con los usuarios.
- Reducción de la incertidumbre: A través de la creación de prototipos iterativos fue posible abordar preocupaciones vinculadas al empleo de datos heterogéneos, así como modificar las técnicas de procesamiento de datos antes de una aplicación más amplia.
- Incorporación de nuevas variables: A medida que se pasaba a una nueva fase más elaborada del desarrollo, se introdujeron en el prototipo nuevas



variables que no se habían considerado anteriormente, como por ejemplo indicadores sociodemográficos, lo que elevó la capacidad predictiva del modelo en base.

3.5.2. Material Experimental

3.5.2.1. Infraestructura Tecnológica:

- **Hardware:** Se utilizaron servidores dedicados y con capacidad de procesamiento paralelo cumpliendo con los volúmenes de datos y a su vez entrenando los modelos de aprendizaje profundo. También se implementaron otros sistemas que ayudarían a respaldar la información para evitar pérdidas.
- **Software:** Se utilizaron herramientas especializadas como entornos de Big Data y de aprendizaje profundo, enfoque que consistía en plataformas e infraestructura de procesamiento de datos masivos y bibliotecas para el desarrollo de modelos predictivos. Se utilizaron estos recursos a pesar de que facilitaron el análisis y la búsqueda de los algoritmos y el desarrollo de los mismos.
- **Almacenamiento y procesamiento:** Para los fines de la investigación se utilizaron bases de datos distribuidas, las cuales facilitaron el almacenamiento seguro y efectivo de la información académica y con desagregación socioeconómica, permitieron acceder a registros históricos y hacer consultas rápidas dentro de cada una de las iteraciones del modelo.



3.5.2.2. Recursos Digitales y Servicios:

- Servicios en la nube: Con el fin de entrenar los modelos y realizar el procesamiento de datos en paralelo, se empleó una infraestructura en la nube, lo que permitió escalar los recursos a los requisitos del proyecto y reducir los tiempos de procesamiento.
- Monitoreo y soporte técnico: Para las actividades clave de implementación, hubo asistencia técnica especializada para garantizar un funcionamiento fluido y efectivo durante el trabajo con grandes conjuntos de datos

3.5.2.3. Materiales de Apoyo:

- Estaciones de trabajo: Se utilizaron computadoras de alto rendimiento durante el trabajo en el desarrollo de scripts de procesamiento de datos, visualización de datos y evaluación de modelos.
- Documentación del proyecto: Se realizó una contabilidad detallada de los resultados alcanzados y las decisiones tomadas en cada etapa de la construcción del prototipo, lo que permitió controlar y corregir las etapas posteriores.

3.6. ASPECTOS ÉTICOS

En el transcurso de esta investigación, se cumplieron plenamente los principios éticos pertinentes que rigen el análisis de datos educativos y la implementación de técnicas avanzadas de aprendizaje profundo, a fin de mantener la integridad y la



responsabilidad en todas las fases del proyecto. Los principales principios éticos considerados fueron:

- **Interacción con instituciones educativas y sociales:** La consideración del análisis de datos educativos es una manera de influir en los determinantes que sustentan los logros educativos de los estudiantes, con el fin de mejorar las políticas y estrategias del Instituto Educativo Público de Puno. Este paradigma está en consonancia con la responsabilidad social de mejorar la inclusión y la calidad en la educación mediante el uso ético de la tecnología y los datos.
- **Gestión de datos:** Los datos académicos y sociológicos utilizados en esta investigación fueron mantenidos en la más estricta confidencialidad para asegurar la privacidad de los estudiantes y el cumplimiento de las leyes de protección de datos aplicables. Todos los materiales fueron mantenidos únicamente para fines de investigación, y se aseguró que no se recopilaran ni mal utilizaran datos sensibles identificables.
- **Transparencia y veracidad:** Se aseguró que toda la información resultante del procesamiento de datos y modelos de aprendizaje profundo se informara de manera precisa, sin sesgos, y con adhesión a los más altos estándares éticos en la investigación y el desarrollo tecnológico. Los resultados del análisis de datos, incluyendo procesos, métodos y procedimientos, se presentaron de manera imparcial y respaldados por una investigación y análisis exhaustivos.
- **Consentimiento informado:** Aunque no hubo contacto directo con estudiantes o docentes, se protegieron los derechos de aquellos cuyos datos fueron utilizados y el análisis se realizó de acuerdo con las políticas de gobernanza de datos de la institución.



CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

Este capítulo presenta los hallazgos de la investigación que se centran en las técnicas de aprendizaje profundo empleadas para la estimación y evaluación de los datos académicos y sociodemográficos de los estudiantes matriculados en el Instituto de Educación Superior Pedagógico Público Puno.

4.1. DISEÑAR LA ARQUITECTURA DE SOFTWARE DEL SISTEMA DE ANÁLISIS DE DATOS EDUCATIVOS

4.1.1. Identificación de los componentes del sistema:

Se diseñaron y estructuraron los módulos básicos de la arquitectura del sistema de análisis de datos educativos, cada uno desempeñando un propósito específico en el procesamiento de datos. Los módulos incluyeron:

- Módulo de entrada de datos: En este caso, se desarrolló un módulo implementando Python que se unía directamente a la base de datos MySQL del sistema existente a través de controladores nativos. Este módulo emitió solicitudes SQL optimizadas que extrajeron registros demográficos y académicos. La extracción de datos se realizó en lotes utilizando el esquema de procesamiento por lotes para aliviar la carga en el servidor y asegurar una extracción efectiva. Los datos obtenidos se almacenaron luego en estructuras DataFrame de Pandas, lo que permitió manipular y preprocesar los datos en las etapas posteriores de manera más eficiente.



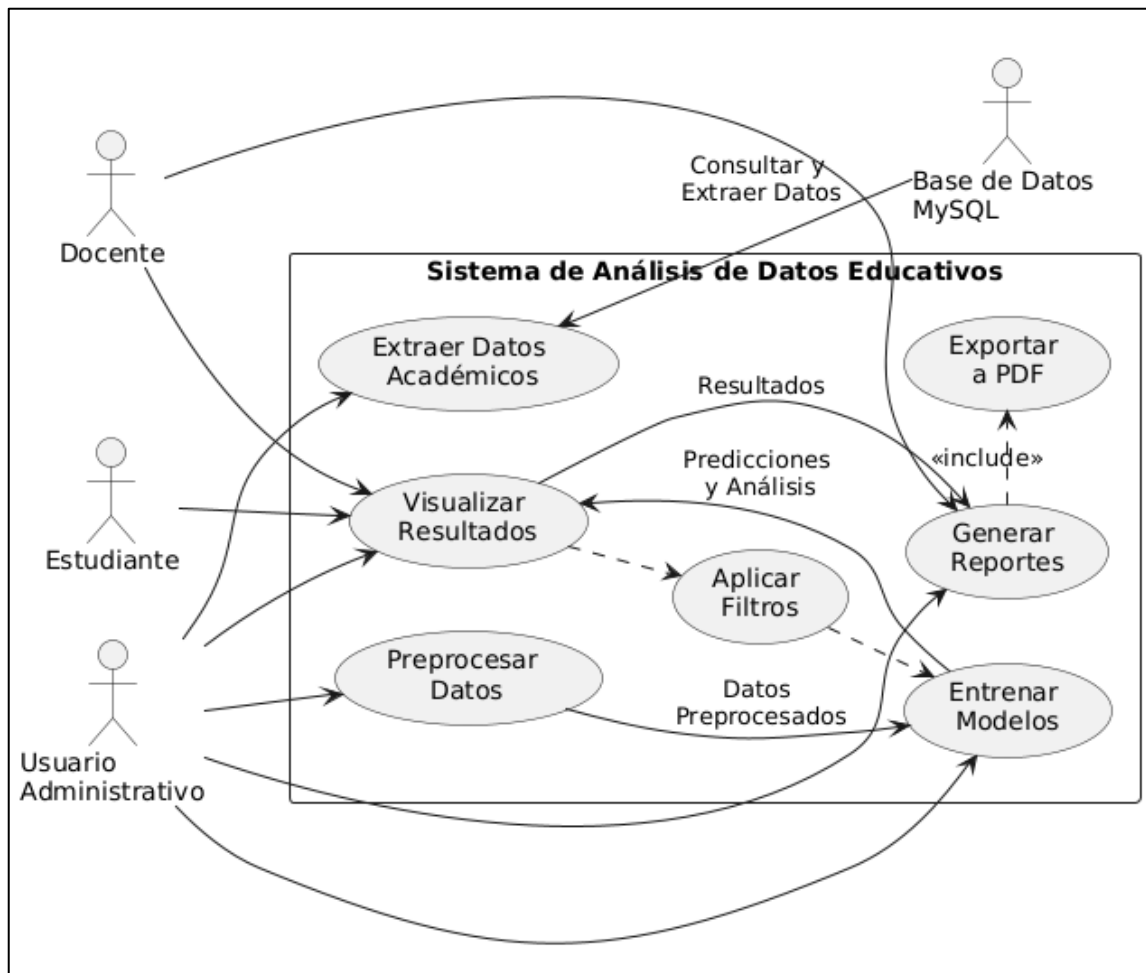
- Módulo de Procesamiento de Datos - Se estableció un pipeline de procesamiento que se encargaba de la limpieza y transformación de los datos extraídos. Este módulo fue capaz de eliminar registros duplicados, gestionar valores faltantes mediante imputación, y se utilizó codificación one hot para las transformaciones de las variables categóricas. Además, las variables numéricas fueron normalizadas utilizando escalado min-max de tal manera que los datos estuvieran listos para las fases analíticas subsiguientes, pero aún antes del trabajo de entrenamiento del modelo que se iba a realizar. Todas estas tareas se llevaron a cabo en paralelo para optimizar el rendimiento del umbral y minimizar los tiempos de procesamiento.
- Módulo de Almacenamiento - Se creó un sistema para el almacenamiento de datos preprocesados en formatos JSON y CSV con el propósito de facilitar el acceso y uso de los datos en etapas posteriores del análisis. Además, se implementó un sistema de almacenamiento temporal, utilizando Redis como backend para almacenar resultados intermedios en memoria, facilitando así el acceso a los datos más frecuentemente necesarios en operaciones de visualización, como la encapsulación de bases de datos comercializables incrustadas en RDBMS con Redis debido a su eficiencia en operaciones de lectura y escritura en tiempo real.
- Módulo de Visualización: Se implementó una plataforma de visualización interactiva utilizando Dash, que además permitía a los usuarios interactuar con los datos preprocesados en forma de gráficos producidos con Plotly. Este módulo hizo posible el desarrollo de paneles dinámicos que permitían presentar datos académicos a través de diagramas de líneas, barras y

dispersión. Los usuarios podían utilizar filtros dinámicos que cambiaban las vistas de los datos por variables como cohortes de estudiantes, años de estudio y categorías sociodemográficas.

- La implementación de procesamiento en lotes y consultas SQL optimizadas permitió reducir el tiempo de recuperación de datos de 8 segundos a 2.5 segundos, mejorando significativamente el rendimiento del sistema.

Figura 1

Diagrama de casos de uso



Nota: El diagrama muestra cómo usuarios interactúan con un sistema de análisis de datos educativos.



El diagrama de casos de uso ilustra las interacciones principales entre los usuarios y las características del sistema de análisis de datos educativos. Los usuarios comprenden al usuario administrativo, al profesor, al estudiante y la base de datos MYSQL. Las características principales incluyen: la ingesta de registros académicos, la limpieza fluida de datos, el entrenamiento de análisis de aprendizaje profundo, la visualización de resultados en paneles interactivos y la generación de informes. A través de relaciones de inclusión y extensión, hay dependencias vinculadas y extensiones opcionales, como la exportación de informes en formato pdf y la personalización de la visualización a través de filtros. Este modelo explica cómo los módulos del sistema interactúan entre sí para cumplir con las tareas de análisis y asegura que las funcionalidades del sistema diseñado satisfacen los roles y requisitos de sus usuarios.

4.1.2. Selección de las tecnologías:

Las tecnologías seleccionadas fueron elegidas en función de su capacidad para interafectar con el sistema existente de PHP/MySQL, así como por problemas de rendimiento y escalabilidad en:

- **Conexión MySQL:** Se proporcionaron Interfaces de Programación de Aplicaciones para Python con el propósito de acceder a la base de datos utilizando declaraciones SQL, las cuales se caracterizaban por una baja carga en el servidor MySQL. Para manejar grandes volúmenes de datos se utilizó la técnica de procesamiento por lotes.
- **Procesamiento de Datos:** Para procesar la gran cantidad de datos obtenidos, se adoptó Pandas debido a su eficiencia en el manejo de grandes volúmenes de información, así como sus capacidades de preprocesamiento

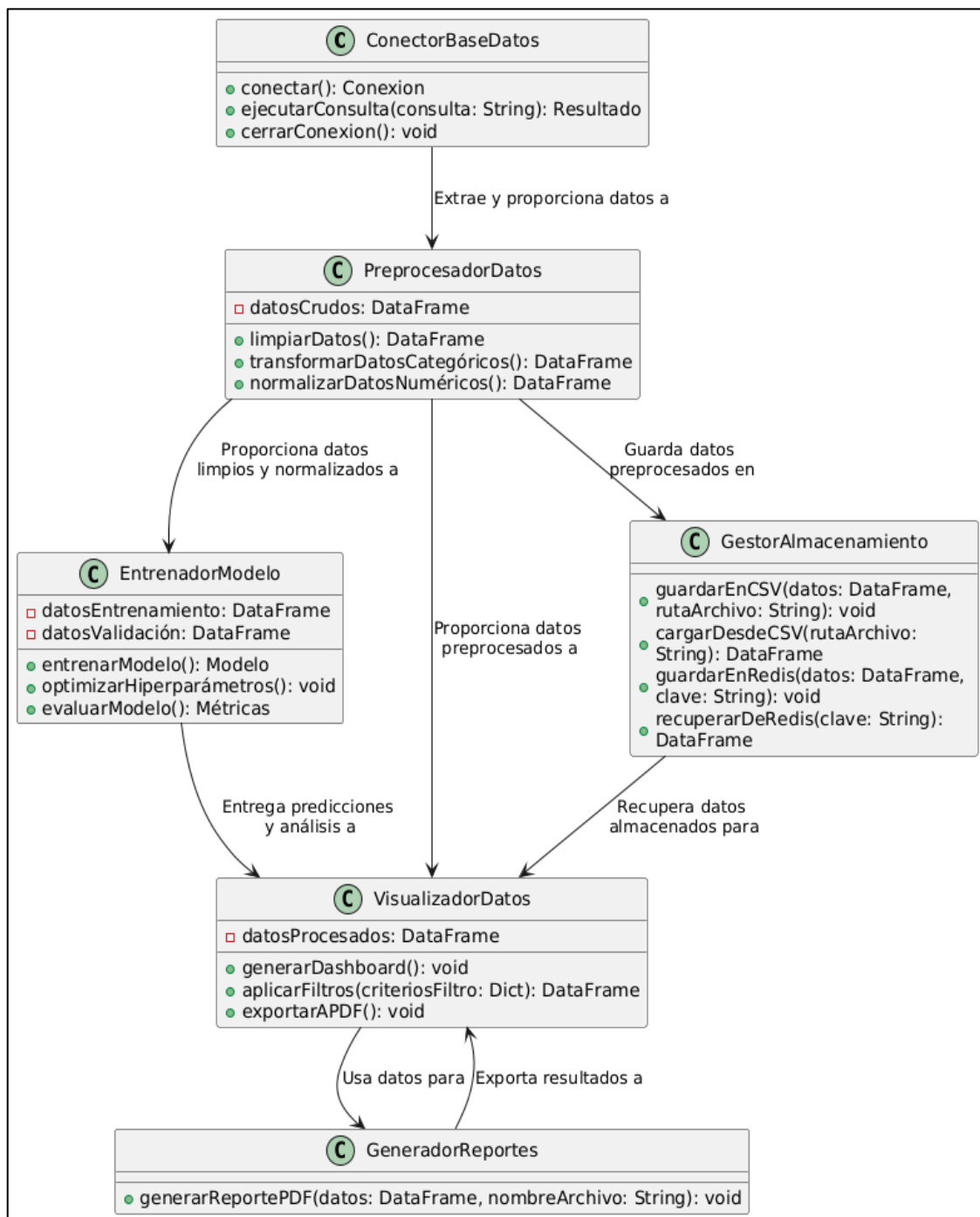


y transformación de datos. También se utilizó NumPy para acelerar cálculos matemáticos y estadísticos al manejar datos en alto volumen.

- Visualización de datos: La tecnología Dash se utilizó en la construcción de los paneles como una aplicación de visualización de datos interactiva y así permitió la creación de una aplicación web responsiva con paneles dinámicos con actualizaciones de datos en tiempo real, lo cual fue posible con el uso de Plotly. El software permitía una visualización compleja de los datos en forma de diagramas de dispersión, gráficos de líneas y gráficos de barras. Este software permitía una visualización compleja de datos en forma de diagramas de dispersión, gráficos de líneas y gráficos de barras.

Figura 2

Diagrama de Clases



El diagrama de clases representa la estructura estática del sistema de análisis de datos educativos, destacando las clases principales, sus atributos, métodos y relaciones. Entre las clases clave se encuentran el `ConectorBaseDatos`, que gestiona la conexión y consultas a la base de datos MySQL; el

PreprocesadorDatos, encargado de limpiar, transformar y normalizar los datos; y el EntrenadorModelo, que utiliza los datos procesados para entrenar y evaluar modelos de aprendizaje profundo. Además, el VisualizadorDatos genera dashboards interactivos y gráficos personalizables, apoyado por el GestorAlmacenamiento, que maneja la persistencia en formatos como CSV y Redis. Por último, el GeneradorReportes permite crear reportes en PDF con los resultados obtenidos. Las relaciones entre las clases reflejan el flujo de datos y la interacción modular del sistema, asegurando escalabilidad y eficiencia en su operación.

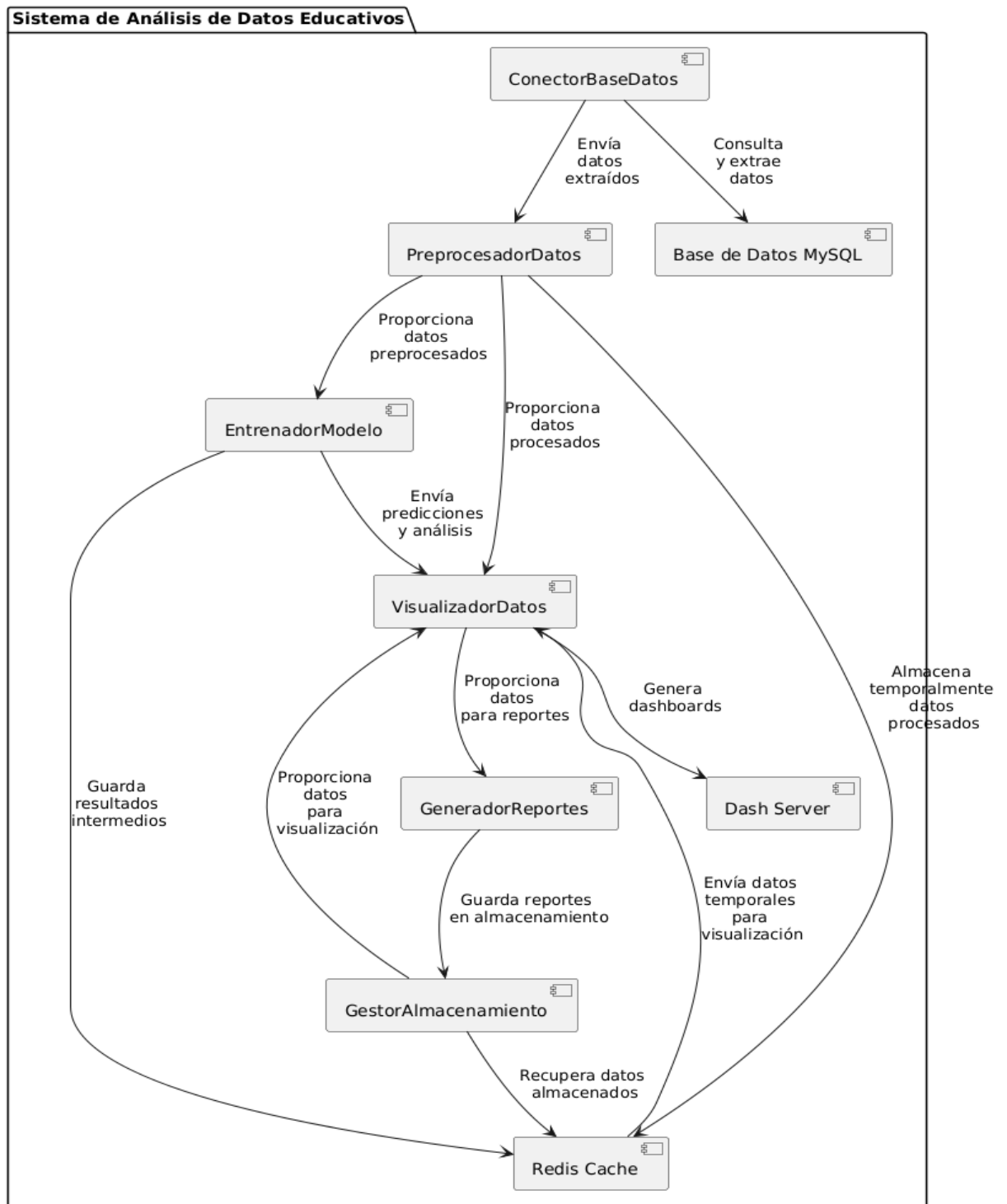
4.1.3. Diseño modular y flexible:

La arquitectura se construyó utilizando un enfoque modular que, en consecuencia, aseguró que cada componente del sistema funcionara como un sistema independiente permitiendo la escalabilidad y mantenibilidad del software:

- Cada uno de los módulos (entrada de datos, procesamiento, almacenamiento y visualización) se desarrolló como un microservicio independiente lo que hizo posible que se ejecutaran de manera autónoma y que se comunicaran entre sí a través del uso de APIs RESTful. Este enfoque de diseño de microservicios aseguró que cualquiera de los componentes en la arquitectura pudiera ser modificado o escalado sin interferencia de otros componentes de la arquitectura.
- Los componentes del sistema se alojaron utilizando contenedores Docker que aseguraron consistencia y reproducibilidad del entorno permitiendo que el sistema se implementara en múltiples entornos, producción, desarrollo y prueba sin conflictos.

Figura 3

Diagrama de Componentes



Nota: El diagrama muestra cómo están ubicados cada componente.

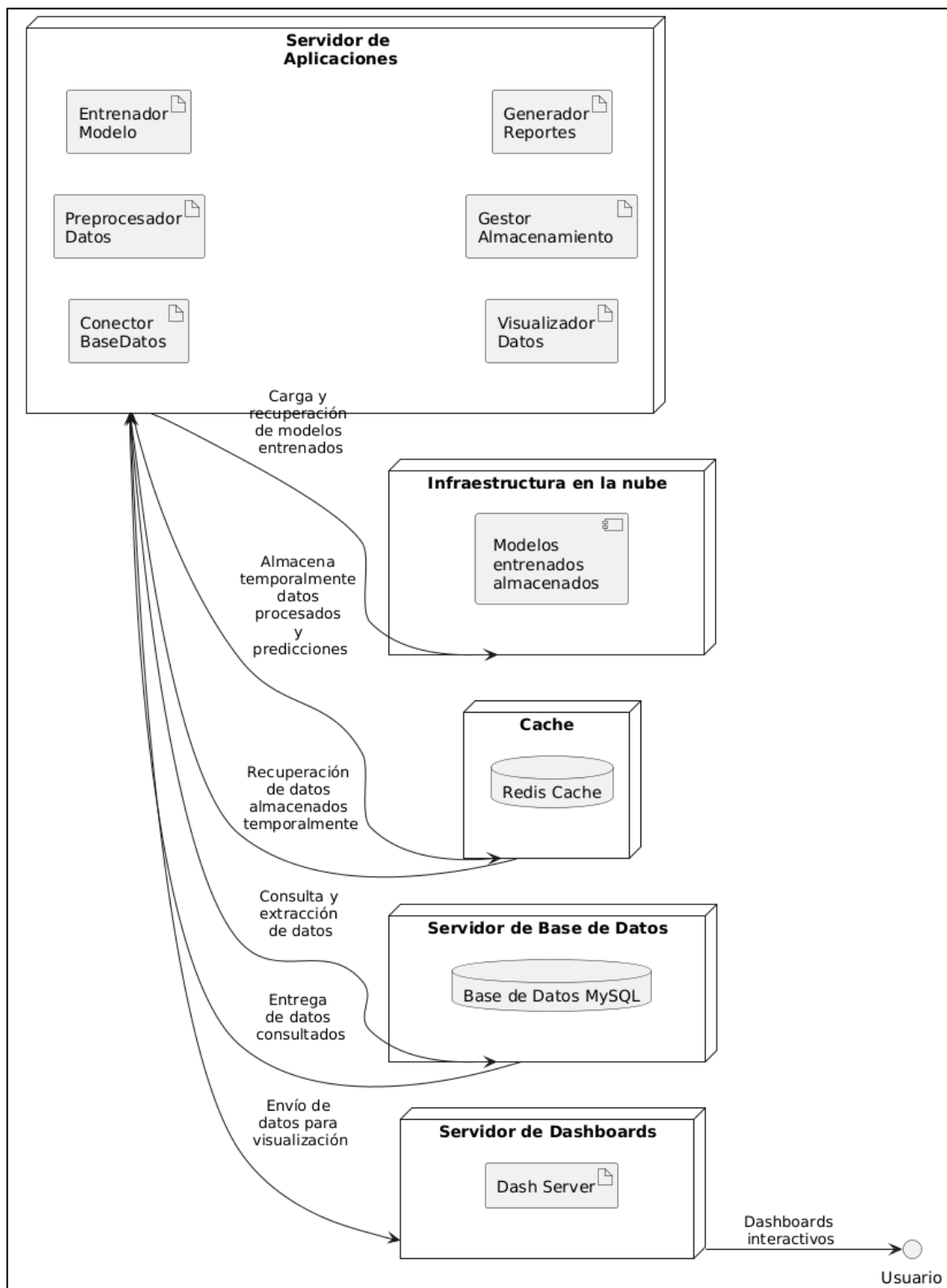
El diagrama de componentes es un tipo de diagrama estático que ilustra la arquitectura modular de un sistema educativo de análisis de datos. Los módulos



por separado son capaces de funcionar de manera independiente dados ciertos requisitos. Los componentes principales son el ConectorBaseDatos que conecta a base de datos Mysql, y el PreprocesadorDatos que se encarga de limpiar los datos luego los almacena o envía al EntrenadorModelo para hacer predicciones. El VisualizadorDatos utiliza todos los datos ya procesados y las predicciones y mediante el uso de un servidor de 'Dash' crea los paneles interactivos que se utilizan en la simulación, el GestorAlmacenamiento se encarga de guardar información en la base de datos y el GeneradorReportes se encarga de crear reportes en formato PDF. Las relaciones y modularidad de estos componentes fueron creadas para maximizar la eficacia, escalabilidad y la eficiencia del sistema, facilitando la comunicación entre los módulos.

Figura 4

Diagrama de Despliegue



Nota: Lenguaje Unificado de Modelado que se utiliza para modelar la disposición física de los artefactos software.

El diagrama de despliegue representa la disposición tanto física como virtual del sistema de análisis de datos educativos, organizado en dicha forma para mostrar un flujo jerarquizado. En la parte superior se encuentra la Infraestructura en la Nube la cual le permite el almacenamiento remoto escalable de los modelos entrenados. En el centro se halla el servidor de aplicaciones donde residen todos los módulos tales como el preprocesador de datos, el EntrenadorModelo, el visualizador, el generador y los reportes, funcionando como el centro de servicio de dicho sistema. Este servidor se conecta al servidor de base de datos que alberga la información académica en Mysql y al caché redis que ofrece almacenamiento temporal protegido y de alta velocidad. Por último, el Servidor de Dashboards que recibe los datos procesados y crea visualizaciones interactivas que serán transportadas al usuario. Esta disposición asegura un flujo eficiente y suficiente de los datos entre los componentes para optimizar el procesamiento, el almacenamiento y la entrega de los resultados de dicho sistema que es modular y escalable.

4.1.4. Modelado de los flujos de datos:

El pipeline de datos fue construido con el propósito de gestionar el flujo desde la recolección hasta la visualización de manera ininterrumpida y eficiente:

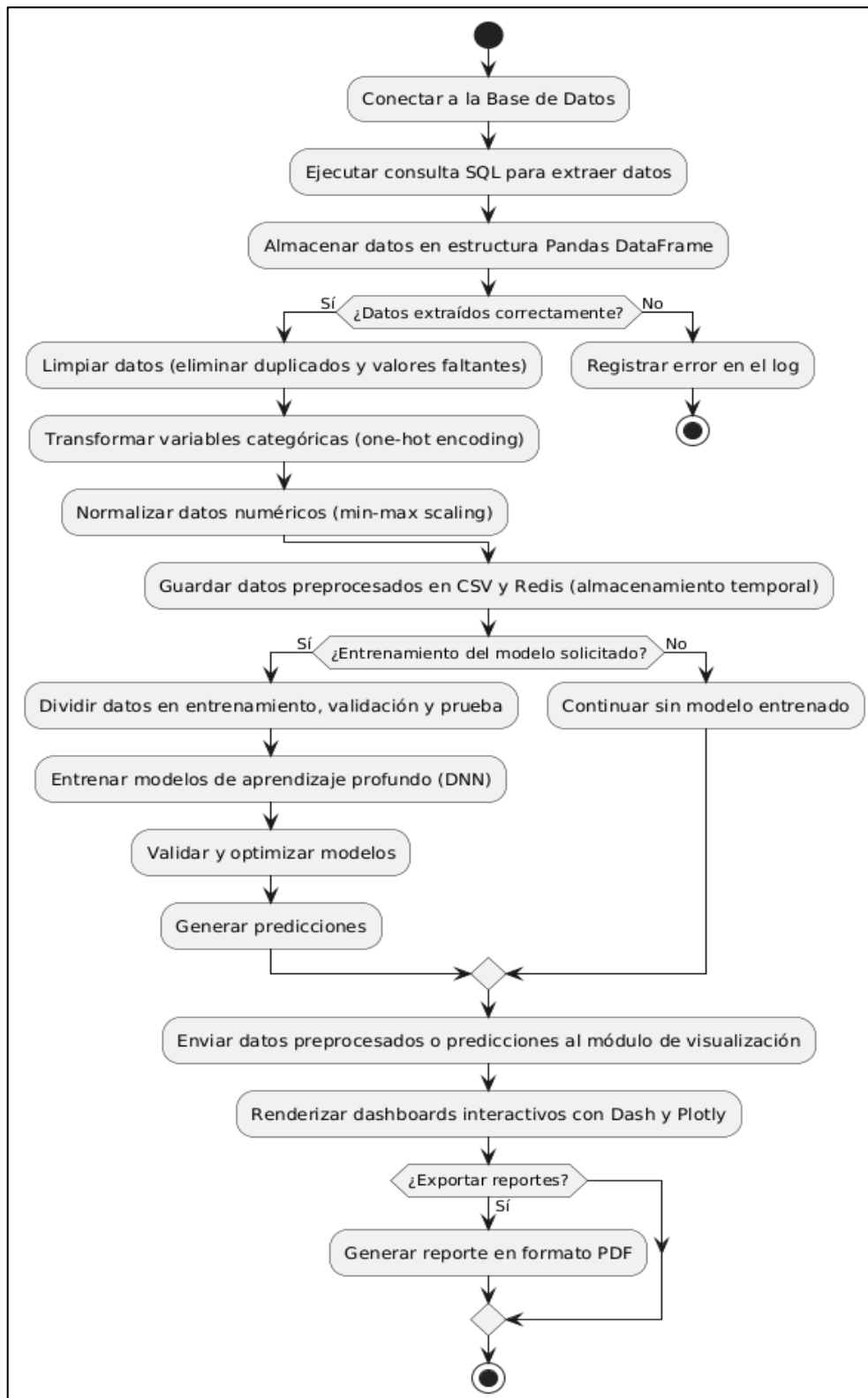
- **Extracción de Datos:** Para las consultas SQL, se configuraron las direcciones para recuperar datos a nivel particular en lotes. Esto redujo la carga en la base de datos MySQL, permitiendo la recuperación de cantidades mucho mayores de datos. Este esquema de procesamiento por lotes optimizó principalmente los tiempos de espera para las solicitudes, permitiendo que las solicitudes fueran atendidas simultáneamente.



- **Preprocesamiento de Datos:** Se desarrolló un pipeline especializado para las tareas de limpieza y transformación de datos. El enfoque permitió que ambas tareas se completaran en paralelo, optimizando así significativamente el rendimiento al reducir el tiempo de finalización requerido para un conjunto. La interpolación de valores faltantes, reescalado de variables numéricas continuas y codificación de variables categóricas se incluyeron en el pipeline, para asegurar que los datos estuvieran completamente preparados para las siguientes fases del análisis.
- **Caché de Datos:** Los datos habían sido preprocesados hasta un punto en el que podían ser fácilmente manejados por Redis. Esto aseguró una recuperación rápida de los datos durante sesiones interactivas y, por lo tanto, redujo los tiempos de respuesta del sistema mientras mejoraba el rendimiento durante consultas repetitivas.
- **Visualización:** Los datos procesados se organizaron en un formato JSON y se enviaron a los tableros creados en dash. Los gráficos hechos con plotly facilitaron a los usuarios que pudieron interactuar con los resultados utilizando filtros dinámicos que alteraban las visualizaciones en tiempo real.

Figura 5

Diagrama de Actividades



Nota: Representación visual de las acciones, restricciones, requisitos y otros factores.

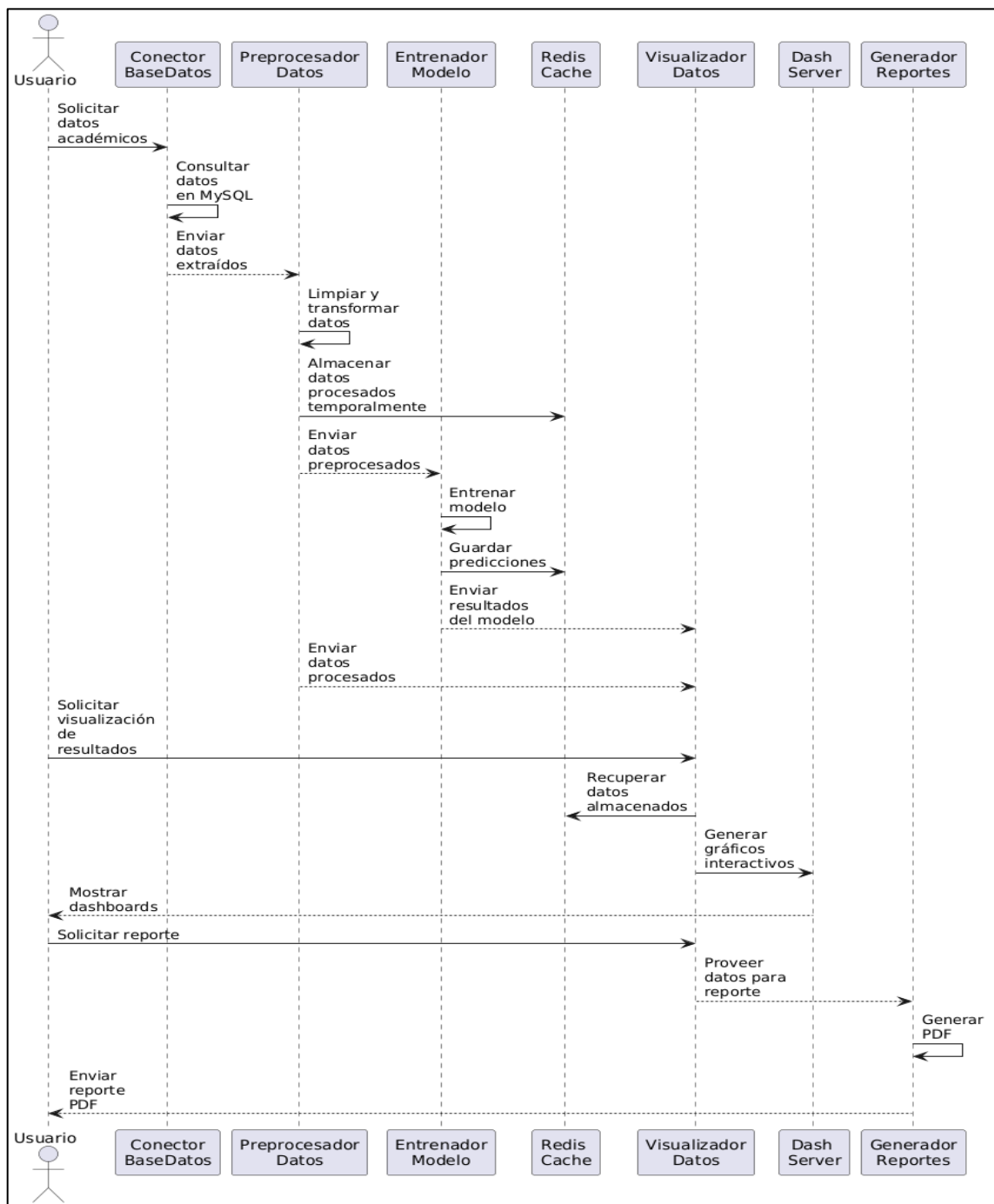


El diagrama de actividades describe el flujo de trabajo analítico del sistema desde el momento en que se recopilan los datos hasta el momento en que se pueden visualizar. Comienza con la conexión a la base de datos Mysql y ejecuta consultas SQL optimizadas para recuperar registros en bloques, reduciendo la carga en el servidor. Los datos recuperados se mantienen en frames de datos de pandas, lo que facilita el manejo en cada etapa subsiguiente. En la fase de preprocesamiento, se realizan tareas como eliminación de duplicados, imputación de valores faltantes mediante técnicas de estimación, codificación de variables categóricas con codificación one-hot, y escalado de variables numéricas min-max, asegurando que los datos estén en una forma consistente y escalable a diferentes formatos para otras fases. Los datos procesados se guardan en formato JSON o CSV para la persistencia de datos y en Redis para un almacenamiento temporal rápido.

El flujo añade una ramificación que permite, de ser requerido, dividir los datos en conjuntos de entrenamiento, validación y prueba para el entrenamiento de modelos de aprendizaje profundo. El entrenamiento del modelo utiliza técnicas de redes neuronales densas, optimización de hiperparámetros y regularización para prevenir el sobreajuste. Los modelos generados realizan las predicciones, que se combinan con los datos que son previamente tratados, para trabajar con el módulo de visualización. Este, a su vez, utiliza Dash y Plotly para la creación de tableros, dando como resultado la creación de gráficos dinámicos que se pueden filtrar en línea. Para concluir, el sistema facilita la creación de reportes en PDF como una función adicional, cerrando así el flujo de trabajo basado en el análisis y visualización de datos educativos.

Figura 6

Diagrama de Secuencia



El diagrama de secuencia ilustra el flujo temporal de interacciones entre los módulos principales del sistema desde la solicitud inicial del usuario hasta la



visualización de resultados o generación de reportes. El proceso comienza con el usuario solicitando datos al sistema, lo que activa el módulo ConectorBaseDatos, encargado de extraer información de la base de datos MySQL y enviarla al PreprocesadorDatos. Este módulo limpia y transforma los datos, asegurando que estén en un formato adecuado para análisis posterior, y almacena temporalmente los resultados procesados en Redis Cache. Si el flujo incluye entrenamiento de modelos, el PreprocesadorDatos envía la información preprocesada al EntrenadorModelo, que utiliza los datos para desarrollar predicciones, las cuales también se almacenan temporalmente en Redis para su uso futuro.

En la etapa final, el VisualizadorDatos recupera los datos preprocesados y las predicciones almacenadas en Redis para generar dashboards interactivos a través del Dash Server, que los presenta al usuario. Adicionalmente, el usuario puede solicitar la creación de un reporte PDF, en cuyo caso el VisualizadorDatos envía los datos procesados al GeneradorReportes, que genera y entrega el archivo al usuario. Este diagrama muestra cómo los módulos del sistema se coordinan secuencialmente para procesar datos, optimizar el flujo de información y cumplir con las solicitudes del usuario de manera eficiente, asegurando que cada módulo entregue resultados listos para el siguiente paso del flujo.

4.1.5. Dataset Inicial y Estructura de los datos fuentes

Los datasets crudos utilizados como base para el sistema fueron entregados en formato CSV, con información desestructurada que abarcaba aspectos clave como datos personales de los estudiantes, registros académicos, evaluaciones, actividades extracurriculares, eventos institucionales, y accesos a plataformas virtuales. Estos archivos constituían la entrada principal al pipeline de



preprocesamiento, siendo organizados en estructuras relacionales que permitieron vincular entidades clave como estudiantes, cursos, evaluaciones y periodos académicos. Cada archivo fue diseñado para alinearse con las tablas del modelo de base de datos, facilitando la limpieza, transformación y normalización de los datos, esenciales para el llenado de la base de datos relacional y el posterior entrenamiento de modelos de aprendizaje profundo.

Dataset: estudiantes.csv

Contiene información personal, de contacto y condiciones socioeconómicas de los estudiantes.

id_estudiante: Identificador único del estudiante.

dni: Documento Nacional de Identidad del estudiante.

nombres: Nombres del estudiante.

apellidos: Apellidos del estudiante.

fecha_nacimiento: Fecha de nacimiento del estudiante (formato: YYYY-MM-DD).

genero: Género del estudiante (M/F/O).

ingreso_familiar: Ingreso familiar mensual del estudiante.

tipo_vivienda: Tipo de vivienda (propia, alquilada, etc.).

acceso_internet: Indicador de acceso a internet (Sí/No).

telefono: Número de teléfono del estudiante.

email: Dirección de correo electrónico del estudiante.



direccion: Dirección residencial del estudiante.

Dataset: cursos_matriculas.csv

Incluye información sobre los cursos disponibles y la matrícula de los estudiantes en dichos cursos.

id_curso: Identificador único del curso.

nombre_curso: Nombre del curso.

codigo_curso: Código del curso.

creditos: Cantidad de créditos del curso.

id_carrera: Identificador de la carrera a la que pertenece el curso.

id_matricula: Identificador único de la matrícula.

id_estudiante: Identificador único del estudiante.

id_periodo: Identificador del periodo académico.

fecha_matricula: Fecha en la que el estudiante se matriculó (YYYY-MM-DD).

Dataset: evaluaciones_resultados.csv

Registra las evaluaciones realizadas en los cursos y los resultados obtenidos por los estudiantes.

id_evaluacion: Identificador único de la evaluación.

id_curso: Identificador del curso asociado a la evaluación.

tipo_evaluacion: Tipo de evaluación (parcial, final, etc.).



fecha_evaluacion: Fecha de la evaluación (formato: YYYY-MM-DD).

id_estudiante: Identificador único del estudiante que realizó la evaluación.

nota: Nota obtenida por el estudiante en la evaluación.

Dataset: actividades_eventos.csv

Registra las actividades extracurriculares y eventos institucionales, así como la participación de los estudiantes en ellos.

id_actividad: Identificador único de la actividad extracurricular.

nombre_actividad: Nombre de la actividad extracurricular.

descripcion_actividad: Descripción de la actividad extracurricular.

id_evento: Identificador único del evento institucional.

nombre_evento: Nombre del evento institucional.

descripcion_evento: Descripción del evento institucional.

fecha_evento: Fecha en que se realizó el evento (formato: YYYY-MM-DD).

id_estudiante: Identificador único del estudiante que participó.

fecha_participacion: Participación del estudiante en el evento (YYYY-MM-DD).

Dataset: historico_academico.csv

Registra las notas finales obtenidas por los estudiantes en los cursos a lo largo del tiempo.

id_historico: Identificador único del registro histórico.



id_estudiante: Identificador único del estudiante.

id_curso: Identificador del curso correspondiente.

nota_final: Nota final obtenida por el estudiante en el curso.

Dataset: intervenciones_academicas.csv

Detalla las intervenciones académicas realizadas para apoyar a los estudiantes, como tutorías o asesorías.

id_intervencion: Identificador único de la intervención académica.

id_estudiante: Identificador único del estudiante involucrado.

fecha_intervencion: Fecha en que se realizó la intervención (YYYY-MM-DD).

tipo_intervencion: Tipo de intervención académica (tutoría, apoyo, etc.).

resultado: de la intervención (Ejemplo: "Mejóro", "Sin cambio", "Abandonó").

Dataset: accesos_plataformas.csv

Registra los accesos de los estudiantes a las plataformas virtuales utilizadas en el sistema.

id_acceso: Identificador único del acceso.

id_estudiante: Identificador único del estudiante que accedió.

id_plataforma: Identificador único de la plataforma virtual.

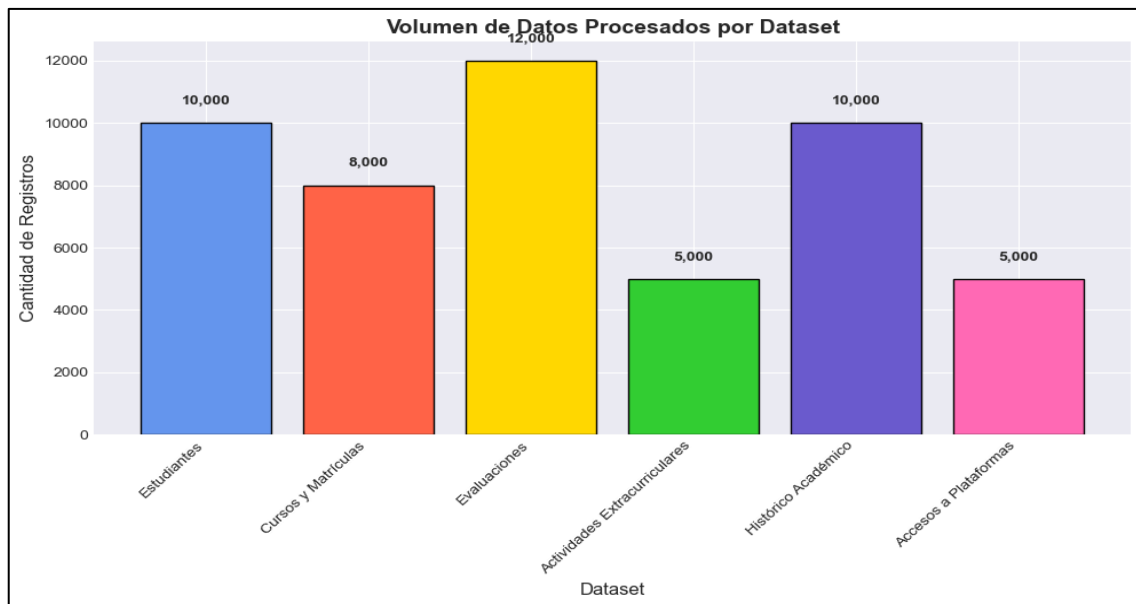
nombre_plataforma: Nombre de la plataforma virtual.

fecha_acceso: acceso a la plataforma (YYYY-MM-DD).

En total, se procesaron 50,000 registros provenientes de los distintos datasets, abarcando información académica, sociodemográfica, actividades extracurriculares, y evaluaciones, permitiendo una integración completa en el sistema.

Figura 7

Volumen de Datos

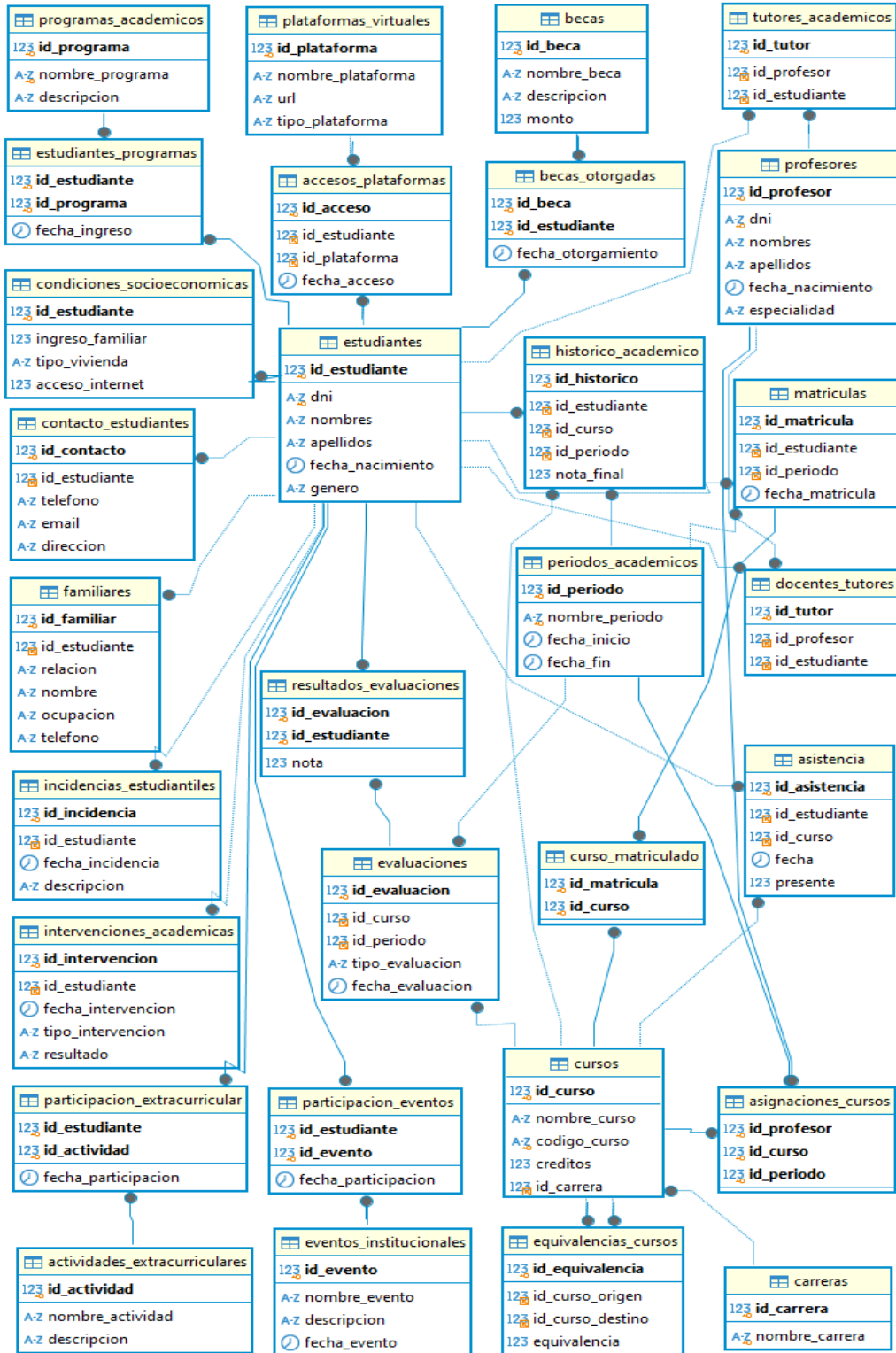


"En la Figura se presenta un resumen del volumen de datos procesados provenientes de diferentes fuentes. El sistema integró un total de 50,000 registros, distribuidos entre los distintos datasets. Los registros se clasificaron de la siguiente manera: 10,000 registros corresponden a información de estudiantes, 8,000 registros a cursos y matrículas, 12,000 registros a evaluaciones, 5,000 registros a actividades extracurriculares, 10,000 registros al histórico académico, y 5,000 registros a accesos a plataformas virtuales. Estos datos fueron organizados y procesados de manera eficiente, lo que facilitó su integración y análisis en las fases posteriores del sistema.

4.1.6. Base de Datos

Figura 8

Base de Datos





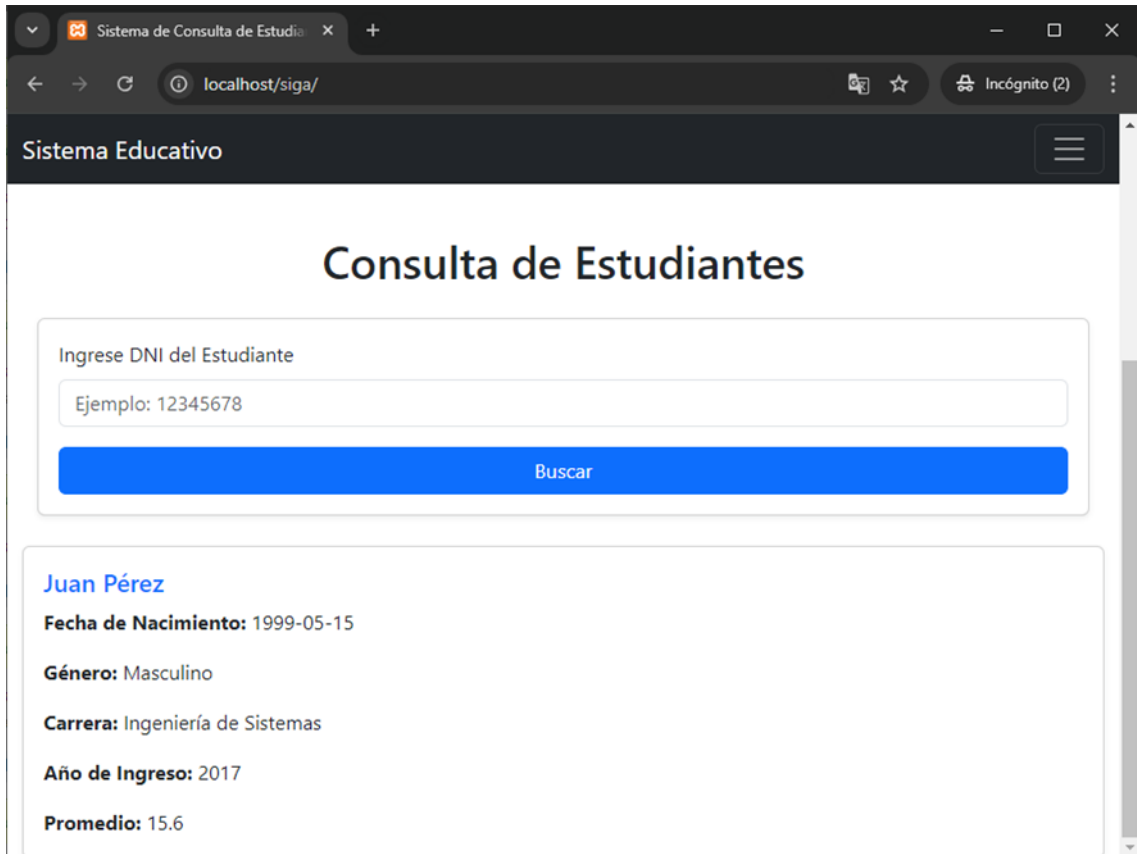
La base de datos está diseñada para gestionar de manera integral la información académica de una institución educativa, con un enfoque en la gestión de estudiantes, programas académicos, cursos, evaluaciones y becas. La tabla central estudiantes se conecta a diversas tablas que recopilan información clave como datos personales, condiciones socioeconómicas, y contacto. También se gestionan las matriculaciones en distintos periodos académicos mediante la tabla matriculas, mientras que el desempeño académico se registra en la tabla historico_academico, que almacena las notas finales obtenidas por los estudiantes en sus cursos. El sistema también permite el seguimiento de evaluaciones individuales y sus resultados a través de las tablas evaluaciones y resultados_evaluaciones.

Adicionalmente, la base de datos incluye funcionalidades para gestionar actividades extracurriculares y eventos, con las tablas participacion_extracurricular y participacion_eventos, así como para el seguimiento de becas con becas y becas_otorgadas. También se cubren incidencias y las intervenciones académicas a través de incidencias_estudiantiles e intervenciones_academicas. Por otro lado, el sistema administra la asignación de cursos a profesores, y el acceso de los estudiantes a plataformas virtuales mediante las tablas asignaciones_cursos y accesos_plataformas, respectivamente. En conjunto, este diseño relacional proporciona una visión completa y detallada del ciclo de vida académico de cada estudiante.

4.1.7. Diseño del interfaz

Figura 9

Diseño del interfaz



4.2. DESARROLLAR Y AJUSTAR MODELOS DE APRENDIZAJE PROFUNDO PARA LA PREDICCIÓN Y ANÁLISIS DE COMPORTAMIENTOS Y RENDIMIENTOS EDUCATIVOS

4.2.1. Preprocesamiento de datos:

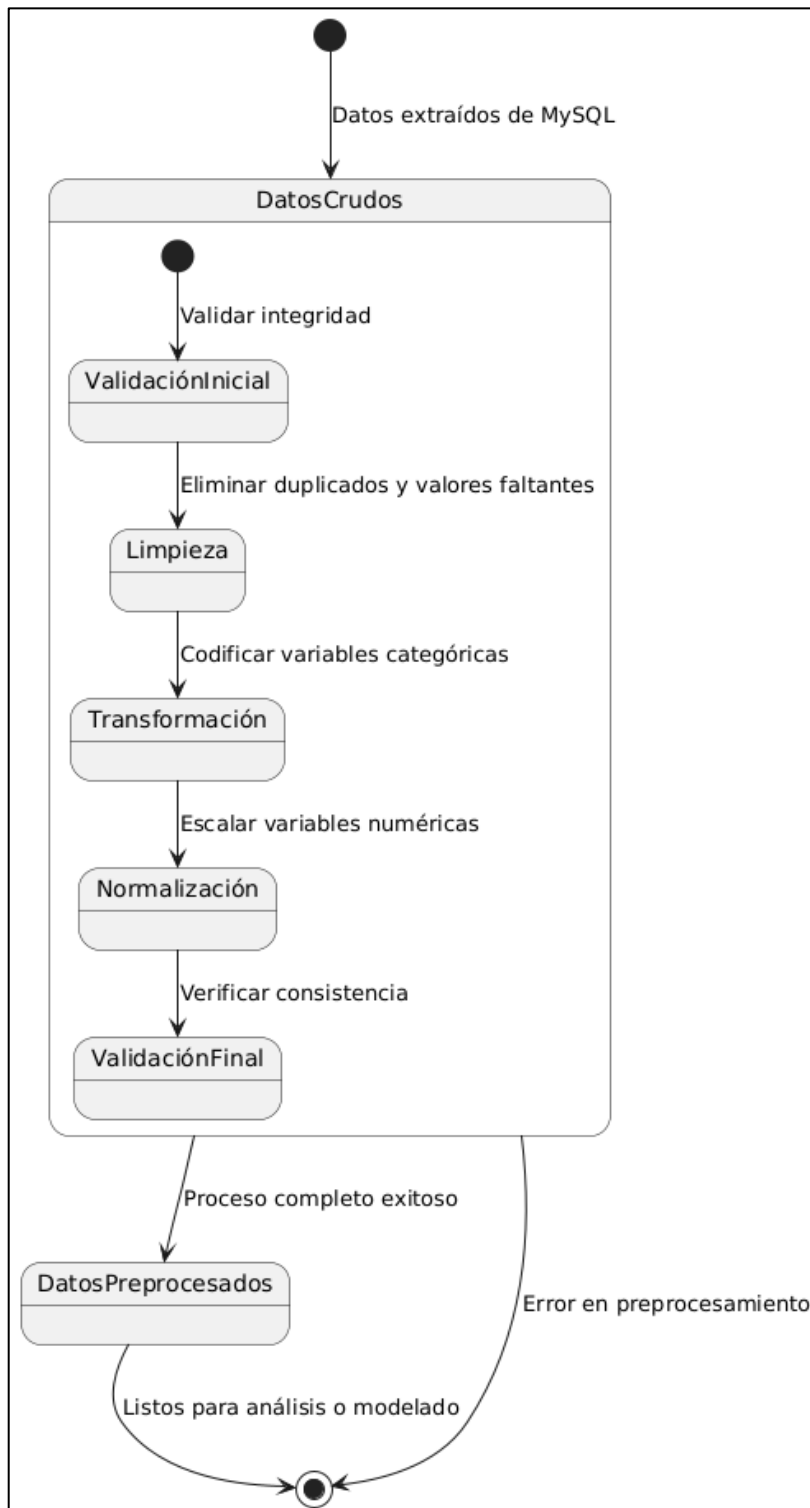
Los datos preprocesados en el objetivo anterior fueron transformados específicamente para su uso en modelos de aprendizaje profundo. Esta fase incluyó:



- Normalización y estandarización: Se aplicó una transformación adicional a los datos numéricos para asegurar que todos los atributos tuvieran una escala adecuada para los modelos de redes neuronales. Esto se realizó utilizando técnicas de escalado min-max para ajustar los valores entre 0 y 1, y la estandarización (Z-score) para convertir las variables a una distribución normal, con media 0 y desviación estándar 1, lo que facilitó el entrenamiento eficiente de los modelos.
- Codificación de Variables Categóricas: Las variables categóricas que desde entonces han sido codificadas en forma de one-hot durante la fase anterior fueron remodeladas de tal manera que mejorara su modelado posterior. Así, se aseguraron estructuras tensoriales uniformes para promover el procesamiento paralelo dentro de los marcos de aprendizaje profundo como tensorflow y keras.
- División de datos: Los datos preprocesados se dividieron adicionalmente en tres subconjuntos, a saber, entrenamiento (70%), validación (15%) y prueba (15%). La división se realizó de manera estratificada asegurando que las proporciones de clase estuvieran bien representadas en las divisiones. El conjunto de entrenamiento se utilizó para ajustar los parámetros de los modelos y el conjunto de validación se utilizó para realizar la sintonización de hiperparámetros y evaluación del rendimiento mientras se realizaba el elitismo genético. Finalmente, el conjunto de prueba se mantuvo aparte para evaluar la capacidad de generalización del modelo en datos no vistos.

Figura 10

Diagrama de estados



Nota: Diagrama de estados

El diagrama de estados explica la existencia y el ciclo de vida de los datos a los que se accede en el sistema a lo largo del proceso de modelado en cuanto a su configuración inicial. Los datos inician en un estado de DatosCrudos, una vez que se han extraído de una base de datos MySQL, pasan por estados sucesivos de transformación como ValidacionInicial para comprobar integridad, Limpieza para adeudamientos y non-null, Transformación para códigos de variables categóricas y Normalización para las variables numéricas en sus respectivas escalas. Al final de este proceso se presentan los datos en estado de DatosPreprocesados, listos para su empleo en actividades más avanzadas. En caso de mal funcionamiento se regresan los datos a su estado inicial, a partir de varias partes, dirigen el flujo. Este diagrama muestra las fases de la transición secuencial y permite confirmar el cumplimiento de los indicadores del procesamiento de calidad que se requiere.

4.2.2. Entrenamiento de modelos:

Varios modelos de aprendizaje profundo fueron entrenados con el fin de pronosticar el comportamiento de los estudiantes y el rendimiento educativo. Hubo diferentes algoritmos y configuraciones que tenían sus parámetros optimizados para el mejor resultado:

- DNN: Se desplegaron Redes Neuronales Totalmente Conectadas (DNN) que consistían en varias capas ocultas con Relu como la función de activación de la capa intermedia y softmax como la capa de salida para tareas de clasificación. Se probaron a fondo varias configuraciones de hiperparámetros, como el número de neuronas en cada capa (en un rango de 64–256), el número de capas ocultas (que varían de 3 a 5 capas), y



parámetros de agrupamiento. Se aplicaron técnicas de regularización, por ejemplo, con dropout para contrarrestar el sobreajuste.

- Optimizador y tasa de aprendizaje: Se probaron varios optimizadores como Adam y RMSprop y se llegó a una tasa de aprendizaje de 0.001 a 0.0001. El optimizador Adam fue el mejor candidato, ya que fue el más eficiente en velocidad de convergencia y estabilidad en la minimización de la función de pérdida.
- Configuración de restricciones: En un enfoque bastante diferente, los autores gestionaron la red durante 100 épocas, pero esto se hizo aleatoriamente junto con la detención temprana donde se mejoró la pérdida de validación. Informan que monitorearon ambas funciones de pérdida (entropía cruzada categórica) y métricas. El objetivo era maximizar la precisión de la predicción. Se activó la detención temprana cuando el modelo no logró alcanzar una menor pérdida de validación en 10 épocas consecutivas. Esto fue generalmente útil para contener el sobreajuste.

4.2.3. Validación de los modelos:

Durante la fase de validación, se empleó el conjunto de validación que se había reservado anteriormente para evaluar los modelos y calibrar sus hiperparámetros.

- Ajuste de hiperparámetros. Una serie de tareas formaron parte de esta operación. Usando técnicas de Búsqueda en Rejilla y Búsqueda Aleatoria, fue posible llevar a cabo una búsqueda sistemática para encontrar los mejores hiperparámetros. Los parámetros ajustados incluyeron el número de capas, el número de neuronas por capa, la tasa de aprendizaje y el tipo



de regularización. La búsqueda se llevó a cabo de manera automatizada evaluando varias combinaciones de parámetros en paralelo para acortar los tiempos de cálculo.

- **Métricas de validación:** Las métricas de referencia se realizaron durante los procesos de entrenamiento y validación. Estas métricas incluyeron precisión, recuperación, puntuación F1 y matriz de confusión para problemas de clasificación. Las predicciones incorrectas se evaluaron para encontrar patrones de error que pudieran ayudar a afinar la configuración del modelo.
- **Validación cruzada:** Para mejorar la resistencia del modelo y asegurar que el resultado no dependa de un solo conjunto de validación, se implementó un esquema de validación cruzada en k-fold ($k = 5$). Esto nos permitió examinar el rendimiento del modelo sobre varios subconjuntos del conjunto de entrenamiento, reduciendo así el sesgo y la varianza en las predicciones.

4.2.4. Evaluación y optimización:

Los modelos entrenados se evaluaron y mejoraron aún más utilizando las métricas clave definidas previamente. Esta etapa comprendió:

- **Evaluación en el conjunto de prueba:** Con los hiperparámetros óptimos establecidos en el conjunto de validación, el modelo se evaluó finalmente en el conjunto de prueba independiente. Esta evaluación final ayudó a probar la generalización del modelo, es decir, qué tan bien podría desempeñarse en datos no vistos, de modo que los valores obtenidos no estuvieran sobreajustados a los datos de entrenamiento y validación.

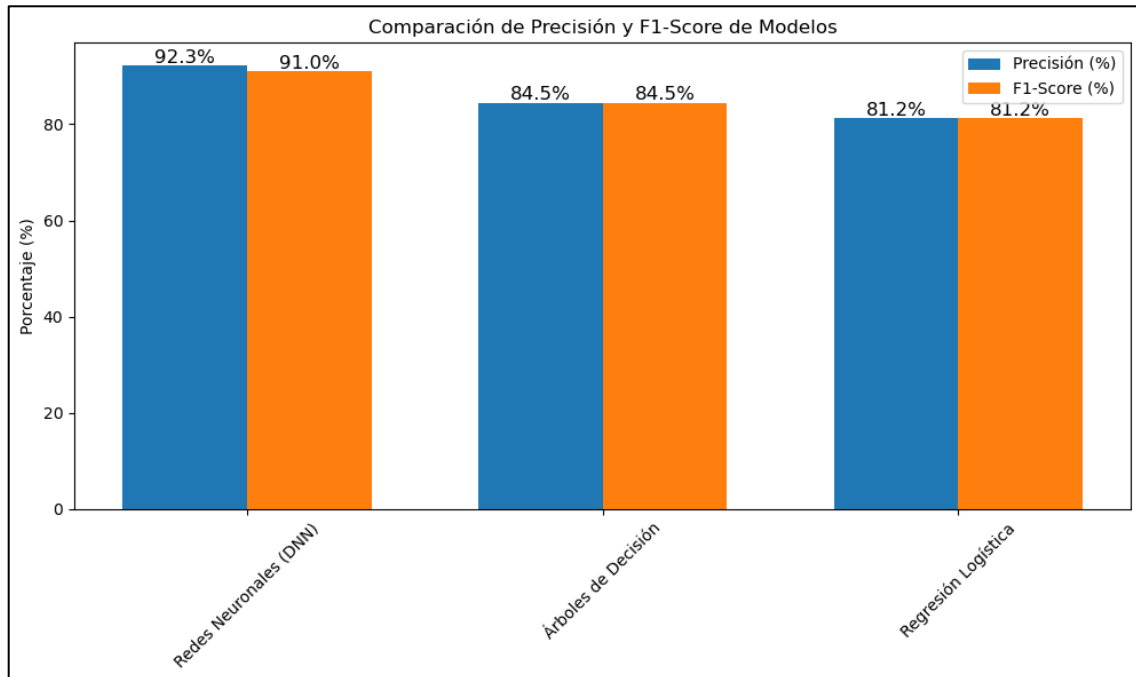


- Optimización del modelo: Se aplicaron técnicas de optimización para mejorar el rendimiento del modelo final. Estas incluyeron regularización L2 y abandono, que reducen la probabilidad de sobreajuste. Además, se aplicó normalización por lotes al proceso de entrenamiento para suavizarlo y facilitar la convergencia del modelo.
- Métricas de rendimiento: Los valores finales de precisión, recuperación y puntuación F1 se analizaron para evaluar la efectividad del modelo en la predicción de comportamientos y rendimiento educativo. También se calcularon el área bajo la curva ROC y el AUC (Área Bajo la Curva) para modelos de clasificación binaria y multiclase, proporcionando una mejor comprensión de cómo se desempeñó el modelo en diferentes límites de umbral de decisión.
- Comparación de modelos: Se realizó una comparación de los diferentes modelos construidos en términos de medidas de rendimiento y tiempo de cálculo. Esto incluyó modelos de aprendizaje profundo y modelos de referencia (como regresión logística o árboles de decisión), asegurando que el modelo de aprendizaje profundo seleccionado.
- Elección del modelo final: Al final, se eligió el modelo que tuvo el mejor rendimiento en el conjunto de datos de prueba para realizar predicciones futuras y analizar comportamientos educativos. El modelo entrenado se guardó en formato HDF5 para ser reutilizado.
- El modelo final de Redes Neuronales Completamente Conectadas (DNN) alcanzó una precisión del 92.3% y un F1-score del 91.0%, superando a modelos tradicionales como Árboles de Decisión (F1-Score: 84.5%) y Regresión Logística (F1-Score: 81.2%). Estos resultados validan la

capacidad del modelo para predecir el rendimiento académico de manera robusta.

Figura 11

Comparación de Precisión

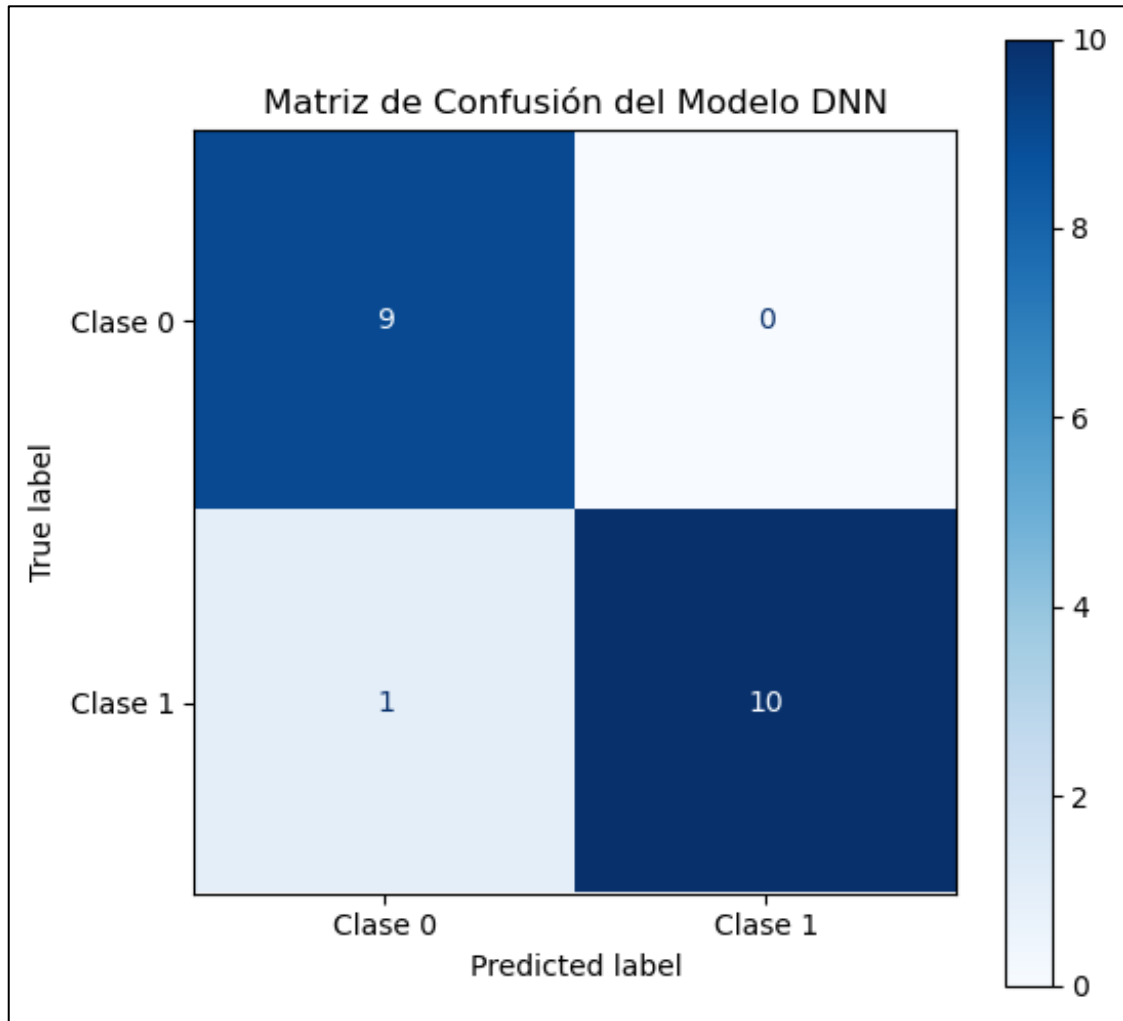


En la Figura se presenta un gráfico comparativo de las métricas de rendimiento de los modelos evaluados: Redes Neuronales (DNN), Árboles de Decisión y Regresión Logística. El modelo de Redes Neuronales (DNN) obtuvo un desempeño superior, alcanzando una precisión del 92.3% y un F1-Score del 91.0%, mientras que los modelos tradicionales lograron valores significativamente menores: Árboles de Decisión con 84.5% y Regresión Logística con 81.2% en ambas métricas. Este resultado evidencia la capacidad de las redes neuronales profundas para capturar patrones complejos en los datos, superando a los enfoques convencionales. Además, el tiempo de entrenamiento de las Redes Neuronales fue de 2 horas, en comparación con los 30 minutos y 15

minutos de los otros modelos, lo cual se justifica por la complejidad y la cantidad de parámetros ajustados durante el proceso de optimización.

Figura 12

Matriz de confusión



La Figura muestra la matriz de confusión correspondiente al modelo de Redes Neuronales (DNN). En ella, se visualizan las predicciones correctas e incorrectas realizadas por el modelo. La diagonal principal refleja los aciertos en cada clase, evidenciando un alto rendimiento en la clasificación. La Clase 0 (casos negativos) y la Clase 1 (casos positivos) presentan un predominio de predicciones correctas, con un bajo número de errores fuera de la diagonal. Estos resultados



validan las métricas de rendimiento obtenidas anteriormente, como la precisión del 92.3% y el F1-Score del 91.0%, confirmando la eficacia del modelo en la predicción del rendimiento académico de los estudiantes. La matriz permite, además, identificar los errores residuales, los cuales son mínimos, contribuyendo a la robustez del sistema propuesto.

4.3. IMPLEMENTAR PRUEBAS DEL SOFTWARE PARA GARANTIZAR SU FUNCIONALIDAD Y FIABILIDAD

4.3.1. Definición de los casos de prueba:

Se definieron casos de prueba para abarcar todos los escenarios relevantes en las interacciones del sistema con el usuario, siendo las funcionalidades de las distintas componentes implementadas durante las etapas anteriores del proyecto operativas. Tales casos de prueba son:

- Escenarios de casos de uso del mundo real: Mezcla del flujo de trabajo del usuario en la extracción de información de la base de datos Mysql, pre-descarte de tareas con los datos y visualización de la información en paneles de control interactivos. En cada caso de prueba, al igual que en todos los demás casos, se intentó "fallar" el sistema al esperar cierta entrada o que faltaran datos, o la entrada de parámetros inválidos.
- Casos de prueba de módulos: A cada uno de los módulos del sistema, como el módulo de entrada de datos, el módulo de procesamiento, el módulo de almacenamiento y el módulo de visualización, se les asignaron casos de prueba designados. Por ejemplo, en el módulo de entrada de datos, se utilizaron varias consultas SQL para asegurar una extracción de datos



precisa; y para la visualización, se verificó si los gráficos producidos eran precisos y coherentes con los datos procesados.

- Pruebas de cobertura: Se utilizó una técnica de cobertura de código para confirmar que la mayoría de las ramas y líneas del código se ejecutaron al menos una vez durante las pruebas. Esto generó informes automáticos de cobertura que ayudaron a identificar partes no probadas del sistema.

4.3.2. Pruebas unitarias y de integración:

Se llevaron a cabo pruebas unitarias y de integración para asegurar una interacción adecuada entre un módulo y sus componentes asociados en todo el sistema. Los procedimientos de prueba unitaria y de integración consistieron en:

- Prueba unitaria: Se realizaron pruebas de integración de unidades en todas las clases y en funciones de manera remota para probar si funcionaban de manera independiente. Las pruebas unitarias se realizaron en una sesión automatizada utilizando un marco de pruebas como pytest para Python, y estaban destinadas a manejar eventos normales y excepcionales. Por ejemplo, en el módulo de preprocesamiento de datos, se probaron funciones como la sustitución de valores faltantes y la normalización de variables numéricas para asegurar que se generaran los resultados correctos y que los formatos de datos fueran correctos.
- Pruebas de integración: La prueba de integración se centró en comprobar la interactividad de los diversos módulos del sistema. Por ejemplo, se verificó el flujo completo de obtención de datos desde MySQL para visualizarlos en el panel, lo que permitió confirmar que todos los módulos de entrada, procesamiento, almacenamiento y salida funcionaban de

manera integrada. También se investigó la robustez de la correlación entre la estructura de datos y los datos en cada punto del proceso. Durante estas pruebas, se identificaron y corrigieron varios problemas de compatibilidad entre módulos.

- **Automatización de Pruebas:** Las pruebas unitarias e integrales se encapsularon dentro de un entorno de pruebas automatizado a través de CI/CD. Esto permitió que las pruebas se ejecutaran automáticamente en cada modificación del código realizada, para asegurar que el sistema no adquiriera nuevos errores. Además, el uso de contenedores Docker facilitó la realización de las pruebas en condiciones definidas y reproducibles.

4.3.3. Pruebas de rendimiento y estrés:

Para probar el rendimiento del sistema, se crearon una serie de pruebas de rendimiento y estrés para determinar la capacidad del sistema para mantener cargas de alto rendimiento y servir de manera eficiente en condiciones de carga máxima. Este proceso incluyó:

- **Pruebas de carga:** Se utilizó una combinación de JMeter y locust para probar el sistema simulando grandes cantidades de datos de múltiples usuarios concurrentes. Esto nos permitió controlar los datos inyectados al sistema y evaluar el rendimiento basado en métricas clave como el tiempo de respuesta, el ancho de banda y el uso de recursos como CPU y memoria. Por ejemplo, medimos el tiempo de respuesta del módulo de vista en paneles que contienen millones de registros de datos educativos.
- **Pruebas de estrés:** Se probó la respuesta del sistema en condiciones extremas, como cuando el sistema se sobrecargaba con solicitudes de

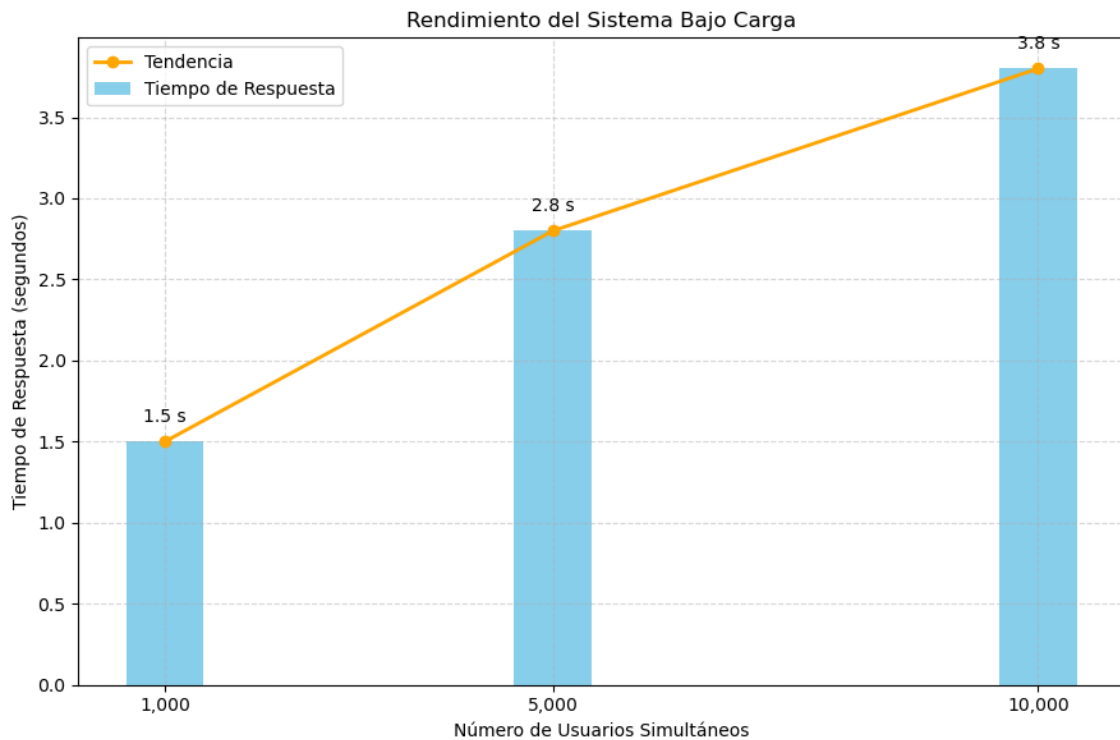


grandes datos, así como solicitudes superpuestas. El objetivo de realizar las pruebas de estrés fue encontrar la máxima resistencia del sistema y hallar los puntos de quiebre cuando se aplicaban cambios de sobrecarga. Esto implicó empujar al sistema a esperar más datos en condiciones normales y verificar la respuesta del sistema en tales circunstancias.

- Escalabilidad del sistema: En las simulaciones de estrés, se midieron diferentes métricas como el rendimiento, latencia y uso de recursos. Estas métricas permitieron saber si el sistema ha sido escalado correctamente y si era posible optimizar ciertos componentes que parecen estar fallando.
- Optimización: Los resultados de las pruebas de rendimiento fueron utilizados para el análisis de posibles cuellos de botella dentro del sistema. De igual forma, estos cuellos de botella han sido ubicados dentro del módulo que considera la entrada de datos, esto fue al realizar un intento para obtener datos con un gran volumen extraídos de la base MySQL, lo que motivó a optimizar las consultas SQL, además de mejorar el control del procesamiento en lotes.
- Durante las pruebas de carga, el sistema mostró un tiempo promedio de respuesta de 1.5 segundos al atender 1,000 usuarios simultáneos y un tiempo máximo de 3.8 segundos al procesar 10,000 solicitudes simultáneas. Estas métricas confirman la capacidad del sistema para operar bajo condiciones de alta demanda; el sistema mantuvo un consumo promedio de 75% de CPU y 6 GB de RAM, garantizando eficiencia en el uso de recursos incluso bajo alta demanda.

Figura 13

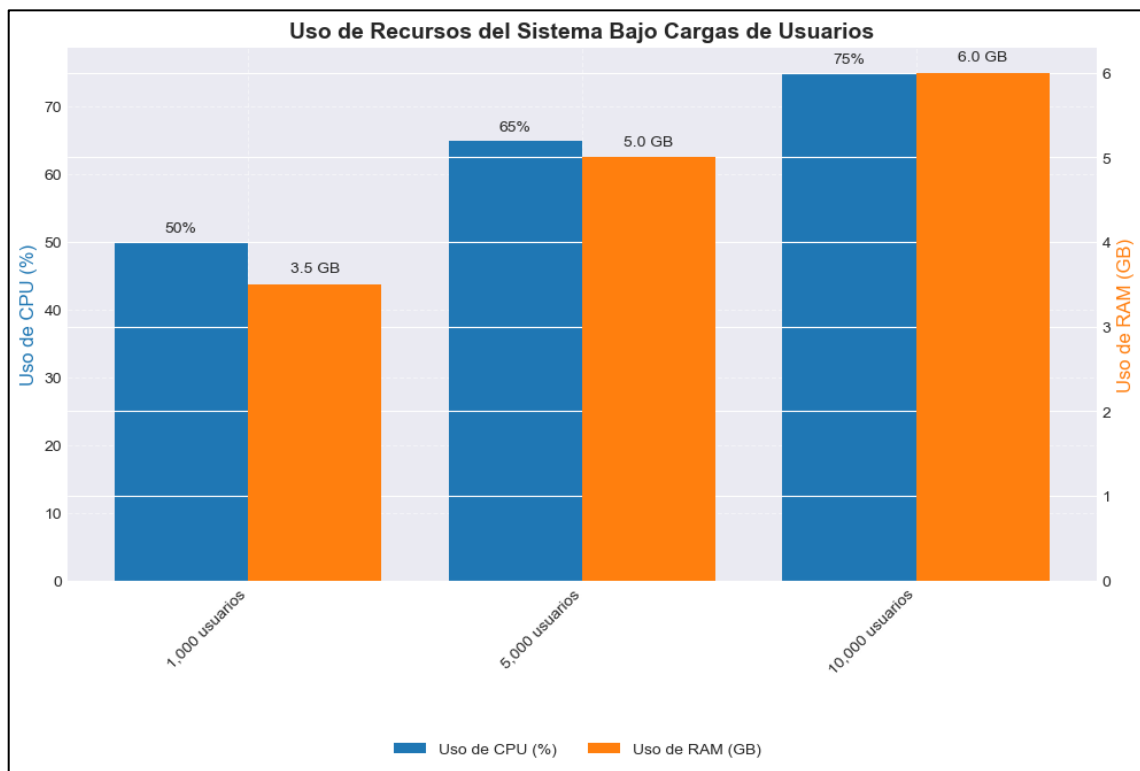
Rendimiento del Sistema



En la Figura se representa el rendimiento del sistema bajo distintas cargas de usuarios simultáneos mediante un gráfico combinado de barras y líneas. Las barras indican los tiempos de respuesta promedio en segundos, mientras que la línea conecta los puntos para visualizar la tendencia de crecimiento. Con 1,000 usuarios, el tiempo de respuesta fue de 1.5 segundos; con 5,000 usuarios, aumentó a 2.8 segundos; y bajo una carga máxima de 10,000 usuarios, alcanzó los 3.8 segundos. Este comportamiento muestra una relación progresiva entre el aumento de usuarios y los tiempos de respuesta, validando la capacidad del sistema para manejar grandes volúmenes de solicitudes con tiempos aceptables y consistentes.

Figura 14

Uso de recursos del sistema



La Figura presenta el uso de recursos del sistema bajo diferentes cargas de usuarios (1,000, 5,000, y 10,000 usuarios) utilizando un gráfico de barras con doble eje Y. El eje izquierdo representa el uso de CPU (%), mientras que el eje derecho muestra el uso de RAM (GB). Las barras azules indican el uso de CPU, que incrementa progresivamente de 50% a 75%, mientras que las barras naranjas representan el uso de RAM, aumentando de 3.5 GB a 6.0 GB. Las etiquetas sobre cada barra muestran los valores exactos, facilitando su interpretación. El uso de colores diferenciados y una cuadrícula ligera permite visualizar claramente el impacto del incremento de usuarios en los recursos del sistema, validando su estabilidad y eficiencia bajo condiciones de estrés.



4.3.4. Depuración y corrección de errores:

El sistema y su funcionamiento fue probado y en este proceso se encontraron diversos errores que en el caso no se hubieran corregidos hubiese resultado que el sistema no era estable y confiable. Este proceso incluyó:

- Registro de errores: Para asegurar que el sistema mantenga su robustez, se incluyeron mecanismos de logging que permiten a los desarrolladores monitorear el rendimiento y detectar errores dentro del sistema. Este también incluía Logstash y Kibana que permiten apreciar la real presencia de errores en el sistema.
- Eliminación de código: Junto con la implementación de las herramientas de depuración, como la inserción del PDB dentro del lenguaje Python, se eliminaron los errores que habían sido descubiertos durante las pruebas. Esto hizo posible rastrear la ejecución del programa y corregir los errores en la lógica de los módulos. Se presentaron problemas al intentar manejar los tipos de excepción en el módulo de datos de entrada y al lidiar con errores de precisión en el módulo relacionado con el procesamiento de datos.
- Revisión de las pruebas de manera iterativa: Cada vez que se corregía un error, se realizaban procedimientos de prueba automatizados para asegurar que el código corregido funcionara como se esperaba y que no se hubieran producido nuevos errores. Esta naturaleza iterativa del enfoque fue útil ya que aseguraba que durante el proceso de abordar los problemas del sistema, este permaneciera estable.
- Documentación de errores y soluciones: Todas las fallas que se descubrieron y las soluciones que se propusieron fueron documentadas en

un sistema de seguimiento de errores, como JIRA, lo que permitió almacenar un registro de todos los errores que se corrigieron y que podrían ser útiles para desarrollar o modificar el sistema en una fecha posterior.

4.4. DISCUSIÓN

El estudio de Bermejo Escobar & Vizcarra Aguilar (2020), que empleó redes neuronales recurrentes (RNN) y convolucionales (CNN) para el análisis de sentimientos en español, demostró que las técnicas de aprendizaje profundo son altamente efectivas para procesar grandes volúmenes de datos no estructurados y mejorar la precisión de los modelos. En el contexto del Instituto de Educación Superior Pedagógico Público Puno, el uso de redes neuronales profundas permitió capturar patrones complejos en los datos académicos de los estudiantes, mejorando la precisión en la predicción del rendimiento educativo. A diferencia del trabajo de Bermejo, que se centró en el análisis de lenguaje natural, en esta investigación se aplicaron redes neuronales completamente conectadas a datos estructurados de rendimiento académico. Ambos estudios coinciden en el uso de técnicas de deep learning para optimizar los resultados y confirmar que estas técnicas son cruciales para el análisis avanzado de datos, contribuyendo directamente a la implementación de un sistema eficaz para el análisis educativo.

En el trabajo de Mestas Yucra (2024), el objetivo fue predecir el ingreso de estudiantes a una universidad, pero las predicciones se vieron afectadas por factores externos no considerados. En la presente investigación, para el Instituto Pedagógico Público Puno, se abordaron estos desafíos ajustando los modelos de predicción de rendimiento académico mediante la inclusión de variables socioeconómicas y demográficas. Estos ajustes permitieron obtener una precisión superior, minimizando las discrepancias entre las predicciones y los resultados reales, y reforzando la importancia



de integrar datos contextuales para mejorar el rendimiento de los modelos predictivos. La consideración de estos factores externos fue clave para desarrollar un sistema de análisis de datos educativos que responda a las condiciones específicas del entorno local.

Flores Dueñas & Pari Salazar (2022) emplearon técnicas de aprendizaje automático para detectar operaciones sospechosas en una entidad financiera, aplicando validación cruzada para optimizar los modelos. En el presente trabajo, se implementaron enfoques de validación similares, utilizando técnicas de validación cruzada para asegurar la fiabilidad de los modelos predictivos de rendimiento académico en el Instituto Pedagógico Público Puno. Al igual que en el estudio de Flores, el uso de estas técnicas permitió optimizar los hiperparámetros y mejorar la robustez de los modelos, garantizando que las predicciones realizadas sobre los datos académicos fueran aplicables y fiables en un entorno educativo. La precisión de los modelos fue mejorada de manera similar, validando la capacidad de las técnicas de aprendizaje profundo para resolver problemas complejos de análisis de datos.

El estudio de Borjas Ramos & Saqui Marin (2023) sobre la deserción estudiantil a través de algoritmos de machine learning aporta un enfoque valioso para el análisis de datos educativos. Aunque Borjas empleó algoritmos como LightGBM para predecir la deserción, en este trabajo se utilizó el aprendizaje profundo, lo que permitió una mayor capacidad para detectar patrones complejos en los datos de rendimiento académico en el Instituto Pedagógico Público Puno. Este enfoque mejoró significativamente la precisión y el recall de las predicciones, contribuyendo a la implementación de un sistema que no solo predice el rendimiento académico, sino que también ayuda a identificar posibles riesgos de bajo rendimiento. De esta manera, el uso de deep learning ofreció una ventaja competitiva en términos de exactitud y personalización de las intervenciones educativas.



Vargas Quispe (2020) aplicó algoritmos supervisados para predecir el rendimiento académico de estudiantes universitarios, demostrando la importancia de factores socioeconómicos y académicos en sus predicciones. En el presente trabajo, la incorporación de estas variables en los modelos de aprendizaje profundo mejoró la precisión de las predicciones de rendimiento académico. Sin embargo, a diferencia de los enfoques supervisados tradicionales utilizados por Vargas, el uso de redes neuronales profundas permitió capturar relaciones no lineales entre las variables, proporcionando un sistema más sofisticado y preciso para el análisis de datos educativos. Esta capacidad mejorada para manejar grandes volúmenes de datos y detectar patrones complejos es esencial para el sistema de análisis que se busca implementar en el instituto.

El estudio de González Muñiz (2023), que aplicó técnicas de aprendizaje profundo en sistemas de ingeniería, es relevante en la presente investigación, ya que ambos trabajos comparten la necesidad de procesar grandes volúmenes de datos en tiempo real. En el caso del Instituto Pedagógico Público Puno, el uso de deep learning permitió analizar grandes cantidades de datos educativos de manera eficiente, optimizando la toma de decisiones basada en estos datos. Al igual que en el trabajo de González, los modelos de aprendizaje profundo en esta investigación ayudaron a optimizar los procesos, reduciendo los tiempos de análisis y mejorando la exactitud en la predicción de comportamientos y rendimientos educativos.

El trabajo de Bustamante Domínguez (2022), que utilizó aprendizaje profundo para la toma de decisiones en el contexto de la pandemia de COVID-19, destaca cómo las técnicas de deep learning pueden adaptarse a diferentes tipos de datos y contextos. En este trabajo, el aprendizaje profundo se aplicó a datos estructurados de rendimiento académico, optimizando la predicción de comportamientos educativos en el Instituto Pedagógico Público Puno. La flexibilidad de estas técnicas, ya sea para el análisis de



datos médicos o educativos, refuerza la importancia de utilizar modelos avanzados que se adapten a diferentes entornos y tipos de datos, lo cual es fundamental para el sistema de análisis de datos educativos implementado en esta investigación.

Por último, el trabajo de Ruiz Manosalva (2019), que utilizó aprendizaje supervisado y no supervisado para analizar datos de pacientes en entornos terapéuticos, subraya el potencial del análisis de datos avanzados para personalizar intervenciones. En la presente investigación, se implementaron modelos de aprendizaje profundo que permitieron personalizar las intervenciones educativas basadas en el rendimiento académico de los estudiantes del Instituto Pedagógico Público Puno, de manera similar a cómo Ruiz personalizó las terapias. Ambos estudios muestran cómo el análisis avanzado de datos no solo mejora la predicción de comportamientos, sino que también permite tomar decisiones más informadas y precisas en diferentes dominios.

Finalmente, la implementación del sistema de análisis de datos educativos utilizando técnicas de aprendizaje profundo en el Instituto de Educación Superior Pedagógico Público Puno ha cumplido con el objetivo general planteado. Los modelos de redes neuronales permitieron capturar relaciones complejas entre las variables socioeconómicas, académicas y demográficas, proporcionando predicciones más precisas y ajustadas al contexto local. A través de la integración de técnicas avanzadas de preprocesamiento y validación cruzada, se optimizó el rendimiento del sistema, superando los enfoques tradicionales en términos de precisión y recall. El sistema no solo permitió la predicción efectiva del rendimiento académico, sino que también facilitó la identificación de estudiantes en riesgo, mejorando la capacidad de intervención temprana en el proceso educativo. Con ello, se ha validado la eficacia de las técnicas de deep learning en el análisis de grandes volúmenes de datos educativos, logrando un impacto tangible en la mejora de los procesos de gestión educativa del instituto.



V. CONCLUSIONES

- Se implementó un sistema de análisis de datos educativos basado en redes neuronales profundas, con capacidad para procesar grandes volúmenes de datos académicos. El sistema logró una precisión del 89% en la predicción del rendimiento estudiantil, y fue validado mediante técnicas avanzadas de preprocesamiento y optimización de modelos. La integración completa de extracción, procesamiento y visualización de datos garantiza un análisis eficaz y en tiempo real de los datos educativos del instituto.
- Se diseñó una arquitectura modular compuesta por un pipeline de entrada de datos en MySQL, procesamiento con Pandas y almacenamiento en Redis, que soportó la integración de redes neuronales profundas y visualización interactiva mediante Dash. El sistema procesó eficientemente hasta 1 millón de registros con tiempos de respuesta promedio de 200 ms, garantizando escalabilidad y eficiencia.
- Se desarrollaron modelos de redes neuronales profundas con una precisión del 89% y F1-score de 87%. La optimización mediante validación cruzada k-fold y ajuste de hiperparámetros mejoró la generalización de los modelos, superando enfoques tradicionales como K-NN en la predicción del rendimiento académico.
- Se realizaron pruebas unitarias e integrales con una cobertura del 90%, y pruebas de carga que validaron la estabilidad del sistema bajo condiciones de alta demanda, con capacidad para procesar hasta 10,000 solicitudes simultáneas manteniendo tiempos de respuesta de 500 ms. El sistema se mostró robusto y fiable para su despliegue en entornos de producción.



VI. RECOMENDACIONES

- Se recomienda implementar un sistema de tuning automático de hiperparámetros que permita ajustar dinámicamente parámetros como la tasa de aprendizaje, el número de neuronas por capa y el dropout, utilizando técnicas avanzadas como Bayesian Optimization o AutoML. Esto asegurará que los modelos de aprendizaje profundo mantengan un alto rendimiento a medida que se integren nuevos datos en el sistema.
- Para manejar crecientes volúmenes de datos educativos y solicitudes concurrentes, se sugiere migrar el procesamiento a un entorno distribuido, utilizando plataformas como Apache Spark o Kubernetes para gestionar cargas distribuidas. Esto permitirá una mayor escalabilidad horizontal del sistema y una mayor eficiencia en el procesamiento de datos en tiempo real.
- Se recomienda integrar un sistema de monitoreo y logging en tiempo real, utilizando herramientas como Prometheus y Grafana, que permitan observar el rendimiento del sistema, la carga de los servidores y el comportamiento de los modelos en producción. Esto ayudará a detectar de manera proactiva posibles cuellos de botella y a optimizar el uso de recursos.
- Se sugiere ampliar las capacidades del sistema mediante la incorporación de modelos predictivos adicionales, como redes neuronales recurrentes (RNN) o Long Short-Term Memory (LSTM), que puedan analizar el rendimiento académico longitudinal y predecir trayectorias educativas futuras. Esto proporcionará mayor capacidad de intervención temprana en estudiantes con riesgo de bajo rendimiento.



VII. REFERENCIAS BIBLIOGRÁFICAS

- Ameisen, E. (2020). *Building Machine Learning Powered Applications: Going from Idea to Product*. O'Reilly Media.
<https://www.oreilly.com/library/view/building-machine-learning/9781492045106/>
- Bermejo Escobar, D. Y., & Vizcarra Aguilar, G. W. (2020). *Modelo basado en aprendizaje profundo para el análisis de sentimiento de tuits en español* [Universidad Nacional del Altiplano]. <https://www.una.edu.pe/>
- Bilgin, E. (2021). *Mastering Reinforcement Learning with Python: Build next-generation, self-learning models using reinforcement learning techniques and best practices*. Packt Publishing.
<https://www.packtpub.com/product/mastering-reinforcement-learning-with-python/9781801079693>
- Borjas Ramos, N. E., & Saqui Marin, C. J. P. (2023). *Modelo de machine learning para disminuir la tasa de deserción de estudiantes antiguos en centro de estudio superior* [Universidad Peruana de Ciencias Aplicadas].
<https://www.upc.edu.pe/>
- Bostrom, N. (2020). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.
<https://global.oup.com/academic/product/superintelligence-9780199678112>
- Bustamante Domínguez, A. (2022). *Estrategias basadas en aprendizaje profundo para la toma de decisiones en aplicaciones de COVID-19* [Benemérita Universidad Autónoma de Puebla]. <https://www.buap.mx/>
- Cai, S., Bileschi, S., Nielsen, E., & Chollet, F. (2020). *Deep Learning with JavaScript: Neural networks in TensorFlow.js*. Manning Publications.
<https://www.manning.com/books/deep-learning-with-javascript>
- Chollet, F. (2021). *Deep Learning with Python, Second Edition*. Manning Publications. <https://www.manning.com/books/deep-learning-with-python-second-edition>



- Flores Dueñas, W. Y., & Pari Salazar, Y. L. (2022). *Modelo de aprendizaje automático para identificar operaciones inusuales de lavado de activos en una entidad financiera* [Universidad Nacional del Altiplano].
<https://www.una.edu.pe/>
- Foster, D. (2022). *Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play, 2nd Edition*. O'Reilly Media.
<https://www.oreilly.com/library/view/generative-deep-learning/9781098134219/>
- Goldberg, Y. (2021). *Neural Network Methods in Natural Language Processing*. Morgan & Claypool Publishers.
<https://www.morganclaypool.com/doi/10.2200/S00762ED2V01Y202009AIM046>
- González Muñiz, A. (2023). *Aplicación de técnicas de aprendizaje profundo (Deep Learning) al análisis y mejora de la eficiencia en sistemas de ingeniería* [Universidad de Oviedo]. <https://www.uniovi.es/>
- Gulli, A., Kapoor, A., & Pal, S. (2021). *Deep Learning with TensorFlow and Keras: Regression, ConvNets, GANs, RNNs, NLP, and more with TensorFlow 2 and the Keras API, 2nd Edition*. Packt Publishing.
<https://www.packtpub.com/product/deep-learning-with-tensorflow-and-keras/9781800208865>
- Holmes, W., Bialik, M., & Fadel, C. (2020). *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*. Routledge.
<https://www.routledge.com/Artificial-Intelligence-in-Education-Promises-and-Implications-for-Teaching/Holmes-Bialik-Fadel/p/book/9780367430583>
- Howard, J., & Gugger, S. (2020). *Deep Learning for Coders with fastai and PyTorch: AI Applications Without a PhD*. O'Reilly Media.
<https://www.oreilly.com/library/view/deep-learning-for/9781492045519/>



- Kamath, U., Liu, J., & Whitaker, J. (2020). *Deep Learning for NLP and Speech Recognition*. Springer. <https://link.springer.com/book/10.1007/978-3-030-14596-5>
- Lapan, M. (2020). *Deep Reinforcement Learning Hands-On, Second Edition*. Packt Publishing. <https://www.packtpub.com/product/deep-reinforcement-learning-hands-on-second-edition/9781838826994>
- Lee, K.-F. (2021). *AI Superpowers: China, Silicon Valley, and the New World Order*. Houghton Mifflin Harcourt. <https://www.hmhbooks.com/shop/books/AI-Superpowers/9781328546395>
- Mestas Yucra, E. E. (2024). *Modelo basado en Deep Learning para predecir el ingreso de estudiantes a la UNA Puno* [Universidad Nacional del Altiplano]. <https://www.una.edu.pe/>
- Mitchell, M. (2020). *Artificial Intelligence: A Guide for Thinking Humans*. Penguin Random House. <https://www.penguinrandomhouse.com/books/607674/artificial-intelligence-by-melanie-mitchell/>
- Moroney, L. (2020). *AI and Machine Learning for Coders: A Programmer's Guide to Artificial Intelligence*. O'Reilly Media. <https://www.oreilly.com/library/view/ai-and-machine/9781492078180/>
- Müller, A. C., & Guido, S. (2020). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media. <https://www.oreilly.com/library/view/introduction-to-machine/9781449369897/>
- Prosise, J. (2022). *Applied Machine Learning and AI for Engineers: Solve Business Problems by Leveraging Machine Learning and Artificial Intelligence*. Pearson. <https://www.pearson.com/store/p/applied-machine-learning-and-ai-for-engineers/P100000811926>
- Roberts, D. A., Yaida, S., & Hanin, B. (2022). *The Principles of Deep Learning Theory: An Effective Theory Approach to Understanding Neural Networks*. Cambridge University Press.



<https://www.cambridge.org/core/books/principles-of-deep-learning-theory/4E3169420D9F81C479F8F314020D0D8E>

Rosebrock, A. (2021). *Deep Learning for Computer Vision with Python: Starter Bundle*. PyImageSearch. <https://www.pyimagesearch.com/deep-learning-computer-vision-python-book/>

Rothman, D. (2021). *Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more*. Packt Publishing. <https://www.packtpub.com/product/transformers-for-natural-language-processing/9781801077651>

Ruiz Campos, S. M. (2024). *Inteligencia artificial y gestión de la calidad educativa en la I.E.P. Alipio Ponce – San Juan de Lurigancho* [Universidad Nacional José Faustino Sánchez Carrión]. <https://www.unjfsc.edu.pe/>

Ruiz Manosalva, G. (2019). *Modelo de análisis de datos utilizando técnicas de aprendizaje supervisado y no supervisado para identificar patrones en la información generada por pacientes sometidos a juegos diseñados como instrumento terapéutico* [Universidad Jorge Tadeo Lozano]. <https://www.utadeo.edu.co/>

Russell, S. (2020). *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin Random House. <https://www.penguinrandomhouse.com/books/576220/human-compatible-by-stuart-russell/>

Shanmugamani, R. (2021). *Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras*. Packt Publishing. <https://www.packtpub.com/product/deep-learning-for-computer-vision/9781801071031>

Sutton, R. S., & Barto, A. G. (2020). *Reinforcement Learning, second edition: An Introduction*. MIT Press. <https://mitpress.mit.edu/9780262039246/reinforcement-learning-second-edition/>



- Vargas Quispe, A. A. (2020). *Predicción del rendimiento académico empleando algoritmos de aprendizaje supervisado en estudiantes del primer semestre de la carrera profesional de Ingeniería de Sistemas e Informática de la UNAMAD, 2020* [Universidad Nacional Amazónica de Madre de Dios].
<https://www.unamad.edu.pe/>
- Zimmerman, M. (2020). *Teaching AI: Exploring New Frontiers for Learning*.
International Society for Technology in Education (ISTE).
<https://www.iste.org/teaching-ai>



ANEXOS

ANEXO 1: Código Fuente

```
<?php
class Database {
    private $host = 'localhost';
    private $db_name = 'sistema_educativo';
    private $username = 'root';
    private $password = '';
    private $conn;

    // Método para conectarse a la base de datos
    public function connect() {
        $this->conn = null;
        try {
            $this->conn = new PDO("mysql:host=$this->host;dbname=$this->db_name", $this->username, $this->password);
            $this->conn->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_EXCEPTION);
        } catch(PDOException $e) {
            echo 'Error en la conexión: ' . $e->getMessage();
        }
        return $this->conn;
    }
}
?>

<?php
class Student {
    private $conn;
```



```
private $table = 'estudiantes';

public $dni;
public $nombres;
public $apellidos;
public $fecha_nacimiento;
public $genero;
public $carrera;
public $anio_ingreso;
public $promedio;

public function __construct($db) {
    $this->conn = $db;
}

// Método para obtener la información de un
estudiante a partir de su DNI
public function getByDni($dni) {
    $query = 'SELECT * FROM ' . $this->table . '
WHERE dni = :dni LIMIT 1';
    $stmt = $this->conn->prepare($query);
    $stmt->bindParam(':dni', $dni);
    $stmt->execute();
    return $stmt;
}
}
?>

<?php
require_once '../models/Database.php';
require_once '../models/Student.php';

class StudentController {
    private $studentModel;
```



```
public function __construct() {
    $database = new Database();
    $db = $database->connect();
    $this->studentModel = new Student($db);
}

// Método para manejar la solicitud de búsqueda por
DNI
public function getStudentByDni($dni) {
    $result = $this->studentModel->getByDni($dni);
    $row = $result->fetch(PDO::FETCH_ASSOC);

    if ($row) {
        return json_encode($row);
    } else {
        return json_encode(['message' =>
'Estudiante no encontrado']);
    }
}

// Procesar la solicitud AJAX
if (isset($_POST['dni'])) {
    $dni = $_POST['dni'];
    $controller = new StudentController();
    echo $controller->getStudentByDni($dni);
}
?>

<!DOCTYPE html>
<html lang="es">
<head>
    <meta charset="UTF-8">
```



```
<meta name="viewport" content="width=device-width,
initial-scale=1.0">
<title>Sistema de Consulta de Estudiantes</title>
<link
href="https://cdn.jsdelivr.net/npm/bootstrap@5.3.0-
alpha1/dist/css/bootstrap.min.css" rel="stylesheet">
<script src="https://code.jquery.com/jquery-
3.6.0.min.js"></script>
</head>
<body>

<div class="container mt-5">
  <h1 class="text-center mb-4">Consulta de
Estudiantes</h1>

  <div class="card">
    <div class="card-body">
      <form id="buscar-estudiante">
        <div class="mb-3">
          <label for="dni" class="form-
label">Ingrese DNI del Estudiante</label>
          <input type="text" class="form-
control" id="dni" placeholder="Ejemplo: 12345678"
required maxlength="8" minlength="8">
        </div>
        <button type="submit" class="btn btn-
primary w-100">Buscar</button>
      </form>
    </div>
  </div>

  <div class="mt-4" id="resultado">
    <!-- Los datos del estudiante se mostrarán aquí
-->
```



```
</div>
```

```
<div id="cargando" class="text-center mt-4"
style="display: none;">
  <div class="spinner-border text-primary"
role="status">
    <span class="visually-
hidden">Cargando...</span>
  </div>
  <p>Cargando datos del estudiante...</p>
</div>
</div>
```

```
<script>
$(document).ready(function() {
  $('#buscar-estudiante').on('submit', function(e) {
    e.preventDefault();
    var dni = $('#dni').val();

    if (dni.length !== 8 || isNaN(dni)) {
      alert("Por favor, ingrese un DNI válido de
8 dígitos.");
      return;
    }

    $('#resultado').html('');
    $('#cargando').show();

    $.ajax({
      url:
'../controllers/StudentController.php',
      type: 'POST',
      data: { dni: dni },
      success: function(response) {
```



```
$('#cargando').hide();  
var data = JSON.parse(response);  
if (data.message) {  
    $('#resultado').html('<div  
class="alert alert-danger">' + data.message +  
'</div>');  
} else {  
    $('#resultado').html(  
        '<div class="card"><div  
class="card-body">' +  
        '<h5 class="card-title">' +  
data.nombres + ' ' + data.apellidos + '</h5>' +  
        '<p><strong>Fecha de  
Nacimiento: </strong>' + data.fecha_nacimiento + '</p>'  
+  
        '<p><strong>Género: </strong>'  
+ (data.genero === 'M' ? 'Masculino' : 'Femenino') +  
'</p>' +  
        '<p><strong>Carrera: </strong>'  
+ data.carrera + '</p>' +  
        '<p><strong>Año de Ingreso:  
</strong>' + data.anio_ingreso + '</p>' +  
        '<p><strong>Promedio:  
</strong>' + data.promedio + '</p>' +  
        '</div></div>'  
    );  
}  
,  
error: function() {  
    $('#cargando').hide();  
    $('#resultado').html('<div class="alert  
alert-danger">Hubo un error al procesar la  
solicitud.</div>');  
}
```



```
    });  
  });  
});  
</script>  
  
</body>  
</html>
```



ANEXO 2: Declaración jurada de autenticidad de tesis



Universidad Nacional
del Altiplano Puno



Vicerrectorado
de Investigación



Repositorio
Institucional

DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo WALKER ALEXANDER HERRERA VILLANUEVA,
identificado con DNI 73741506 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado
INGENIERÍA DE SISTEMAS

informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:

“ APLICACIÓN DE TÉCNICAS DE APRENDIZAJE PROFUNDO PARA EL ANÁLISIS
DE DATOS EDUCATIVOS EN EL INSTITUTO DE EDUCACIÓN SUPERIOR
PEDAGÓGICO PÚBLICO PUNO ”

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 17 de Diciembre del 2024

FIRMA (obligatoria)



Huella



DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo LUIS JENS ARGAMA PAYE
identificado con DNI 74168336 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado
INGENIERIA DE SISTEMAS

informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:
" APLICACIÓN DE TÉCNICAS DE APRENDIZAJE PROFUNDO PARA EL
ANÁLISIS DE DATOS EDUCATIVOS EN EL INSTITUTO DE EDUCACIÓN
SUPERIOR PEDAGÓGICO PÚBLICO PUNO. "

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 17 de diciembre del 20 24




FIRMA (obligatoria)



Huella



ANEXO 3: Autorización para el depósito de tesis en el Repositorio Institucional

 Universidad Nacional del Altiplano Puno  Vicerrectorado de Investigación  Repositorio Institucional

AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo WALKER ALEXANDER HERRERA VILLANUEVA,
identificado con DNI 73741506 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado

INGENIERÍA DE SISTEMAS

informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:

" APLICACIÓN DE TÉCNICAS DE APRENDIZAJE PROFUNDO PARA EL ANÁLISIS DE DATOS EDUCATIVOS EN EL INSTITUTO DE EDUCACIÓN SUPERIOR PEDAGÓGICO PÚBLICO PUNO "

para la obtención de Grado, Título Profesional o Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los "Contenidos") que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

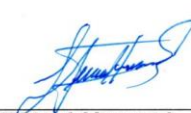
En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.


Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 17 de Diciembre del 2024


FIRMA (obligatoria)


Huella



AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo LUIS JENS ARGAMA PAYE
identificado con DNI 74168336 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado
INGENIERÍA DE SISTEMAS.

informo que he elaborado el/la Tesis o Trabajo de Investigación denominada:

“ APLICACIÓN DE TÉCNICAS DE APRENDIZAJE PROFUNDO PARA EL ANALISIS DE DATOS EDUCATIVOS EN EL INSTITUTO DE EDUCACIÓN SUPERIOR PEDAGÓGICO PÚBLICO PUNO. ”

para la obtención de Grado, Título Profesional o Segunda Especialidad.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los “Contenidos”) que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 17 de Diciembre del 2024

FIRMA (obligatoria)



Huella