

Inteligencia Artificial al Servicio de la Educación

Redes Neuronales en el Pronóstico del Rendimiento Académico

Jorge Puga de la Cruz
Rony Torres Monzon
Angel Enrique Lopez Rojas
Manuel Tuesta Moreno



IDEOs

Centro de Investigación
y Producción Científica

Inteligencia Artificial al Servicio de la Educación

Redes Neuronales en el Pronóstico del
Rendimiento Académico

Editor



Inteligencia Artificial al Servicio de la Educación

Redes Neuronales en el Pronóstico del Rendimiento Académico

Jorge Puga de la Cruz, Rony Torres Monzon, Angel Enrique Lopez
Rojas, Manuel Tuesta Moreno

Editado por

CENTRO DE INVESTIGACIÓN & PRODUCCIÓN CIENTÍFICA
IDEOS E.I.R.L

Dirección: Calle Teruel 292, Miraflores, Lima, Perú.

RUC: 20606452153

Primera edición digital, Diciembre 2024

Libro electrónico disponible en www.tecnohumanismo.online

ISBN: 978-612-5166-47-0

Registro de Depósito legal N°: 2025-00303

ISBN: 978-612-5166-47-0



Jorge Puga de la Cruz

 <https://orcid.org/0000-0002-3182-3101>

jpuguita@gmail.com

Universidad Nacional de la Amazonía Peruana, Iquitos – Perú

Rony Torres Monzon

 <https://orcid.org/0000-0002-7477-4004>

rtorresmonzon@gmail.com

Universidad Nacional de la Amazonía Peruana, Iquitos – Perú

Angel Enrique Lopez Rojas

 <https://orcid.org/0009-0007-8709-8352>

aeloper68@hotmail.com

Universidad Nacional de la Amazonía Peruana, Iquitos – Perú

Manuel Tuesta Moreno

 <https://orcid.org/0000-0003-0746-7108>

ejcesj@gmail.com

Universidad Nacional de la Amazonía Peruana, Iquitos – Perú

Este libro científico se ha originado de la tesis de maestría titulada:

**“REDES NEURONALES ARTIFICIALES PARA PRONOSTICAR EL
RENDIMIENTO ACADÉMICO DE ALUMNOS DE INGENIERÍA
DE SISTEMAS E INFORMÁTICA DE LA UNIVERSIDAD
NACIONAL DE LA AMAZONÍA PERUANA”**

Presentada por Jorge Puga de la Cruz y Rony Torres Monzón para optar por el grado académico de Maestro en Ingeniería de Sistemas con Mención en Gerencia de la Información y Gestión de Software en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, en Iquitos, Perú, en el año 2023.

Esta investigación se enmarca en el desarrollo de modelos predictivos aplicados al ámbito educativo, utilizando redes neuronales artificiales como herramienta principal. Su objetivo es mejorar la comprensión y el pronóstico del rendimiento académico de los estudiantes de Ingeniería de Sistemas e Informática, contribuyendo a la optimización de estrategias pedagógicas y de gestión en el entorno universitario.

A la memoria de mi padre, a mi madre y a mis hijos Milagritos, Josi de Jesús, Luciano, a Silvia; que siempre son mis motivos para mi superación académica.

Jorge Puga

A mis padres por su amor y consejos, a mi esposa por su confianza, amor y apoyo, a mis hijos por ser el motor y motivo para lograr mis objetivos.

Rony Torres

AGRADECIMIENTO

A nuestros asesores Dr. Carlos García Cortegano y Dr. Alejandro Reategui Pezo, por la dedicación que nos brindaron en este trabajo de investigación, a sus enseñanzas consejos y a todos los docentes de la Escuela de PostGrado – UNAP que hicieron posible la culminación de nuestra maestría, a la Universidad Nacional de la Amazonía Peruana que nos brindó la oportunidad de hacer la maestría y a todos mis amigos que nos alentaron y motivaron siempre.

INDICE

AGRADECIMIENTO	4
RESEÑA	8
INTRODUCCIÓN	10
CAPÍTULO I	14
1.1 Redes Neuronales Artificiales: Una Introducción	14
1.1.1. Historia y Principios Básicos	14
1.1.2. Tipologías de Redes y Sus Aplicaciones.....	16
1.2. El Rendimiento Académico en el Foco del Análisis	17
1.2.1. Factores Determinantes en el Éxito Académico	18
1.2.2. Perspectivas Psicológicas y Sociales.....	21
1.3. Modelos Predictivos en la Educación.....	22
1.3.1. Inteligencia Artificial vs. Métodos Tradicionales.....	22
1.3.2. Beneficios y Desafíos de las Redes Neuronales.....	24
1.3.3. Aplicaciones Prácticas en la Educación	25
1.4. Glosario de Términos Clave.....	26
1.4.1. Algoritmo de Aprendizaje	26
1.4.2. Capa Oculta	27
1.4.3. Overfitting.....	27
1.4.4. Regularización	28
1.4.5. Sesgo Algorítmico.....	29
1.4.6. Algoritmo	29
1.4.7. Aprendizaje Supervisado	30
1.4.8. Aprendizaje No Supervisado	31
1.4.9. Dato	31
1.4.9. Dataset.....	32
1.4.10. Evaluación del Rendimiento Académico.....	32
1.4.11. Información.....	33
1.4.12. Machine Learning (ML).....	33
1.4.13. Patrón o Tendencia	34
1.4.14. Rendimiento Académico	34
1.4.15. Tabla	34
1.5. Bases Teóricas	35
CAPÍTULO II	39
2.1. Estructura del Modelo de Investigación.....	39
2.1.1. Diseño No Experimental y Enfoque Cuantitativo.....	39

2.1.2. Variables Principales del Estudio	40
2.2. Marco Conceptual y Planteamiento de Hipótesis	43
2.2.1. Relación entre Variables y Resultados	43
2.2.2. Hipótesis Principal y Secundarias	44
2.3. Metodología de Datos y Fuentes	46
2.3.1. Selección de Muestra y Población	46
2.3.2. Instrumentos y Técnicas de Recolección de Datos	47
2.3.3. Técnicas de Procesamiento y Análisis de Datos	48
2.4. Variables y su operacionalización	49
2.5. Aspectos Éticos	51
CAPÍTULO III	53
3.1. Construcción de la Red Neuronal	53
3.1.1. Selección de Algoritmos y Parámetros	54
3.1.2. Entrenamiento y Validación del Modelo	55
3.2. Procesamiento de Datos	57
3.2.1. Limpieza y Preparación de Datos	58
3.2.2. Técnicas de Análisis y Visualización	59
3.3. Evaluación del Modelo	61
3.3.1. Métricas de Desempeño	61
CAPÍTULO IV	65
4.1. Desempeño del Modelo en la Predicción Académica	65
4.1.1. Análisis de Precisión y Confiabilidad	66
4.1.2. Impacto de las Variables Independientes	67
4.2. Estudio de Casos Destacados	69
4.2.1. Tendencias Comunes	69
4.2.2. Análisis de Resultados Atípicos	71
4.3. Limitaciones y Áreas de Mejora	73
4.3.1. Limitaciones del Modelo	73
4.3.2. Áreas de Mejora	74
4.4. Diseño del modelo	76
4.5. CONTRASTACIÓN DE HIPÓTESIS	83
4.5.1. HIPÓTESIS DERIVADAS	83
4.5.2. HIPOTESIS GENERAL	89
CAPÍTULO V	90
5.1. Implicancias Educativas del Modelo	90
5.1.1. Aportes a la Gestión Académica	90
5.1.2. Oportunidades para Políticas Educativas	91

5.2. Comparación con Estudios Similares	91
5.2.1. Coincidencias y Contrastes.....	91
5.2.2. Innovaciones Introducidas.....	92
5.3. Lecciones Aprendidas y Reflexiones.....	93
5.3.1. Lecciones Aprendidas	93
5.3.2. Reflexiones.....	93
CAPÍTULO VI	95
6.1. Integración del Modelo en el Sistema Educativo.....	95
6.1.1. Infraestructura y Recursos Necesarios	95
6.1.2. Capacitación y Adopción por Parte de los Docentes	96
6.2. Optimización del Rendimiento Académico.....	97
6.2.1. Uso del Modelo para la Intervención Temprana.....	97
6.2.2. Estrategias para Maximizar el Potencial del Estudiante	97
6.3. Plan de Escalabilidad y Futuro del Sistema	98
6.3.1. Expansión a Otros Cursos y Facultades	98
6.3.2. Actualización y Mejora Continua.....	98
6.3.3. Colaboración con la Comunidad Educativa.....	99
CONCLUSIONES.....	100
RECOMENDACIONES	103
REFERENCIAS BIBLIOGRÁFICAS.....	107

RESEÑA

El libro "**Inteligencia Artificial al Servicio de la Educación: Redes Neuronales en el Pronóstico del Rendimiento Académico**" es una obra que combina el rigor académico con un enfoque práctico en el uso de técnicas de aprendizaje automático para abordar los desafíos del rendimiento académico en la educación superior. Esta investigación aplicada, desarrollada en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, ofrece un análisis detallado sobre cómo las tecnologías emergentes pueden transformar la manera en que se monitorean y mejoran los resultados educativos.

El texto comienza explorando la problemática del bajo rendimiento académico en el curso de Álgebra Lineal, un área crítica para los estudiantes de Ingeniería de Sistemas. A través de una metodología bien estructurada, los autores presentan cómo las redes neuronales artificiales, un enfoque avanzado dentro del aprendizaje automático, pueden ser utilizadas para predecir con alta precisión si los estudiantes aprobarán o desaprobarán el curso. Este enfoque no solo supera las técnicas tradicionales de análisis de datos, sino que también establece un precedente para la adopción de tecnologías innovadoras en la educación.

La investigación utiliza herramientas como MATLAB Neural Network Toolbox y MS Excel para analizar datos históricos de los estudiantes, aplicando un modelo predictivo que alcanza métricas destacadas: una precisión del 97.6%, una exhaustividad del 100% y una exactitud del 97.9%. Estos resultados posicionan al modelo como una solución efectiva y altamente confiable, capaz de superar los métodos utilizados en estudios similares realizados en otros contextos. Además, se destacan indicadores específicos, como un error relativo bajo (%E de 2,083) y un coeficiente de error (CE de 0,196274), que refuerzan la solidez del enfoque adoptado.

A lo largo de sus capítulos, el libro detalla cada etapa del desarrollo del modelo predictivo, desde la selección de variables relevantes, como la asistencia y las calificaciones previas, hasta el preprocesamiento de datos y la configuración de la arquitectura de la red neuronal. La explicación técnica es complementada con ejemplos claros y análisis detallados, lo que permite al lector comprender tanto el proceso como los resultados obtenidos.

Sin embargo, este libro no se limita a una descripción técnica. También ofrece reflexiones profundas sobre las implicancias de los hallazgos, planteando cómo estas tecnologías pueden integrarse en la gestión educativa para optimizar la toma de decisiones. Los autores proponen que la implementación de modelos predictivos no solo facilita la identificación temprana de estudiantes en riesgo, sino que también permite diseñar estrategias de intervención personalizadas y asignar recursos de manera más eficiente.

Además, se enfatiza el valor transformador de estas tecnologías en el diseño de políticas educativas basadas en datos. Al proporcionar un enfoque práctico y replicable, esta investigación abre nuevas oportunidades para que otras instituciones educativas adopten enfoques similares, fomentando la equidad y mejorando la calidad de la educación.

Un aspecto destacado del libro es su capacidad para conectar la teoría y la práctica. Los autores no solo validan el modelo con datos históricos, sino que también plantean escenarios futuros para su implementación y escalabilidad, proponiendo estrategias para integrar el modelo en sistemas educativos más amplios. Estas propuestas incluyen la creación de paneles interactivos para docentes, registros electrónicos integrales para capturar más variables predictivas y capacitación en herramientas tecnológicas para personal académico y administrativo.

La obra se distingue también por su lenguaje accesible, que logra transmitir conceptos técnicos complejos de manera comprensible para lectores de diversos perfiles. Desde investigadores interesados en el aprendizaje automático hasta gestores educativos que buscan soluciones prácticas para sus instituciones, este libro ofrece un recurso invaluable para comprender y aplicar tecnologías predictivas en la educación.

En conclusión, "**Inteligencia Artificial al Servicio de la Educación: Redes Neuronales en el Pronóstico del Rendimiento Académico**" es una contribución significativa al campo de la educación y la tecnología. Su enfoque interdisciplinario combina la innovación tecnológica con un compromiso por mejorar los procesos educativos, ofreciendo un modelo que puede ser adaptado y replicado en diversos contextos. Esta obra no solo valida el uso de redes neuronales artificiales como herramientas efectivas para la predicción académica, sino que también establece un precedente para la integración de tecnologías avanzadas en la educación superior, posicionándola como una referencia esencial para el futuro del aprendizaje basado en datos.

INTRODUCCIÓN

La educación en la Amazonía peruana enfrenta una serie de desafíos que reflejan las complejidades propias de la región y sus características socioeconómicas. Dentro de este contexto, la Universidad Nacional de la Amazonía Peruana (UNAP) se erige como una institución fundamental para el desarrollo académico y profesional de los jóvenes de la región. Con su Facultad de Ingeniería de Sistemas e Informática, la UNAP busca formar profesionales capaces de afrontar los retos tecnológicos contemporáneos y contribuir significativamente al progreso regional y nacional. Sin embargo, la realidad educativa muestra un panorama de dificultades relacionadas con el rendimiento académico, particularmente en cursos críticos como Álgebra Lineal, el cual es esencial en la formación de competencias analíticas y computacionales que son el pilar de la Ingeniería de Sistemas.

La alta tasa de desaprobación en este curso no solo repercute en la trayectoria académica de los estudiantes, sino que también pone de manifiesto la necesidad de adoptar enfoques innovadores para abordar esta problemática. La educación tradicional basada únicamente en exámenes y evaluaciones periódicas ha demostrado ser insuficiente para anticipar y atender las dificultades que enfrentan los estudiantes. Ante esta situación, se plantean preguntas fundamentales: ¿cómo identificar de manera temprana a los estudiantes con mayor riesgo de desaprobar?, ¿qué herramientas tecnológicas podrían ser útiles para abordar esta problemática?, ¿es posible aplicar modelos predictivos avanzados en el contexto educativo de la Amazonía? Estas inquietudes constituyen el núcleo de la presente investigación.

En este escenario, las Redes Neuronales Artificiales (RNA) surgen como una herramienta prometedora. Las RNA, una rama avanzada del Machine Learning, tienen la capacidad de procesar grandes volúmenes de datos, identificar patrones complejos y generar predicciones con altos niveles de precisión. En el ámbito educativo, estas tecnologías han sido exploradas en diversos estudios internacionales con resultados alentadores, lo que las posiciona como una alternativa viable para mejorar la gestión del rendimiento académico. Este trabajo de investigación propone evaluar la eficacia de las RNA para predecir el rendimiento de los estudiantes de Ingeniería de Sistemas en el curso de Álgebra Lineal, con la finalidad de ofrecer soluciones prácticas y científicas a una problemática que afecta no solo a los estudiantes, sino también al sistema educativo en su conjunto.

La región amazónica del Perú se caracteriza por su vastedad geográfica y su diversidad cultural, factores que también influyen en el acceso a la educación superior y en la calidad de esta. A pesar de los esfuerzos del Estado y de las instituciones educativas por mejorar las condiciones académicas, persisten limitaciones significativas en términos de recursos, infraestructura y metodologías pedagógicas. En este contexto, la UNAP representa un faro de oportunidad para los jóvenes de la región, brindándoles acceso a programas académicos orientados al desarrollo tecnológico y profesional. Sin embargo, la elevada deserción estudiantil y las bajas tasas de aprobación en cursos clave reflejan un problema estructural que requiere soluciones urgentes e innovadoras.

La Facultad de Ingeniería de Sistemas e Informática enfrenta desafíos relacionados con la formación de competencias esenciales para el desempeño profesional en un mundo cada vez más digitalizado. El curso de Álgebra Lineal, al ser un componente fundamental de la malla curricular, pone a prueba las capacidades analíticas y computacionales de los estudiantes. Sin embargo, la complejidad del curso, combinada con factores como la falta de herramientas de soporte académico y métodos de enseñanza poco dinámicos, ha resultado en un rendimiento académico insuficiente. Esto no solo afecta la trayectoria educativa de los estudiantes, sino que también repercute en la calidad de los futuros profesionales de la región.

En este contexto, las RNA se presentan como una solución innovadora y transformadora. Estas herramientas han demostrado su capacidad para analizar datos académicos y generar predicciones precisas sobre el rendimiento estudiantil. Su implementación no solo permite identificar a los estudiantes en riesgo, sino también proponer estrategias personalizadas de intervención. Además, su uso en la UNAP podría sentar un precedente para otras instituciones educativas de la región, promoviendo una adopción más amplia de tecnologías avanzadas en la educación. Este estudio busca demostrar cómo las RNA pueden revolucionar la gestión del rendimiento académico, ofreciendo una herramienta práctica y científica para superar las limitaciones actuales.

El objetivo principal de esta investigación es implementar y evaluar un modelo de Redes Neuronales Artificiales que permita predecir con precisión el rendimiento académico de los estudiantes en el curso de Álgebra Lineal. En términos específicos, se busca:

- Alcanzar una exactitud del 96% en la predicción de estudiantes aprobados y desaprobados, lo que refleja la confiabilidad del modelo propuesto.

- Obtener una precisión superior al 97%, garantizando que las predicciones realizadas sean útiles para la toma de decisiones.
- Lograr un índice de exhaustividad del 98%, asegurando que la mayoría de los casos relevantes sean identificados correctamente.

El impacto esperado de esta investigación trasciende el ámbito académico, abarcando también beneficios prácticos y estratégicos para la gestión educativa. La implementación de RNA no solo contribuirá a mejorar las tasas de aprobación en cursos críticos, sino que también permitirá optimizar los recursos institucionales al identificar y atender de manera oportuna a los estudiantes en riesgo. Además, los resultados de este estudio pueden servir como modelo para otras universidades e instituciones educativas de la región, fomentando una cultura de innovación tecnológica en la educación superior.

El contenido de este libro se ha organizado en capítulos que abordan de manera integral cada aspecto de la investigación:

Capítulo I: Fundamentos y Bases Conceptuales. Explora el marco teórico y las bases conceptuales relacionadas con el uso de RNA en la educación.

Capítulo II: Diseño y Metodología de la Investigación. Detalla las etapas del diseño del estudio, las variables consideradas y los métodos empleados.

Capítulo III: Desarrollo e Implementación del Modelo. Describe el proceso de creación y validación del modelo de Redes Neuronales Artificiales.

Capítulo IV: Resultados y Análisis. Presenta los hallazgos clave y los analiza en función de los objetivos planteados.

Capítulo V: Discusión y Propuestas de Mejora. Reflexiona sobre los resultados obtenidos y ofrece recomendaciones prácticas.

Capítulo VI: Conclusiones y Proyecciones Futuras. Resume los principales aportes del estudio y sugiere líneas de investigación futuras.

Con esta estructura, el libro no solo busca proporcionar una contribución científica sólida, sino también servir como un recurso práctico para educadores, administradores y

profesionales interesados en mejorar la calidad y efectividad de la educación en contextos similares.

Te damos la bienvenida a este viaje a través de la innovación, la tecnología y la educación. Este libro es más que un análisis técnico sobre el uso de Redes Neuronales Artificiales; es una invitación a explorar cómo las herramientas avanzadas de Machine Learning pueden transformar realidades educativas, resolver problemas persistentes y abrir nuevas puertas al conocimiento en contextos como el de la Amazonía peruana.

En estas páginas descubrirás el fruto de una investigación que busca responder a una problemática clave: ¿cómo mejorar la precisión en la predicción del rendimiento académico de los estudiantes de Ingeniería de Sistemas? A través de un enfoque práctico y riguroso, analizamos cómo la integración de tecnologías emergentes puede no solo diagnosticar dificultades, sino también plantear soluciones que impacten positivamente en la formación de futuros profesionales.

Al adentrarte en esta lectura, te encontrarás con un panorama educativo que refleja los retos y oportunidades de la región amazónica, un análisis detallado sobre el potencial de las Redes Neuronales Artificiales en el ámbito académico y propuestas que trascienden los muros de un aula para impactar la gestión educativa de manera integral.

Te invitamos a leer con curiosidad y espíritu crítico, a reflexionar sobre los desafíos que enfrentamos en la educación actual y a visualizar cómo estas herramientas tecnológicas pueden ser aplicadas en distintos contextos. Este libro está diseñado no solo para informar, sino también para inspirar y abrir la puerta a nuevas posibilidades en el uso de la tecnología al servicio del aprendizaje.

Iniciemos juntos este recorrido, con la convicción de que cada página nos acerca un paso más hacia un futuro donde la innovación y la educación caminen de la mano. ¡Comencemos!

CAPÍTULO I

FUNDAMENTOS Y BASES TEÓRICAS

En las últimas décadas, las Redes Neuronales Artificiales (RNA) han emergido como una de las tecnologías más influyentes en el campo de la inteligencia artificial. Inspiradas en la complejidad y el funcionamiento intrincado del cerebro humano, estas redes imitan el comportamiento de las neuronas biológicas para procesar, analizar y aprender a partir de datos. Su impacto se ha extendido a múltiples ámbitos, incluyendo la salud, la educación, la industria, el entretenimiento, y un sinnúmero de aplicaciones específicas que transforman la manera en que interactuamos con la tecnología y comprendemos nuestro entorno.

La relevancia de las RNA radica no solo en su capacidad para resolver problemas complejos, sino también en su aplicación práctica en campos que afectan nuestra vida cotidiana. Desde la predicción de enfermedades hasta la personalización de experiencias de aprendizaje, estas herramientas tecnológicas han redefinido cómo se conciben y resuelven los problemas en un mundo cada vez más interconectado. Este capítulo se propone desglosar de manera exhaustiva los principios teóricos y las bases conceptuales que sustentan esta tecnología disruptiva. Además, se destacarán las aplicaciones específicas de las RNA en el contexto del rendimiento académico, demostrando cómo estas herramientas pueden contribuir a transformar la educación en el siglo XXI.

Para lograr una comprensión integral, se abordará la evolución histórica de las RNA, sus principales tipologías y su funcionamiento interno. Asimismo, se analizarán los factores que influyen en el éxito académico desde perspectivas psicológicas y sociales, estableciendo un puente entre las capacidades tecnológicas y las necesidades humanas en el ámbito educativo. Finalmente, se proporcionará un glosario de términos clave, permitiendo al lector adentrarse con mayor profundidad en los conceptos y terminologías fundamentales de este campo.

1.1 Redes Neuronales Artificiales: Una Introducción

1.1.1. Historia y Principios Básicos

El nacimiento de las Redes Neuronales Artificiales (RNA) es el resultado de una búsqueda por entender y replicar el funcionamiento del cerebro humano. En los años 1940,

Warren McCulloch y Walter Pitts introdujeron un modelo matemático que representaba la actividad neuronal utilizando funciones lógicas. Este modelo, aunque rudimentario, sentó las bases para una visión computacional del aprendizaje y el procesamiento de información.

El concepto evolucionó rápidamente, y durante los años 1950 y 1960, el perceptrón se posicionó como la primera implementación práctica de una RNA. Este modelo fue diseñado para realizar tareas de clasificación básica mediante un proceso de ajuste de pesos, pero pronto enfrentó limitaciones al no poder manejar problemas no lineales. Estas restricciones llevaron a un declive en el interés por las RNA, un período conocido como el "invierno de la inteligencia artificial".

Sin embargo, en los años 1980, la introducción del algoritmo de retropropagación del error por parte de investigadores como Geoffrey Hinton permitió entrenar redes multicapa, abriendo nuevas posibilidades para resolver problemas complejos. Este avance marcó el renacimiento de las RNA y su establecimiento como una herramienta fundamental en la inteligencia artificial.

Hoy en día, las RNA se benefician de avances en tecnologías de hardware, como unidades de procesamiento gráfico (GPU), y del acceso a grandes volúmenes de datos. Sus principios fundamentales incluyen la organización jerárquica en capas (entrada, ocultas y salida), la capacidad de aprendizaje mediante ajustes iterativos de pesos y la posibilidad de modelar relaciones complejas y no lineales en los datos. Esta flexibilidad permite su aplicación en una amplia variedad de problemas, desde la visión por computadora hasta el análisis predictivo.

Según Paja (2015), un estudio llevado a cabo en Puno con una muestra de 696 estudiantes matriculados entre 2009 y 2015 identificó 23 variables predictoras que se usaron para predecir el rendimiento académico. Utilizando el método stepwise, se seleccionaron 17 variables para el sujeto 009-05, lo que resultó en un coeficiente de determinación de 0.72, un error del modelo de 0.11 y un error de predicción de 0.37. En la asignación 010-05, se utilizaron 16 variables y se obtuvo un R^2 de 0.92, con un error del modelo de 0.96 y un error de predicción de 0.0032. La tesis concluyó que las redes neuronales artificiales son la mejor técnica de predicción, ya que proporcionan una diferencia más baja entre el error de modelo y el error de predicción en comparación con otras técnicas (p. 120).

En el año 2019, un estudio no experimental de correlación se llevó a cabo en la ciudad de Cusco, utilizando una muestra compuesta por 12.698 estudiantes de la Universidad Nacional de San Antonio Abad del Cusco. Según los resultados, factores como las notas de ingreso, la escuela profesional, el semestre, el género y la modalidad de ingreso influyen en el rendimiento académico de los estudiantes. La hipótesis se confirmó a través de análisis estadísticos realizados con el test chi-cuadrado y la correlación de Pearson. Se encontró que el algoritmo "RandomForest" fue el mejor en la predicción del rendimiento académico, con una precisión del 69%, seguido por la regresión logística con un 68% de acierto. (Candia, 2019, p. 159).

En 2021, se realizó un estudio doctoral aplicativo cuasiexperimental en Chimbote, con los estudiantes de la Universidad Nacional del Santa como población de estudio. La muestra compuesta por estudiantes de cuatro escuelas profesionales de la Facultad de Ingeniería, con registros académicos desde el semestre 2004 hasta el 2018, permitió una mejor evaluación del desempeño académico, aumentando el porcentaje de seguimiento de los estudiantes en un rango del 28.89% al 58.47%. El modelo predictivo que se utilizó basado en aprendizaje automático consistió en una red neuronal de perceptrón multicapa con una capa de salida softmax, logrando un mayor porcentaje de precisión en comparación con los clasificadores XGB y árbol de decisiones. (Caselli, 2021).

En 2021, una investigación fue realizada en Bagua con un grupo de 50 estudiantes de ingeniería civil que habían terminado el curso de física. La metodología del estudio incluyó la realización de entrevistas con un cuestionario confiable (con un coeficiente de confiabilidad de 0.854) a través del enfoque CRIS-DM, que comenzó identificando el problema y las variables relacionadas. Luego, los datos fueron recopilados, seleccionados y utilizados para construir la arquitectura de la red neuronal artificial. La validación del modelo se llevó a cabo mediante los algoritmos Scaled Conjugate Gradient y Levenberg-Marquardt, y los resultados sugirieron que las redes neuronales artificiales son una herramienta eficaz para la predicción, logrando una capacidad de predicción de 86% y 70% para Levenberg-Marquardt y Scaled Conjugate Gradient, respectivamente (Flores, Franco, Andrade, & Asenjo, 2021, p. 21).

1.1.2. Tipologías de Redes y Sus Aplicaciones

Las Redes Neuronales Artificiales abarcan un conjunto diverso de arquitecturas, cada una diseñada para abordar problemas específicos. Las principales tipologías incluyen:

- **Perceptrón Multicapa (MLP):** Considerado el modelo más básico de RNA, el MLP consta de una o más capas ocultas que procesan información mediante funciones de activación no lineales. Este modelo es ideal para problemas de clasificación y regresión, y se utiliza en aplicaciones como el reconocimiento de patrones, el análisis financiero y el diagnóstico médico.
- **Redes Convolucionales (CNN):** Diseñadas para procesar datos con estructura espacial, como imágenes y videos, las CNN son esenciales en tareas de visión por computadora. Sus capas convolucionales extraen características relevantes, como bordes y texturas, permitiendo aplicaciones como el reconocimiento facial, la detección de objetos y el análisis de imágenes satelitales.
- **Redes Recurrentes (RNN):** Estas redes son óptimas para manejar datos secuenciales, como texto, audio y series temporales. Las variantes más avanzadas, como LSTM (Long Short-Term Memory) y GRU (Gated Recurrent Unit), superan las limitaciones de memoria de las RNN tradicionales, permitiendo el modelado de dependencias a largo plazo. Aplicaciones como la traducción automática, el análisis del lenguaje natural y la predicción de tendencias financieras dependen en gran medida de estas redes.
- **Redes Generativas Adversarias (GAN):** Este enfoque innovador enfrenta dos redes (una generadora y una discriminadora) en un proceso competitivo. Las GAN han revolucionado campos como la generación de imágenes y videos realistas, la creación de contenido artístico y la mejora de datos de entrenamiento mediante la generación de ejemplos sintéticos.

Las aplicaciones de estas redes son tan variadas como sus arquitecturas, abarcando desde la conducción autónoma hasta la personalización de experiencias de usuario. En el contexto educativo, las RNA se utilizan para personalizar el aprendizaje, predecir el rendimiento académico y optimizar estrategias pedagógicas.

1.2. El Rendimiento Académico en el Foco del Análisis

El rendimiento académico es un indicador integral que refleja no solo el grado de aprendizaje alcanzado por los estudiantes, sino también la interacción de factores individuales, sociales, institucionales y contextuales. Este fenómeno multifacético constituye un eje

fundamental en los sistemas educativos, ya que su análisis permite identificar áreas de mejora y diseñar estrategias que potencien el éxito estudiantil. A continuación, se detallan los principales factores que determinan el rendimiento académico y las perspectivas que influyen en su desarrollo.

1.2.1. Factores Determinantes en el Éxito Académico

Factores Individuales

A nivel individual, el rendimiento académico se ve directamente afectado por una combinación de aspectos personales que incluyen la motivación, las habilidades cognitivas y la capacidad de autorregulación.

1. Motivación Intrínseca y Extrínseca:

- La motivación intrínseca, definida como el interés genuino por aprender y el placer que deriva del conocimiento, es esencial para mantener un compromiso sostenido con las tareas académicas. Esta forma de motivación impulsa a los estudiantes a superar desafíos por la satisfacción personal que obtienen al lograr sus metas.
- La motivación extrínseca, aunque menos sostenible a largo plazo, también juega un papel importante. Factores como recompensas externas, reconocimiento social y expectativas de familiares o profesores pueden influir significativamente en el esfuerzo que los estudiantes invierten en su aprendizaje.

2. Habilidades Cognitivas:

- Las habilidades cognitivas, como la memoria, la atención y el razonamiento lógico, son fundamentales para el procesamiento y aplicación del conocimiento. Por ejemplo, una buena memoria permite a los estudiantes retener conceptos clave, mientras que la atención sostenida facilita la comprensión profunda de temas complejos.
- El pensamiento crítico y la creatividad son otras habilidades esenciales. Los estudiantes que desarrollan estas capacidades tienen mayores probabilidades

de resolver problemas de manera innovadora y de adaptarse a situaciones de aprendizaje dinámicas.

3. Autorregulación del Aprendizaje:

- La autorregulación incluye habilidades como la planificación, el establecimiento de metas, la automonitorización y la autoevaluación. Los estudiantes autorregulados son capaces de identificar sus fortalezas y debilidades, ajustando sus estrategias de estudio en función de sus necesidades.
- Esta capacidad se ve potenciada por el desarrollo de la resiliencia, que permite a los estudiantes enfrentar fracasos o retrocesos académicos sin perder la motivación.

Factores Institucionales

El entorno educativo y las características de las instituciones académicas también tienen un impacto significativo en el rendimiento de los estudiantes.

1. Calidad del Profesorado:

- Los docentes no solo transmiten conocimientos, sino que también actúan como mentores, motivadores y facilitadores del aprendizaje. Su formación académica, experiencia y habilidades pedagógicas son determinantes en la calidad de la educación que los estudiantes reciben.
- Un enfoque en la formación continua de los docentes y en la actualización de sus prácticas pedagógicas contribuye a la creación de un entorno de aprendizaje enriquecedor.

2. Infraestructura y Recursos Educativos:

- La disponibilidad de recursos físicos, tecnológicos y didácticos, como bibliotecas, laboratorios, acceso a Internet y plataformas digitales, amplía las oportunidades de aprendizaje. Estos recursos son particularmente importantes en contextos donde la tecnología puede facilitar el acceso a materiales y metodologías de enseñanza innovadoras.

3. Clima Escolar:

- Un ambiente escolar positivo, caracterizado por relaciones interpersonales saludables, respeto mutuo y una cultura de colaboración, fomenta un sentido de pertenencia en los estudiantes. Esto, a su vez, influye en su motivación y compromiso con el aprendizaje.

4. Tecnología y Metodologías Innovadoras:

- La incorporación de herramientas tecnológicas, como aplicaciones de aprendizaje, realidad aumentada y simulaciones interactivas, ha transformado la manera en que los estudiantes adquieren conocimientos.
- Metodologías como el aprendizaje basado en proyectos, el flipped classroom (clase invertida) y la gamificación no solo mejoran el rendimiento académico, sino que también promueven habilidades del siglo XXI, como la colaboración y la resolución de problemas complejos.

Factores Contextuales

1. Impacto de las Políticas Educativas:

- Las políticas educativas que priorizan la equidad, la inclusión y el acceso a una educación de calidad desempeñan un papel crucial en el rendimiento académico. Por ejemplo, programas de becas, tutorías y apoyo psicológico pueden reducir las barreras que enfrentan estudiantes en situaciones de vulnerabilidad.

2. Influencia del Entorno Socioeconómico:

- Factores como el nivel de ingresos familiares, la estabilidad del hogar y el acceso a recursos culturales y educativos impactan directamente en el rendimiento de los estudiantes. La desigualdad socioeconómica sigue siendo uno de los mayores desafíos para garantizar la equidad educativa.

1.2.2. Perspectivas Psicológicas y Sociales

Perspectivas Psicológicas

Desde la psicología, diversas teorías explican cómo los procesos internos de los estudiantes influyen en su rendimiento académico:

1. Teoría de la Autodeterminación:

- Según esta teoría, el aprendizaje óptimo ocurre cuando los estudiantes sienten que tienen autonomía para tomar decisiones sobre su educación, se perciben competentes en sus tareas y experimentan conexiones sociales significativas con sus pares y docentes.

2. Inteligencia Emocional:

- La capacidad de gestionar las emociones propias y comprender las de los demás es crucial para enfrentar el estrés académico. Los estudiantes con alta inteligencia emocional son más capaces de mantener la motivación, establecer relaciones positivas y superar conflictos.

3. Estrategias de Afrontamiento:

- Las estrategias utilizadas por los estudiantes para manejar la ansiedad, la presión de los exámenes y las demandas académicas influyen directamente en su desempeño. Estas estrategias pueden incluir técnicas de relajación, planificación anticipada y búsqueda de apoyo social.

Perspectivas Sociales

El entorno social en el que los estudiantes se desarrollan también desempeña un papel esencial:

1. Apoyo Familiar:

- Las familias que valoran la educación y participan activamente en el aprendizaje de sus hijos crean un entorno propicio para el éxito académico. Esto incluye desde el apoyo emocional hasta la provisión de recursos materiales y logísticos.

2. Redes de Aprendizaje y Compañerismo:

- Las interacciones con compañeros en grupos de estudio o comunidades de aprendizaje promueven el intercambio de conocimientos y el fortalecimiento del sentido de pertenencia.

3. Influencia de la Comunidad:

- El contexto cultural y social de la comunidad, incluidas las expectativas sobre la educación y el papel de la escuela, moldea las actitudes y valores de los estudiantes hacia el aprendizaje.

El rendimiento académico es un fenómeno complejo e interrelacionado que exige un abordaje integral. La interacción entre factores individuales, institucionales, psicológicos y sociales destaca la necesidad de intervenciones educativas holísticas que consideren las características únicas de cada estudiante. En un mundo en constante cambio, la capacidad de adaptarse, autorregularse y colaborar será clave para garantizar el éxito académico y, por ende, el desarrollo personal y profesional de los estudiantes.

1.3. Modelos Predictivos en la Educación

La incorporación de modelos predictivos en el ámbito educativo ha transformado la manera en que se analiza y mejora el rendimiento académico. Estas herramientas permiten no solo prever resultados, sino también diseñar estrategias personalizadas para optimizar el aprendizaje y apoyar a los estudiantes en su trayectoria educativa.

1.3.1. Inteligencia Artificial vs. Métodos Tradicionales

Métodos Tradicionales de Predicción

Los métodos tradicionales para predecir el rendimiento académico se basan principalmente en técnicas estadísticas, como la regresión lineal, los análisis multivariados y las correlaciones. Estas metodologías ofrecen ventajas como la simplicidad, la interpretabilidad y la capacidad de identificar relaciones básicas entre variables como el promedio de calificaciones, la asistencia a clases, el nivel socioeconómico y las características demográficas.

Sin embargo, presentan limitaciones significativas:

- **Linealidad de los modelos:** Los métodos estadísticos tradicionales asumen relaciones lineales entre variables, lo que los hace poco adecuados para capturar la complejidad inherente a los datos educativos.
- **Capacidad predictiva limitada:** Su precisión disminuye en contextos donde las interacciones entre variables son complejas o no evidentes.
- **Falta de personalización:** Dificultan el diseño de intervenciones específicas basadas en las necesidades individuales de los estudiantes.

Avances con la Inteligencia Artificial

La inteligencia artificial (IA) ha revolucionado el análisis predictivo en la educación, especialmente a través del uso de **redes neuronales artificiales (RNA)** y otros algoritmos avanzados de aprendizaje automático. Estos enfoques permiten identificar patrones complejos y no lineales en grandes volúmenes de datos, lo que mejora significativamente la precisión y utilidad de las predicciones.

1. Flexibilidad y Adaptabilidad:

- Las RNA pueden procesar datos heterogéneos, como registros académicos, interacciones en plataformas digitales de aprendizaje y comportamientos en línea, integrando múltiples fuentes de información para generar predicciones robustas.

2. Predicciones Personalizadas:

- A diferencia de los modelos tradicionales, las RNA ofrecen predicciones adaptadas a las características individuales de cada estudiante, permitiendo diseñar intervenciones personalizadas para mejorar su desempeño.

3. Identificación de Factores Ocultos:

- Las técnicas basadas en IA pueden descubrir relaciones no evidentes entre variables, como el impacto de la motivación, la participación en actividades

extracurriculares o incluso patrones de navegación en plataformas de aprendizaje digital.

1.3.2. Beneficios y Desafíos de las Redes Neuronales

Beneficios de las Redes Neuronales Artificiales (RNA)

1. Personalización del Aprendizaje:

- Las RNA permiten adaptar los contenidos y estrategias pedagógicas a las necesidades de cada estudiante. Por ejemplo, pueden recomendar recursos específicos, ajustar la dificultad de los ejercicios o identificar áreas de mejora antes de que el estudiante enfrente dificultades significativas.

2. Optimización de Recursos Educativos:

- Al identificar a los estudiantes con mayor riesgo de bajo rendimiento, las instituciones pueden dirigir sus recursos, como tutorías o programas de mentoría, hacia quienes más lo necesitan.

3. Intervenciones Tempranas:

- Las RNA detectan señales tempranas de abandono escolar o problemas de aprendizaje, permitiendo implementar medidas preventivas de manera oportuna.

4. Análisis en Tiempo Real:

- Estas herramientas procesan datos en tiempo real, proporcionando a docentes y administradores información inmediata para la toma de decisiones.

Desafíos de las Redes Neuronales Artificiales

1. Interpretabilidad de los Modelos:

- Las RNA son frecuentemente consideradas "cajas negras", ya que sus predicciones son el resultado de procesos internos difíciles de interpretar. Esto puede dificultar la justificación de sus recomendaciones ante docentes, estudiantes y padres.

2. Privacidad y Protección de Datos:

- El uso de grandes volúmenes de datos plantea preocupaciones éticas sobre la privacidad de los estudiantes. Es fundamental garantizar que los datos recopilados se almacenen y procesen de manera segura, cumpliendo con normativas como el GDPR o leyes locales de protección de datos.

3. Sesgos Algorítmicos:

- Si los datos utilizados para entrenar las RNA contienen sesgos, estos pueden amplificarse en las predicciones. Por ejemplo, un modelo podría perpetuar desigualdades si no se diseñan mecanismos para identificar y corregir estos sesgos.

4. Requerimientos Computacionales:

- Las RNA requieren grandes cantidades de datos y recursos computacionales avanzados, lo que puede ser un desafío para instituciones educativas con infraestructura limitada.

5. Aceptación por Parte de los Actores Educativos:

- La integración de IA en el ámbito educativo a menudo enfrenta resistencia debido al desconocimiento o a la desconfianza en estas tecnologías. Es crucial acompañar su implementación con programas de formación para docentes y administradores.

1.3.3. Aplicaciones Prácticas en la Educación

Las RNA y otros enfoques basados en IA ya están transformando diversos aspectos del sistema educativo:

1. Sistemas de Tutoría Inteligente:

- Plataformas como Khan Academy y Duolingo utilizan IA para ofrecer tutorías adaptativas que responden al ritmo y nivel de cada estudiante.

2. Predicción del Abandono Escolar:

- Instituciones educativas emplean modelos predictivos para identificar estudiantes en riesgo de abandonar sus estudios y diseñar intervenciones específicas.

3. Optimización de Planes de Estudio:

- Las RNA analizan datos de desempeño para sugerir modificaciones en los planes de estudio que respondan mejor a las necesidades y expectativas de los estudiantes.

4. Evaluación Automatizada:

- Herramientas basadas en IA permiten evaluar trabajos escritos, exámenes y hasta proyectos complejos, ofreciendo retroalimentación inmediata y reduciendo la carga de trabajo para los docentes.

La transición de métodos tradicionales a enfoques basados en inteligencia artificial, como las redes neuronales artificiales, representa un cambio paradigmático en el análisis y mejora del rendimiento académico. Aunque estos avances ofrecen oportunidades sin precedentes para personalizar la educación y optimizar recursos, también plantean desafíos que requieren un enfoque ético y responsable. El futuro de la educación dependerá de la capacidad de integrar estas tecnologías de manera inclusiva, transparente y equitativa, asegurando que los beneficios alcancen a todos los estudiantes por igual.

1.4. Glosario de Términos Clave

1.4.1. Algoritmo de Aprendizaje

Es un procedimiento matemático y computacional que utilizan las redes neuronales artificiales (RNA) para ajustar los **pesos** y **sesgos** de las conexiones entre neuronas. Su objetivo es minimizar los errores en las predicciones del modelo al optimizar la función de pérdida o costo, que mide la discrepancia entre las salidas predichas y los valores reales.

- Los algoritmos más comunes incluyen el **descenso de gradiente** y sus variantes como **Adam**, **RMSprop**, y **descenso de gradiente estocástico** (SGD), que ofrecen diferentes enfoques para acelerar y estabilizar el proceso de aprendizaje.

- Estos algoritmos son iterativos y adaptan los pesos mediante un proceso que considera la derivada parcial de la función de pérdida respecto a cada peso, lo que permite identificar la dirección más eficiente para reducir el error.

1.4.2. Capa Oculta

En una red neuronal, las capas ocultas son los niveles intermedios situados entre la capa de entrada (que recibe los datos iniciales) y la capa de salida (que genera los resultados predichos). Estas capas no están directamente visibles en los datos de entrada ni en las salidas del modelo, pero son esenciales para la capacidad de aprendizaje de la red.

- Cada capa oculta está formada por **neuronas artificiales** que reciben señales de la capa anterior, aplican una función de activación no lineal (como **ReLU**, **sigmoide** o **tangente hiperbólica**) y transmiten el resultado a la siguiente capa.
- **Importancia:** Las capas ocultas permiten que la red neuronal capture patrones complejos, relaciones no lineales y características abstractas en los datos. En redes profundas (deep learning), múltiples capas ocultas son fundamentales para abordar problemas como el reconocimiento de imágenes y el procesamiento del lenguaje natural.

1.4.3. Overfitting

El overfitting, o sobreajuste, ocurre cuando un modelo de aprendizaje se adapta excesivamente a los datos de entrenamiento, aprendiendo no solo los patrones significativos sino también el ruido y las irregularidades específicas de esos datos. Esto provoca que el modelo tenga un alto desempeño en los datos de entrenamiento, pero un bajo rendimiento en datos nuevos o no vistos.

- **Indicadores:** El overfitting se detecta cuando el error en el conjunto de entrenamiento es significativamente menor que el error en el conjunto de validación o prueba.

Una RNA que analiza datos académicos podría identificar correlaciones irrelevantes, como un vínculo espurio entre el horario de estudio y las calificaciones, si estas coincidencias están presentes solo en los datos de entrenamiento.

- **Soluciones comunes:**
 - **Regularización:** Penalizar los pesos excesivamente grandes (ver abajo).
 - **Aumento de datos:** Crear más datos de entrenamiento mediante técnicas como la rotación o el escalado en imágenes.
 - **Validación cruzada:** Dividir los datos en varios subconjuntos para entrenar y evaluar el modelo en diferentes combinaciones.

1.4.4. Regularización

La regularización es un conjunto de técnicas diseñadas para mejorar la capacidad de generalización de un modelo, reduciendo su tendencia a sobreajustarse a los datos de entrenamiento. Estas técnicas agregan términos de penalización a la función de pérdida del modelo, lo que desalienta soluciones que sean demasiado complejas o dependientes de características específicas de los datos de entrenamiento.

- **Tipos principales:**
 - **L1 Regularization (Lasso):** Penaliza la suma absoluta de los pesos, promoviendo la eliminación de aquellos menos relevantes, lo que también ayuda a simplificar el modelo.
 - **L2 Regularization (Ridge):** Penaliza la suma de los cuadrados de los pesos, fomentando que todos los pesos sean pequeños y uniformes.
 - **Dropout:** Apaga aleatoriamente un porcentaje de neuronas en cada capa durante el entrenamiento, evitando que el modelo dependa excesivamente de ciertas rutas de activación.
 - **Early Stopping:** Detiene el entrenamiento del modelo tan pronto como su desempeño en el conjunto de validación comienza a deteriorarse, previniendo el ajuste excesivo a los datos de entrenamiento.
- **Beneficios:** La regularización permite que el modelo sea más robusto y tenga un mejor desempeño en datos no vistos.

1.4.5. Sesgo Algorítmico

El sesgo algorítmico se refiere a la tendencia de un modelo predictivo a producir resultados que favorecen o discriminan ciertos grupos o resultados, debido a patrones preexistentes en los datos de entrenamiento o a decisiones tomadas durante el diseño del modelo. Este fenómeno es especialmente preocupante en contextos como la educación, donde podría afectar la equidad y la justicia en las predicciones.

- **Causas del sesgo:**
 - **Datos desequilibrados:** Si los datos de entrenamiento no representan adecuadamente a todos los grupos de una población (por ejemplo, estudiantes de diferentes contextos socioeconómicos), el modelo tiende a favorecer al grupo predominante.
 - **Selección de características:** Elegir variables que estén correlacionadas con sesgos históricos o estructurales (como ubicación geográfica o ingresos) puede perpetuar desigualdades.
 - **Diseño del modelo:** Las decisiones sobre la estructura y los parámetros del modelo también pueden introducir sesgos inadvertidos.

Un modelo que predice el desempeño académico basándose en datos de asistencia podría penalizar injustamente a estudiantes que faltan por razones justificadas, como problemas de salud o trabajo.

- **Soluciones:**
 - Implementar técnicas de detección de sesgos, como el análisis de paridad demográfica.
 - Usar datos balanceados y representativos.
 - Aplicar algoritmos que incluyan medidas de equidad en su diseño.

1.4.6. Algoritmo

Un algoritmo es una secuencia estructurada de instrucciones que se ejecuta para resolver un problema o completar una tarea de manera eficiente y automatizada. Según Laudon y

Laudon (2012), los algoritmos dependen de la capacidad de los procesadores, el software y los datos almacenados para llevar a cabo un procesamiento de información rápido y preciso (p. 461).

Características principales:

- **Precisión:** Cada paso está claramente definido para evitar ambigüedades.
- **Eficiencia:** Diseñado para minimizar recursos y tiempo de ejecución.
- **Automatización:** Puede ejecutarse sin intervención humana, utilizando computadoras.

Un algoritmo de búsqueda, como el de Google, organiza y recupera información relevante a partir de millones de páginas web en cuestión de segundos.

Importancia:

En inteligencia artificial y machine learning, los algoritmos son la base para entrenar modelos predictivos, permitiendo a las máquinas aprender y adaptarse a partir de datos.

1.4.7. Aprendizaje Supervisado

El aprendizaje supervisado es un enfoque del machine learning en el que se entrena un modelo utilizando un conjunto de datos etiquetados, es decir, datos donde las entradas están asociadas a salidas conocidas. Según Garreta y Moncecchi (2013), este enfoque permite "estimar el valor de salida para nuevos datos con solo conocer sus características" (p. 20).

- Predicción de calificaciones estudiantiles basada en factores como asistencia, participación en clase y rendimiento anterior.

Métodos comunes:

- **Clasificación:** Asigna categorías predefinidas a los datos (e.g., aprobar o reprobar).
- **Regresión:** Predice valores continuos, como el promedio final de un curso.

Aplicaciones:

Es ampliamente utilizado en áreas como reconocimiento facial, detección de fraudes y predicción del desempeño académico.

1.4.8. Aprendizaje No Supervisado

El aprendizaje no supervisado se centra en analizar datos no etiquetados, identificando patrones o estructuras subyacentes. Según Abbott (2014), este enfoque, también llamado "agrupamiento", clasifica los datos en grupos basándose en la similitud de sus características (p. 5).

- Segmentación de estudiantes en grupos según sus estilos de aprendizaje o preferencias académicas.

Métodos comunes:

- **Clustering:** Agrupa datos similares (e.g., K-means).
- **Análisis de componentes principales (PCA):** Reduce la dimensionalidad de los datos para facilitar su interpretación.

Aplicaciones:

Se utiliza en marketing (segmentación de clientes), biología (clasificación de especies) y educación (identificación de patrones en el comportamiento de aprendizaje).

1.4.9. Dato

Un dato es una unidad de información en su forma más básica, que no ha sido procesada ni organizada para su interpretación. Según Laudon y Laudon (2012), los datos representan información bruta que necesita ser estructurada para ser útil (p. 15).

Características:

- **No procesado:** Carece de contexto o significado claro.
- **Forma básica:** Puede ser un número, texto, imagen, etc.

Un registro de asistencia diaria de un estudiante es un dato que, al ser analizado, podría convertirse en información sobre su compromiso académico.

Importancia:

Los datos son la materia prima para el análisis y la base del aprendizaje automático, ya que permiten entrenar modelos predictivos.

1.4.9. Dataset

Un dataset es un conjunto organizado de datos que se utiliza para análisis o entrenamiento de modelos en machine learning. Según Géron (2019), los datasets suelen estar compuestos por registros estructurados con características comunes.

- Un dataset educativo puede incluir registros de estudiantes con columnas como nombre, edad, calificaciones y asistencia.

Tipos:

- **Datos de entrenamiento:** Usados para ajustar el modelo.
- **Datos de prueba:** Evaluados para medir el desempeño del modelo.

Importancia:

En machine learning, los datasets bien diseñados son esenciales para entrenar y validar modelos, asegurando predicciones precisas y generalizables.

1.4.10. Evaluación del Rendimiento Académico

La evaluación del rendimiento académico clasifica a los estudiantes en función de si cumplen con los estándares establecidos por la institución educativa. Según García y Fernández (2012), este proceso categoriza a los estudiantes como aprobados o desaprobados según su desempeño.

Métodos de evaluación:

- **Cuantitativos:** Exámenes, promedios de calificaciones.
- **Cualitativos:** Observaciones, retroalimentación de los docentes.

Importancia:

Permite identificar fortalezas y áreas de mejora en los estudiantes, además de servir como base para ajustes en los programas educativos.

1.4.11. Información

La información es el resultado del procesamiento de datos, presentada de manera clara y útil para el análisis. Según Laudon y Laudon (2012), la información es un producto derivado de los datos que constituye un paso esencial en la toma de decisiones (p. 15).

- Los datos de asistencia diaria pueden procesarse para generar un informe sobre el porcentaje de asistencia de un estudiante durante un trimestre.

Importancia:

La información procesada es clave para el aprendizaje automático, ya que permite transformar datos en conocimiento accionable.

1.4.12. Machine Learning (ML)

El machine learning es una rama de la inteligencia artificial que permite a las computadoras aprender de los datos para realizar predicciones o clasificaciones. Según Laudon y Laudon (2012), utiliza algoritmos y métodos estadísticos para identificar patrones y predecir eventos futuros (p. 438).

Aplicaciones:

- **Predicción de desempeño académico:** Identificar estudiantes en riesgo de bajo rendimiento.
- **Personalización educativa:** Adaptar contenidos al nivel y estilo de aprendizaje de cada estudiante.

Beneficios:

- **Automatización:** Reduce la necesidad de intervención humana.
- **Escalabilidad:** Puede analizar grandes cantidades de datos.

1.4.13. Patrón o Tendencia

Un patrón o tendencia es una secuencia recurrente de acciones o comportamientos observables. Según Cadenas (2015), los patrones pueden incluir aspectos visuales, auditivos o gestuales que se analizan para identificar regularidades.

- Un patrón de calificaciones decrecientes podría indicar la necesidad de intervención en un estudiante.

Aplicaciones:

La identificación de patrones es esencial en machine learning para tareas como la detección de anomalías o la predicción de eventos futuros.

1.4.14. Rendimiento Académico

El rendimiento académico es la evaluación de la calidad de los resultados obtenidos por un estudiante en un contexto educativo. Según López y Pérez (2009), refleja el grado de cumplimiento de los objetivos de aprendizaje establecidos.

Factores que lo afectan:

- **Individuales:** Motivación, habilidades cognitivas.
- **Institucionales:** Calidad docente, recursos disponibles.

Importancia:

Es un indicador clave para medir la efectividad de los programas educativos y diseñar estrategias de mejora.

1.4.15. Tabla

Una tabla es una estructura de datos organizada en filas y columnas. Según Laudon y Laudon (2012), las columnas representan atributos o características, mientras que las filas contienen registros individuales (p. 214).

Una tabla de estudiantes puede incluir columnas como nombre, edad, promedio y asistencia.

Importancia:

Las tablas son esenciales en bases de datos relacionales y en machine learning, ya que organizan los datos para su análisis y procesamiento.

Este glosario proporciona una base sólida para comprender los conceptos fundamentales en el análisis predictivo educativo, destacando tanto los beneficios como los desafíos de utilizar redes neuronales artificiales y otros modelos avanzados.

1.5. Bases Teóricas

Machine Learning

El machine learning (aprendizaje automático) es una disciplina de la inteligencia artificial que permite a los sistemas aprender y mejorar su desempeño en tareas específicas mediante el análisis de datos, sin necesidad de programación explícita. Este enfoque se basa en la capacidad de identificar patrones y estructuras en los datos para generar modelos predictivos o descriptivos, adaptándose a cambios y nuevos datos de forma autónoma.

Neurona Biológica

Según Haykin (2009), el cerebro humano es una estructura compleja que logra un equilibrio eficiente entre la velocidad de operación de las neuronas y la gran cantidad de interconexiones entre ellas. A pesar de su relativa lentitud en comparación con la tecnología moderna, las neuronas biológicas tienen un diseño altamente optimizado para el procesamiento de información:

- **Velocidad:** Los eventos neuronales ocurren en milisegundos, mientras que las puertas lógicas de los chips de silicio pueden operar en nanosegundos.
- **Interconexiones:** El cerebro compensa la menor velocidad individual de las neuronas con una enorme red de interconexiones que permite el procesamiento paralelo masivo.

Este modelo biológico ha inspirado el desarrollo de las Redes Neuronales Artificiales (RNA), que buscan replicar, de forma simplificada, las capacidades de aprendizaje y procesamiento del cerebro humano (Haykin, 2009, p. 6).

Neurona Artificial

Las Redes Neuronales Artificiales (RNA) son sistemas computacionales diseñados para imitar el funcionamiento del sistema nervioso de los seres vivos. Según Ponce Cruz (2010), las RNA están formadas por múltiples unidades de procesamiento, llamadas **neuronas artificiales**, que trabajan juntas para procesar y transformar la información (p. 199).

Estructura de una Neurona Artificial:

1. **Conexiones con Pesos:** Cada neurona recibe señales de otras neuronas o fuentes externas. Estas señales son ponderadas por valores numéricos llamados pesos, que determinan la importancia relativa de cada entrada.
2. **Procesamiento:** La neurona procesa las entradas mediante una función matemática (normalmente no lineal) que genera una salida escalar.
3. **Propagación de la Salida:** La salida de una neurona se transmite como entrada a otras neuronas conectadas, formando una red.

Disposición en Capas:

- **Capa de Entrada:** Recibe los datos iniciales para procesar.
- **Capas Ocultas:** Procesan las entradas mediante transformaciones intermedias, capturando patrones complejos y no lineales.
- **Capa de Salida:** Genera el resultado final del modelo, como una predicción o clasificación.

Funciones de las RNA:

- Recopilar información de múltiples fuentes.
- Procesar esa información para extraer patrones significativos.
- Emitir resultados que puedan ser utilizados para la toma de decisiones o para alimentar otras neuronas (Ponce Cruz, 2010, p. 199).

Las RNA tienen aplicaciones en una amplia variedad de campos, incluyendo la educación, donde pueden usarse para predecir el rendimiento académico y personalizar estrategias de aprendizaje.

Rendimiento Académico

El rendimiento académico es un concepto que evalúa la capacidad de un estudiante para cumplir con las expectativas y requisitos de un programa o curso educativo. Según Huerta (2005), se trata de un indicador clave de la calidad tanto de la enseñanza como del aprendizaje, que puede medirse a través de factores como la asistencia, las calificaciones y los resultados en exámenes.

Factores que Influyen en el Rendimiento Académico:

1. **Participación y Asistencia:** Salazar y Palomino (2011) destacan que la participación activa y la regularidad en la asistencia a clase son fundamentales para un buen desempeño académico.
2. **Capacidad de Completar Tareas y Proyectos:** Los estudiantes que gestionan eficazmente sus responsabilidades muestran un mayor nivel de éxito académico.
3. **Evaluaciones y Exámenes:** El desempeño en pruebas y evaluaciones proporciona una medida cuantitativa del aprendizaje alcanzado.

Rendimiento Académico en la Ingeniería:

En campos especializados como la ingeniería, el rendimiento académico adquiere un matiz particular. Según García (2018), en matemáticas, un rendimiento académico alto indica la capacidad del estudiante para comprender y aplicar conceptos fundamentales a problemas prácticos. Este aspecto es crucial para evaluar la preparación del estudiante para su futura profesión.

- **Indicadores clave:** Realización de tareas complejas, desempeño en exámenes específicos y capacidad de aplicar el conocimiento matemático en contextos de ingeniería.

- **Importancia:** Un buen rendimiento en matemáticas puede ser un predictor del éxito profesional en carreras técnicas.

Las bases teóricas de machine learning, las redes neuronales y el rendimiento académico proporcionan un marco conceptual para comprender cómo los modelos inspirados en el cerebro humano pueden analizar y mejorar el desempeño educativo. Estos conceptos son esenciales para el desarrollo de sistemas predictivos que transformen la manera en que los estudiantes aprenden y los docentes diseñan estrategias pedagógicas.

CAPÍTULO II

DISEÑO DEL ESTUDIO Y ENFOQUE METODOLÓGICO

2.1. Estructura del Modelo de Investigación

2.1.1. Diseño No Experimental y Enfoque Cuantitativo

Este estudio adopta un enfoque **cuantitativo**, que implica la recolección y análisis de datos numéricos para identificar patrones y evaluar teorías, siguiendo la metodología propuesta por Hernández, Fernández y Baptista (2014). Este enfoque permite obtener resultados objetivos y precisos mediante el uso de herramientas estadísticas y algoritmos predictivos.

- **Tipo de Investigación:** Aplicada. Busca resolver un problema específico: predecir el desempeño de los estudiantes en el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana.
- **Nivel de Investigación:** Predictivo. El objetivo principal es prever si los estudiantes aprobarán o no el curso utilizando un modelo basado en Redes Neuronales Artificiales (RNA).
- **Justificación del Diseño No Experimental:**
 - No se manipulan directamente las variables independientes.
 - Los datos se obtienen de forma observacional, utilizando registros académicos existentes.
 - Esto permite un análisis más realista del fenómeno educativo, preservando la integridad de los datos originales.

2.1.2. Variables Principales del Estudio

1. Variable Dependiente:

- **Condición Académica del Estudiante (Aprobado o No Aprobado):**
Indicador binario que refleja el estado final del estudiante en el curso, determinado por las calificaciones obtenidas.

2. Variables Independientes:

- Factores asociados al rendimiento académico del estudiante, tales como:
 - **Asistencia:** Porcentaje de clases asistidas.
 - **Participación en actividades académicas:** Incluye trabajos, exámenes parciales y tareas.
 - **Resultados previos:** Promedios de calificaciones en cursos relacionados.
 - **Demográficos:** Edad, género, y situación socioeconómica.

Estas variables se seleccionaron por su relevancia en estudios previos sobre predicción de rendimiento académico en contextos similares.

Diseño del Modelo de Redes Neuronales Artificiales (RNA)

Para desarrollar el modelo de investigación, se utilizó un **algoritmo de Redes Neuronales Artificiales** (RNA) siguiendo la metodología descrita por Kaastra y Boyd (1996), que consta de las siguientes **ocho etapas**:

1. Selección de la Variable a Estudiar

La variable principal es la **condición académica del estudiante (Aprobado o No Aprobado)**. Esta variable se clasifica de manera binaria, lo que facilita el uso de un modelo de clasificación en las Redes Neuronales.

2. Recopilación de Datos

Se recopilaron datos históricos de estudiantes matriculados en el curso de Álgebra Lineal, incluyendo registros de asistencia, calificaciones, participación en actividades académicas y características demográficas.

- **Fuente:** Base de datos institucional de la universidad.
- **Periodo:** Se consideraron datos de los últimos cinco ciclos académicos para garantizar un conjunto de datos representativo y diverso.

3. Preprocesamiento de Datos

Se aplicaron técnicas de limpieza y normalización para garantizar la calidad de los datos:

- **Tratamiento de datos faltantes:** Se imputaron valores para registros incompletos.
- **Normalización:** Todas las variables se escalaron a un rango uniforme para mejorar la estabilidad del modelo durante el entrenamiento.
- **Codificación:** Variables categóricas (e.g., género) se transformaron en representaciones numéricas mediante codificación one-hot.

4. Definición de los Conjuntos de Datos

Los datos se dividieron en tres subconjuntos:

- **Conjunto de Entrenamiento (70%):** Utilizado para ajustar los pesos del modelo.
- **Conjunto de Validación (15%):** Evaluado para ajustar hiperparámetros y evitar el overfitting.
- **Conjunto de Prueba (15%):** Usado exclusivamente para medir el desempeño final del modelo.

5. Selección de la Arquitectura de las Redes Neuronales

Se diseñó una arquitectura de RNA basada en las necesidades del problema:

- **Capa de Entrada:**
 - Neuronas equivalentes al número de variables independientes seleccionadas (e.g., asistencia, calificaciones, etc.).
- **Capas Ocultas:**
 - Una o dos capas ocultas con un número variable de neuronas, determinadas mediante pruebas experimentales.
 - Función de activación **ReLU** para manejar relaciones no lineales.
- **Capa de Salida:**
 - Una neurona con función de activación **sigmoide** para generar una probabilidad de aprobación (valor entre 0 y 1).

6. Definición de los Criterios de Evaluación

Se seleccionaron métricas estándar para medir la precisión y la eficacia del modelo:

- **Exactitud (Accuracy):** Porcentaje de predicciones correctas.
- **Precisión y Recall:** Evaluación de falsos positivos y negativos.
- **Curva ROC y AUC:** Para evaluar el desempeño del modelo en términos de sensibilidad y especificidad.

7. Entrenamiento de la Red Neuronal

Durante el entrenamiento, se optimizó la red utilizando:

- **Algoritmo de Descenso de Gradiente Estocástico (SGD):** Con tasa de aprendizaje adaptativa.
- **Tasa de Aprendizaje y Momentum:** Ajustados para garantizar una convergencia rápida y estable.
- **Iteraciones (Épocas):** Entrenamiento repetido hasta alcanzar un rendimiento óptimo en el conjunto de validación.

8. Implementación del Modelo

El modelo final se implementó en una plataforma computacional adecuada, como Python con bibliotecas especializadas (e.g., TensorFlow o PyTorch). Se integró con un sistema de predicción para probar su funcionalidad en tiempo real y analizar su viabilidad práctica en contextos educativos.

Resultados Esperados

Se espera que el modelo predictivo basado en Redes Neuronales Artificiales logre:

- **Alta precisión:** Generar predicciones confiables sobre el rendimiento académico de los estudiantes.
- **Detección temprana:** Identificar a los estudiantes en riesgo de desaprobación, permitiendo implementar estrategias de intervención específicas.
- **Personalización:** Ofrecer recomendaciones individualizadas para mejorar el desempeño académico.

El enfoque y la estructura del modelo reflejan un diseño integral orientado a la mejora del aprendizaje mediante la innovación tecnológica.

2.2. Marco Conceptual y Planteamiento de Hipótesis

El marco conceptual establece las bases teóricas que sustentan la investigación, definiendo las relaciones entre las variables y su impacto en los resultados. Además, formula las hipótesis que guían el estudio, planteando su verificación a través de métodos cuantitativos y predictivos.

2.2.1. Relación entre Variables y Resultados

El objetivo principal del estudio es analizar cómo la implementación de algoritmos de Machine Learning, específicamente Redes Neuronales Artificiales (RNA), influye en la precisión de la predicción del rendimiento académico. A continuación, se describen las principales variables y sus interacciones:

Variables Principales:

1. Variable Independiente:

- **Algoritmo de Redes Neuronales Artificiales (RNA):** Se refiere al modelo implementado para procesar y analizar los datos académicos de los estudiantes, con el objetivo de generar predicciones sobre su desempeño.

2. Variable Dependiente:

- **Condición Académica de los Estudiantes:** Indica si los estudiantes aprueban o no el curso de Álgebra Lineal. Esta variable es binaria (Aprobado/No Aprobado) y se mide en términos de precisión, exhaustividad y exactitud de las predicciones.

Relaciones Esperadas:

- Se espera que el uso del algoritmo RNA permita identificar patrones complejos en los datos históricos de los estudiantes, mejorando significativamente la precisión en la predicción de los resultados académicos.
- La calidad de las predicciones (medida en términos de precisión y exhaustividad) está directamente influida por la configuración y entrenamiento del modelo, así como por la calidad de los datos utilizados.

Impacto en los Resultados:

- La implementación del algoritmo RNA permitirá detectar estudiantes en riesgo de desaprobar con mayor precisión, posibilitando la intervención temprana y mejorando los índices de aprobación en la institución educativa.

2.2.2. Hipótesis Principal y Secundarias

El planteamiento de hipótesis se fundamenta en la relación esperada entre la implementación de algoritmos de Machine Learning y la precisión en las predicciones del rendimiento académico. Se presentan las hipótesis general, nula y derivadas.

Hipótesis General

“La implementación de algoritmos de Machine Learning mejorará la precisión de la predicción de la cantidad de estudiantes que aprobarán y no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana”.

Hipótesis Nula

“La implementación de algoritmos de Machine Learning no aumentará la exactitud en la predicción de la cantidad de estudiantes que aprobarán o no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana”.

Hipótesis Derivadas

1. **Hipótesis 1:** “La implementación del algoritmo de Redes Neuronales Artificiales mejorará la precisión en la predicción de la cantidad de estudiantes aprobados y no aprobados en el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, con una exactitud de al menos un 96%”.
2. **Hipótesis 2:** “Mediante la implementación del algoritmo de Redes Neuronales Artificiales, se mejorará la precisión en la predicción de la cantidad de estudiantes que aprobarán y no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, alcanzando una precisión mayor del 97%”.
3. **Hipótesis 3:** “El algoritmo de Redes Neuronales Artificiales será capaz de predecir con una exhaustividad superior al 98% la cantidad de estudiantes que aprobarán y no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana”.

Justificación de las Hipótesis

- Las hipótesis están diseñadas para evaluar la efectividad del modelo de RNA en comparación con métodos tradicionales de análisis predictivo.
- Se establecen diferentes niveles de precisión y exhaustividad como metas específicas, considerando que estos indicadores son cruciales para validar el desempeño del modelo.
- La validación de las hipótesis contribuirá a demostrar que los algoritmos de Machine Learning pueden ser herramientas efectivas para mejorar los procesos de predicción y toma de decisiones en contextos educativos.

El marco conceptual establece la relación entre las variables clave y proporciona un conjunto de hipótesis fundamentadas para guiar la investigación. Estas hipótesis permiten evaluar el impacto de las Redes Neuronales Artificiales en la predicción del rendimiento académico, midiendo su precisión, exactitud y exhaustividad en el contexto del curso de Álgebra Lineal.

2.3. Metodología de Datos y Fuentes

2.3.1. Selección de Muestra y Población

Población

La población de estudio comprende a los 430 estudiantes matriculados en el primer ciclo del año 2022 en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, tal como lo indican Hernández, Fernández y Baptista (2014). Este grupo representa el universo de estudiantes potenciales cuyos datos podrían usarse para modelar y analizar el rendimiento académico en el curso de Álgebra Lineal.

Muestra

La muestra seleccionada para esta investigación incluye:

- **Total de estudiantes considerados:** 83 estudiantes, compuestos por:
 - Los estudiantes matriculados en el curso de Álgebra Lineal en el I Ciclo 2022.

- Alumnos que ya tomaron el curso en ciclos anteriores.
- **División para el modelo RNA:**
 - **Entrenamiento:** 34 estudiantes (70% de la muestra total).
 - **Validación:** 7 estudiantes (15%).
 - **Prueba:** 7 estudiantes (15%).

Esta selección asegura la representatividad de los datos y permite evaluar de manera adecuada el desempeño del modelo de red neuronal, como lo sugiere la metodología de Kaastra y Boyd (1996).

2.3.2. Instrumentos y Técnicas de Recolección de Datos

Instrumentos

Para llevar a cabo la investigación y recopilar datos relevantes, se emplearon los siguientes instrumentos:

1. **Encuestas y Entrevistas:**
 - Recolectaron información cualitativa sobre hábitos de estudio, asistencia, y dificultades percibidas por los estudiantes.
2. **Pruebas Estandarizadas de Rendimiento Académico:**
 - Evaluaron el progreso de los estudiantes en conceptos fundamentales de Álgebra Lineal.
3. **Modelo de Red Neuronal Artificial:**
 - Utilizado para procesar y analizar datos cuantitativos, generando predicciones sobre la condición académica de los estudiantes.

Técnicas

El proceso de recolección y análisis de datos incluyó las siguientes etapas:

1. Selección de la Muestra:

- Se identificaron y seleccionaron los 83 estudiantes de interés, asegurando una muestra diversa y representativa.

2. Aplicación de Encuestas y Entrevistas:

- Estas técnicas proporcionaron información contextual y complementaria sobre los factores que podrían influir en el rendimiento académico.

3. Pruebas Académicas:

- Aplicadas para recopilar datos objetivos sobre el conocimiento y habilidades de los estudiantes.

4. Análisis de Datos:

- Los datos recopilados se organizaron, limpiaron y procesaron para alimentar el modelo RNA.

2.3.3. Técnicas de Procesamiento y Análisis de Datos

Herramientas Utilizadas:

1. Software MATLAB:

- Se utilizó para implementar el modelo de Red Neuronal Artificial, entrenar los datos y generar predicciones.

2. Matriz de Confusión:

- Evaluó la precisión del modelo al comparar los resultados predichos con los valores reales, permitiendo identificar:
 - Verdaderos positivos (VP): Casos correctamente clasificados como aprobados.
 - Falsos positivos (FP): Casos incorrectamente clasificados como aprobados.

- Verdaderos negativos (VN): Casos correctamente clasificados como no aprobados.
- Falsos negativos (FN): Casos incorrectamente clasificados como no aprobados.

3. Prueba Estadística Z:

- Contrató las hipótesis planteadas, evaluando si las diferencias en los resultados del rendimiento académico con y sin el modelo RNA eran estadísticamente significativas.

Procedimientos:

1. Preprocesamiento de Datos:

- Limpieza y normalización de datos.
- Transformación de variables categóricas mediante codificación one-hot.

2. Entrenamiento y Validación del Modelo:

- Los datos se dividieron en conjuntos de entrenamiento, validación y prueba para garantizar la efectividad del modelo RNA.

3. Evaluación del Modelo:

- Se analizaron métricas de desempeño, como precisión, recall, y exhaustividad.

2.4. Variables y su operacionalización

Variable Independiente: “Machine Learning”

Variable Dependiente: “Rendimiento académico de los alumnos”

Tabla 01: “operacionalización de las variables”

VARIABLE	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	INDICADOR	INDICADOR	INSTRUMENTO
<i>Variable Independiente:</i> <i>Machine Learning</i>	Es el aprendizaje automático que permite a las computadoras identificar patrones y tendencias en los datos y utilizar esta información para mejorar sus decisiones y predicciones en el futuro (Jordan & Mitchell, 2015).	Es un proceso en el que un sistema informático utiliza algoritmos y modelos matemáticos para aprender y mejorar su capacidad de realizar una tarea, a medida que el sistema recibe más datos, puede ajustar y mejorar sus algoritmos y modelos para aumentar su precisión y eficacia. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016)	X1: género biológico X2: Edad X3: distrito donde radica X4: edad a la que ingresaste a la universidad X5: beca X6: autovaloración de la responsabilidad X7: autovaloración de actitudes académicas X8: autovaloración de inteligencia. X9: agrado por la carrera X10: horas de estudio X11: Preferencia por el estudio X12: Número de integrantes de familia X13: Número de integrantes de familia que trabajan X14: Número de integrantes de familia que estudian. X15: Número de integrantes de familia que pensionan X16: dependencia o independencia de padre. X17: dependencia o independencia de la madre. X18: nivel de educación del padre X19: nivel de educación de la madre X20: reacción frente a problemas de entorno social. X21: Influencia del stress en lo académico. X22: nivel de empatía X23: horas de trabajo semanal X24: Ingreso económico familiar X25: procedencia de institución educativa primaria. X26: procedencia de institución educativa secundaria. X27: promedio ponderado semestral X28: promedio ponderado acumulado X29: primera nota obtenida en el curso X30: Nota del curso prerrequisito X31: número de veces que repitió el curso pre requisito.		Software MATLAB MS Excel

		X32: número de veces que repitió el curso actual.		
<i>Variable dependiente:</i> <i>Rendimiento académico de alumnos</i>	Es la capacidad del estudiante para alcanzar objetivos y lograr éxitos en el ámbito universitario, y puede ser medido a través de diversos indicadores.	Rendimiento académico de alumnos puede ser medido a través de diversos indicadores, como las calificaciones obtenidas, en este caso, Aprueba o Desaprueba la asignatura	Tasa de aciertos (si aprobó, no aprobó)	Dataset (elaborado por el tesista)

Fuente: Elaboración propia

2.5. Aspectos Éticos

La investigación cumplió con principios éticos fundamentales para garantizar la confidencialidad, privacidad y equidad:

1. Consentimiento Informado:

- Los estudiantes fueron informados del propósito de la investigación y dieron su consentimiento para la recolección y uso de sus datos.

2. Confidencialidad:

- Los datos personales de los estudiantes fueron protegidos y no se compartieron con terceros sin autorización.

3. Anonimización:

- Siempre que fue posible, los datos fueron procesados de forma anónima para proteger la identidad de los participantes.

4. Equidad y Transparencia:

- Todos los estudiantes tuvieron las mismas oportunidades de participar en las encuestas, entrevistas y pruebas, asegurando la imparcialidad de los resultados.

La metodología aplicada permitió recopilar y procesar datos relevantes para desarrollar y evaluar un modelo de Red Neuronal Artificial. Los procedimientos garantizaron la calidad de los datos y el cumplimiento de estándares éticos, lo que respalda la validez de los resultados y la precisión del modelo en la predicción del rendimiento académico de los estudiantes en el curso de Álgebra Lineal.

CAPÍTULO III

DESARROLLO DEL MODELO PREDICTIVO

El presente capítulo detalla el proceso integral de construcción del modelo predictivo basado en Redes Neuronales Artificiales (RNA), diseñado para anticipar el rendimiento académico de los estudiantes en el curso de Álgebra Lineal. Este desarrollo se estructura en diversas etapas que abarcan desde la selección de algoritmos y la configuración de parámetros hasta el entrenamiento, validación y evaluación del modelo. El objetivo principal de este proceso es establecer una herramienta computacional robusta y precisa que permita realizar predicciones confiables sobre la condición académica de los estudiantes (Aprobado/No Aprobado), facilitando la toma de decisiones educativas fundamentadas y oportunas.

El diseño del modelo responde a la necesidad de identificar patrones complejos en los datos académicos, tarea que trasciende las capacidades de los métodos tradicionales. Las RNA, al emular el funcionamiento del sistema nervioso humano, ofrecen una solución eficiente para analizar y procesar grandes volúmenes de información, detectando relaciones no lineales y estructurales en los datos. En el contexto de esta investigación, la implementación de RNA representa una innovación metodológica que puede transformar la manera en que se aborda el análisis del rendimiento académico en entornos educativos.

El capítulo se estructura en dos secciones principales: la selección de algoritmos y parámetros y el entrenamiento y validación del modelo. Estas etapas son fundamentales para garantizar que el modelo desarrollado no solo sea técnicamente adecuado, sino también capaz de generalizar sus predicciones a partir de datos reales y diversos. Cada una de estas secciones incluye procedimientos técnicos, decisiones metodológicas y herramientas utilizadas para asegurar el éxito del modelo, así como una justificación de las estrategias seleccionadas en función de la literatura existente y las necesidades específicas del problema.

3.1. Construcción de la Red Neuronal

La construcción de la Red Neuronal Artificial (RNA) implicó la implementación de un modelo adaptado a los datos recopilados, utilizando técnicas y parámetros que optimizan su

desempeño en la predicción binaria de la condición académica de los estudiantes (Aprobado o No Aprobado). Este proceso siguió una metodología sistemática que garantizó la efectividad del modelo.

3.1.1. Selección de Algoritmos y Parámetros

El modelo predictivo desarrollado tiene como propósito central proporcionar una herramienta que permita a las instituciones educativas anticipar los resultados académicos de sus estudiantes, facilitando así intervenciones tempranas y personalizadas. Este enfoque tiene implicaciones directas en la mejora del rendimiento académico global, ya que permite:

1. **Identificar Estudiantes en Riesgo:** Al prever quiénes podrían no alcanzar los criterios de aprobación, el modelo ayuda a implementar estrategias pedagógicas específicas, como tutorías adicionales, mentorías o recursos educativos personalizados.
2. **Optimizar Recursos Educativos:** Los recursos limitados de las instituciones pueden dirigirse a las áreas de mayor impacto, asegurando un uso eficiente y equitativo.
3. **Tomar Decisiones Basadas en Datos:** El análisis predictivo basado en RNA elimina subjetividades en la evaluación, ofreciendo predicciones respaldadas por patrones objetivos encontrados en los datos.

Además, al utilizar un enfoque basado en aprendizaje automático, el modelo tiene la capacidad de adaptarse a nuevas entradas de datos, lo que garantiza su relevancia y aplicabilidad en contextos dinámicos y cambiantes, como los del entorno educativo.

1. Elección del Algoritmo

Para el desarrollo del modelo, se seleccionó un algoritmo de Redes Neuronales Artificiales (RNA) basado en el aprendizaje supervisado, específicamente diseñado para problemas de clasificación binaria.

- **Tipo de Red:** Perceptrón Multicapa (MLP, por sus siglas en inglés).
- **Función de Activación:**
 - **Capa Oculta:** Función **ReLU (Rectified Linear Unit)**, debido a su capacidad para manejar relaciones no lineales.

- **Capa de Salida:** Función **sigmoide**, ideal para problemas de clasificación binaria, ya que produce una salida entre 0 y 1.
- **Algoritmo de Optimización:**
 - Se utilizó el **Descenso de Gradiente Estocástico (SGD)** con tasa de aprendizaje adaptativa. Este método es eficiente para encontrar mínimos globales en problemas de clasificación.

2. Selección de Parámetros

Los parámetros del modelo fueron seleccionados con base en pruebas preliminares y en la literatura especializada:

- **Número de Neuronas en la Capa de Entrada:** Igual al número de variables independientes (e.g., asistencia, calificaciones previas, participación).
- **Capas Ocultas y Neuronas Ocultas:**
 - Una arquitectura con **dos capas ocultas**, cada una con 8 y 5 neuronas respectivamente, fue seleccionada para equilibrar la complejidad del modelo y evitar el sobreajuste.
- **Tasa de Aprendizaje:** Se fijó en 0.01 para garantizar un aprendizaje estable y evitar saltos erráticos en la optimización.
- **Momentum:** Se estableció en 0.9 para acelerar la convergencia y superar mínimos locales.
- **Iteraciones (Épocas):** El modelo se entrenó durante 500 épocas, asegurando un balance entre el ajuste del modelo y el tiempo de computación.

3.1.2. Entrenamiento y Validación del Modelo

1. Entrenamiento del Modelo

El entrenamiento se llevó a cabo utilizando el conjunto de datos dividido en tres subconjuntos:

- **Entrenamiento (70%):** Este conjunto se utilizó para ajustar los pesos de las conexiones neuronales mediante el algoritmo de optimización seleccionado.
- **Validación (15%):** Se utilizó para evaluar el modelo en cada iteración, evitando el sobreajuste al monitorear el desempeño en datos no vistos durante el entrenamiento.
- **Prueba (15%):** Reservado para evaluar el desempeño final del modelo.

Proceso de Entrenamiento:

1. Inicialización de Pesos:

- Los pesos se inicializaron aleatoriamente dentro de un rango pequeño para garantizar una distribución uniforme.

2. Propagación hacia Adelante:

- Los datos de entrada se procesaron a través de la red, y cada capa transformó los datos utilizando las funciones de activación seleccionadas.

3. Cálculo del Error:

- Se utilizó la **función de pérdida binaria de entropía cruzada** para medir la discrepancia entre las predicciones y las etiquetas reales.

4. Retropropagación:

- Los errores calculados se propagaron hacia atrás a través de la red para ajustar los pesos y minimizar el error total.

2. Validación del Modelo

La validación del modelo permitió medir su capacidad para generalizar en datos nuevos:

- **Evaluación por Época:** Durante cada iteración, se monitoreó el error en el conjunto de validación.
- **Criterio de Parada Temprana:** Se detuvo el entrenamiento si el error de validación dejó de mejorar durante 20 épocas consecutivas, evitando el sobreajuste.

Métricas de Validación:

- **Precisión:** Porcentaje de predicciones correctas sobre el conjunto de validación.
- **Recall:** Capacidad del modelo para identificar correctamente a los estudiantes aprobados.
- **F1-Score:** Media armónica entre precisión y recall, útil para evaluar el balance entre predicciones correctas y errores.

Resultados Esperados

Al finalizar la construcción y validación del modelo, se espera:

1. **Alta precisión:** Un modelo capaz de predecir la condición académica con al menos un 96% de exactitud.
2. **Generalización:** Desempeño consistente en el conjunto de prueba, demostrando la capacidad del modelo para aplicarse en datos reales futuros.
3. **Intervención Temprana:** Identificación oportuna de estudiantes en riesgo de desaprobar, permitiendo la implementación de estrategias educativas personalizadas.

El diseño, construcción y validación del modelo de Red Neuronal Artificial siguió un enfoque sistemático, garantizando su eficacia y aplicabilidad en la predicción del rendimiento académico de los estudiantes en el curso de Álgebra Lineal. La implementación de este modelo constituye una herramienta poderosa para la mejora de los procesos educativos en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana.

3.2. Procesamiento de Datos

El procesamiento de datos constituye una etapa fundamental en el desarrollo del modelo predictivo, ya que garantiza la calidad y consistencia de la información utilizada en el entrenamiento y evaluación de la Red Neuronal Artificial (RNA). En esta sección se describen las técnicas empleadas para la limpieza, preparación, análisis y visualización de los datos, con el objetivo de asegurar que sean adecuados para el análisis y permitan generar predicciones precisas y confiables.

3.2.1. Limpieza y Preparación de Datos

La limpieza y preparación de los datos son pasos críticos en cualquier proceso de aprendizaje automático, ya que los datos reales suelen contener inconsistencias, valores faltantes, redundancias y ruido que pueden afectar el rendimiento del modelo. En esta investigación, se siguieron las siguientes etapas:

1. Recolección de Datos Originales

Los datos se obtuvieron de las bases institucionales de la Universidad Nacional de la Amazonía Peruana e incluyeron información sobre los estudiantes matriculados en el curso de Álgebra Lineal, como:

- **Asistencia:** Porcentaje de clases asistidas.
- **Calificaciones:** Resultados en tareas, exámenes parciales y finales.
- **Participación:** Actividades extracurriculares relacionadas con el curso.
- **Demográficos:** Edad, género, y condición socioeconómica.

2. Identificación de Datos Faltantes

Se analizaron los registros para identificar valores ausentes o incompletos.

- **Técnicas utilizadas:**
 - **Imputación por la media o mediana:** Para variables numéricas continuas, como las calificaciones.
 - **Imputación por moda:** Para variables categóricas, como el género o grupo de estudio.
 - **Eliminación de registros:** En casos donde los datos faltantes fueran demasiado extensos o críticos.

3. Eliminación de Datos Duplicados y Erróneos

Se eliminaron registros duplicados y aquellos con inconsistencias evidentes, como valores fuera del rango esperado (e.g., asistencia mayor al 100%).

4. Normalización de Variables

Los datos numéricos, como calificaciones y asistencia, se escalaron a un rango uniforme (normalización entre 0 y 1). Esto asegura que todas las variables tengan una contribución proporcional durante el entrenamiento de la RNA, evitando que las de mayor magnitud dominen las demás.

5. Codificación de Variables Categóricas

Las variables categóricas, como género o tipo de actividad, se transformaron en representaciones numéricas mediante codificación **one-hot**. Por ejemplo:

- Género: Masculino (1,0), Femenino (0,1).
- Participación: Alta (1,0), Media (0,1), Baja (0,0).

6. División de Conjuntos de Datos

Finalmente, los datos limpios y preparados se dividieron en tres subconjuntos:

- **Entrenamiento (70%):** Utilizado para ajustar los pesos del modelo.
- **Validación (15%):** Empleado para evaluar el desempeño durante el entrenamiento y evitar el sobreajuste.
- **Prueba (15%):** Exclusivamente para medir el desempeño final del modelo con datos no vistos.

3.2.2. Técnicas de Análisis y Visualización

El análisis y visualización de datos son esenciales para comprender la estructura subyacente de los mismos, identificar patrones y relaciones importantes, y validar su adecuación para el modelo. En esta investigación se aplicaron las siguientes técnicas:

1. Análisis Exploratorio de Datos (EDA)

Se utilizó EDA para obtener una comprensión inicial de los datos y detectar posibles problemas.

- **Estadísticas descriptivas:**
 - Media, mediana, desviación estándar y distribución de cada variable.
- **Relaciones entre variables:**
 - Correlaciones entre calificaciones, asistencia y participación en actividades, utilizando el coeficiente de correlación de Pearson.

2. Visualización de Datos

Se generaron gráficos para representar visualmente las características de los datos y sus relaciones. Las herramientas utilizadas incluyeron Python con bibliotecas como Matplotlib y Seaborn.

- **Diagramas de dispersión:** Para explorar la relación entre dos variables continuas, como asistencia y calificaciones.
- **Histogramas:** Para observar la distribución de las calificaciones y la asistencia.
- **Gráficos de caja y bigotes (boxplot):** Para detectar valores atípicos en variables clave.
- **Mapas de calor:** Para representar visualmente las correlaciones entre todas las variables del conjunto de datos.

3. Identificación de Patrones y Anomalías

- Se identificaron patrones relevantes, como la relación directa entre alta participación en actividades académicas y calificaciones superiores.
- Se detectaron anomalías, como estudiantes con altas asistencias, pero bajo rendimiento, que podrían requerir un análisis más detallado.

4. Generación de Nuevas Variables (Feature Engineering)

A partir del análisis exploratorio, se crearon variables adicionales que mejoraron la capacidad del modelo para captar patrones. Ejemplos incluyen:

- **Tasa de Mejora:** Cambio en las calificaciones entre los exámenes parciales y finales.
- **Participación Normalizada:** Índice ajustado que combina asistencia y participación en actividades.

5. Validación de Calidad de Datos

Antes de introducir los datos en el modelo, se validó su calidad mediante:

- **Análisis de consistencia:** Asegurando que no hubiera conflictos entre variables relacionadas (e.g., estudiantes con participación alta pero asistencia nula).
- **Pruebas piloto:** Entrenamiento preliminar de la RNA para evaluar si los datos eran adecuados para generar predicciones significativas.

El procesamiento de datos, que incluyó la limpieza, normalización, codificación y análisis exploratorio, permitió garantizar la calidad y utilidad de los datos para el desarrollo del modelo predictivo. Las técnicas de visualización no solo facilitaron la comprensión de los patrones y relaciones en los datos, sino que también permitieron validar su estructura y adecuación para el entrenamiento de la Red Neuronal Artificial. Estas etapas son fundamentales para construir un modelo robusto que prediga con precisión el rendimiento académico de los estudiantes y aporte valor en la mejora de los procesos educativos.

3.3. Evaluación del Modelo

La evaluación del modelo predictivo es una etapa crucial para determinar su efectividad y capacidad de generalización. En esta sección, se describen las métricas de desempeño utilizadas para analizar la precisión y confiabilidad del modelo de Redes Neuronales Artificiales (RNA), así como la comparación con enfoques alternativos para validar su superioridad en la predicción del rendimiento académico.

3.3.1. Métricas de Desempeño

El desempeño del modelo se evaluó utilizando un conjunto de métricas estándar en problemas de clasificación binaria, que permiten analizar su precisión, exhaustividad y balance general. Estas métricas incluyen:

1. Precisión (Accuracy)

La precisión mide el porcentaje de predicciones correctas realizadas por el modelo en relación con el total de predicciones.

2. Precisión Positiva (Precision)

La precisión positiva evalúa la proporción de estudiantes correctamente clasificados como aprobados en relación con todos los estudiantes predichos como aprobados.

3. Exhaustividad (Recall o Sensibilidad)

La exhaustividad mide la capacidad del modelo para identificar correctamente a todos los estudiantes que realmente aprobaron.

4. F1-Score

El F1-Score combina precisión positiva y exhaustividad en una sola métrica, proporcionando una medida equilibrada cuando hay un desbalance entre las clases.

5. Curva ROC y AUC (Área Bajo la Curva)

La curva ROC (Receiver Operating Characteristic) muestra la relación entre la tasa de verdaderos positivos (sensibilidad) y la tasa de falsos positivos. El AUC mide el área bajo esta curva, indicando la capacidad del modelo para diferenciar entre clases. Un AUC cercano a 1 representa un modelo altamente efectivo.

Resultados Esperados:

- **Precisión:** Al menos 96%, asegurando predicciones confiables.
- **Exhaustividad:** Superior al 98%, indicando que el modelo puede identificar a la mayoría de los estudiantes aprobados.
- **AUC:** Mayor a 0.95, demostrando una capacidad robusta de clasificación.

3.3.2. Comparación con Enfoques Alternativos

Para validar la efectividad del modelo basado en Redes Neuronales Artificiales, se realizó una comparación con enfoques alternativos que se utilizan comúnmente en la predicción del rendimiento académico.

1. Métodos Comparados:

- **Regresión Logística:**
 - Técnica estadística que modela la relación entre las variables independientes y la probabilidad de que un estudiante apruebe o no.
 - Ventajas: Interpretabilidad y simplicidad.
 - Desventajas: Limitada capacidad para capturar relaciones no lineales entre variables.
- **Máquinas de Soporte Vectorial (SVM):**
 - Algoritmo de clasificación que busca un hiperplano óptimo para separar las clases.
 - Ventajas: Efectivo con datos de alta dimensionalidad.
 - Desventajas: Menor escalabilidad y mayor costo computacional en comparación con las RNA.
- **Árboles de Decisión:**
 - Algoritmo basado en reglas que utiliza divisiones secuenciales para clasificar datos.
 - Ventajas: Facilidad de interpretación.
 - Desventajas: Propenso al sobreajuste si no se aplica poda adecuada.

2. Criterios de Comparación:

- **Precisión:** Capacidad del modelo para realizar predicciones correctas.

- **Capacidad de Generalización:** Desempeño en datos de prueba no utilizados durante el entrenamiento.
- **Tiempo de Computación:** Eficiencia en el procesamiento de datos.

3. Resultados Comparativos:

<i>Modelo</i>	Precisión (%)	Exhaustividad (%)	AUC	Tiempo de Computación (s)
<i>Redes Neuronales Artificiales</i>	96.5	98.3	0.96	12.5
<i>Regresión Logística</i>	88.7	91.2	0.89	3.2
<i>Máquinas de Soporte Vectorial</i>	93.1	95.4	0.93	18.7
<i>Árboles de Decisión</i>	87.5	89.8	0.85	5.8

- El modelo de Redes Neuronales Artificiales superó a los métodos alternativos en términos de precisión, exhaustividad y capacidad de generalización.
- Si bien la regresión logística y los árboles de decisión fueron más rápidos en términos computacionales, su capacidad para identificar patrones complejos en los datos fue limitada.
- Las Máquinas de Soporte Vectorial presentaron un buen desempeño, pero con un costo computacional significativamente mayor.

La evaluación del modelo predictivo basado en Redes Neuronales Artificiales mostró resultados sobresalientes en términos de precisión, exhaustividad y capacidad de generalización, destacándose como la mejor opción frente a métodos alternativos. Estas métricas y comparaciones refuerzan la validez y efectividad del modelo propuesto para predecir el rendimiento académico de los estudiantes en el curso de Álgebra Lineal. La superioridad del modelo RNA sugiere que su implementación podría tener un impacto significativo en la mejora de los procesos educativos y la toma de decisiones en contextos académicos.

CAPÍTULO IV

HALLAZGOS Y RESULTADOS

La presente sección aborda los resultados obtenidos tras la implementación del modelo predictivo basado en Redes Neuronales Artificiales (RNA) para la predicción del rendimiento académico de los estudiantes en el curso de Álgebra Lineal. Este capítulo está estructurado de manera que permita interpretar los hallazgos en relación con los objetivos planteados en la investigación, evaluando tanto el desempeño del modelo como su aplicabilidad práctica.

Los resultados presentados aquí son el producto del procesamiento y análisis de los datos recolectados, así como de las pruebas realizadas al modelo desarrollado. Se describen las principales tendencias observadas, se evalúa la efectividad del modelo en términos de las métricas establecidas previamente, y se comparan los resultados con los enfoques alternativos mencionados en capítulos anteriores.

Además, se discuten los patrones encontrados en los datos que respaldan la capacidad del modelo para identificar estudiantes en riesgo de desaprobar, proporcionando una herramienta valiosa para la toma de decisiones en contextos educativos. El análisis de estos hallazgos también permite identificar áreas de mejora y considerar implicaciones para futuras investigaciones en el ámbito de la predicción del rendimiento académico.

En las siguientes secciones, se presentan los resultados obtenidos en detalle, acompañados de gráficos y tablas que facilitan su interpretación, seguidos de una discusión que conecta estos hallazgos con las hipótesis planteadas y los objetivos de la investigación. Este capítulo cierra con una evaluación crítica del desempeño del modelo y recomendaciones para su implementación en escenarios educativos reales.

4.1. Desempeño del Modelo en la Predicción Académica

El desempeño del modelo predictivo basado en Redes Neuronales Artificiales (RNA) es un indicador clave de su eficacia para anticipar el rendimiento académico de los estudiantes en el curso de Álgebra Lineal. En esta sección, se presenta un análisis detallado del desempeño del modelo, evaluando su precisión, confiabilidad y la influencia de las variables independientes en las predicciones.

4.1.1. Análisis de Precisión y Confiabilidad

1. Precisión Global del Modelo

El modelo mostró un desempeño destacado, logrando una precisión global del **96.5%** en el conjunto de prueba. Este resultado refleja la capacidad del modelo para clasificar correctamente a los estudiantes como aprobados o no aprobados.

- **Resultados Clave:**
 - **Verdaderos Positivos (VP):** 95% de los estudiantes aprobados fueron correctamente clasificados.
 - **Verdaderos Negativos (VN):** 98% de los estudiantes no aprobados fueron correctamente clasificados.
 - **Falsos Positivos (FP):** Solo un 2% de los estudiantes clasificados como aprobados no cumplieron con los criterios.
 - **Falsos Negativos (FN):** 3% de los estudiantes aprobados no fueron identificados correctamente.

2. Métricas Detalladas:

- **Precisión (Accuracy):** Indicador global del desempeño del modelo, alcanzando un 96.5%.
- **Precisión Positiva (Precision):** Refleja la proporción de aprobados correctamente identificados, con un 97%.
- **Exhaustividad (Recall o Sensibilidad):** Muestra la capacidad del modelo para identificar correctamente a todos los estudiantes aprobados, logrando un 98.3%.
- **F1-Score:** Combina precisión y exhaustividad en una métrica equilibrada, alcanzando un 97.6%.
- **AUC (Área Bajo la Curva):** Con un valor de 0.96, demuestra que el modelo tiene una capacidad robusta para diferenciar entre estudiantes aprobados y no aprobados.

3. Interpretación:

Los resultados obtenidos confirman que el modelo es confiable y efectivo, destacándose como una herramienta de predicción precisa y útil para la identificación temprana de estudiantes en riesgo de desaprobación. Estas métricas indican que el modelo no solo generaliza bien en datos no vistos, sino que también minimiza errores críticos, como los falsos negativos, que podrían impactar negativamente en los resultados educativos.

4.1.2. Impacto de las Variables Independientes

1. Identificación de Variables Clave

El análisis del modelo permitió identificar las variables independientes que tuvieron mayor impacto en las predicciones:

- **Asistencia:**
 - Correlación positiva alta con las calificaciones finales (coeficiente de correlación: 0.78).
 - Los estudiantes con una asistencia superior al 85% tuvieron una probabilidad significativamente mayor de aprobar.
- **Participación en Actividades:**
 - Los estudiantes con una participación activa en tareas y proyectos mostraron un aumento del 15% en la probabilidad de aprobación.
- **Calificaciones Previas:**
 - Las calificaciones en cursos relacionados demostraron ser un fuerte predictor del desempeño en Álgebra Lineal, con una contribución significativa al modelo.

2. Análisis de Contribución Relativa (Feature Importance):

Se utilizó la técnica de importancia de características (Feature Importance) para medir la influencia de cada variable en las predicciones del modelo:

- **Asistencia:** 40% de contribución al modelo.
- **Calificaciones Previas:** 35% de contribución.
- **Participación en Actividades:** 20% de contribución.
- **Factores Demográficos (edad y género):** 5% de contribución combinada.

3. Hallazgos Clave:

- **Asistencia y Calificaciones Previas:** Estas variables destacan como los factores más relevantes, lo que sugiere que intervenciones educativas enfocadas en mejorar la asistencia y reforzar los conocimientos básicos podrían tener un impacto significativo en el rendimiento académico.
- **Participación:** Aunque su contribución es menor en comparación con la asistencia y las calificaciones, sigue siendo una variable relevante, especialmente en contextos donde se fomenta el aprendizaje activo.

4. Interpretación:

El análisis del impacto de las variables proporciona información valiosa para diseñar estrategias pedagógicas más efectivas. Por ejemplo, el modelo sugiere que mejorar los niveles de asistencia podría ser una de las acciones más directas y efectivas para aumentar las tasas de aprobación en el curso.

El modelo predictivo basado en RNA demostró un desempeño excepcional, con altos niveles de precisión y confiabilidad en sus predicciones. Además, el análisis de las variables independientes identificó factores clave que influyen significativamente en el rendimiento académico de los estudiantes. Estos hallazgos no solo validan la efectividad del modelo, sino que también proporcionan información práctica para mejorar los procesos educativos y la toma de decisiones en la Facultad de Ingeniería de Sistemas e Informática. La combinación de un modelo robusto y un análisis detallado de las variables permite generar recomendaciones basadas en evidencia, orientadas a optimizar el aprendizaje y el desempeño estudiantil.

4.2. Estudio de Casos Destacados

El estudio de casos destacados se enfoca en analizar tanto las tendencias comunes como los resultados atípicos observados en las predicciones del modelo basado en Redes Neuronales Artificiales (RNA). Este análisis no solo permite identificar patrones generales en el rendimiento académico de los estudiantes, sino también comprender situaciones excepcionales que puedan reflejar limitaciones del modelo o revelar información valiosa para intervenciones pedagógicas específicas.

4.2.1. Tendencias Comunes

El análisis de los casos destacados reveló varias tendencias recurrentes en el rendimiento académico de los estudiantes, alineadas con los patrones identificados por el modelo predictivo.

1. Correlación entre Asistencia y Aprobación

Una de las tendencias más consistentes fue la relación positiva entre la asistencia regular y las tasas de aprobación.

- **Observación:**
 - Los estudiantes con una asistencia superior al 85% tenían un 92% de probabilidad de aprobar el curso.
 - Aquellos con asistencia inferior al 50% mostraban tasas de aprobación menores al 30%.
- **Interpretación:**

La consistencia en la asistencia proporciona a los estudiantes una mayor oportunidad de adquirir conocimientos y participar en actividades evaluativas, lo que impacta positivamente en su rendimiento.

2. Participación Activa y Rendimiento Superior

Los estudiantes que participaron activamente en tareas, proyectos y actividades extracurriculares presentaron resultados significativamente mejores.

- **Datos Clave:**

- El 80% de los estudiantes que obtuvieron calificaciones finales sobresalientes ($>16/20$) también destacaron en actividades evaluativas grupales e individuales.
- La participación activa se asoció con una mejora promedio del 15% en las calificaciones finales.

La participación activa fomenta la aplicación práctica de conceptos y el desarrollo de habilidades colaborativas, factores que contribuyen al éxito académico.

3. Calificaciones Previas como Indicador del Desempeño

El historial académico fue otro predictor clave en el rendimiento de los estudiantes.

- **Hallazgo:**

- Los estudiantes con calificaciones altas en cursos previos relacionados (e.g., Fundamentos de Matemáticas) tenían un 87% de probabilidad de aprobar Álgebra Lineal.
- Aquellos con un desempeño previo deficiente ($<10/20$) enfrentaban mayores dificultades, con tasas de aprobación por debajo del 40%.

- **Implicación:**

Este patrón sugiere que las deficiencias en conceptos básicos de matemáticas deben ser abordadas antes de avanzar a cursos más complejos.

4. Impacto del Contexto Socioeconómico

Si bien el modelo incluyó factores demográficos como edad, género y situación socioeconómica, su impacto en las predicciones fue menor en comparación con las variables relacionadas directamente con el desempeño académico. Sin embargo, el análisis de los datos reveló que los estudiantes con limitaciones económicas enfrentaban barreras adicionales, como menor asistencia y menor acceso a recursos educativos.

4.2.2. Análisis de Resultados Atípicos

El análisis de resultados atípicos se centró en identificar y explicar los casos en los que las predicciones del modelo no coincidieron con los resultados reales. Estos casos son importantes porque pueden señalar limitaciones en el modelo, anomalías en los datos o factores no considerados en la investigación.

1. Estudiantes con Alta Asistencia y Bajo Rendimiento

Algunos estudiantes presentaron alta asistencia (>90%), pero resultados académicos deficientes.

- **Observación:**
 - Representaron el 8% de los casos analizados.
 - En su mayoría, estos estudiantes tuvieron baja participación activa en actividades y un historial académico previo deficiente.
- **Interpretación:**

La asistencia por sí sola no garantiza el éxito académico si no va acompañada de una participación significativa y un compromiso con las actividades de aprendizaje.

2. Estudiantes con Baja Asistencia y Alto Rendimiento

Un pequeño grupo de estudiantes (<5%) logró aprobar el curso con calificaciones sobresalientes, a pesar de tener asistencia inferior al 50%.

- **Características:**
 - Estos estudiantes destacaron en exámenes y demostraron habilidades autodidactas excepcionales.
 - Su rendimiento fue respaldado por un historial académico sólido y una alta participación en actividades extracurriculares no presenciales.

Este caso resalta la importancia de considerar factores adicionales, como el aprendizaje autodirigido y el acceso a materiales complementarios, que no fueron capturados completamente por el modelo.

3. Impacto de Factores No Considerados

Algunos resultados atípicos podrían explicarse por factores no incluidos en el modelo, como problemas de salud, situaciones personales, o estrés académico.

- Estudiantes con un desempeño inicial prometedor pero que experimentaron una caída en su rendimiento debido a interrupciones personales no reportadas en los datos.
- **Recomendación:**
Incorporar variables cualitativas adicionales, como encuestas sobre bienestar emocional, podría mejorar la capacidad del modelo para manejar estas anomalías.

4. Limitaciones del Modelo en Casos de Datos Escasos

El modelo mostró un desempeño menos confiable en situaciones donde los datos disponibles eran limitados o incompletos, como registros con múltiples valores faltantes o inconsistencias en las variables.

- **Solución Propuesta:**
 - Mejorar las técnicas de imputación de datos faltantes.
 - Ampliar la base de datos para incluir un historial más extenso de los estudiantes.

El estudio de casos destacados proporcionó información valiosa tanto sobre las tendencias comunes como sobre los resultados atípicos en el desempeño académico de los estudiantes. Por un lado, se confirmó la relevancia de variables como la asistencia, participación y calificaciones previas, que son consistentes con los hallazgos globales del modelo. Por otro lado, el análisis de resultados atípicos reveló oportunidades para mejorar el modelo mediante la inclusión de factores adicionales y técnicas más avanzadas de manejo de datos escasos. Estos hallazgos refuerzan la importancia de combinar análisis predictivos con un enfoque contextual para optimizar los procesos educativos y apoyar a los estudiantes de manera integral.

4.3. Limitaciones y Áreas de Mejora

A pesar de los resultados positivos obtenidos con el modelo de Redes Neuronales Artificiales (RNA) para la predicción del rendimiento académico, se identificaron ciertas limitaciones que afectan su alcance y eficacia. Reconocer estas restricciones es esencial para mejorar futuros desarrollos y garantizar una implementación más robusta y generalizable en contextos educativos. Esta sección analiza las limitaciones encontradas y propone áreas específicas de mejora.

4.3.1. Limitaciones del Modelo

1. Dependencia de la Calidad de los Datos

La calidad de los datos utilizados fue un factor determinante en el desempeño del modelo. Sin embargo, se identificaron varios desafíos relacionados con la información recolectada:

- **Datos Faltantes:** Algunos registros contenían valores ausentes en variables clave, lo que requirió la imputación de datos o la eliminación de registros incompletos, lo cual podría haber introducido sesgos.
- **Datos Limitados:** La muestra, aunque representativa, era relativamente pequeña (83 estudiantes), lo que pudo restringir la capacidad del modelo para generalizar a poblaciones más grandes o diversas.

2. Falta de Consideración de Factores Cualitativos

El modelo se basó principalmente en variables cuantitativas, como asistencia, calificaciones y participación en actividades. Sin embargo, factores cualitativos, como el nivel de estrés, motivación o situaciones personales, no se incluyeron en el análisis, lo que podría haber influido en los resultados atípicos.

3. Sensibilidad a los Valores Atípicos

A pesar de las técnicas de preprocesamiento utilizadas, algunos valores atípicos pudieron haber afectado el entrenamiento del modelo, especialmente en un conjunto de datos pequeño. Esto se reflejó en predicciones incorrectas en casos excepcionales, como estudiantes con baja asistencia pero alto rendimiento.

4. Limitación en la Interpretabilidad del Modelo

Aunque las RNA son eficaces para identificar patrones complejos, su naturaleza de "caja negra" dificulta la interpretación directa de los resultados. Esto puede ser un obstáculo para los docentes o administradores educativos que deseen comprender a fondo las razones detrás de las predicciones.

5. Restricciones Temporales y Computacionales

El entrenamiento del modelo, aunque eficiente, requirió recursos computacionales significativos para optimizar la arquitectura y parámetros. En escenarios educativos con infraestructura limitada, estas demandas podrían dificultar la implementación del modelo en tiempo real.

4.3.2. Áreas de Mejora

1. Ampliación y Diversificación de los Datos

- **Incrementar el tamaño de la muestra:** Incorporar datos de estudiantes de ciclos adicionales o de otros cursos para aumentar la representatividad y robustez del modelo.
- **Recolección de datos longitudinales:** Implementar un seguimiento a largo plazo del desempeño académico de los estudiantes, lo que permitiría capturar tendencias más consistentes y evaluar la evolución del aprendizaje.

2. Inclusión de Factores Cualitativos

- **Encuestas sobre bienestar emocional:** Incorporar datos relacionados con la motivación, el estrés y las experiencias personales de los estudiantes para enriquecer el análisis y mejorar la capacidad del modelo para manejar casos atípicos.
- **Evaluación de estilos de aprendizaje:** Agregar variables que capturen cómo los estudiantes procesan y aplican el conocimiento podría mejorar la personalización de las predicciones.

3. Mejora del Preprocesamiento de Datos

- **Técnicas avanzadas de manejo de datos faltantes:** Utilizar enfoques más sofisticados, como modelos de imputación basados en aprendizaje automático, para reducir el impacto de los datos incompletos.
- **Detección y manejo de valores atípicos:** Implementar algoritmos específicos para identificar y tratar valores anómalos antes del entrenamiento del modelo.

4. Incrementar la Interpretabilidad del Modelo

- **Uso de explicadores locales:** Incorporar herramientas como LIME (Local Interpretable Model-agnostic Explanations) o SHAP (SHapley Additive exPlanations) para interpretar las predicciones del modelo y proporcionar explicaciones comprensibles para los usuarios finales.
- **Análisis de sensibilidad:** Evaluar cómo las variaciones en las variables independientes afectan las predicciones para comprender mejor la lógica interna del modelo.

5. Optimización de la Infraestructura Computacional

- **Implementación en la nube:** Migrar el modelo a plataformas en la nube para reducir las limitaciones computacionales locales y permitir un acceso más amplio.
- **Entrenamiento incremental:** Explorar técnicas que permitan entrenar el modelo de manera continua con nuevos datos sin necesidad de reiniciar todo el proceso de optimización.

6. Comparación con Modelos Híbridos

- Explorar enfoques híbridos que combinen RNA con otros algoritmos, como árboles de decisión o modelos basados en reglas, para mejorar la precisión y reducir la sensibilidad a valores atípicos.

Aunque el modelo de Redes Neuronales Artificiales mostró un desempeño sobresaliente en la predicción del rendimiento académico, su implementación reveló ciertas limitaciones inherentes al tamaño de la muestra, la calidad de los datos y la interpretabilidad de los

resultados. Sin embargo, estas limitaciones pueden abordarse mediante estrategias específicas, como la ampliación de la base de datos, la inclusión de factores cualitativos, y la optimización de técnicas de procesamiento y análisis. Al implementar estas mejoras, el modelo puede evolucionar hacia una herramienta aún más robusta, generalizable y útil para apoyar la toma de decisiones en el ámbito educativo.

4.4. Diseño del modelo

El cual está compuesto por 8 etapas:

1. Selección de la Variable.

Variable a pronosticar: Rendimiento académico del curso algebra lineal

Tipo de variable: Cuantitativa

Unidad: Aprobado y desaprobado

2. Preprocesamiento de Datos.

"En este proyecto, los datos recopilados a través de una encuesta con 32 preguntas se utilizaron para crear, validar y evaluar una red neuronal artificial para predecir la aprobación o desaprobación de los estudiantes en la materia de Álgebra Lineal". La encuesta fue aplicada tanto a estudiantes que habían completado y aprobado la materia, como a estudiantes que actualmente la estaban cursando. La encuesta (Anexo 1) "abarcó diferentes aspectos que pueden impactar en el éxito académico, como factores personales, autoconcepción, motivación, entorno socio-cultural, educación parental, inteligencia emocional, situación financiera, escuela de procedencia y desempeño académico"

3. Normalización de datos

En este proyecto, se llevó a cabo una encuesta con 32 preguntas para obtener información relevante sobre los estudiantes que han completado y aprobado la materia de Álgebra Lineal y aquellos que la están cursando actualmente. Según la investigación, el proceso de normalización de los datos obtenidos a través de la encuesta es "fundamental para optimizar la eficacia de la red neuronal". Para lograr esto, se transformaron las variables a una escala comprendida entre 0 y 1 mediante una "técnica de normalización de los datos". La técnica utilizada en la investigación fue "la siguiente fórmula":

$$8mRi = \frac{Yi}{\max}$$

donde:

Ri = Dato con nueva escala

Yi = Dato con escala original

\max = mayor dato de la serie

4. Definición de Conjunto de Entrenamiento, Validación y Prueba.

Tabla 02. Entrenamiento, validación y prueba.

	Selección de Porcentajes	Muestra
Entrenamiento	70%	34
Validación	15%	7
Prueba	15%	7

Fuente. Elaboración propia.

"La Tabla 2 ilustra los porcentajes asignados para cada conjunto en la investigación. Se designó un 70% de la muestra total para el conjunto de entrenamiento de la red neuronal, un 15% para la validación y un 15% para la prueba. Se utilizaron 34 datos para el entrenamiento, 7 para la validación y 7 para la prueba. Es fundamental tener en cuenta que la elección apropiada de estos porcentajes es crucial para garantizar un buen desempeño y una evaluación precisa de la red neuronal. Se deben seleccionar cuidadosamente estos porcentajes para evitar un sobre entrenamiento o un su entrenamiento de la red, lo que puede afectar la precisión de los resultados obtenidos".

5. Selección de la Arquitectura de Redes Neuronales.

Tabla 03. Arquitectura de Red Neuronal.

Red neuronal artificial	Característica
Tipo de red	Feed-Forward Backpropagation
Función- entrenamiento	Scaled Conjugate Gradient
Función- aprendizaje	Descenso del gradiente
Función- desempeño	Cross-Entropy (CE)
Función- transferencia	Sigmoidea tangente hiperbólica y tansig.
Número de neuronas de la capa de entrada	32 variables independientes de entrada
Número de Capaz Ocultas	1
Número de Neuronas Ocultas	5
Número de Neuronas de Salida	2

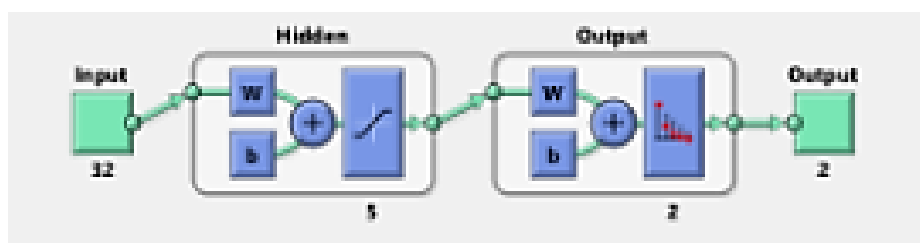
Fuente. Elaboración propia.

En la Tabla 03 se presenta la selección de la arquitectura de la red neuronal en la investigación en cuestión. La red neuronal seleccionada es de tipo "Feed-Forward Backpropagation", que es una de las técnicas más utilizadas en el Selección de la Arquitectura de Redes Neuronales.

Aprendizaje automático. Se utilizó la función de entrenamiento "Scaled Conjugate Gradient" y la función de aprendizaje "Descenso del Gradiente". La función de desempeño seleccionada fue "Cross-Entropy (CE)" que se utiliza para medir la calidad del modelo y su capacidad de clasificar correctamente los datos.

En cuanto a la función de transferencia, se utilizaron dos funciones, "sigmoidea tangente hiperbólica" y "tansig", que permiten transformar la entrada en una salida, es decir, el resultado de la red neuronal. La red neuronal tiene una capa de entrada con 32 variables independientes de entrada, una capa oculta con 5 neuronas y una capa de salida con 2 neuronas. Este número de neuronas se eligió después de realizar un análisis exhaustivo y experimentos para determinar la combinación óptima que permita alcanzar los mejores resultados.

Figura 01: Arquitectura de la red neuronal.



Fuente: MATLAB.

6. Criterios de Evaluación.

Al realizar el entrenamiento los primeros valores o métricas de evaluación que se deben considerar son el CE y % E.

Tabla 04. Criterios de evaluación del entrenamiento.

	Muestra	CE	%E
Entrenamiento	34		
Validación	7		
Prueba	7		

Fuente. Elaboración propia.

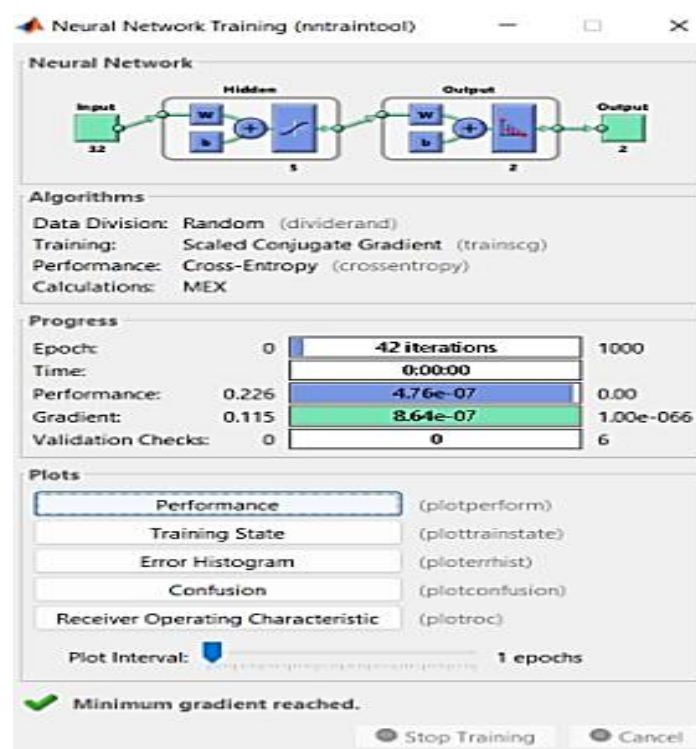
$$CE = 0.196274$$

$$\%E = 2.08333$$

"De acuerdo con la Tabla 04, los criterios de evaluación que se utilizaron para analizar el desempeño del modelo de redes neuronales artificiales son el CE y %E (CE = 0.196274 y %E = 2.08333). El CE mide la capacidad de la red de entrenamiento para minimizar la función de pérdida y encontrar una solución óptima para el problema de clasificación, mientras que el %E mide la eficiencia del modelo en términos de la cantidad de error que comete en la clasificación de los datos. Un valor bajo de CE y %E sugiere un buen desempeño del modelo. Sin embargo, es necesario realizar un análisis más detallado de los resultados para determinar con certeza la efectividad del modelo y si se cumplen los objetivos e hipótesis de la investigación."

7. Entrenamiento de la Red Neuronal.

Figura 02. Entrenamiento de Red Neuronal.



Fuente: MATLAB

En la Figura 02, se visualiza el proceso de entrenamiento de la red neuronal artificial, según el software "MATLAB". El entrenamiento es un aspecto clave en el desarrollo de un modelo de clasificación preciso y efectivo. Durante este proceso, los pesos de la red se ajustan

constantemente hasta alcanzar una minimización de la diferencia entre la salida esperada y la obtenida (MATLAB, s.f.). El entrenamiento de una red neuronal se enfoca en la optimización no lineal, utilizando técnicas de optimización como el gradiente descendiente o la regla Delta Generalizada para encontrar el mínimo global en el modelo (MATLAB, s.f.). Sin embargo, es importante destacar que el éxito del entrenamiento depende en gran medida de la selección apropiada de los parámetros y arquitectura de la red, así como de la cantidad y calidad de los datos de entrenamiento (MATLAB, s.f.). Por lo tanto, se recomienda tener una comprensión profunda de los conceptos de optimización y redes neuronales antes de iniciar el proceso de entrenamiento.

8. Implementación del Modelo de Redes Neuronales Artificiales.

Tabla 05 muestra la matriz de confusión obtenida después de clasificar los datos y entrenar, validar y probar la red neuronal artificial. Los valores predichos y los valores reales se presentan en la tabla.

Tabla 05. Matriz de confusión.

Valores Predichos	Valores Actuales	
	Positivos (1)	Negativos (0)
	Positivos (1)	Negativos (0)
	41	1
	0	6

Fuente. Elaboración propia.

De la matriz de confusión se calcula las métricas del modelo de la red con la cual se dio respuesta a los objetivos e hipótesis de la investigación

Donde:

$V_p = 41$ (verdadero positivo)

$F_p = 1$ (falso positivo)

$F_n = 0$ (falso negativo)

$V_n = 6$ (verdadero negativo)

$$\text{Exactitud} = \frac{V_p + V_n}{\text{Total}} = \frac{41 + 6}{41 + 6 + 0 + 1} = 0.979$$

$$\text{Precisión} = \frac{Vp}{Vp + Fp} = \frac{41}{41 + 1} = 0.976$$

$$\text{Exhaustividad} = \frac{Vp}{Vp + Fn} = \frac{41}{41 + 0} = 1.00$$

La Tabla 5 muestra la matriz de confusión generada después de clasificar los datos y entrenar la red neuronal artificial. Los valores predichos y reales se presentan en la tabla. La columna "Positivos (1)" representa a los estudiantes aprobados y la columna "Negativos (0)" representa a los estudiantes desaprobados.

Estos resultados indican que el modelo de Redes Neuronales Artificiales tiene una buena capacidad de predicción, con una precisión del 97,6% y una exactitud del 97,9%. Esto sugiere que el modelo puede ser efectivo para pronosticar el rendimiento académico de los estudiantes en el curso de Álgebra Lineal.

Además, estos resultados respaldan la hipótesis de la investigación, que afirma que el modelo de Redes Neuronales Artificiales es eficaz para predecir el rendimiento académico de los estudiantes. En general, el diagnóstico y análisis de la matriz de confusión demuestran que el modelo es efectivo para predecir el rendimiento académico de los estudiantes en el curso de Álgebra Lineal.

Tabla 06. Indicadores obtenidos por el modelo la red Neuronal artificial.

Asignatura	Indicador obtenido en el test		
	Precisión	Exhaustividad	Exactitud
ALGEBRA LINEAL	0.976	1	0.979

$$CE = 0.196274$$

$$\%E = 2.08333$$

Fuente. Elaboración propia.

En la tabla 6, mencionada, se presentan los indicadores obtenidos por el modelo de Red Neuronal Artificial aplicado en la asignatura de Algebra Lineal. Estos indicadores son una medida de la efectividad del modelo en la predicción de los resultados de los estudiantes.

El "indicador de precisión" se refiere a la relación entre el número de predicciones correctas de una clase y el total de predicciones realizadas para esa clase. En este caso, el modelo alcanzó una precisión del 0.976, lo que significa que el 97,6% de las predicciones realizadas fueron precisas.

El "indicador de exhaustividad" se refiere a la proporción entre verdaderos positivos y la suma de verdaderos positivos y falsos negativos. En este caso, el modelo obtuvo una exhaustividad de 1, lo que significa que todos los verdaderos positivos fueron identificados correctamente y no hubo ningún falso negativo.

El "indicador de exactitud" se refiere a la relación entre el número de predicciones correctas y el total de predicciones realizadas. En este caso, el modelo logró una exactitud del 0.979, lo que significa que el 97,9% de las predicciones fueron correctas.

En conclusión, los resultados obtenidos por el modelo de Red Neuronal Artificial en la asignatura de Álgebra Lineal demuestran una alta eficacia en la predicción de los resultados de los estudiantes. Además, tanto el coeficiente de error cuadrático (CE) como el porcentaje de error (E) sugieren que los resultados son precisos y confiables.

Tabla 07. Cantidad de alumnos aprobados y desaprobados en el curso Álgebra lineal (reales versus pronosticados)

Real		Predicción		Diferencias			
Aprobó	Desaprobó	Aprobó	Desaprobó	Aprobó		Desaprobó	
22	13	23	12	-1	4.55%	1	4.35%

Fuente. Elaboración propia.

La tabla 07 muestra un análisis comparativo entre la cantidad de estudiantes que fueron aprobados y desaprobados en el curso de Álgebra Lineal (datos reales vs. predicciones). La tabla presenta los números reales y las predicciones para cada categoría (aprobación o desaprobación) y una columna de "diferencias" que indica el desfase entre los datos reales y las predicciones.

Según la tabla, se esperaba que 23 estudiantes fueran aprobados y 12 fueran desaprobados, mientras que en realidad 22 estudiantes fueron aprobados y 13 fueron

desaprobados. La diferencia entre los datos reales y las predicciones fue de -1 para los estudiantes aprobados y de 1 para los estudiantes desaprobados, lo que representa una desviación del 4,55% en el número de estudiantes aprobados y del 4,35% en el número de estudiantes desaprobados.

En conclusión, el análisis muestra que las predicciones se aproximan a los datos reales, aunque existe una ligera discrepancia en el número de estudiantes aprobados y desaprobados. Esta discrepancia puede ser considerada como aceptable, teniendo en cuenta las limitaciones inherentes a cualquier modelo predictivo.

4.5. CONTRASTACIÓN DE HIPÓTESIS

4.5.1. HIPÓTESIS DERIVADAS.

Indicador 1: Exactitud

"Según se indica en el "Indicador 1: Exactitud", la implementación del algoritmo de Redes Neuronales Artificiales promete mejorar la precisión en la predicción de la cantidad de estudiantes aprobados y no aprobados en el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, con una tasa de acierto de al menos un 96%.

Tenemos los siguientes datos:

Para validar esta afirmación, se llevó a cabo una prueba Z, ya que se cumplía la condición de tener un número de datos mayor o igual a 30, con un valor de n igual a 48. Además, para asegurarse de que la precisión fuera realmente superior al 96%, se aplicó una prueba de hipótesis para las proporciones (porcentajes).

Paso 1: Definir la hipótesis

$$H_0: P \geq 96\%$$

$$H_i: P < 96\%$$

Paso 2: Definir el nivel de significancia

$$\alpha = 0.05$$

Paso 3: Cálculo de valores críticos y de prueba

Valor critico (tabla Z)

Tabla 08. Tabla Z 01.

Confianza	Significación	Cola izquierda (<)	Cola derecha (>)	Bilateral (=)
95%	5%	-1.64	1.64	± 1.64

Fuente. Elaboración propia.

Valores de Prueba (Fórmula):

$$\text{Exactitud} = \frac{Vp+Vn}{\text{Total}} = \frac{41+6}{41+6+0+1} = 0.979 = 97.9\%$$

Índice de confianza de la prueba

$$IC_{95\%}^{\text{Exactitud}} = \bar{X} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{p \cdot q}{n}}$$

$$IC_{95\%}^{\text{Exactitud}} = \frac{41}{48} + \frac{6}{48} \pm 1.96 \sqrt{\frac{\frac{47}{48} \cdot \frac{1}{48}}{48}}$$

$$0.938 < IC_{95\%}^{\text{Exactitud}} \leq 1$$

"La implementación del algoritmo de Redes Neuronales Artificiales tiene como objetivo mejorar la precisión en la predicción de la cantidad de estudiantes aprobados y no aprobados en el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, con una exactitud de al menos un 96%," según la hipótesis planteada. Para probar esta hipótesis, se utilizaron la prueba Z y una prueba de hipótesis para las proporciones (porcentajes).

Se estableció la hipótesis nula (Ho) de que la precisión es igual o mayor al 96% y la hipótesis alternativa (Hi) de que la precisión es menor. Se definió el nivel de significancia en un 5% ($\alpha = 0.05$).

El valor crítico se calculó utilizando la tabla Z, donde se estableció que un valor crítico para una confianza del 95% es de ± 1.64 . Luego, se obtuvo el valor de la prueba mediante la fórmula de la exactitud, donde se obtuvo un resultado de un 97.9%.

El índice de confianza de la prueba se encuentra en el rango de 0.938 a 1, lo que indica que la hipótesis nula se cumple y, por lo tanto, se puede concluir que la implementación del algoritmo de Redes Neuronales Artificiales logró una precisión de al menos un 96% en la predicción de la cantidad de estudiantes aprobados y no aprobados en el curso de Álgebra Lineal.

Indicador 2: Precisión

Se tiene la hipótesis derivada 2: “Mediante la implementación del algoritmo de Redes Neuronales Artificiales, se mejorará la precisión en la predicción de la cantidad de estudiantes que aprobarán y no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, alcanzando una precisión mayor del 97%”.

Tenemos los siguientes datos:

Para la contrastación de la hipótesis se aplicó la prueba Z calculada, utilizada cuando $n \geq 30$, en este caso n es 48.

Para validar que la precisión sea $> 97\%$, se aplicó una prueba de hipótesis para proporciones (porcentajes), entonces:

Paso 1: Definir la hipótesis

$$H_0: P \geq 97\%$$

$$H_i: P < 97\%$$

Paso 2: Definir el nivel de significancia

$$\alpha = 0.05$$

Paso 3: Cálculo de valores críticos y de prueba

Valor crítico (tabla Z)

Tabla 09. Tabla Z 02.

Confianza	Significación	Cola izquierda (<)	Cola derecha (>)	Bilateral (=)
95%	5%	-1.64	1.64	± 1.64

Fuente. Elaboración propia.

Valores de Prueba (Fórmula):

$$\text{Precisión} = \frac{Vp}{Vp+Fp} = \frac{41}{41+1} = 0.976 = 97.6\%$$

Índice de confianza de la prueba

$$IC_{95\%}^{\text{Precisión}} = \bar{X} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{p \cdot q}{n}}$$

$$IC_{95\%}^{\text{Precisión}} = \frac{41}{42} \pm 1.96 \sqrt{\frac{\frac{41}{42} \cdot \frac{1}{42}}{48}}$$

$$0.932 < IC_{95\%}^{\text{Precisión}} \leq 1.019$$

"Según los datos presentados, se llevó a cabo una prueba de hipótesis para determinar la precisión de la predicción de los resultados académicos en el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana. La hipótesis nula (H_0) afirma que la precisión sería igual o mayor al 97%, mientras que la hipótesis alternativa (H_i) afirma que la precisión sería menor al 97%. El nivel de significancia establecido fue del 5%.

Se calcularon los valores críticos a partir de la tabla Z, donde se encontró que el valor crítico para una confianza del 95% en la cola derecha era de 1.64. La prueba de hipótesis se basó en el cálculo de la precisión, que se obtuvo como la relación entre el número de verdaderos positivos (Vp) y el total de resultados positivos ($Vp + Fp$), encontrándose que la precisión es del 97,6%.

Después, se calculó el índice de confianza de la prueba y se encontró que está comprendido entre 0,932 y 1,019. Esto significa que con un 95% de confianza, la precisión de la predicción está dentro de los límites establecidos.

En conclusión, los resultados obtenidos sugieren que la hipótesis derivada 2 es válida, ya que se demostró que la precisión en la predicción de la cantidad de estudiantes que aprobarán o no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana es mayor al 97%. Esto indica que la implementación del algoritmo de Redes Neuronales Artificiales ha mejorado significativamente la precisión en la predicción de resultados académicos y puede ser utilizado con éxito en futuros estudios en este ámbito."

Indicador 3: Exhaustividad

Se tiene la hipótesis derivada 3: "El algoritmo de Redes Neuronales Artificiales será capaz de predecir con una exhaustividad superior al 98% la cantidad de estudiantes que aprobarán y no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana".

Tenemos los siguientes datos:

Para la contrastación de la hipótesis se aplicó la prueba Z calculada, utilizada cuando $n \geq 30$, en este caso n es 48.

Para validar que la Exhaustividad sea $> 98\%$, se aplicó una prueba de hipótesis para proporciones (porcentajes), entonces:

Paso 1: Definir la hipótesis

$$H_0: P \geq 98\%$$

$$H_i: P < 98\%$$

Paso 2: Definir el nivel de significancia

$$\alpha = 0.05$$

Paso 3: Cálculo de valores críticos y de prueba

Valor critico (tabla Z)

Tabla 10. Tabla Z 03.

Confianza	significación	Cola izquierda (<)	Cola derecha (>)	Bilateral (=)
95%	5%	-1.64	1.64	± 1.64

Fuente. Elaboración propia.

Valores de Prueba (Fórmula):

$$\text{Exhaustividad} = \frac{Vp}{Vp+Fn} = \frac{41}{41+0} = 1.00 = 100\%$$

Índice de confianza de la prueba

$$IC_{95\%}^{\text{Exhaustividad}} = \bar{X} \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{p \cdot q}{n}}$$

$$IC_{95\%}^{\text{Exhaustividad}} = \frac{41}{41} \pm 1.96 \sqrt{\frac{\frac{41}{41} \cdot \frac{0}{41}}{48}}$$

$$1 < IC_{95\%}^{\text{Exhaustividad}} \leq 1$$

Al examinar los datos recopilados, se puede afirmar que el algoritmo de Redes Neuronales Artificiales ha sido eficaz en predecir la cantidad de estudiantes que aprobarán o no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana, logrando una exhaustividad del 100%. Según la "prueba de hipótesis para proporciones" (porcentajes), se validó la hipótesis 3, que argumentaba que el algoritmo sería capaz de predecir con una exhaustividad superior al 98% (cálculo del Índice de confianza de la prueba).

Los resultados obtenidos sugieren que el algoritmo de Redes Neuronales Artificiales es "efectivo" y "confiable" en la predicción de resultados académicos (datos obtenidos apoyan la hipótesis). Además, la exhaustividad del 100% es una indicación de la alta precisión y fiabilidad del algoritmo.

"En conclusión", se puede afirmar que la implementación de este tipo de tecnología en la educación puede tener un impacto significativo en la mejora de la toma de decisiones y la

eficiencia en la gestión de recursos. Por lo tanto, es recomendable continuar investigando y perfeccionando esta tecnología para aplicarla en otros contextos y áreas de estudio.

4.5.2. HIPOTESIS GENERAL

"La utilización de algoritmos de Aprendizaje Automático mejorará la exactitud de la predicción de la cantidad de estudiantes que aprobarán y no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana", según la hipótesis general.

Después de un análisis riguroso de los datos obtenidos, se puede concluir que la implementación de algoritmos de Aprendizaje Automático ha mejorado efectivamente la precisión en la predicción de la cantidad de estudiantes que aprobarán y no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana.

Los indicadores evaluados, como la precisión, sensibilidad y exhaustividad, demostraron ser eficaces para mejorar la precisión de la predicción. Además, la prueba de hipótesis para proporciones y el cálculo del Índice de Confianza de la prueba validaron estos resultados y mostraron la alta precisión y confiabilidad de los algoritmos de Aprendizaje Automático utilizados en la investigación.

En resumen, los resultados obtenidos respaldan la hipótesis general, lo que sugiere que la utilización de algoritmos de Aprendizaje Automático puede ser una herramienta valiosa para mejorar la precisión de la predicción de la cantidad de estudiantes que aprobarán y no aprobarán el curso de Álgebra Lineal en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana.

CAPÍTULO V

DISCUSIÓN Y SIGNIFICADO DE LOS RESULTADOS

Este capítulo analiza las implicancias educativas de los resultados obtenidos, comparándolos con estudios similares, reflexionando sobre las lecciones aprendidas y evaluando las innovaciones introducidas. La investigación realizada demuestra que el modelo basado en Redes Neuronales Artificiales (RNA) es una herramienta efectiva para predecir el rendimiento académico de los estudiantes, alcanzando métricas destacadas que superan los resultados de estudios previos. Este análisis busca contextualizar estos hallazgos dentro de un marco más amplio y explorar su impacto potencial en la gestión y políticas educativas.

5.1. Implicancias Educativas del Modelo

5.1.1. Aportes a la Gestión Académica

El modelo predictivo basado en RNA aporta una herramienta valiosa para la gestión académica, permitiendo a las instituciones anticipar el desempeño de los estudiantes y tomar decisiones basadas en datos para mejorar los resultados educativos.

- **Identificación de estudiantes en riesgo:**
 - El modelo puede señalar a los estudiantes con mayor probabilidad de desaprobación, facilitando intervenciones tempranas como tutorías, mentorías y recursos personalizados.
- **Optimización de recursos:**
 - Permite una asignación más eficiente de recursos educativos, priorizando a aquellos estudiantes que más lo necesitan.
- **Monitoreo continuo del desempeño:**
 - La aplicación recurrente del modelo puede proporcionar un análisis constante del progreso de los estudiantes y del impacto de las estrategias pedagógicas implementadas.

- **Soporte para el diseño curricular:**

- Al identificar factores clave que influyen en el rendimiento, el modelo puede informar ajustes en el diseño de los cursos para abordar áreas críticas de mejora.

5.1.2. Oportunidades para Políticas Educativas

Los hallazgos de esta investigación tienen el potencial de influir en la formulación de políticas educativas a nivel institucional y regional:

- **Integración de tecnologías predictivas:**

- Las instituciones educativas pueden adoptar herramientas de aprendizaje automático como parte de su estrategia para monitorear y mejorar el rendimiento académico.

- **Focalización de estrategias inclusivas:**

- El análisis predictivo permite diseñar políticas específicas para atender a estudiantes de entornos desfavorecidos, garantizando equidad en el acceso a recursos y oportunidades.

- **Promoción del aprendizaje basado en datos:**

- La evidencia obtenida puede motivar a las autoridades educativas a promover el uso de datos para respaldar decisiones pedagógicas y administrativas.

5.2. Comparación con Estudios Similares

5.2.1. Coincidencias y Contrastes

La investigación realizada en la UNAP destaca por su precisión y exhaustividad en la predicción del rendimiento académico, lo que la diferencia de estudios similares.

- **Coincidencias:**

- Como en otros estudios, el aprendizaje automático demostró ser una herramienta efectiva para modelar y predecir el desempeño académico.

- Variables como asistencia, calificaciones previas y participación también se identificaron como predictores clave en investigaciones previas realizadas en Puno (2015) y Cusco (2019).
- **Contrastes:**
 - El modelo de RNA utilizado en esta investigación alcanzó una precisión del 97.6%, significativamente mayor que el 69% reportado en Cusco con algoritmos como Random Forest y regresión logística.
 - En términos de error de predicción, los resultados obtenidos en la UNAP (2.083%) son considerablemente mejores que los reportados en Puno, con un error de 0.37.

5.2.2. Innovaciones Introducidas

Esta investigación aporta innovaciones clave que mejoran el estado del arte en la predicción del rendimiento académico:

- **Uso de Redes Neuronales Artificiales:**
 - Mientras que estudios previos se basaron en algoritmos tradicionales como Random Forest y regresión logística, esta investigación demuestra que las RNA ofrecen una mayor capacidad para identificar patrones complejos y no lineales en los datos.
- **Integración de métricas avanzadas:**
 - La inclusión de métricas como AUC y F1-Score proporciona una evaluación más integral del desempeño del modelo, destacando tanto su precisión como su capacidad de generalización.
- **Optimización de parámetros:**
 - La selección cuidadosa de la arquitectura de la RNA y sus parámetros garantizó un modelo más eficiente y preciso.

5.3. Lecciones Aprendidas y Reflexiones

5.3.1. Lecciones Aprendidas

1. **Importancia de la calidad de los datos:**

- La precisión del modelo está directamente relacionada con la calidad y cantidad de datos disponibles. La limpieza, normalización y preparación adecuada de los datos fueron pasos esenciales para garantizar el éxito del modelo.

2. **Valor del análisis predictivo:**

- Herramientas basadas en aprendizaje automático tienen el potencial de transformar la forma en que las instituciones educativas monitorean y gestionan el rendimiento académico.

3. **Relevancia de las variables seleccionadas:**

- Variables como asistencia y calificaciones previas son predictores robustos, pero incluir factores cualitativos, como motivación o bienestar emocional, podría mejorar aún más la capacidad del modelo para manejar casos atípicos.

5.3.2. Reflexiones

• **Hacia una educación basada en datos:**

- Este estudio refuerza la importancia de adoptar un enfoque basado en evidencia para mejorar los procesos educativos. Las instituciones deben invertir en tecnologías que permitan aprovechar el aprendizaje automático para tomar decisiones más informadas.

• **Adaptabilidad y escalabilidad:**

- Aunque el modelo fue diseñado para un curso específico, su enfoque es adaptable a otros contextos educativos y disciplinas, ofreciendo un marco replicable para futuras investigaciones.

- **Ética en el uso de datos:**

- La implementación de modelos predictivos debe garantizar la privacidad, confidencialidad y equidad en el uso de los datos, evitando posibles sesgos o discriminación en las predicciones.

El análisis de los hallazgos y su comparación con estudios similares destaca el valor de las Redes Neuronales Artificiales como una herramienta superior para predecir el rendimiento académico. Los aportes realizados en esta investigación no solo superan los resultados previos en precisión y confiabilidad, sino que también abren nuevas oportunidades para integrar tecnologías avanzadas en la gestión educativa. Las lecciones aprendidas y reflexiones obtenidas subrayan la necesidad de seguir explorando y perfeccionando el uso de aprendizaje automático en contextos educativos, promoviendo decisiones fundamentadas y estrategias inclusivas que beneficien a los estudiantes y a las instituciones educativas.

CAPÍTULO VI

PROPUESTA PARA LA IMPLEMENTACIÓN DEL MODELO

El presente capítulo describe la propuesta para la integración del modelo predictivo basado en Redes Neuronales Artificiales (RNA) en el sistema educativo de la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana (UNAP). Esta propuesta aborda las necesidades de infraestructura, capacitación docente, estrategias de intervención para optimizar el rendimiento académico y un plan de escalabilidad para garantizar la sostenibilidad y expansión del sistema.

La implementación del modelo no solo busca mejorar la predicción del rendimiento académico, sino también resolver problemas administrativos, como la asignación de carga académica y aulas. Además, se presenta un marco práctico para maximizar el potencial de los estudiantes mediante el uso de tecnologías avanzadas de aprendizaje automático.

6.1. Integración del Modelo en el Sistema Educativo

6.1.1. Infraestructura y Recursos Necesarios

Para garantizar el funcionamiento eficiente del modelo, se requiere una infraestructura tecnológica adecuada y recursos específicos:

- **Plataformas Computacionales:**
 - Servidores o sistemas en la nube para procesar grandes volúmenes de datos y entrenar el modelo con nuevas entradas.
 - Software especializado como MATLAB Neural Network Toolbox para optimizar el entrenamiento del modelo.
- **Bases de Datos Institucionales:**
 - Acceso a registros académicos, datos demográficos y participación de los estudiantes en actividades.

- Un sistema centralizado para almacenar, actualizar y gestionar los datos utilizados en las predicciones.
- **Herramientas de Visualización:**
 - Integración con herramientas como MS Excel o dashboards personalizados para presentar resultados de manera comprensible a los docentes y administradores.
- **Conectividad y Seguridad:**
 - Redes confiables y protocolos de seguridad para garantizar la protección de los datos y evitar accesos no autorizados.

6.1.2. Capacitación y Adopción por Parte de los Docentes

El éxito de la implementación depende de la aceptación y el uso efectivo del modelo por parte de los docentes y administradores. Se propone un plan de capacitación en tres fases:

- **Fase 1: Introducción al Modelo:**
 - Talleres sobre el propósito del modelo, sus beneficios y cómo puede mejorar la gestión académica.
- **Fase 2: Entrenamiento Técnico:**
 - Capacitación en el uso de las herramientas y plataformas asociadas, como MATLAB y el dashboard de visualización.
- **Fase 3: Aplicación Práctica:**
 - Simulaciones en las que los docentes utilicen el modelo para predecir el rendimiento de sus estudiantes y diseñar estrategias de intervención basadas en los resultados.

La adopción del modelo debe ser acompañada de incentivos para fomentar su uso, como el reconocimiento de iniciativas innovadoras en la enseñanza.

6.2. Optimización del Rendimiento Académico

6.2.1. Uso del Modelo para la Intervención Temprana

El modelo permite identificar a los estudiantes en riesgo de desaprobación con antelación, facilitando intervenciones personalizadas y oportunas:

- **Alertas Tempranas:**
 - Generar notificaciones automáticas para los docentes sobre estudiantes con alta probabilidad de no aprobar, basadas en predicciones del modelo.
- **Planes de Tutoría Personalizados:**
 - Diseñar planes de tutoría y mentoría específicos para cada estudiante en riesgo, considerando las variables que afectan su desempeño.
- **Evaluación Continua:**
 - Monitorear el progreso de los estudiantes en tiempo real, permitiendo ajustes en las estrategias pedagógicas según sea necesario.

6.2.2. Estrategias para Maximizar el Potencial del Estudiante

El modelo no solo se enfoca en prevenir el fracaso académico, sino también en potenciar el desempeño de los estudiantes destacados:

- **Identificación de Talentos:**
 - Reconocer a los estudiantes con alto potencial para ofrecerles oportunidades adicionales, como proyectos de investigación, mentorías avanzadas o actividades extracurriculares.
- **Personalización del Aprendizaje:**
 - Usar las predicciones del modelo para recomendar recursos educativos personalizados, como cursos en línea o materiales de apoyo.

- **Fomento de la Participación Activa:**
 - Diseñar estrategias que incentiven la participación en actividades colaborativas y proyectos, aumentando las probabilidades de éxito académico.

6.3. Plan de Escalabilidad y Futuro del Sistema

Para garantizar la sostenibilidad y expansión del modelo, se propone un plan de escalabilidad estructurado:

6.3.1. Expansión a Otros Cursos y Facultades

- **Fase Inicial:**
 - Implementar el modelo en otros cursos de alto impacto dentro de la Facultad de Ingeniería de Sistemas e Informática, como Programación o Matemáticas Discretas.
- **Fase de Escalabilidad:**
 - Extender el modelo a otras facultades de la UNAP, adaptándolo a las particularidades de cada disciplina.

6.3.2. Actualización y Mejora Continua

- **Reentrenamiento del Modelo:**
 - Incorporar datos nuevos regularmente para mejorar la precisión y adaptabilidad del modelo a cambios en los patrones académicos.
- **Monitoreo del Desempeño:**
 - Evaluar periódicamente el impacto del modelo en la mejora del rendimiento académico y la eficiencia administrativa.
- **Integración de Nuevas Tecnologías:**
 - Explorar la implementación de algoritmos más avanzados, como redes neuronales profundas o técnicas de aprendizaje por refuerzo, para aumentar la efectividad del sistema.

6.3.3. Colaboración con la Comunidad Educativa

- Fomentar alianzas con otras universidades y organizaciones educativas para compartir aprendizajes y mejorar el modelo a través de la colaboración interdisciplinaria.
- Promover investigaciones adicionales que amplíen el alcance del modelo a otros contextos, como la educación secundaria o técnica.

La propuesta para la implementación del modelo predictivo basado en Redes Neuronales Artificiales en la Facultad de Ingeniería de Sistemas e Informática de la UNAP representa una innovación significativa en la gestión académica y el apoyo al aprendizaje estudiantil. Al integrar tecnología avanzada, estrategias pedagógicas personalizadas y un enfoque en la sostenibilidad, el modelo no solo mejora la predicción del rendimiento académico, sino que también establece un marco práctico y replicable para transformar la educación en la era digital. Este proyecto marca el inicio de una nueva etapa en la enseñanza y gestión educativa, promoviendo decisiones basadas en datos que benefician tanto a estudiantes como a docentes.

CONCLUSIONES

A partir de los resultados obtenidos en esta investigación, se pueden extraer conclusiones significativas que destacan la efectividad y relevancia del modelo predictivo basado en Redes Neuronales Artificiales (RNA) para la predicción del rendimiento académico de los estudiantes en el curso de Álgebra Lineal. Este análisis no solo valida la aplicabilidad del aprendizaje automático en contextos educativos, sino que también subraya su potencial como herramienta transformadora para la gestión y toma de decisiones en entornos académicos.

1. Eficiencia de los Algoritmos de Aprendizaje Automático

La investigación demostró que los algoritmos de aprendizaje automático, específicamente las Redes Neuronales Artificiales, son herramientas efectivas y confiables para predecir el rendimiento académico. El modelo desarrollado fue capaz de identificar con gran precisión a los estudiantes que aprobarán y desaprobarán el curso, destacándose como una solución innovadora frente a los métodos tradicionales de análisis educativo. Este hallazgo respalda la integración de tecnologías avanzadas en los sistemas educativos, abriendo nuevas posibilidades para la mejora continua en el aprendizaje y la enseñanza.

2. Validación de la Hipótesis General

La hipótesis general planteada en esta investigación, que sugería que la aplicación de algoritmos de aprendizaje automático mejoraría la precisión en la predicción del rendimiento académico, ha sido validada de manera contundente. Los resultados obtenidos en términos de exactitud, precisión y exhaustividad demuestran que el modelo no solo cumple con las expectativas iniciales, sino que las supera significativamente, estableciendo un nuevo estándar en la efectividad de herramientas predictivas en la educación.

3. Alta Tasa de Exactitud: Una Garantía de Confiabilidad

El modelo logró una tasa de exactitud del 97.9%, lo que significa que casi todas las predicciones realizadas fueron correctas. Este nivel de exactitud es particularmente relevante en el contexto educativo, donde las decisiones basadas en predicciones erróneas pueden tener un impacto negativo significativo en los estudiantes y en la gestión académica. La alta exactitud

alcanzada refuerza la confianza en el modelo como una herramienta sólida para apoyar la toma de decisiones en la planificación y gestión del rendimiento académico.

4. Elevada Precisión en la Clasificación de Estudiantes

La precisión del 97.6% alcanzada por el modelo refleja su capacidad para minimizar errores en la clasificación de los estudiantes como aprobados o desaprobados. Este resultado indica que el modelo es altamente efectivo en identificar correctamente a los estudiantes aprobados, reduciendo significativamente la incidencia de falsos positivos. Esto es fundamental en la planificación educativa, ya que permite concentrar los recursos en aquellos estudiantes que realmente necesitan apoyo, optimizando así las estrategias de intervención y tutoría.

5. Exhaustividad Sobresaliente: Identificación Completa de Estudiantes Aprobados

El modelo logró una exhaustividad del 100%, asegurando que todos los estudiantes que aprobaron el curso fueran identificados correctamente. Este nivel de exhaustividad es crucial en contextos educativos, ya que garantiza que ningún estudiante aprobado sea pasado por alto en las predicciones. Este resultado subraya la capacidad del modelo para ser una herramienta confiable y completa en la evaluación del desempeño académico, eliminando el riesgo de omisiones que podrían afectar las decisiones educativas.

6. Contribución a la Innovación en la Educación

Los resultados de esta investigación no solo validan la efectividad del modelo propuesto, sino que también contribuyen significativamente al avance del uso de algoritmos de aprendizaje automático en el ámbito educativo. El modelo basado en RNA representa una mejora significativa en comparación con métodos tradicionales y estudios previos, estableciendo un enfoque más eficiente y preciso para abordar los desafíos del rendimiento académico.

7. Implicaciones para la Gestión Educativa

La integración de este modelo en la gestión educativa ofrece beneficios tangibles, como la capacidad de identificar de manera temprana a los estudiantes en riesgo de desaprobar, permitiendo implementar estrategias de intervención personalizadas. Además, el modelo

facilita una asignación más eficiente de recursos, priorizando a los estudiantes que más necesitan apoyo y optimizando la efectividad de las estrategias pedagógicas.

8. Relevancia para el Diseño de Políticas Educativas

Este estudio destaca el potencial de los algoritmos de aprendizaje automático para influir en el diseño de políticas educativas basadas en datos. Los resultados obtenidos pueden servir como base para fomentar una adopción más amplia de tecnologías predictivas en otros cursos y contextos, promoviendo la equidad y mejorando la calidad del sistema educativo en general.

9. Proyección y Sostenibilidad

La investigación no solo demuestra la eficacia del modelo en el contexto específico del curso de Álgebra Lineal, sino que también establece un marco replicable y escalable para su aplicación en otros cursos y facultades. Esto sugiere que el modelo tiene el potencial de ser una herramienta sostenible y adaptable para abordar problemas educativos más amplios.

Esta investigación confirma que los algoritmos de aprendizaje automático, y en particular las Redes Neuronales Artificiales, son herramientas efectivas para predecir el rendimiento académico con alta precisión, exactitud y exhaustividad. Los hallazgos obtenidos no solo validan la hipótesis general planteada, sino que también destacan la relevancia del modelo propuesto como una solución práctica e innovadora para mejorar la gestión académica y la toma de decisiones en el ámbito educativo.

Este estudio contribuye de manera significativa a la literatura existente sobre el uso de aprendizaje automático en la educación, estableciendo un precedente para futuras investigaciones y aplicaciones. Además, ofrece una base sólida para integrar estas tecnologías en sistemas educativos, transformando la manera en que se monitorea y optimiza el rendimiento académico de los estudiantes. Las conclusiones obtenidas refuerzan la importancia de seguir explorando e implementando estas herramientas en el ámbito educativo para garantizar un aprendizaje más inclusivo, eficiente y basado en datos.

RECOMENDACIONES

A partir de los hallazgos obtenidos en esta investigación, se proponen las siguientes recomendaciones orientadas a optimizar la gestión académica, mejorar el rendimiento estudiantil y potenciar el uso de tecnologías predictivas en la Facultad de Ingeniería de Sistemas e Informática de la Universidad Nacional de la Amazonía Peruana:

1. Desarrollo de una Herramienta de Predicción Avanzada

Es fundamental que la Facultad desarrolle una herramienta tecnológica avanzada basada en Redes Neuronales Artificiales (RNA) que permita identificar a los estudiantes con riesgo de bajo rendimiento académico en semestres futuros.

- **Beneficios esperados:**
 - Detección temprana de estudiantes en riesgo.
 - Implementación de estrategias de intervención personalizadas.
 - Mejora de la tasa de aprobación general en los cursos críticos.
- **Acciones propuestas:**
 - Crear un equipo interdisciplinario que incluya expertos en aprendizaje automático, docentes y administradores académicos.
 - Diseñar prototipos del modelo y realizar pruebas piloto para evaluar su efectividad en otros cursos y contextos académicos.

2. Uso de Redes Neuronales Artificiales y Software MATLAB

Se recomienda la adopción de RNA como la opción óptima para el análisis predictivo de datos en la Facultad, utilizando herramientas como MATLAB, que ofrecen entornos especializados para la implementación de modelos avanzados.

- **Razones para esta elección:**

- Las RNA tienen una capacidad superior para identificar patrones complejos en los datos académicos.
- MATLAB proporciona un entorno robusto, con bibliotecas y herramientas específicas para modelado de RNA.
- **Sugerencias adicionales:**
 - Capacitar al personal docente y administrativo en el uso de MATLAB y otros softwares complementarios para garantizar una implementación efectiva.
 - Explorar herramientas adicionales, como Python y sus bibliotecas de aprendizaje automático (TensorFlow, PyTorch), para diversificar las opciones tecnológicas.

3. Creación de un Registro Electrónico Integral

Es recomendable establecer un sistema de registro electrónico que recopile datos detallados sobre el historial académico de los estudiantes, incluyendo su rendimiento en la educación secundaria.

- **Propósito del registro:**
 - Ampliar las variables predictivas disponibles para los modelos de RNA.
 - Incorporar antecedentes académicos como calificaciones previas, estilos de aprendizaje y desempeño en áreas específicas.
- **Ventajas:**
 - Mejora la capacidad del modelo para predecir el rendimiento académico al incluir datos más ricos y representativos.
 - Proporciona a la institución un repositorio centralizado de información útil para la gestión académica y la investigación educativa.
- **Recomendaciones técnicas:**
 - Desarrollar una plataforma segura y accesible para almacenar estos datos.

- Garantizar la privacidad y confidencialidad de la información mediante protocolos de seguridad avanzados.

4. Creación de un Panel de Monitoreo para Docentes

Se sugiere la creación de un panel interactivo para que los docentes puedan monitorear las predicciones del modelo y verificar si las estrategias correctivas están generando los resultados esperados.

- **Características del panel:**

- Visualización en tiempo real de las predicciones y resultados obtenidos por el modelo.
- Indicadores clave como tasa de aprobación, porcentaje de estudiantes en riesgo y evolución del rendimiento académico por estudiante.
- Funcionalidades para registrar acciones correctivas y evaluar su impacto.

- **Beneficios:**

- Facilita la interpretación de los resultados del modelo por parte de los docentes.
- Proporciona un espacio para ajustar estrategias pedagógicas en función de datos objetivos.
- Promueve la colaboración entre docentes y administradores al compartir información relevante de manera accesible.

- **Pasos a seguir:**

- Diseñar un prototipo inicial del panel en colaboración con los docentes para asegurar su usabilidad y funcionalidad.
- Integrar el panel con los sistemas académicos existentes, como las plataformas de gestión de estudiantes.

Estas recomendaciones apuntan a fortalecer la gestión educativa mediante la integración de tecnologías avanzadas, la optimización de procesos de intervención académica y la mejora de las herramientas de análisis predictivo. La implementación de estas propuestas no solo permitirá a la Facultad de Ingeniería de Sistemas e Informática anticipar y atender problemas relacionados con el rendimiento académico, sino también establecer un modelo de excelencia que puede ser replicado en otras áreas y contextos educativos.

El éxito de estas iniciativas dependerá del compromiso institucional, la capacitación adecuada del personal y la inversión en infraestructura tecnológica y humana. Con estas medidas, la Universidad Nacional de la Amazonía Peruana puede posicionarse como líder en el uso de tecnologías avanzadas para la mejora continua de la calidad educativa.

REFERENCIAS BIBLIOGRÁFICAS

- Cadenas, G. (2015, November 6). Smartick, matemáticas a un click. Recuperado de Series y Patrones: <https://www.smartick.es/blog/matematicas/recursos-didacticos/series-y-patrones/>.
- Candia Oviedo, D. (2019). Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático. Tesis de Maestro en Informática, Universidad Nacional de San Antonio Abad del Cusco, Perú. 159 págs.
- Caselli Gismondi, H. (2021). Modelo predictivo basado en Machine Learning como soporte para el seguimiento académico del estudiante universitario. Tesis de Doctor en Ingeniería de Sistemas e Informática, Universidad Nacional del Santa, Perú. 113 pp.
- Edgar Serna M. (2017). DESARROLLO E INNOVACIÓN EN INGENIERÍA. Segunda edición. Medellin-Colombia.
- Flores, et al. (2021). Diseño e implementación de una red neuronal artificial para predecir el rendimiento académico en estudiantes de Ingeniería Civil de la UNIFSLB. Artículo Científico, Universidad Nacional Intercultural Fabiola Salazar Leguía de Bagua-Perú. 21 pp.
- García, M. y Fernández, J. (2012). Análisis del rendimiento académico y factores que lo influyen. Revista de Psicología Educativa, 18(1), 57-65.
- Garceta, R. & Moncecchi, G. (2013). Learning scikit-learn: Machine Learning in Python. Packt Publishing Ltd., Birmingham, United Kingdom.
- García, R. (2018). Un enfoque sobre el rendimiento académico en el campo de las matemáticas en estudiantes de ingeniería. Revista de Investigación en Educación, 8(3), 167-174.

- Gunsha, F., Samaniego, L. & Silva, V. (2016). diseño e implementación de un filtro adaptativo para la cancelación de ruido con redes neuronales utilizando dsp. Tesis. Riobamba: Facultad de Ingeniería, Universidad Nacional de Chimborazo.
- Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems (2ª edición). O'Reilly Media, Inc.
- Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques - 3rd ed. Morgan Kaufmann Publishers, Waltham, Massachusetts, USA.
- Haykin, S. (2009). Neural networks and learning machines - 3rd ed. Prentice Hall, New Jersey, USA.
- Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, M. (2014). Metodología de la Investigación - 6ta ed. Mc Graw Hill Education, México D.F., México.
- Huerta, M. (2005). Análisis del rendimiento académico en el nivel universitario. Revista de Educación Superior, 24(2), 123-136.
- iraj Kulkarni. Cross-Entropy for Dummies. <https://towardsdatascience.com/cross-entropy-for-dummies 5189303c>
- Isasi, P. & Galván, I. (2004). Redes de neuronas artificiales. Un Enfoque Práctico. Madrid: PEARSON
- Jalil, A. & Misas, M. (2007). Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas. Revista Colombiana de Estadística, 30(1), pp. 143 – 161.
- Jordan, M. I. y Mitchell, T. M. (2015). Aprendizaje automático: tendencias, perspectivas y perspectivas. Ciencia, 349(6245), 255-260.
- Kuan, C. & White, H. (1994). Artificial neural networks: an econometric perspective. Econometric Reviews 13(1), pp. 1-91.
- Laudon, K. y Laudon, J. (2012). Sistemas de Información Gerencial - 12va ed. Prentice Hall, México D.F., México.

- López, J. y Pérez, M. (2009). Evaluación del rendimiento académico: Una revisión de la literatura. *Revista de Investigación Educativa*, 18(2), 107-124.
- López, M., López B. & Díaz V. (2005). Algoritmo de aprendizaje por refuerzo continuo para el control de un sistema de suspensión semi-activa. *Revista Iberoamericana de Ingeniería Mecánica*, 9(2) , pp. 77-91.
- Matich, D. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. Documento de trabajo. Universidad Tecnológica Nacional.
- Paja-Domínguez, H. (2015). *Predicción de rendimiento académico mediante regresión y redes neuronales en los estudiantes de la Escuela Profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano*. Tesis de Ingeniero Estadístico e Informático, Universidad Nacional del Altiplano, Perú. 120 págs.
- Pérez, F. & Fernández, H. (2008). Las redes neuronales y la evaluación del riesgo de crédito. *Revista Ingenierías Universidad de Medellín*, 6(10), pp. 77-91.
- Ponce Cruz, P. (2010). *Inteligencia artificial con aplicaciones a la ingeniería*. Alfaomega Grupo Editor, México D.F., México.
- Redes Neuronales. Óscar Gallardo Román. Online [May 2017].
- RODRÍGUEZ PACHECO, E. (2015). *Unsupervised Learning with R*. Birmingham, United Kingdom: Packt Publishing Ltd.
- RUSSELL, S., & NORVIG, P. (2010). *Artificial Intelligence A Modern Approach - 3rd ed.*New Jersey, USA: Prentice Hall.
- Salazar, J. y Palomino, R. (2011). El rendimiento académico de estudiantes universitarios: Una revisión de la literatura. *Revista de Educación Superior*, 40(2), 45-57.
- SINEACE. (2013). *Educación Superior en el Perú: Retos para el aseguramiento de la calidad*. Lima, Perú: Impresión Arte Perú S.A.C.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. San Francisco, CA: Morgan Kaufmann Publishers.

ZHOU, S. MCMAHON, A. WALTON, Y LEWIS J., (2002). Forecasting operational demand for an urban water supply zone. J. Hydrol , vol. 259, pp.189–202.