

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS
FACULTAD DE INGENIERÍA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS



Presentado por

GERALDINE NILA FLORES MOSCOSO

**PREDICCIÓN DEL NIVEL DE ESTRÉS ACADÉMICO EN
ESTUDIANTES DE LA UNIVERSIDAD NACIONAL JOSÉ
MARÍA ARGUEDAS APLICANDO UN MODELO DE
INTELIGENCIA ARTIFICIAL**

Asesor:

IVÁN SORIA SOLÍS

**TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
INGENIERO DE SISTEMAS**

ANDAHUAYLAS – APURIMAC – PERÚ

2024



APROBACIÓN DEL ASESOR

Quien suscribe:

M.Sc. Iván Soria Solís por la presente:

CERTIFICA,

Que, el Bachiller en Ingeniería de Sistemas, GERALDINE NILA FLORES MOSCOSO ha culminado satisfactoriamente el informe final de tesis intitulado: "PREDICCIÓN DEL NIVEL DE ESTRÉS ACADÉMICO EN ESTUDIANTES DE LA UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS APLICANDO UN MODELO DE INTELIGENCIA ARTIFICIAL" para optar el Título Profesional de Ingeniero de Sistemas.

Andahuaylas, 16 de enero del 2024

M.Sc. Iván Soria Solís
Asesor

Geraldine Nila Flores Moscoso
Tesista



FACULTAD DE INGENIERIA

**ACTA DE SUSTENTACIÓN
DE TESIS**

En la Av. José María Arguedas del Local Académico SL01 (Ccoyahuacho) en el auditorio de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Nacional José María Arguedas ubicado en el distrito de San Jerónimo de la Provincia de Andahuaylas, siendo las 11:00 horas del día 09 de enero del año 2024, se reunieron los docentes: MSc. Richard Carrión Abollaneda, MSc. Jesús Farfán Inca Roca, MSc. Neptalí Menejes Palomino, en condición de integrantes del Jurado Evaluador del Informe Final de Tesis intitulado: "PREDICCIÓN DEL NIVEL DE ESTRÉS ACADÉMICO EN ESTUDIANTES DE LA UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS APLICANDO UN MODELO DE INTELIGENCIA ARTIFICIAL", cuya autora es la Bachiller en Ingeniería de Sistemas **GERALDINE NILA FLORES MOSCOSO** y el asesor Msc. Iván Soria Solís, con el propósito de proceder a la sustentación y defensa de dicha tesis.

Luego de la sustentación y defensa de la tesis, el Jurado Evaluador **ACORDÓ: APROBAR** por **UNANIMIDAD** a la Bachiller en Ingeniería de Sistemas **GERALDINE NILA FLORES MOSCOSO**, obteniendo la siguiente calificación y mención:

Nota escala vigesimal		Mención
Números	Letras	
16	DIECISEIS	MUY BUENO

En señal de conformidad, se procedió a la firma de la presente acta en 03 ejemplares.


MSc. Richard Carrión Abollaneda
Presidente del Jurado Evaluador


MSc. Jesús Farfán Inca Roca
Primer Miembro del Jurado Evaluador


MSc. Neptalí Menejes Palomino
Segundo Miembro del Jurado Evaluador



APROBACIÓN DEL JURADO DICTAMINADOR

LA TESIS: "Predicción del Nivel de Estrés Académico en Estudiantes de la Universidad Nacional José María Arguedas Aplicando un Modelo de Inteligencia Artificial"; para optar el Título Profesional de Ingeniero de Sistemas, ha sido evaluada por el Jurado Dictaminador conformado por:

PRESIDENTE: M.Sc. Richard Carrión Abollaneda

PRIMER MIEMBRO: M.Sc. Jesús Farfán Inca Roca

SEGUNDO MIEMBRO: M.Sc. Neptalí Menejes Palomino

Habiendo sido aprobado por UNANIMIDAD/MAYORIA, en la ciudad de Andahuaylas el día 09 del mes de enero del 2024.

Andahuaylas, 16 de enero del 2024.

M.Sc. Richard Carrión Abollaneda
Presidente del Jurado Evaluador

M.Sc. Jesús Farfán Inca Roca
Primer Miembro del Jurado Evaluador

M.Sc. Neptalí Menejes Palomino
Segundo Miembro del Jurado Evaluador



Universidad Nacional José María Arguedas

Identidad y Excelencia para el Trabajo Productivo y el Desarrollo

Unidad de Investigación de la Facultad Ingeniería
C011-2024

Andahuaylas, 20 de mayo de 2024

La Unidad de Investigación de la Facultad de Ingeniería, expide la:

Constancia

De porcentaje de similitud (04%) según el software Turnitin, al informe final de investigación: Predicción del nivel de estrés académico en estudiantes de la Universidad Nacional José María Arguedas aplicando un modelo de inteligencia artificial. Presentado por el Bach. Geraldine Nila Flores Moscoso cuyo asesor es el MSc. Iván Soria Solís.

Dra. María del Carmen Delgado Laime
Presidente de la Unidad de Investigación de la
Facultad de Ingeniería

MSc. Fidelia Tapia Tadeo
Miembro de la
Unidad de Investigación de la Facultad de Ingeniería

M.Sc. Richard Carrión Abollaneda
Miembro de la
Unidad de Investigación de la Facultad de Ingeniería

NOMBRE DEL TRABAJO

PREDICCIÓN DEL NIVEL DE ESTRÉS ACADÉMICO EN ESTUDIANTES DE LA UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS APLICANDO UN MODELO DE INTELIGENCIA ARTIFICIAL

AUTOR

GERALDINE NILA FLORES MOSCOSO

RECUENTO DE PALABRAS

17905 Words

RECUENTO DE CARACTERES

102626 Characters

RECUENTO DE PÁGINAS

116 Pages

TAMAÑO DEL ARCHIVO

3.3MB

FECHA DE ENTREGA

Jun 17, 2024 12:36 PM GMT-5

FECHA DEL INFORME

Jun 17, 2024 12:37 PM GMT-5

● **4% de similitud general**

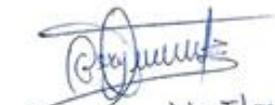
El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 4% Base de datos de Internet
- Base de datos de Crossref
- 1% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

● **Excluir del Reporte de Similitud**

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 15 palabras)


UNIVERSIDAD NACIONAL
JOSÉ MARÍA ARGUEDAS
M.Sc. Ing. Iván Soria Salis
DOCENTE


Geraldine Nila Flores Moscoso
70420760

● 4% de similitud general

Principales fuentes encontradas en las siguientes bases de datos:

- 4% Base de datos de Internet
- Base de datos de Crossref
- 1% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

FUENTES PRINCIPALES

Las fuentes con el mayor número de coincidencias dentro de la entrega. Las fuentes superpuestas no se mostrarán.

1	repositorio.unajma.edu.pe Internet	<1%
2	doczz.net Internet	<1%
3	repositorio.continental.edu.pe Internet	<1%
4	Moh. Fatkuroji, Fajrizal, Taslim, Eka Sabna, Kursiah Warti Ningsih. "O... Crossref	<1%
5	assets.researchsquare.com Internet	<1%
6	redie.mx Internet	<1%
7	repositorio.uta.edu.ec Internet	<1%
8	repositorio.ug.edu.ec Internet	<1%

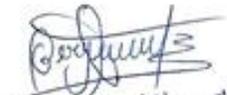

 UNIVERSIDAD NACIONAL
 JOSÉ MARÍA ARGÜEDAS
 I.Sc. Ing. Iván Sofia Solis
 DOCENTE


 Geraldine Nika Flores Moscoso
 70420760

9	1library.co Internet	<1%
10	medicinaysaludpublica.com Internet	<1%
11	imf.org Internet	<1%
12	repositorio.unsaac.edu.pe Internet	<1%
13	neurona.ciens.ucv.ve Internet	<1%
14	repositorio.usmp.edu.pe Internet	<1%
15	hdl.handle.net Internet	<1%
16	repositorio.unac.edu.pe Internet	<1%
17	bbva.com Internet	<1%
18	Ribeiro, Lisete Inês Queiroz. "Categorização e Determinação dos Fator..." Publication	<1%
19	digibuo.uniovi.es Internet	<1%
20	repositorio.usm.cl Internet	<1%

21	ba.unibo.it Internet	<1%
22	gob.pe Internet	<1%
23	researchgate.net Internet	<1%


UNIVERSIDAD I. C. S. J.
JOSE MARÍA TORO
M.Sc. Ing. Ivan Soria Sot
DOCENTE


Gerakline Nila Flores Moscoso
70420760

DEDICATORIA

Esta tesis está dedicada a mi mamá, Petronila Moscoso Romero; por el apoyo y amor que me brinda siempre. A mi papá, Vidal Flores Mezares y por supuesto a mi hermano Brunito, por motivarme para el término de mi investigación y quererme como lo hace.

AGRADECIMIENTO

Agradezco a Dios, por darme salud y felicidad en la vida.

A mi asesor M.Sc. Iván Soria Solis por su apoyo y dedicación para que esta tesis sea realizada de la mejor manera.

A mis padres y hermano, por empujarme hacia el éxito de alguna u otra forma, gracias por estar siempre para mí.

A mis amigos y a Samuel Gutiérrez por el cariño y apoyo que me brindaron.

A los docentes y estudiantes de la Universidad Nacional José María Arguedas, por haber colaborado para que esta investigación pueda desarrollarse.

Índice

Introducción	1
Capítulo I: Problema de Investigación.....	3
1.1 Realidad Problemática	3
1.2 Formulación del Problema.....	5
1.2.1 Problema General.....	5
1.2.2 Problemas específicos	5
1.3 Objetivos	6
1.3.1 General	6
1.3.2 Específicos	6
1.4 Justificación	6
1.5 Limitaciones.....	7
Capítulo II: Antecedentes	9
2.1 Antecedentes a Nivel Nacional.....	9
2.2 Antecedentes a Nivel Internacional	9
Capítulo III: Marco Teórico.....	13
3.1 Inteligencia Artificial	13
3.1.1 Subcampos de la Inteligencia Artificial	14
3.1.2 Problemas en la Inteligencia Artificial	23
3.1.3 Metodología Cross-Industry Standard Process for Data Mining (CRISP-DM)	24
3.2 Estrés Académico.....	27
3.2.1 Niveles de Estrés Académico.....	27
3.2.2 Inventario SISCO del Estrés Académico	29
3.3 Definición de Variables	31
Capítulo IV: Metodología de Investigación.....	32
4.1 Tipo y Nivel de Investigación.....	32
4.1.1 Tipo de Investigación.....	32
4.1.2 Nivel de Investigación	32

4.1.3	Diseño de Investigación	32
4.2	Operacionalización de Variables	33
4.3	Población y Muestra	34
4.3.1	Población.....	34
4.3.2	Muestra	34
4.4	Desarrollo de la investigación.....	35
Capítulo V: Resultados		61
5.1.	Modelo de Árbol de Decisión.....	61
5.2.	Modelo de Máquina de Vectores de Soporte.....	65
5.3.	Modelo de K-Vecinos más Cercanos.....	66
5.4.	Modelo de Redes Neuronales	67
5.5.	Modelo de Bosques Aleatorios	70
Capítulo VI: Discusión		74
Conclusiones		76
Bibliografía		78
Anexos		82

Lista de tablas

Tabla 1 <i>Operacionalización de variables</i>	33
Tabla 2 <i>Ejemplos de modificación del encabezado en el dataset</i>	38
Tabla 3 <i>Puntuación de las dimensiones del inventario SISCO</i>	42
Tabla 4 <i>Baremos del estrés académico</i>	43
Tabla 5 <i>Factores considerados para la limpieza y normalización de datos</i>	48
Tabla 6 <i>Instancias antes y después del balanceo de datos</i>	53
Tabla 7 <i>Reglas establecidas por el modelo de árboles de decisión para predecir el nivel de estrés académico de un estudiante de la UNAJMA</i>	63
Tabla 8 <i>Árboles de decisión - Resultados con validación cruzada</i>	64
Tabla 9 <i>MVS - Resultados con validación cruzada</i>	66
Tabla 10 <i>KNN - Resultados con validación cruzada</i>	67
Tabla 11 <i>Redes neuronales - Resultados con validación cruzada</i>	69
Tabla 12 <i>Bosques aleatorios - Resultados con validación cruzada</i>	72
Tabla 13 <i>Evaluación de modelos</i>	73

Lista de figuras

Figura 1 <i>Relación entre la inteligencia artificial, el aprendizaje automático y aprendizaje profundo</i>	13
Figura 2 <i>Esquema de redes neuronales</i>	18
Figura 3 <i>Partes del árbol de decisión</i>	19
Figura 4 <i>Diagrama Metodología CRISP -DM</i>	25
Figura 5 <i>Dataset inicial con columnas por defecto de Google Forms</i>	37
Figura 6 <i>Visualización de la data antes de aplicar el mapeo</i>	42
Figura 7 <i>Visualización de la data después de aplicar el mapeo</i>	43
Figura 8 <i>Visualización de datos faltantes</i>	44
Figura 9 <i>Codificación One Hot a variables nominales</i>	45
Figura 10 <i>Codificación de la característica "Ciclo"</i>	46
Figura 11 <i>Codificación ordinal de 'Habitos alimenticios'</i>	46
Figura 12 <i>Codificación ordinal de 'Motivacion academica'</i>	47
Figura 13 <i>Codificación ordinal de 'Consumo_Alcohol'</i>	47
Figura 14 <i>Valores de Num_hijos después del tratamiento de datos faltantes</i>	50
Figura 15 <i>Matriz de correlación</i>	54
Figura 16 <i>Puntuación de información mutua</i>	55
Figura 17 <i>Importancia de características</i>	56
Figura 18 <i>Gráfico de barras de distribución</i>	57
Figura 19 <i>Conjunto de datos final 'df_reducido'</i>	59
Figura 20 <i>Árbol de decisión para la segunda fase</i>	62

Figura 21 Métricas obtenidas en la fase 2 del modelo de árboles de decisión	64
Figura 22 Matriz de confusión con MVS	65
Figura 23 Métricas obtenidas del modelo de MVS	66
Figura 24 Métricas obtenidas del modelo K-Vecinos más cercanos	67
Figura 25 Matriz de confusión con redes neuronales	68
Figura 26 Métricas obtenidas del modelo de redes neuronales	69
Figura 27 Gráfica de exactitud y pérdida de entrenamiento y pruebas - RN.....	70
Figura 28 Matriz de confusión con bosques aleatorios	71
Figura 29 Métricas obtenidas del modelo de bosques aleatorios.....	71

Resumen

La presente investigación se enfoca en la predicción del nivel de estrés académico en estudiantes de la Universidad Nacional José María Arguedas mediante la aplicación de un modelo de inteligencia artificial. El objetivo principal es desarrollar un modelo capaz de predecir el nivel de estrés académico en los estudiantes, utilizando técnicas de inteligencia artificial. Para lograr este objetivo, se recopiló datos de una muestra representativa de estudiantes de la UNAJMA, incluyendo variables relacionadas con el estrés académico, como la carga de trabajo, la presión académica, la calidad del sueño y la aplicación del inventario SISCO. Estos datos se utilizaron para entrenar y validar un modelo de inteligencia artificial, específicamente un modelo de aprendizaje automático. El modelo de inteligencia artificial fue diseñado siguiendo las etapas de la metodología CRISP-DM (Cross Industry Standard Process for Data Mining), para analizar los datos recopilados y generar predicciones lo más exactas posible sobre el nivel de estrés académico de los estudiantes. Además, se llevaron a cabo pruebas y validaciones exhaustivas para evaluar la exactitud del modelo. Los resultados obtenidos mostraron que el modelo de inteligencia artificial fue capaz de predecir el nivel de estrés académico en los estudiantes de la UNAJMA con una exactitud significativa. Estos hallazgos tienen importantes implicaciones para el desarrollo de estrategias de intervención temprana y programas de apoyo dirigidos a reducir el estrés académico en los estudiantes universitarios, brindando además una base sólida para futuras investigaciones.

Palabras clave: Predicción, nivel de estrés académico, estudiantes, modelo de inteligencia artificial.

Abstract

The present research focuses on the prediction of the level of academic stress in students at the José María Arguedas National University through the application of an artificial intelligence model. The main objective is to develop a model capable of predicting the level of academic stress in students, using artificial intelligence techniques. To achieve this objective, data were collected from a representative sample of UNAJMA students, including variables related to academic stress, such as workload, academic pressure, sleep quality, and the application of the SISCO inventory. This data was used to train and validate an artificial intelligence model, specifically a machine learning model. The artificial intelligence model was designed following the stages of the CRISP-DM (Cross Industry Standard Process for Data Mining) methodology, to analyze the data collected and generate the most accurate predictions possible about the level of academic stress of the students. Additionally, extensive testing and validation was carried out to evaluate the accuracy of the model. The results obtained showed that the artificial intelligence model was able to predict the level of academic stress in UNAJMA students with significant accuracy. These findings have important implications for the development of early intervention strategies and support programs aimed at reducing academic stress in college students, also providing a solid foundation for future research.

Keywords: *Prediction, academic stress level, students, artificial intelligence model.*

Chumasqa

Kay investigacionqa, Universidad Nacional José María Arguedas nisqapi yachakuqkunapa nivel de estrés académico nisqamanta willakuymi, chaypaqmi ruwakun modelo de inteligencia artificial nisqawan. Hatun yuyayqa, huk modelo ruwaymi, chaywanmi yachakuqkunapa nivel de estrés académico nisqa yachayta atinman, chaypaqmi técnicas de inteligencia artificial nisqawan. Kay ruwayman chayanapaq, UNAJMA yachakuqkunamanta muestra representativa nisqamanta willakuykunata huñukurqa, chaypim karqa variables nisqakuna estrés académico nisqawan tupaq, carga de trabajo, presión académica, calidad de puñuy, chaynallataq inventario SISCO nisqapa aplicacionninpas. Kay willakuykunaqa huk modelo de inteligencia artificial nisqatam yachachinapaq hinaspa validanapaq, específicamente huk modelo de aprendizaje máquina nisqamanta. Chay modelo de inteligencia artificial nisqa ruwasqa karqan CRISP-DM (Proceso Estándar de la Industria Cruzada para Minería de Datos) nisqa metodología nisqa etapas nisqakunata qatispa, chaywanmi huñusqa willakuykunata t'aqwirirqanku, hinallataq aswan chiqan willakuykunata ruwanankupaq, yachakuqkunapa nivel de estrés académico nisqamanta. Chaymantapas, hatun pruebakuna, validación nisqakunam ruwakurqa, chaynapi chay modelopa chiqap kayninta chaninchanapaq. Chay ruwasqakunam qawarichirqa chay modelo de inteligencia artificial nisqa UNAJMA yachakuqkunapa nivel de estrés académico nisqamanta willakuyta ancha chiqap kasqanwan. Kay tarikuykuna ancha chaniyuq implicacionkunayuq kanku estrategias de intervención temprana ruwanapaq chaymanta programakuna yanapakuypaq, chaykunam munanku universidadpi estudiantekunapi estrés académico nisqa pisiyachinapaq, chaynallataqmi hamuq investigacionpaq allin cimientota qun.

Sapaq simikuna: *Willakuy, nivel de estrés académico, yachakuqkuna, modelo de inteligencia artificial.*

Términos y Abreviaturas

CRISP-DM	Cross industry standard process for data mining y en español, Proceso estándar inter-industrias para minería de datos
IA	Inteligencia artificial
IBM	International Business Machines Corporation
LR	Logistic Regression y en español, regresión logística
LSTM	Long-short term memory y en español, memoria a corto plazo
SVM	Support Vector Machine y en español, máquina de vectores de soporte
UNAJMA	Universidad Nacional José María Arguedas

Introducción

La educación superior es un período crucial en la vida de los estudiantes, pero también puede ser una etapa de gran presión y estrés. El estrés académico puede tener un impacto significativo en el rendimiento académico y en el bienestar general de los estudiantes. Por lo tanto, es importante identificar y abordar este problema para promover un ambiente de aprendizaje saludable y exitoso.

En esta investigación, se propone utilizar un modelo de inteligencia artificial para predecir el nivel de estrés académico en estudiantes de la UNAJMA. La inteligencia artificial ha demostrado ser una herramienta poderosa en diversos campos, y su aplicación en el ámbito de la psicología y la educación puede brindar nuevas oportunidades para comprender y abordar el estrés académico.

El objetivo principal de esta investigación es desarrollar un modelo de inteligencia artificial que pueda analizar diferentes variables socioeconómicas, para predecir el nivel de estrés académico en los estudiantes.

La presente investigación consta de seis capítulos, como se muestra a continuación:

CAPÍTULO I: Problema de investigación: En este capítulo se realizó el desarrollo de la realidad problemática, los objetivos, las hipótesis, justificación y las limitaciones.

CAPÍTULO II: Antecedentes: En este capítulo se encuentran los antecedentes a nivel nacional e internacional, para poder visualizar los resultados y demás de investigaciones afines con esta.

CAPÍTULO III: Marco teórico: En este capítulo se desarrolló todos los conceptos y definiciones relacionados con la variable independiente y dependiente, modelo de IA y estrés académico respectivamente.

CAPÍTULO IV: Metodología de la investigación: Aquí, se realizaron los procedimientos según la metodología CRISP-DM, además de definiciones del tipo y nivel de investigación.

CAPÍTULO V: Resultados: Se realizó el desarrollo y presentación de resultados que se obtuvieron al utilizar las diferentes técnicas de IA.

CAPÍTULO VI: Discusión: Finalmente, en este capítulo se realizó la discusión de los resultados obtenidos en la presente investigación con los resultados obtenidos por otros autores con las investigaciones presentadas en los antecedentes.

Los resultados de esta investigación pueden tener implicaciones significativas para la comunidad académica y educativa. La capacidad de predecir el nivel de estrés académico en los estudiantes puede permitir la implementación de intervenciones tempranas y personalizadas para apoyar su bienestar emocional y su éxito académico. Además, la aplicación de un modelo de inteligencia artificial en este contexto puede abrir nuevas vías de investigación y desarrollo en el campo de la psicología y la educación.

Capítulo I: Problema de Investigación

En este capítulo se presenta la idea estructurada de la investigación, consta de apartados como realidad problemática, formulación del problema, objetivos, justificación y limitaciones.

1.1 Realidad Problemática

El estrés, hoy en día se ha convertido en un tema con bastante importancia, debido al interés y concientización que últimamente se ha creado sobre la salud mental en las personas a nivel mundial. En una entrevista a la Dra. Thaymis Báez y el Dr. Héctor Hernández, Ochoa (2023) escribe que, según las proyecciones de la Organización Mundial de la Salud, se prevé que la salud mental llegaría a ser una preocupación global significativa para el año 2030 y esto fue agravado por la llegada del COVID-19, el cual ha acelerado el deterioro de la salud mental de las personas. De acuerdo a estimaciones realizadas por la OMS, la depresión y la ansiedad experimentaron un aumento del 25% durante el primer año de pandemia y se menciona que la población con mayor vulnerabilidad en lo que se relaciona a salud mental son los adolescentes y adultos jóvenes como son los universitarios.

En adición, durante los últimos años se ha incrementado el número de personas que sufren de un cuadro de estrés alto, según LaMotte (2021), el 2020 fue un año donde se obtuvo un récord en cuanto a los niveles de estrés, el 40% de los adultos de todo el mundo mencionaron que pasaron por cuadros de estrés, esta cifra es la más alta en los últimos 15 años. El porcentaje aumenta aún más si nos centramos en el Perú; en el 2019 el Ministerio de Educación (MINEDU) y el Ministerio de Salud (MINSA) en preocupación referente a los porcentajes presentes dentro de la comunidad universitaria los cuales son: Ansiedad (82%), estrés (79%) y violencia (52%), realizaron la inauguración del Primer Centro de Salud Comunitario Universitario en la Universidad Nacional

Mayor de San Marcos con la visión de poder trabajar de la misma manera con otras 21 universidades dentro del Perú. MINEDU (2019). Sin embargo, en el 2021, el Ministerio de Salud (MINSA) alertó con un alto porcentaje del 52.2% de los limeños sufre de estrés. (EL PERUANO, 2022) ;en donde no se menciona si fueron estudiantes universitarios y al parecer las cifras no mejoraron en comparación a las cifras proporcionadas en el 2019.

Viendo este panorama, el estrés no es ajeno a los estudiantes universitarios. El estrés académico es considerado una de las causas principales para la deserción universitaria, bajo rendimiento académico, problemas de inclusión social e incluso el suicidio. En la provincia de Andahuaylas, hasta junio del 2023 se ha suscitado una ola de más de 7 suicidios, de los cuales 3 fueron jóvenes universitarios y 1 de ellos perteneciente a la Universidad Nacional José María Arguedas (UNAJMA). Contacto Directo (2019); estos lamentables sucesos hacen que tomemos el tema de salud mental como prioridad, además de actuar con responsabilidad social para que de alguna manera podamos aportar en la prevención de estos.

Si bien es cierto existen niveles del estrés, que son causados por distintos eventos de los cuales el cuerpo reconoce como amenazas, por ejemplo: Una exposición, un examen, sobrecarga académica, falta de tiempo para cumplir con los deberes, etc. La detección inicial es esencial para no llegar a tener mayores problemas en un futuro cercano. La UNAJMA, es un centro de estudios importante en la región Apurímac en donde se forman gran cantidad de jóvenes para llegar a ser profesionales, es por ello que es de vital importancia cuidar de la salud mental de los estudiantes.

En lo que concierne la determinación de los niveles de estrés académico en estudiantes de la UNAJMA, no existe ningún sistema o modelo que pueda ayudar en la predicción de estos niveles a través de datos demográficos y reacciones físicas, psicológicas y comportamentales, esto conlleva a que día a día se deteriore más la salud mental de los estudiantes además de que podría

causar enfermedades físicas y emocionales como la indiferencia, pérdida de las ganas de vivir, o incluso podría conllevar a obtener conductas autodestructivas y comportamientos evasivos, drogadicción, abandono, etc. (Vergara, 2022)

En ese sentido, la detección temprana de niveles de estrés y la toma de decisiones son cruciales para que los estudiantes de la UNAJMA puedan gozar de una buena salud mental, además, recibir apoyo a tiempo si fuese necesario. Adicionando a ello, la predicción temprana de los niveles de estrés académico beneficia tanto a los estudiantes como a la universidad, en consecuencia, existiría un mejor desempeño por parte de los estudiantes universitarios en el ámbito académico. Por esa razón, se ha optado por realizar un modelo predictivo de los niveles de estrés académico en estudiantes universitarios aplicando Inteligencia Artificial (IA).

1.2 Formulación del Problema

1.2.1 Problema General

¿Cómo se puede predecir el nivel de estrés académico en estudiantes de la Universidad Nacional José María Arguedas?

1.2.2 Problemas específicos

- ❖ ¿Cuál es el nivel de exactitud que se obtiene de un modelo de Inteligencia Artificial en la predicción del nivel del estrés académico en estudiantes de la UNAJMA?
- ❖ ¿Cuál es el algoritmo de Inteligencia Artificial más adecuado para realizar la predicción del nivel del estrés académico en estudiantes de la UNAJMA en base a un conjunto de datos proporcionado?
- ❖ ¿Cuál es el nivel de exactitud que se obtiene de un modelo de Inteligencia Artificial en la predicción del nivel de estrés académico en estudiantes de la UNAJMA al variar el número de características utilizadas?

1.3 Objetivos

1.3.1 General

Predecir los niveles de estrés académico de los estudiantes de la UNAJMA con un modelo predictivo de Inteligencia Artificial.

1.3.2 Específicos

- ❖ Determinar el nivel de exactitud que se obtiene de un modelo de Inteligencia Artificial en la predicción del nivel del estrés académico en estudiantes de la UNAJMA.
- ❖ Determinar el algoritmo de Inteligencia Artificial más adecuado para realizar la predicción del nivel del estrés académico en estudiantes de la UNAJMA en base a un conjunto de datos proporcionado.
- ❖ Determinar el nivel de exactitud que se obtiene de un modelo de Inteligencia Artificial en la predicción del nivel de estrés académico en estudiantes de la UNAJMA al variar el número de características utilizadas.

1.4 Justificación

El modelo de IA realizó predicciones con un porcentaje de precisión aceptable sobre el nivel de estrés académico, de esta manera se puede mejorar en la toma de decisiones con respecto a los estudiantes de la Universidad Nacional José María Arguedas e incluso se podría aplicar en estudiantes de otras universidades. Esto en el menor tiempo posible por el uso de la IA porque si algo debemos de tener en cuenta cuando hablamos de salud mental, es el tiempo.

La importancia de velar por el bienestar de los estudiantes universitarios y su salud mental, permitirá que muchas otras universidades empleen este modelo y realicen las predicciones para mejorar paulatinamente en cuanto a la calidad educativa, rendimiento académico, etc. En consecuencia, si un estudiante se encuentra con buena salud mental y bajos niveles de estrés, este

podrá responder de mejor manera ante los diferentes agentes estresores que se presentan en el día a día de un estudiante.

La investigación tiene un alto impacto social, las universidades que usen el modelo de predicción, obtendrán una buena imagen como un plus ante las demás, podrán mejorar en la toma de decisiones y en consecuencia mejorará el rendimiento académico que es la principal preocupación en cuanto a las universidades; mejorará también la calidad de estudiantes que tengan en sus aulas, por ende, serán excelentes profesionales gozando de buena salud mental.

El impacto económico que se generará será principalmente para el estudiante y la institución, debido a que, si les dan la debida atención a los estudiantes con altos niveles de estrés, no sería necesario realizar un diagnóstico con un profesional. Además, la implementación del modelo no requiere de gastos altos lo cual es una gran ventaja a diferencia de otras propuestas.

De esta forma, la presente investigación beneficiará principalmente a la población estudiantil de la UNAJMA.

1.5 Limitaciones

- ❖ Una de las principales limitaciones fue la escasa bibliografía en español, además del enfoque; debido a que la mayoría de trabajos de investigación de este tipo o tema son de la carrera de psicología, por lo que ellos realizan la evaluación con cuestionarios y no así mediante modelos de IA.
- ❖ Teniendo en referencia la primera limitación, los cuestionarios para determinar el nivel del estrés académico son demasiado largos por lo que para los estudiantes que llenaron las encuestas realizadas, fue bastante tedioso realizarlo.
- ❖ Otra limitación importante la cual hizo que el proyecto de investigación tardara fue en la etapa de recolección de datos, ya que la mayoría de los estudiantes no responden

voluntariamente a las encuestas, fue necesario motivarlos o aplicar las mismas con el apoyo de los docentes de la universidad para que puedan responder a la encuesta brindada.

Capítulo II: Antecedentes

En este capítulo, se presenta los antecedentes nacionales e internacionales como son los estudios previos similares al tema de investigación presentado.

2.1 Antecedentes a Nivel Nacional

La presente investigación cuenta con el siguiente antecedente nacional de estudio.

- ❖ En el proyecto de investigación titulado “Predicción del nivel del estrés en estudiantes universitarios utilizando técnicas de Machine Learning” el autor Vilca (2019) trabajó con algoritmos de árboles de decisión para determinar el nivel de estrés académico en estudiantes de una Universidad en Juliaca, se extrajeron los datos de los estudiantes a través de encuestas virtuales y se obtuvo como resultado una precisión del 97% con una profundidad de 4 en cuanto a las predicciones realizadas. Además, aplicó la metodología denominada proceso estándar inter-industrias para minería de datos (CRISP-DM), la cual consta de 5 fases para entrenar su modelo con aprendizaje de máquina y con ello realizar la predicción del nivel de estrés.

2.2 Antecedentes a Nivel Internacional

La presente investigación cuenta con los siguientes antecedentes internacionales de estudio.

- ❖ Según Jaques et al. (2017) en su artículo de investigación “Predicción del estado de ánimo, salud y nivel de estrés del mañana usando aprendizaje multitarea personalizado y adaptación de dominio” mencionan que realizaron el entrenamiento de data robusta de diferentes personas conseguida mediante sensores, encuestas y celulares para que a partir de ello realicen un modelo personalizado con técnicas de adaptación de dominio, esos datos fueron tratados como una tarea y de esta forma realizaron una predicción individualizada

del estado de ánimo, estrés y salud física del mañana basados en datos actuales de hoy. Utilizaron redes neuronales profundas para el entrenamiento del modelo, además de redes Gaussianas, mencionan que estos modelos proporcionan un error menor a los tradicionales en un 13 – 22%.

- ❖ Desde otro punto de vista Elzeiny & Qaraq (2018), en su investigación titulada “Enfoques de aprendizaje automático para la detección automática del estrés: Una revisión” realizan una investigación sobre la detección automática del estrés con diferentes algoritmos y enfoques, tales como la fatiga al conducir , estrés en el trabajo y estudio; esto a través de la voz humana, la detección de parpadeos, temperatura corporal, el comportamiento de las personas, y el uso de sensores y teléfonos celulares. Finalmente llegan a la conclusión que los algoritmos de bosques aleatorios, máquina de vectores de soporte y árboles de decisión son efectivos para la clasificación y predicción del estrés mental.
- ❖ Por otro lado, Gogas, Papadimitriou, & Agrapetidou (2018) en su artículo titulado “Pronóstico de quiebras bancarias y pruebas de estrés: Un enfoque de Machine Learning” encontraron buenos resultados al seleccionar información de bancos que quebraron en Estados Unidos y bancos que lograron permanecer activos. Las variables seleccionadas alimentaron a un modelo de pronóstico de máquina de vectores de soporte, el cual obtuvo como resultado un 99.22% de precisión en el pronóstico general. Si bien es cierto esta investigación se centra en el pronóstico de las quiebras bancarias y no así en el estrés académico, pero, también realizaron una prueba de estrés que informa a los interesados para que estos tomen acciones y se puedan evitar fallas en un futuro cercano.
- ❖ Los autores Ahuja & Banga (2019) en el artículo “Detección del estrés mental en estudiantes universitarios usando algoritmos de Machine Learning”, calcularon el nivel del

estrés de los estudiantes considerando datos una semana antes de los exámenes y durante el tiempo que pasan usando internet. Su objetivo fue el de analizar el estrés en estudiantes en diferentes circunstancias de su vida. Aplicaron también cuatro diferentes algoritmos para el entrenamiento de los datos; regresión lineal, Naive Bayes, Random Forest y Máquina de vectores de soporte, de los cuales evaluaron la sensibilidad, especificidad y precisión del modelo entrenado. De la aplicación de estos algoritmos quien obtuvo el porcentaje de precisión más alto fue el algoritmo de Máquina de Vectores de Soporte con un 85.71%. El conjunto de datos con los que trabajaron fueron tomados de 206 estudiantes del Instituto Jaypee de Tecnologías de Información.

- ❖ Según Umematsu et al. (2019), en “Estudiantes mejorando, pronóstico del estrés en la vida diaria usando LSTM redes neuronales” aplicaron un modelo de redes neuronales de memoria a corto plazo (LSTM) por sus siglas en inglés Long Short-Term Memory, regresión logística (LR) por sus siglas en inglés Logistic Regression y máquina de vectores de soporte (SVM) por sus siglas en inglés Support Vector Machine; en donde la salida o predicción es binaria, alto y bajo nivel de estrés. En esta investigación se usó data de señales fisiológicas, uso de celulares y encuestas de comportamiento a los estudiantes. Como resultados se tiene que LSTM supera en la predicción de niveles de estrés en estudiantes a los algoritmos de LR y SVM alcanzando el 83.6% de precisión, esto utilizando datos de una semana anterior. Y, usando solo el subconjunto de datos fisiológicos, el modelo LSTM alcanza el 81.4% de precisión.
- ❖ Desde otro enfoque, Shafiee & Mutalib (2020) en “Predicción de problemas de salud mental entre estudiantes de educación superior usando Machine Learning”, realiza un estudio más global no centrándose solo en el estrés, sino en los problemas mentales que

puedan tener estudiantes de educación superior. Se realiza una comparación entre diferentes trabajos de investigación previos sobre el tema, como conclusión se obtiene que en la mayoría de investigaciones el modelo que obtiene mayor precisión, además de ser el más popular es SVM. La precisión que se obtiene con este algoritmo es de alta precisión entre el 70% y 96%.

- ❖ En “Machine Learning aplicado en la clasificación y predicción de la depresión: Una revisión sistemática” de Osorio et al. (2021), se realiza una investigación en la que el objetivo principal es el de predecir la depresión a través de neuroimágenes y el uso de celulares, los cuales consideran esenciales para realizar la predicción. Además, usan dos algoritmos principales para la construcción de sus modelos; LR y SVM. Como resultados mencionan a los algoritmos de bosques aleatorios, regresión logística, máquina de vectores de soporte, redes neuronales y árboles de decisión como los mejores para obtener buenos resultados a la hora de realizar la predicción de la depresión.

Capítulo III: Marco Teórico

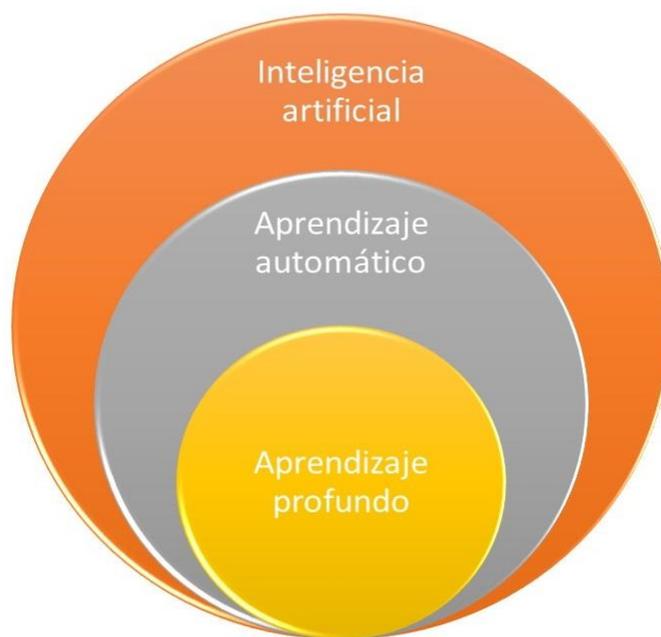
3.1 Inteligencia Artificial

La IA, es un término bastante usado con diferentes conceptos, pero los diferentes autores coinciden en que este término se define como la capacidad que tienen las máquinas o sistemas para poder aprender o imitar al ser humano, las tareas que realiza, el razonamiento y la adaptación; esto a partir de la información que se le proporciona. Con ello se realiza el desarrollo de sistemas o algoritmos, los cuales incluyen diferentes técnicas. (ORACLE, 2023)

La IA posee diversos componentes, pero los principales son el aprendizaje automático y el aprendizaje profundo, en la Figura 1 se muestra la relación de estos.

Figura 1

Relación entre la inteligencia artificial, el aprendizaje automático y aprendizaje profundo



Nota. Tomado de *Relación de pertenencia entre la inteligencia artificial, el aprendizaje automático y el aprendizaje profundo*, por V. Yepes, 2020 (<https://victoryepes.blogs.upv.es/2020/09/15/>).

3.1.1 Subcampos de la Inteligencia Artificial

Los componentes o subcampos de la IA según Russell & Norvig (2020) son: Aprendizaje automático, procesamiento del lenguaje natural, representación del conocimiento e inferencia, razonamiento automático, visión artificial, robótica y computación evolutiva. Se profundiza más en lo que respecta a aprendizaje automático por ser el subcampo más importante y usado en esta investigación.

3.1.1.1. Aprendizaje Automático. Existen múltiples definiciones para el término aprendizaje automático, según Mitchell (1997) el término refiere al campo que se ocupa de la construcción de diferentes programas que mejoran continuamente en el tiempo con la experiencia, en los últimos años se ha desarrollado diferentes aplicaciones, sistemas, programas, en los que está presente este término, por ejemplo a diario en las redes sociales recibimos recomendaciones de contenido que nos gusta ver o de lo que nos parece interesante gracias a la presencia del aprendizaje automático.

Aprendizaje automático es una disciplina de la IA, que a través de diferentes algoritmos es capaz de darle la capacidad de identificar patrones y elaborar predicciones a ordenadores. El ordenador aprende de estos patrones y les permite realizar tareas de forma autónoma sin necesidad de ser programados. (IBERDROLA, 2022)

Según Hurwitz & Kirsch (2018) existen 4 enfoques de aprendizaje automático, los cuales se muestran a continuación:

❖ Aprendizaje supervisado

El aprendizaje supervisado usualmente contiene una gran cantidad de datos en conjunto, pero estos son etiquetados para definir su significado. Pretende encontrar patrones en esos datos a través de un proceso analítico. Por ejemplo: La data sería un millón

de imágenes de frutas y a ello incluir la explicación de lo que es cada fruta, como variable de salida. Al final se crearía un modelo el cual sea capaz de distinguir que fruta es una determinada fotografía. Si la etiqueta es continua, es una regresión; y si los datos provienen de un conjunto contable de valores, esto se convierte en una clasificación. La regresión utilizada para el aprendizaje supervisado ayuda a entender la correlación entre las variables, un ejemplo sería la predicción del clima.

❖ **Aprendizaje no supervisado**

Este tipo de aprendizaje es usado y adecuado cuando se tiene gran cantidad de datos sin etiqueta, por ejemplo: Las redes sociales, como Facebook, Twitter, Instagram, etc. Existe gran cantidad de data en estos medios los cuales están sin etiquetar y sería demasiado engorroso hacerlo.

Comprender el significado de esos datos requiere que los algoritmos clasifiquen la data empezando a entenderlos en función de patrones o grupos que se encuentre, es por ello que este tipo de aprendizaje lleva a cabo un proceso iterativo de análisis de datos sin intervención humana. Adicionando a este concepto IBERDROLA (2022) menciona que este tipo de algoritmos no cuentan con ningún conocimiento previo, por lo que podría producirse un caos de datos con el fin de encontrar patrones y a partir de estos se puedan agrupar y organizarlos de la mejor manera.

❖ **Aprendizaje por refuerzo**

Es un modelo de aprendizaje conductual. Se menciona que el algoritmo se entrena a través de la prueba y error. Recibe una retroalimentación del análisis de datos por lo que los resultados son muy beneficiosos y los mejores para el usuario. El objetivo es que el algoritmo aprenda a partir de la experiencia y con las situaciones de prueba y error es en

donde se recompensan las decisiones correctas. Como ejemplo podemos nombrar que se utiliza para el reconocimiento facial. (ibid.)

❖ **Redes neuronales y aprendizaje profundo**

El aprendizaje profundo es un método específico del aprendizaje automático. Este trabaja con redes neuronales en capas sucesivas para aprender de los datos que se le proporciona con algoritmos supervisados y no supervisados. Es útil cuando aprende patrones a partir de datos no estructurados y no etiquetados. Cuanto más complejo el problema, más capas existirá en el modelo que se esté realizando. (Hurwitz & Kirsch, 2018)

Las redes neuronales son utilizadas para modelar sistemas más complejos, simula el comportamiento del cerebro humano y el trabajo que realizan las neuronas.

También, son definidas como redes interconectadas entre sí, paralelas y a menudo adaptables, están organizadas de manera jerárquica e intentan interactuar con objetos del mundo real de la misma forma que lo hace el sistema nervioso de un ser humano. Las redes neuronales poseen las siguientes ventajas: (Jorge, 2001)

- Aprendizaje adaptativo, es decir, es capaz de aprender a partir de una experiencia inicial.
- Auto organización, es capaz de crear su propia representación de la organización en el aprendizaje.
- Tolerancia a fallos, la red de neuronas puede tener la capacidad de retener información en caso alguna neurona se destruya.
- Operación en tiempo real, las redes neuronales trabajan de forma paralela.

A. Modelo de aprendizaje automático

Según Barrios (2022), un modelo de aprendizaje automático es la expresión de un algoritmo determinado el cual trabaja con grandes cantidades de datos, con ello debe encontrar patrones a través del entrenamiento y una vez que aprenda pueda realizar predicciones.

Así mismo, Microsoft (2022) contrasta la definición anterior mencionando que se trata de un modelo, un archivo el cual ha sido entrenado previamente para poder reconocer, clasificar y agrupar determinados tipos de patrones.

A continuación, se definen los algoritmos más conocidos de aprendizaje automático y los que se usaron en la presente investigación:

❖ Redes Neuronales Artificiales

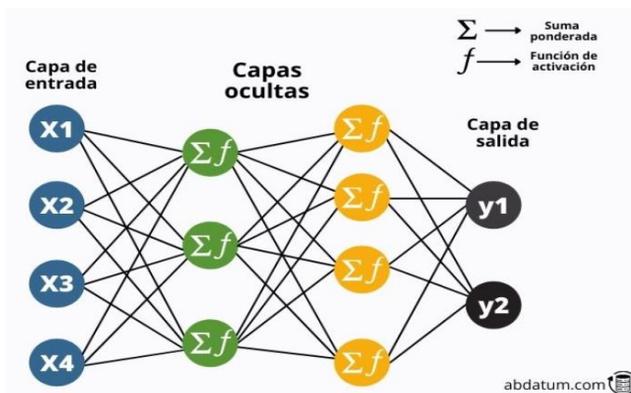
Son sistemas conexionistas que fueron inspirados en las redes neuronales biológicas. Estos sistemas aprenden a realizar diferentes tareas con ejemplos, sin ninguna regla en específico. Transmiten información en cada conexión que se realiza o también llamado sinapsis. Cuando una neurona recibe esa señal, esta es capaz de procesarla y enviar señales a las demás neuronas artificiales. Estas neuronas artificiales suelen tener un peso que se va ajustando según el aprendizaje que van teniendo. El aumento o disminución de este peso significa la fuerza de conexión que están teniendo las neuronas en el proceso de sinapsis. (Lee, Grosse, & Ranganath, 2009)

A continuación, se muestra un esquema en donde se observa la estructura típica de las redes neuronales; donde la información fluye a través de la red mostrada desde la capa de entrada, los cuales son los valores de entrada, después pasa a las capas ocultas en donde se realizan las operaciones matemáticas y finalmente se tiene la capa de salida la cual es el

resultado, el cual se compara con los resultados reales para determinar el coste o grado de error de la red.

Figura 2

Esquema de redes neuronales



Nota. Tomado de *Redes neuronales artificiales*, por Cañadas 2021

Existen 3 tipos de redes neuronales que dependen de las conexiones y operaciones que tengan. (Cañadas, 2021)

Las redes neuronales prealimentadas, son en las que dentro de su arquitectura la información fluye únicamente en una dirección, sin la presencia de conexiones o bucles entre las neuronas de la misma capa. Este tipo de redes neuronales son consideradas como las más simples y fueron las primeras en ser utilizadas para realizar las tareas de IA.

Las redes neuronales convolucionales, también conocidas como Convolutional Neural Networks (CNN), poseen una arquitectura altamente efectiva para el procesamiento de imágenes. Los pesos se organizan en filtros que tienen la capacidad de identificar diversas características visuales en una imagen, como sombras, bordes o la forma de un ojo humano.

Por último, las redes neuronales recurrentes son un tipo de red que permite la formación de ciclos, lo que significa que una neurona puede tener conexiones consigo misma. Estas redes artificiales resultan altamente eficientes al analizar datos con estructura secuencial como es el caso de los textos y un campo de aplicación es el análisis de series temporales. Muchas empresas utilizan esta arquitectura para predecir las ventas de productos basándose en una secuencia de ventas del año anterior.

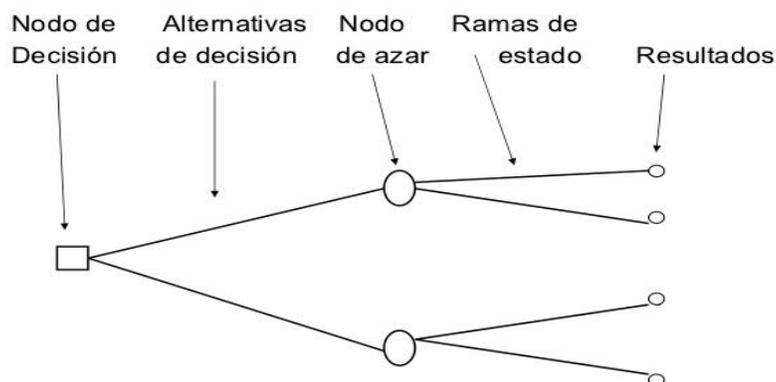
❖ Árboles de Decisión

Según Bouza & Santiago (2012), un árbol de decisión “Es la representación de una función multivariada que fue posible utilizar en la vida práctica a partir del desarrollo de las modernas computadoras”. También se usa como modelo predictivo, el cual representa en las ramas las conjunciones y el valor del objetivo o etiquetas en las hojas. Es más usado para lo que son toma de decisiones.

A continuación, se muestran las partes del árbol de decisión como visión general:

Figura 3

Partes del árbol de decisión



Nota. Tomado de *Árboles de decisión en regresión Machine Learning*, por Veloso 2019

❖ **Máquina de Vectores de Soporte**

Es una máquina de aprendizaje para la clasificación de datos en dos grupos, los vectores de entrada son asignados de forma no lineal a un espacio de características de alta dimensión. Es ahí donde se realiza una decisión lineal. Con este modelo se asegura la alta capacidad de generalización del aprendizaje. (Cortes & Vapnik, 1995).

❖ **Regresión Logística**

El autor Gonzales (2022), menciona que la regresión logística es un método común para resolver problemas de clasificación binaria, ya que la variable independiente está codificada por ejemplo como 1 – 0, si – no, abierto – cerrado, etc. Es simple de aplicar además de ser eficiente. Se obtiene la relación entre una variable dependiente y otras variables independientes ya sean nominales, ordinales, de intervalo o de nivel de razón.

❖ **K Vecinos más cercanos**

Es un método de clasificación supervisada no paramétrico. En el reconocimiento de patrones, este algoritmo es usado para la clasificación de objetos, que se basa en el entrenamiento a través de elementos cercanos de un punto fijo. Usualmente se usa la distancia euclidiana. (Piryonesi & El-Diraby, 2020)

❖ **Naive Bayes**

Es un enfoque de aprendizaje automático supervisado utilizado para la clasificación de tareas, incluyendo la clasificación de texto. Además, forma parte de una familia de algoritmos generativos de aprendizaje, lo que implica que busca modelar la distribución de las entradas pertenecientes a una clase o categoría específica. A diferencia de los clasificadores discriminativos como la regresión logística, este algoritmo no se enfoca en aprender qué características son más relevantes para diferenciar entre clases. (IBM, 2023)

B. Métricas de evaluación de modelos de aprendizaje automático

Existen varias métricas comunes para evaluar el rendimiento que poseen los modelos de aprendizaje automático, los más comunes son los siguientes: (FAYRIX, 2019)

❖ Exactitud

Mide la proporción correcta de predicciones que realiza el modelo. Indica el resultado del número de elementos clasificados correctamente en comparación al total. No funciona bien con clases que estén desequilibradas.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Donde:

TP: True positives

TN: True negatives

FP: False positives

FN: False negatives

❖ Exhaustividad o sensibilidad

También conocida como recall, es usada para evaluar el rendimiento de un modelo de clasificación binaria. Mide la cantidad de verdaderos positivos que clasificó el modelo en función al número total de valores positivos.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Donde:

TP: True positives

TN: True negatives

❖ Precisión

La precisión es una métrica común utilizada para evaluar el rendimiento de un modelo. Mide la proporción de casos positivos verdaderos, los que el modelo haya etiquetado correctamente como positivos en relación al total de casos que el modelo haya etiquetado como positivos. Se simplifica en la ecuación “(3)” a continuación.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Donde:

TP: True positives

FP: False negatives

❖ Puntuación F1

Esta métrica es el resultado de la media armónica entre las métricas de precisión y la sensibilidad o recall como se muestra en la ecuación “(4)”. La mejor puntuación F1 es igual a 1 lo cual significa que la precisión es perfecta y la peor 0.

$$F1 = 2 * \frac{Precision * Sensibilidad}{Precision + Sensibilidad} \quad (4)$$

Es muy importante saber elegir las métricas a usar al evaluar el rendimiento de un modelo de IA, la elección dependerá del tipo de problema, el desequilibrio de las clases y los objetivos de la investigación. También es importante mencionar que utilizar varias métricas puede proporcionar una visión más completa del rendimiento del modelo.

3.1.1.2. Procesamiento del Lenguaje Natural. Permite la comunicación entre humanos y computadoras mediante lenguaje escrito u oral. Implica comprender el lenguaje, extraer significado y generar respuestas coherentes. Es usado por ejemplo en asistentes virtuales.

3.1.1.3. Representación del Conocimiento e inferencia. Se centra en presentar hechos, conceptos, relaciones y reglas sobre el mundo en formatos que una computadora pueda utilizar para hacer deducciones.

3.1.1.4. Razonamiento Automático. Desarrollo de sistemas capaces de hacer deducciones lógicas, resolver problemas y tomar decisiones de forma automática basados en información disponible y reglas establecidas.

3.1.1.5. Visión Artificial. Procesamiento digital de imágenes y video para detectar, clasificar u obtener información relevante sobre objetos presentes en escenas visuales mediante patrones y modelos.

3.1.1.6. Robótica. Se centra en la construcción de máquinas electro-mecánicas con sensores para percibir su entorno físico y algoritmos que guían su comportamiento y acciones de forma autónoma para realizar tareas específicas.

3.1.1.7. Computación Evolutiva. Uso de mecanismos evolutivos como la selección natural para optimizar soluciones de problemas complejos en distintos dominios de aplicación.

3.1.2 Problemas en la Inteligencia Artificial

Es importante determinar el tipo de problema que se quiere solucionar. Los problemas más comunes dentro de la IA que se resuelven utilizando técnicas de aprendizaje automático son los siguientes: (FAYRIX, 2019)

A. Problemas de clasificación

En este tipo de problema, se asigna una etiqueta o clase a la entrada proporcionada. Por ejemplo, en un problema de clasificación de imágenes, se quiere determinar a qué objeto, animal o persona pertenece la imagen proporcionada.

B. Problemas de regresión

En un problema de regresión, el objetivo es predecir un valor numérico según un conjunto de características que se proporcionan. Por ejemplo, en un problema de regresión de precios de carros, el objetivo es predecir el precio que tiene un determinado carro, dado un conjunto de datos de entrada como la marca, modelo, etc.

C. Problemas de clustering o agrupamiento

En este tipo de problema el objetivo es agrupar los datos en diferentes cluster o grupos según ciertas similitudes.

3.1.3 Metodología Cross-Industry Standard Process for Data Mining (CRISP-DM)

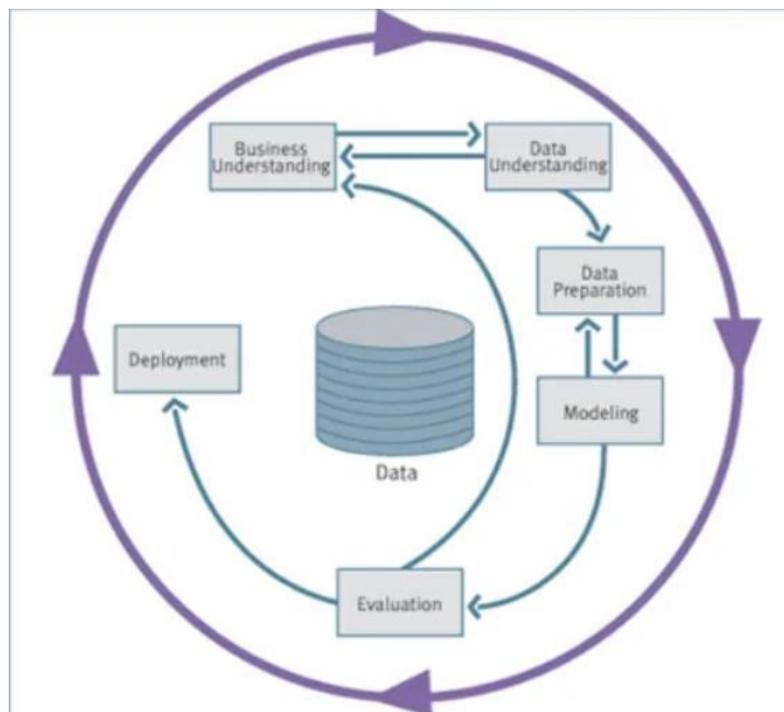
Una metodología, es como la planificación general que se usa como guía para la realizar las actividades o procesos propuestos dentro de un determinado campo de investigación, y de esta manera se pueda obtener resultados. (IBM Analytics, 2015).

La metodología CRISP-DM, desarrollado por International Business Machines Corporation (IBM); es un enfoque utilizado ampliamente para lo que viene a ser la minería de datos, sin embargo, también es aplicable en proyectos de IA. Proporciona un marco estructurado y sistemático para guiar a los profesionales a lo largo de las diferentes etapas.

Esta metodología cuenta con 6 etapas, como se muestra en el diagrama:

Figura 4

Diagrama Metodología CRISP -DM



Nota. Tomado de *Cross-Industry Standard Process for Data Mining*, por Goicochea, 2009

Para el correcto desarrollo de un Modelo, se deben seguir con las etapas de la metodología CRISP – DM IBM Analytics (2015), son las siguientes:

3.1.3.1. Comprensión del Negocio. En esta primera etapa, se debe estudiar los problemas, objetivos y requisitos del proyecto a realizar; después se debe revisar la situación y analizar los beneficios que traerá consigo para que con esta información se tenga éxito en el futuro.

3.1.3.2. Comprensión de los Datos. Aquí se realiza la recolección de los datos iniciales para poder describirlos, analizar si estos datos son confiables, si son suficientes o se necesitan más, siempre basándose en los requisitos con los que se tenía que realizar la recopilación. También es

importante conocer la accesibilidad que se tendrá hacia esos datos. Todo este proceso, evita que nos salgamos del objetivo que marcamos en la primera etapa y consigamos los parámetros indicados y suficientes para lograr el objetivo inicial. (Ramirez, 2021)

3.1.3.3. Preparación de los Datos. Se realizará la limpieza de datos, agruparlos o transformarlos en variables más útiles, normalizar los datos, etc. En resumen, se tienen que preparar los datos para que estos se ajusten al objetivo que estamos buscando.

3.1.3.4. Modelamiento. Teniendo nuestros datos listos para usar, empezamos a elaborar el modelo; siempre considerando qué es lo que queremos realizar, no desviarnos del objetivo. Definir el tipo de modelo, si va a ser predictivo, o descriptivo o quizá un modelo de clasificación. Además, que también se tiene que definir de qué manera se sabrá si estos modelos son válidos o si es lo suficientemente bueno para cumplir el objetivo.

3.1.3.5. Evaluación. Se realiza la evaluación del modelo. Si este es eficiente, eficaz o si tiene un alto nivel de precisión. En caso no se llegue a alcanzar los estándares esperados, la metodología CRISP – DM, permite regresar a etapas anteriores a reevaluar la situación o incluso a volver a alinear el modelo para lograr lo propuesto por el investigador.

3.1.3.6. Implementación. Finalmente, el llegar a esta etapa de la metodología se considera un éxito. En este punto solo queda poner en funcionamiento el Modelo realizado, realizar el reporte final e implementarlo hablando con las partes interesadas ya que el modelo tuvo una evaluación positiva.

3.2 Estrés Académico

Desde que nosotros iniciamos nuestra vida académica ya sea en el jardín de niños o en la escuela, influyen en nosotros ciertas tensiones los cuales están como estresores en nuestro día a día.

El sentirse mal en un estudiante se presenta a debidos factores físicos, emocionales o ya sea de carácter social o interrelacional. Para ello cada estudiante debe de afrontar el contexto en el que se encuentra para resolver dichos problemas, realizar una búsqueda de reconocimiento e identidad y de esta manera pueda volverse algo adaptativo y manejable. (Díaz et al., 2007)

Según Berrío & Zea (2011), los indicadores del estrés académico son los que se mencionan a continuación:

- ❖ Reacciones Físicas: Dolor de cabeza, dificultad para conciliar el sueño, fatiga, sudoración excesiva, etc.
- ❖ Reacciones psicológicas: Desconcentración, bloqueo mental, problemas de memoria, ansiedad, etc.
- ❖ Reacciones comportamentales: Estar ausente de las clases, desgano para cumplir con las tareas académicas, aislamiento, etc.

Con estos indicadores, podríamos identificar fácilmente a una persona que sufre de estrés académico.

3.2.1 Niveles de Estrés Académico

Los niveles de estrés académico pueden variar ampliamente entre diferentes estudiantes y puede depender de distintos factores como las características individuales, el entorno académico, la vida personal y las demandas de la misma, otros.

Existen varios niveles de estrés académico, usualmente se clasifican en tres categorías: (Avila, 2014)

3.2.1.1. Nivel Bajo o Agudo. El estrés agudo es el más común, los estudiantes experimentan bajos niveles de tensión y preocupación relacionada con el rendimiento académico, son capaces de manejar adecuadamente las exigencias académicas.

3.2.1.2. Nivel moderado o agudo episódico. El estrés moderado o también llamado agudo episódico, es la aparición reiterativa del estrés agudo con mucha frecuencia. Los estudiantes experimentan un nivel moderado de tensión en el ámbito académico. En este nivel se tienen dificultades para rendir de manera acertada en el ámbito académico.

3.2.1.3. Nivel alto o crónico. El estrés crónico es un estado preocupante, debido a que surge en el estudiante una alarma constante en la que no observa una salida de la situación que está experimentando. Puede tener efectos negativos en la salud mental y física, además de su rendimiento académico.

En adición, no existe ningún estándar o regla que determine el nivel de estrés en el que se encuentra un estudiante universitario, pero como herramientas se puede considerar la información demográfica, académica, antecedentes de salud mental, el ambiente familiar, ambiente escolar, puntuaciones obtenidas en una escala de estrés académico, horas de sueño, actividad física, etc.

En los diferentes casos el punto de corte o la interpretación de los puntajes deben de ser establecidos de acuerdo a la muestra y los objetivos que tenga la investigación que se está realizando.

3.2.2 Inventario SISCO del Estrés Académico

Existen diversas herramientas para poder medir el nivel de estrés académico, para la presente investigación se utilizó el Inventario SISCO del estrés académico, el cual fue diseñado por profesionales de la salud mental para medir el nivel de estrés académico en los estudiantes.

El inventario SISCO fue diseñado a partir de situaciones y respuestas de estrés, el cual evalúa el estrés percibido en diferentes etapas y ámbitos de la vida, pero enfocado al ámbito académico para poder determinar el nivel de estrés académico en estudiantes universitarios.

El inventario o cuestionario es autoadministrado y se puede llenar de forma individual como también colectiva. La resolución no pasa de los quince minutos.

La primera versión del inventario, está compuesta por 37 ítems, el primer artículo o ítem ayuda a determinar si la persona es un candidato a llenar o no la encuesta. Seguidamente esta un artículo en escala de Lickert de 5 valores numéricos el cual determina el nivel del estrés académico. A continuación, se encuentran 9 artículos o preguntas tipo Lickert de 5 valores categoriales para identificar si las demandas del entorno con estímulos estresores, continúan 18 artículos para identificar la frecuencia de los síntomas y reacciones a los diferentes estresores. Finalmente, 8 artículos que identifican la frecuencia de las estrategias de afrontamiento. (Barraza A. M., 2007)

En adición, Barraza (2018) sacó la segunda versión del inventario SISCO para el estudio del estrés académico, en donde se validaron las siguientes acciones: Se excluyeron 3 elementos de la categoría de síntomas debido a que se detectó que se informaban con menor frecuencia en investigaciones previas; también se incluyeron seis elementos adicionales en la categoría de estresores y siete en la de estrategias de afrontamiento. El criterio que se usó para su inclusión se basó en la frecuencia de aparición en estudios previos realizados por el mismo autor.

Se realizó cambio en la escala de respuestas, aumentando de cinco a seis valores con el fin de evitar el fenómeno de la regresión a la media.

Por último, se eliminó la opción “Otro (especifique)” que aparecía en cada sección como son estresores, síntomas y estrategias de afrontamiento.

Siendo la segunda versión del inventario sistémico cognoscitivista del estrés académico el que se usó en esta investigación, se mencionan los puntos más importantes a continuación:

- ❖ Es una versión extendida del Inventario SISCO original de 31 ítems, creada por Barraza A. M (2007).
- ❖ Esta versión extendida consta de 45 ítems que evalúan los componentes sistémico procesuales del estrés académico; además que se dejó los dos ítems iniciales de la primera versión, el primero que sirve de filtro y pregunta sobre la presencia del estrés académico durante el semestre y el segundo que solicita un valor del nivel de estrés académico con el que se percibió ese estrés en una escala del 1 al 5.
- ❖ Esta versión fue sometida nuevamente a un riguroso proceso de validez y confiabilidad psicométrica, en donde la consistencia interna reportada con el alfa de Cronbach es 0.91, muy alta.
- ❖ La evaluación se lleva a cabo mediante la medición de la dimensión de estresores y síntomas, donde se suman los códigos de respuestas a los ítems. Los valores asignados son: Nunca (0), casi nunca (1), raras veces (2), a veces (3), casi siempre (4) y siempre (5). En la dimensión de estrategias de afrontamiento, los valores se invierten. La calificación final se obtiene sumando todos los ítems, con un rango esperado de 0 a 225 para el instrumento en su totalidad.

- ❖ Los puntos de corte para clasificar los niveles de estrés con esta versión son: Estrés leve (0-78), estrés moderado (79-157) y estrés alto (158-225). Esto validado por Alania (2021) quien menciona también que los baremos o cortes son establecidos según el estudio realizado.

El instrumento completo elaborado por Barraza (2007), se puede visualizar en el Anexo 1, como parte de la validación del cuestionario aplicado para la recolección de datos de la presente investigación.

3.3 Definición de Variables

Variable de estudio 1 (Predicción del nivel de estrés académico en estudiantes de la UNAJMA): Son los resultados que se obtuvieron con el modelo de IA a cerca de la predicción de los niveles de estrés académico, como leve, moderado y alto.

Variable de estudio 2 (Modelo de Inteligencia Artificial): Un modelo de Inteligencia Artificial es construido haciendo uso de un algoritmo con los datos recolectados en el cuestionario aplicado a los estudiantes de la UNAJMA, para poder realizar la predicción.

Capítulo IV: Metodología de Investigación

4.1 Tipo y Nivel de Investigación

A continuación, se presentará el tipo y nivel de investigación a la cual pertenece la presente investigación.

4.1.1 Tipo de Investigación

La presente investigación está considerada dentro de los enfoques de investigación como aplicada por las condiciones que reúne. La investigación aplicada busca resolver un problema práctico, en este caso predecir los niveles de estrés académico en estudiantes universitarios. (Muñoz, 2018)

En consecuencia, en esta investigación se logró predecir el nivel de estrés en estudiantes universitarios con lo cual se podrían realizar futuras predicciones.

4.1.2 Nivel de Investigación

El nivel de investigación es predictivo, puesto que predice el comportamiento de un fenómeno psicosocial: El estrés académico en estudiantes universitarios, específicamente en el ámbito académico y de salud mental. (Muñoz, 2018)

4.1.3 Diseño de Investigación

El diseño de la presente investigación es descriptivo, porque permite caracterizar y describir cuantitativamente el fenómeno del estrés académico en estudiantes universitarios mediante mediciones en un momento específico. (Muñoz, 2018)

4.2 Operacionalización de Variables

Tabla 1

Operacionalización de variables

Variable	Dimensión	Indicadores	Índices/Escala
Variable de estudio 1: Modelo de Inteligencia Artificial	✓ Técnicas de Inteligencia Artificial	✓ Exactitud	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
		✓ Validación cruzada	✓ K, número de Folds ✓ Métrica de evaluación del modelo.
Variable de estudio 2: Predicción del nivel de estrés académico en estudiantes de la UNAJMA.	✓ Primera prueba ✓ Segunda prueba ✓ Tercera prueba	✓ Resultados obtenidos de la predicción	Niveles de estrés académico: ✓ Bajo ✓ Moderado ✓ Alto

En la tabla 1, podemos observar que la variable de estudio 1 posee una dimensión llamada técnicas de inteligencia artificial, la cual se medirá con los indicadores de exactitud y validación cruzada.

Por otro lado, la variable de estudio 2 consta de 3 dimensiones de prueba con el indicador de resultados obtenidos en la predicción, los cuales son estrés bajo, moderado y alto según sea el caso.

4.3 Población y Muestra

4.3.1 Población

La población con la que se trabajó para el desarrollo del estudio estuvo conformada por el total de estudiantes de la Universidad Nacional José María Arguedas de las diferentes escuelas profesionales como son: Ingeniería de sistemas, administración de empresas, ingeniería agroindustrial, contabilidad, educación primaria intercultural e ingeniería ambiental, siendo un total de 2086 estudiantes durante el período académico 2023-I. Comprende un espacio temporal compuesto por los meses de abril a julio del año 2023.

4.3.2 Muestra

La muestra que se obtuvo fue de 322 estudiantes universitarios pertenecientes a las seis escuelas profesionales, estos datos se obtuvieron a través del muestreo no probabilístico aleatorio con un tamaño de muestra fijo.

Expresamos la siguiente ecuación:

$$n = \frac{(N * Z^2 * P * Q)}{(N - 1)E^2 + Z^2 * P * Q}$$

$$n = \frac{(2086 * 1.95^2 * 0.5 * 0.5)}{(2086 - 1)0.05^2 + 1.95^2 * 0.5 * 0.5}$$

$$n = 322$$

Donde:

n = Tamaño de muestra.

N = Tamaño de la población.

Z = Nivel de confianza.

E = Error de estimación máximo aceptado.

P = Probabilidad de éxito.

Q = Probabilidad de que no ocurra el evento estudiado.

Sin embargo, cabe recalcar que el cuestionario debido a que fue a través de Google Forms, fue llenado por 326 personas, por lo que se trabajó con ese número de muestra debido a la naturaleza de la investigación en donde necesitamos la mayor cantidad de datos posible.

4.4 Desarrollo de la Investigación

Para realizar el desarrollo del modelo propuesto se utilizó la metodología CRISP-DM, desarrollada por IMB Analytics (2015) la cual consta de las siguientes etapas:

Etapa I: Comprensión del negocio

En secciones previas de esta investigación, se ha proporcionado información acerca del contexto y problemática en estudio.

Como resultado, en esta sección se consolidará y resumirá esa información.

En el capítulo I, se definió la realidad problemática, el problema general y específicos de la investigación, así como los objetivos. Los principales interesados y beneficiados son la Universidad Nacional José María Arguedas, ya como segundos y terceros se encuentran el Centro de Salud Comunitario de la UNAJMA y en especial los estudiantes pertenecientes a esta misma casa de estudios.

En el capítulo II, se detallan los antecedentes de estudio. Para finalizar, en esta última sección se presentan las métricas de evaluación del modelo los cuales son la exactitud y validación cruzada.

Esta etapa, ha proporcionado una visión completa y detallada del contexto de la investigación, es fundamental conocer la relevancia, resultados y soluciones que se presentarán posteriormente. Con esta base establecida procederemos a las siguientes etapas de investigación, donde se aplicaron métodos y técnicas específicas para abordar el problema y alcanzar los objetivos propuestos.

Etapa II: Comprensión de los datos

En esta etapa, el objetivo principal fue adquirir un conocimiento más profundo acerca de los datos que se utilizaron, cabe recalcar que es importante conocer bien los datos para realizar un buen análisis en las etapas posteriores. Dentro de esta etapa, se realizó la recopilación de datos, exploración de datos, calidad de datos y el entendimiento del contexto, los cuales se desarrollan a continuación:

A. Recopilación de datos

La recolección de datos se realizó a través de un formulario de Google, en el cual se aplicó un cuestionario para obtener información de datos socioeconómicos y otros datos acerca del estrés académico de cada uno de los estudiantes.

El cuestionario se encuentra dividido en dos secciones, la primera parte solicita principalmente datos socioeconómicos y otros considerados importantes para determinar el nivel de estrés académico; esta data para poder realizar la predicción.

En la segunda parte, se colocó una herramienta para medir el nivel de estrés académico, el Inventario Sistemático Cognoscitivista para el estudio del estrés académico (SISCO), con el cual se pudo obtener un puntaje determinado y finalmente el nivel de estrés académico de cada uno de los estudiantes.

Para poder visualizar de mejor manera este cuestionario, ir al Anexo 01.

B. Exploración y calidad de datos

Se realizó una exploración rápida de los datos obtenidos en un inicio, el dataset inicial lo pueden visualizar en el Anexo 03. Cabe recalcar que este dataset fue directamente exportado de Google Forms, por lo que el nombre de las columnas e incluso datos de las preguntas están tal y como se solicitaron en la encuesta como se muestra en la imagen a continuación:

Figura 5

Dataset inicial con columnas por defecto de Google Forms

	1. Marca temporal	2. Indica tu edad	3. Selecciona tu género	4. Selecciona tu lengua materna	5. Indica el ciclo que estas cursando actualmente	6. Indica el número de créditos que estas llevando en el presente semestre	7. ¿Cuál es tu promedio de calificaciones actual?	8. ¿Cuántas horas a la semana trabajas?	9. ¿Cuál es tu ingreso económico familiar?	10. ¿Cuál es tu ingreso económico propio?	11. ¿Con qué frecuencia para enfrentar tu estrés te orientas a:	12. ¿Con qué frecuencia enfrentar tu estrés te orientas a:	13. ¿Con qué frecuencia enfrentar tu estrés te orientas a:	14. ¿Con qué frecuencia enfrentar tu estrés te orientas a:
0	5/06/2023 18:25:16	21	Prefiero no decirlo	Castellano	VIII	24	16.29	0	2000	50	Rara vez	Rara vez	Rara vez	Rara vez
1	5/06/2023 16:26:04	21	Masculino	Castellano	V	22	14	10	0	10	Rara vez	Rara vez	Rara vez	Casi nunca

Se observó que la calidad de datos no es buena, encontramos datos faltantes, el tipo de datos no es homogéneo, los datos están desbalanceados, los datos no se encuentran estandarizados o normalizados, etc.

Una vez identificados los problemas con el dataset obtenido, es necesario realizar la preparación de los datos lo cual se encuentra en la siguiente etapa.

Etapa III: Preparación de los datos

Esta es una de las etapas más importantes de la metodología, se realizó el preprocesamiento de datos, codificación de variables, etiquetado de la variable objetivo, limpieza de datos, tratamiento de datos faltantes, balanceo de datos y selección de características o reducción de la dimensionalidad, los cuales se explican a continuación:

A. Preprocesamiento de datos

Como primer paso se realizó la modificación de los encabezados de las columnas para hacerlos más concisos y manejables. En la Tabla 2, se muestran algunos ejemplos:

Tabla 2*Ejemplos de modificación del encabezado en el dataset*

Sección	Encabezado	Encabezado actualizado
	1. Indica tu edad:	Edad
Primera	5. Indica el número de créditos que estas llevando en el presente semestre:\nSolo responde con un número.	Num_Creditos
	23. Responde, ¿Cuántas horas al día duermes?\nSolo responde con un número	Horas_suenio
	3. A continuación se presentan una serie de aspectos que, en mayor o menor medida, suelen estresar a algunos alumnos. Selecciona la opción que indique ¿con que frecuencia cada uno de esos aspectos te estresa? [3.1.- La competencia con mis compañeros del grupo]	3A
Segunda	4. Indica ¿Con qué frecuencia se te presentan las siguientes reacciones cuando estás estresado? [4.1.- Trastornos en el sueño (insomnio o pesadillas)]	4A
	5. ¿Con qué frecuencia para enfrentar tu estrés te orientas a: [5.1.- Habilidad asertiva (defender nuestras preferencias, ideas o sentimientos sin dañar a otros)]	5A

En la que fue la primera sección de la encuesta realizada, se solicitaron datos socioeconómicos además de otros datos relacionados con el tema de estudio, a estos se les cambió el nombre con algo corto pero representativo para poder visualizar posteriormente de que datos se trata.

Sin embargo, los datos de la segunda sección de la encuesta no son necesarios conocer cuáles son o a que número pertenece, debido a que se procesarán para obtener un único puntaje. Es por ello que según la sección a la que pertenecían se les nombró adjuntando una letra, por ejemplo: 3A, 3B, 3C, 4A, 4B, 4C, 5A, 5B, 5C, etc. Para cada sección se usaron las letras de la A hasta la P.

Después de realizar ese proceso, observamos los encabezados de columnas que quedaron, el número de datos que contiene cada una y el tipo de dato como se muestra a continuación:

Data columns (total 85 columns):

#	Column	Non-Null Count	Dtype
0	Marca temporal	326 non-null	object
1	Edad	326 non-null	object
2	Genero	326 non-null	object
3	Lengua materna	326 non-null	object
4	Ciclo	326 non-null	object
5	Num_Creditos	326 non-null	object
6	Promedio	326 non-null	object
7	Horas_trabajo	326 non-null	object
8	Ingreso_familiar	326 non-null	object
9	Ingreso_propio	326 non-null	object
10	Num_amigos	326 non-null	object
11	Personalidad	326 non-null	object
12	Num_hermanos	326 non-null	object
13	Relacion	326 non-null	object
14	Tipo_relacion	128 non-null	object
15	Hijos	326 non-null	object
16	Num_hijos	87 non-null	object
17	Padres	326 non-null	object
18	Relacion_padres	317 non-null	object
19	Distancia	326 non-null	object
20	Religion	326 non-null	object
21	Habitos alimenticios	326 non-null	object
22	Salud mental_satisfaccion	326 non-null	object
23	Horas_suenio	326 non-null	object
24	Hora de levantarse	326 non-null	object
25	Satisfaccion_carrera	326 non-null	object
26	Continuidad_carrera	326 non-null	object
27	Motivacion academica	326 non-null	object
28	Color	326 non-null	object
29	Horas semanales_RS	326 non-null	object

30	Ataques panico	326 non-null	object
31	Antecedentes	326 non-null	object
32	Visita centros de salud	326 non-null	object
33	Horas de estudio	326 non-null	object
34	Horas_AcFisica	326 non-null	object
35	Consumo_Alcohol	326 non-null	object
36	Momentos de estres	326 non-null	object
37	Escala_estres	326 non-null	int64
38	3A	326 non-null	object
39	3B	326 non-null	object
40	3C	326 non-null	object
41	3D	326 non-null	object
42	3E	326 non-null	object
43	3F	326 non-null	object
44	3G	326 non-null	object
45	3H	326 non-null	object
46	3I	326 non-null	object
47	3J	326 non-null	object
48	3K	326 non-null	object
49	3L	326 non-null	object
50	3M	326 non-null	object
51	3N	326 non-null	object
52	3P	326 non-null	object
53	4A	326 non-null	object
54	4B	326 non-null	object
55	4C	326 non-null	object
56	4D	326 non-null	object
57	4E	326 non-null	object
58	4F	326 non-null	object
59	4G	326 non-null	object
60	4H	326 non-null	object
61	4I	326 non-null	object
62	4J	326 non-null	object
63	4K	326 non-null	object
64	4L	326 non-null	object
65	4M	326 non-null	object
66	4N	326 non-null	object
67	4P	326 non-null	object
68	5A	326 non-null	object
69	5B	326 non-null	object
70	5C	326 non-null	object
71	5D	326 non-null	object
72	5E	326 non-null	object
73	5F	326 non-null	object
74	5G	326 non-null	object

```

75 5H          326 non-null object
76 5I          326 non-null object
77 5J          326 non-null object
78 5K          326 non-null object
79 5L          326 non-null object
80 5M          326 non-null object
81 5N          326 non-null object
82 5P          326 non-null object
83 Carrera    190 non-null object
84 Correo     190 non-null object
dtypes: int64(1), object(84)

```

B. Proceso de etiquetado de datos

La segunda sección del cuestionario, que ahora se encuentra en el dataset en las columnas del 38 al 82, estuvo compuesta básicamente por datos en escala de lickert como: Nunca, casi nunca, rara vez, raras veces, algunas veces, casi siempre y siempre, tal y como indica el Inventario SISCO.

Como no existen datos faltantes en esas columnas, se procedió a reemplazar los valores ordinales por valores numéricos según los baremos establecidos para el inventario que se aplicó.

El inventario SISCO aplicado en esta investigación, posee 3 dimensiones con 15 ítems cada uno, además de dos ítems adicionales al inicio del inventario. Estos dos ítems no son considerados para determinar el puntaje requerido, ya que el primero determina la continuidad de llenar el inventario SISCO y el otro pide la escala de estrés que percibe el estudiante del 1 al 5. (Alania et al., 2021)

En la tabla 3, observamos que para las dimensiones estresores y síntomas del estrés se realizó una puntuación del 0 al 5 del nivel más bajo al más alto. Mientras que para la dimensión de estrategias de afrontamiento se realizó el procedimiento invirtiendo los valores en la puntuación numérica, tal como lo establecen (Alania et al., 2021) en la parte de calificación del artículo de revista.

Tabla 3

Puntuación de las dimensiones del inventario SISCO

Dimensión	Puntuación numérica	Rango o nivel
Estresores y síntomas del estrés	0	Nunca
	1	Casi nunca
	2	Raras veces
	3	A veces
	4	Casi siempre
	5	Siempre
Estrategias de afrontamiento al estrés	5	Nunca
	4	Casi nunca
	3	Raras veces
	2	A veces
	1	Casi siempre
	0	Siempre

Se realizó la puntuación con un mapeo de las columnas involucradas, quedando de la siguiente manera:

Figura 6

Visualización de la data antes de aplicar el mapeo

	3A	3B	3C	3D	3E	3F	3G	3H	3I	3J	3K	3L	3M	3N	3P	4A	4B	4C	4D	4E	4F	4G
	Casi nunca	Rara vez	Casi nunca	Rara vez	Casi nunca	Casi nunca	Algunas veces	Casi nunca	Rara vez	Casi nunca	Algunas veces	Casi nunca	Raras veces	Algunas veces	Casi nunca	Casi siempre	Casi nunca					
	Casi nunca	Rara vez	Rara vez	Casi nunca	Casi nunca	Rara vez	Casi nunca	Casi nunca	Rara vez	Rara vez	Casi nunca	Rara vez	Casi nunca	Casi nunca	Rara vez	Casi nunca	Casi nunca	Casi nunca	Nunca	Raras veces	Casi nunca	Raras veces

En la Figura 6, observamos la data aún con los valores ordinales, esto para todas las dimensiones del inventario SISCO.

Figura 7

Visualización de la data después de aplicar el mapeo

```
data.head()
```

3A_reemplazada	3B_reemplazada	3C_reemplazada	3D_reemplazada	3E_reemplazada	3F_reemplazada	3G_reemplazada	3H_reemplazada	3I_reemplazada	3J_reemplazada	3K_reemplazada	3L_reemplazada
1	2	1	2	1	1	3	1	2	1	1	
1	2	2	1	1	2	1	1	2	2	1	

En la Figura 7, observamos la creación de nuevas columnas ya con los valores numéricos después de realizar el mapeo de datos.

Finalmente se realizó una suma de los valores con respecto a cada instancia con la finalidad de obtener un puntaje entre 0 a 225. Este puntaje se guardó en el data frame “dataV2” el cual contiene también otras columnas con las que se realizará la predicción y tiene el encabezado “puntaje_estres”.

Los baremos establecidos en la tabla 4, se usaron para realizar la discretización de la columna “puntaje_estres”.

Tabla 4

Baremos del estrés académico

Nivel	Puntaje de estrés académico
Bajo o leve	0-78
Moderado	79-157
Alto o crónico	158-225

Según la tabla 4, se realizó a su vez la discretización de la variable objetivo en categorías para facilitar la tarea de clasificación al momento de realizar el entrenamiento del modelo. Se agrupó el puntaje de estrés académico que se obtuvo por estudiante teniendo como referencia los baremos establecidos por Alania et al. (2021), de forma que a los puntajes que se encontraban en el nivel leve, se les asignó el valor numérico ‘0’; a los del nivel moderado, se les asignó el valor numérico ‘1’ y al nivel alto, el valor numérico ‘2’. De esta forma, la variable objetivo creada ‘Nivel_Estres’ contiene los datos etiquetados.

C. Codificación de variables

Para realizar la codificación de variables, primero se buscaron las columnas con valores nominales y sin datos faltantes.

Figura 8

Visualización de datos faltantes

```

▶ datos_faltantes=df_codificado.isnull().sum()
datos_faltantes

```

Edad	0
Ciclo	0
Num_Creditos	0
Promedio	0
Horas_trabajo	0
Ingreso_familiar	0
Ingreso_propio	0
Num_amigos	0
Num_hermanos	0
Tipo_relacion	198
Hijos	0
Num_hijos	239
Relacion_padres	9
Distancia	0
Habitos_alimenticios	0
Salud_mental_satisfaccion	0
Horas_suenio	0
Hora de levantarse	0
Satisfaccion_carrera	0
Motivacion_academica	0
Horas_semanales_RS	0
Ataques_panico	0
Antecedentes	0
Visita centros de salud	0
Horas de estudio	0
Horas_AcFisica	0
Consumo_Alcohol	0
Escala_estres	0
Carrera	136

En la Figura 8 observamos que solo 3 características poseen datos faltantes, los cuales son: Tipo_relacion, Num_hijos y Carrera, por lo que con los demás datos se puede trabajar sin problemas.

Se aplicó la codificación One Hot a las que características nominales: Genero, lengua materna, personalidad, relación, padres, religión, continuidad_carrera, carrera y color. Se considera el tipo de codificación adecuado para este tipo de variables.

Figura 9

Codificación One Hot a variables nominales



The screenshot shows a Jupyter Notebook cell with the command `df_codificado.head(2)` and the resulting DataFrame. The DataFrame has 12 columns representing One-Hot encoded variables. The first two rows of data are as follows:

Relacion_No	Relacion_Sí	Padres_Ninguno	Padres_Solo mamá	Padres_Solo papá	Padres_Sí, ambos	Religion_No	Religion_Soy agnóstico(a)	Religion_Sí	Continuidad_carrera_No	Continuidad_carrera_Sí	Continuidad_carrera_Sí pero r
1	0	0	0	0	1	0	0	1	1	0	
1	0	0	0	0	1	0	0	1	1	0	

En la codificación One Hot se crean columnas usualmente binarias, pero si se tienen por ejemplo 3 categorías diferentes se crean 3 columnas y se le asigna valores binarios, 1 si la observación pertenece a la categoría correspondiente y 0 si no pertenece. Podemos ver el resultado de la aplicación de la codificación en la Figura 9.

Se aplicó codificación de datos numéricos a los datos ordinales, por ejemplo, a la característica “Ciclo” el cual estaba con valores de tipo objeto representado con números romanos, se reemplazó con valores numéricos del 1 al 10 respectivamente. Como resultado se obtiene la característica “CicloEncoded” con los valores únicos numéricos del 1 al 10 tal como se muestra en la Figura número 10.

Figura 10

Codificación de la característica "Ciclo"

```
[ ] df_codificado['CicloEncoded'] = df_codificado['Ciclo'].map( {'I': 0, 'II': 1, 'III': 3, 'IV': 4, 'V': 5, 'VI': 6, 'VII': 7,
                                                                'VIII': 8, 'IX': 9, 'X': 10} ).astype(int)

[ ] #Extraemos los valores unicos de la columna CicloEncoded
df_codificado['CicloEncoded'].unique()

array([ 8,  5,  7,  4,  6, 10,  3,  0,  1,  9])
```

En adición, en lo que respecta a las características: 'Tipo' relacion', 'Relacion_padres', 'Habitos alimenticios', 'Salud mental_satisfaccion', 'Satisfaccion_carrera', 'Motivacion academica', 'Ataques panico', 'Antecedentes', 'Visita centros de salud' y 'Consumo_Alcohol'; también se aplicó la codificación ordinal. Se asignaron números a las diferentes categorías de manera que capturen el orden o la jerarquía entre ellas. En las figuras 11; 12 y 13 se muestran algunos ejemplos de lo realizado.

Figura 11

Codificación ordinal de 'Habitos alimenticios'

```
#Mapeo de las categorias de Habitios alimenticios
df_codificado['Habitios_alimenticios_Encoded'] = df_codificado['Habitios alimenticios'].map(
    {
        'Muy malos': 0,
        'Malos': 1,
        'Regular': 2,
        'Buenos': 3,
        'Excelentes':4
    }).astype(int)
```

Figura 12

Codificación ordinal de 'Motivacion academica'

```
#Mapeo de las categoría de Motivacion academica
df_codificado['MotAcademica_Encoded'] = df_codificado['Motivacion academica'].map(
    {
        'No': 0,
        'Tal vez': 0.5,
        'Sí': 1,
    }).astype(float)
```

Figura 13

Codificación ordinal de 'Consumo_Alcohol'

```
#Mapeo de las categorías de Consumo_Alcohol
df_codificado['ConsumoAlcohol_Encoded'] = df_codificado['Consumo_Alcohol'].map(
    {
        'Nunca': 0,
        'Raramente': 1,
        'A veces': 2,
        'Frecuentemente': 3,
        'Siempre': 4
    }).astype(int)
```

D. Limpieza de datos

Se realizó también la limpieza de datos en cuanto a las características que fueron obtenidas a través de respuestas abiertas o recolectadas de un campo de texto libre. Debido al tipo de datos que fueron no estructurados, los cuales tampoco estaban predefinidos en las características: “Num_credits”, “Promedio”, “Edad”, “Horas_trabajo”, “Ingreso_familiar”, “Ingreso_propio”, “Num_amigos”, “Num_hermanos”, “Distancia”, “Horas_suenio”, “Hora de levantarse”, “Horas semanales_RS”, “Horas de estudio” y “Horas_AcFisica”, se tuvo que aplicar algunas funciones y mapeo de datos para limpiar cada una de las características.

En la tabla 5, se detalla los factores considerados en esta investigación para la limpieza y normalización de datos de este tipo.

Tabla 5

Factores considerados para la limpieza y normalización de datos

Factor considerado	Descripción
Eliminación de signos de puntuación	<p>Se eliminaron signos de puntuación que fueron introducidos como respuesta en los diferentes campos, se reemplazaron con espacios en blanco a los que contenían únicamente signos de puntuación a excepción de los que representaban números decimales.</p> <p>Ejemplo: ‘.,;’...’</p>
Eliminación y reemplazo de valores atípicos	<p>En las características consideradas al realizar la limpieza de datos existen valores no válidos que no contienen información relevante, a esos valores se reemplazaron por valores NAN (valores no numéricos), esto con la finalidad de realizar operaciones futuras con Numpy, Pandas, etc.</p> <p>Ejemplo: ‘Nose’, ‘no recuerdo’, ‘Aún nose’, ‘Regular’, ‘Ni idea’, ‘No tengo un aproximado’, ‘No tengo una cantidad exacta’, etc.</p>
Extracción de valores e información.	<p>Fue necesario realizar la extracción de información relevante, identificando y extrayendo los valores numéricos asociados para asegurar la coherencia y el formato adecuado de los datos para su posterior uso.</p> <p>Ejemplos:</p> <ul style="list-style-type: none"> E. ‘14 a más’: 14 F. ‘6 horas, 4 veces a la semana’: 24 G. ‘18 horas con pago, 25 horas de practica’: 43 H. ‘25 horas a la semana’: 25 I. ‘100 semanal’: 100 J. ‘200👁️’: 200

	<p>K. 'No tengo ingresos mensuales': 0</p> <p>L. '12-15 aprox': 13</p> <p>M. 'amigos? 3': 3</p> <p>N. '20 minutos': 20</p> <p>O. '1 hora': 60</p> <p>P. 'Una hora': 60</p> <p>Q. '15m': 15</p>
--	--

	<p>Para abordar los valores faltantes o indeterminados, se realizó el reemplazo de los mismos asignando valores aleatorios en el intervalo de confianza de la desviación estándar debido a que se realizó este procedimiento con datos de valor numérico.</p> <p>Ejemplo:</p>
--	--

Imputación de datos	<p>En la característica 'PromedioFiltrado', se tenían 18 datos nulos o faltantes por lo que se realizó la imputación de datos con los valores que se obtuvieron aleatoriamente del intervalo de confianza de la desviación estándar, donde los últimos 5 datos de la lista creada fueron: [12, 13, 13, 11, 12].</p> <p>De la misma forma se realizó con las demás características con valores faltantes.</p>
----------------------------	--

	<p>Implicó dividir el rango de valores en "bins" o contenedores, donde cada bin representa un rango o intervalo de valores.</p> <p>Ejemplo:</p>
--	--

Discretización de características	<p>Para obtener la variable objetivo, se tuvo que realizar la discretización tomando en cuenta los siguientes rangos: 'Leve': 0-78; 'Moderado': 79-157; 'Fuerte': 158-225, donde se representó con '0' a los clasificados en 'Leve';</p>
--	--

‘1’ a los clasificados en ‘Moderado’ y ‘2’ a los clasificados en ‘alto’.

E. Tratamiento de datos faltantes

Como se mostró anteriormente, tenemos 3 características con datos faltantes. Para la característica que expresa el Número de hijos “Num_hijos” existen los valores vacíos, además de “1”, “0”, “No tengo”, “2”, “no”, “O”, “Ninguno”, “3” y “NO” los cuales se obtuvieron sacando valores únicos de la columna en mención. Teniendo este panorama y además observando la característica “Hijos”, que indica si un estudiante tiene o no un hijo y la respuesta predominante es “No”, se procedió a rellenar los espacios en blanco con el valor “0”.

Con ello ya no se tienen valores faltantes y los valores únicos ahora de “Num_hijos” son: 0; 1; 2 y 3 como se muestra en la Figura 14.

Figura 14

Valores de Num_hijos después del tratamiento de datos faltantes

```
[ ] #Valores nuevos de la columna Num_hijos
df_codificado['Num_hijos'].unique()

array([0, 1, 2, 3], dtype=int64)
```

En lo que respecta a la característica “Carrera”, existen 136 datos faltantes, los cuales se rellenaron todos con “EPIS – Escuela Profesional de Ingeniería de Sistemas” por la razón de que, en un inicio, la encuesta fue aplicada solo a estudiantes de la EPIS, pero teniendo la necesidad de obtener mayor cantidad de atributos se aplicó la encuesta en las demás escuelas profesionales adicionando el campo de carrera profesional.

Por último, para la característica “Tipo_relacion” la cual hace referencia al tipo de relación amorosa en la que se encuentran, se tenían las opciones “Salientes, pero nada concreto”, “Enamorados” y “Esposos”, además, observando en la característica Tipo_relacion que la mayoría no está en una relación amorosa. Los datos faltantes se rellenaron con “Ninguno”, ya que se asume que esas personas no se encuentran en una relación amorosa.

Con ello no tendríamos datos faltantes en ninguna característica.

Finalmente, después de realizar el preprocesamiento, limpieza y demás procedimientos, se tiene un total de 58 características con 326 atributos entre datos de tipo entero y flotante como se muestra a continuación:

Data columns (total 58 columns):

#	Column	Non-Null Count	Dtype
0	Edad	326 non-null	int64
1	Num_Creditos	326 non-null	int64
2	Num_hijos	326 non-null	int64
3	Escala_estres	326 non-null	int64
4	puntaje_estres	326 non-null	int64
5	Genero_Femenino	326 non-null	int64
6	Genero_Masculino	326 non-null	int64
7	Genero_Prefiero no decirlo	326 non-null	int64
8	Idioma_Castellano	326 non-null	int64
9	Idioma_Otra	326 non-null	int64
10	Idioma_Quechua	326 non-null	int64
11	Personalidad_Ambivertido	326 non-null	int64
12	Personalidad_Extrovertido	326 non-null	int64
13	Personalidad_Introvertido	326 non-null	int64
14	Relacion_No	326 non-null	int64
15	Relacion_Sí	326 non-null	int64
16	Padres_Ninguno	326 non-null	int64
17	Padres_Solo mamá	326 non-null	int64
18	Padres_Solo papá	326 non-null	int64
19	Padres_Sí, ambos	326 non-null	int64
20	Religion_No	326 non-null	int64
21	Religion_Soy agnóstico(a)	326 non-null	int64
22	Religion_Sí	326 non-null	int64
23	Continuidad_carrera_No	326 non-null	int64

24	Continuidad_carrera_Sí	326 non-null	int64
25	Continuidad_carrera_Sí, pero retorné	326 non-null	int64
26	Color_Piel clara	326 non-null	int64
27	Color_Piel media	326 non-null	int64
28	Color_Piel oscura	326 non-null	int64
29	CicloEncoded	326 non-null	int64
30	PromedioFiltrado	326 non-null	float64
31	Horas_trabajo_entero	326 non-null	int32
32	Ingreso_familiarCod	326 non-null	Int64
33	Ingreso_propioEncoded	326 non-null	Int64
34	Num_amigos_Encoded	326 non-null	Int64
35	Num_hermanos_Encoded	326 non-null	Int64
36	Tipo_relacion_Encoded	326 non-null	int32
37	Relacion_padresCod	326 non-null	int32
38	DistanciaCod1	326 non-null	int64
39	Habitos_alimenticios_Encoded	326 non-null	int32
40	SaludMental_Encoded	326 non-null	int32
41	Horas_suenioEncoded	326 non-null	int64
42	HLevantarse_Encoded	326 non-null	float64
43	Satisfaccion_carrera_Encoded	326 non-null	int32
44	MotAcademica_Encoded	326 non-null	float64
45	HorasRS_Encoded	326 non-null	int64
46	AtaquesPanico_Encoded	326 non-null	int32
47	AntecedentesEncoded	326 non-null	float64
48	CSalud_Encoded	326 non-null	int32
49	HEstudio_Encoded	326 non-null	int64
50	AcFisica_Encoded	326 non-null	int64
51	ConsumoAlcohol_Encoded	326 non-null	int32
52	Carrera_EPAA	326 non-null	int32
53	Carrera_EPC	326 non-null	int32
54	Carrera_EPEPI	326 non-null	int32
55	Carrera_EPIA	326 non-null	int32
56	Carrera_EPIAM	326 non-null	int32
57	Carrera_EPIS	326 non-null	int32

F. Balanceo de datos

El balanceo de datos se aplicó al conjunto de datos que se obtuvo después del preprocesamiento de datos.

Se escogió como técnica el sobremuestreo, de forma que no se pierda información como se haría si se usa por ejemplo el submuestreo. La cantidad de instancias que actualmente se posee es

suficiente, sin embargo, al existir una cantidad mínima de datos con respecto a una de las clases es necesario realizar el sobremuestreo, el cual consiste en incrementar el número de instancias a las clases minoritarias replicándose hasta que queden en un grado constante.

Tabla 6

Instancias antes y después del balanceo de datos

Nivel de estrés	Antes del sobre muestreo	Después del sobre muestreo
Estrés leve (0)	45	267
Estrés moderado (1)	267	267
Estrés alto (2)	14	267
Total	326	801

Como se observa en la tabla 6, se realizó el sobre muestreo a todo el conjunto de datos, es por ello que antes de realizar el sobre muestreo se tiene un total de 326 instancias y después un total de 801 con cada instancia igual a 267.

G. Normalización y reducción de la dimensionalidad

Se realizó la normalización de datos para ajustar las escalas de las variables y estas tengan un rango similar, es importante realizar este proceso ya que las variables poseen magnitudes diferentes, se usó la clase ‘StandardScaler’ de scikit-learn. Para visualizar de mejor manera el proceso de normalización de datos, visite el Anexo 2.

Después de la normalización de datos, se realizó el proceso de modelamiento y evaluación del modelo de inteligencia artificial utilizando el conjunto de datos que se obtuvo después del preprocesamiento de datos que se mostró anteriormente, el cual contiene 58 características incluyendo la variable objetivo.

Sin embargo, después se realizó el análisis exploratorio de datos con el fin de realizar la reducción de la dimensionalidad, con lo cual se obtuvo varias versiones del conjunto de datos inicial. Esto con el fin de mejorar los parámetros de rendimiento del modelo.

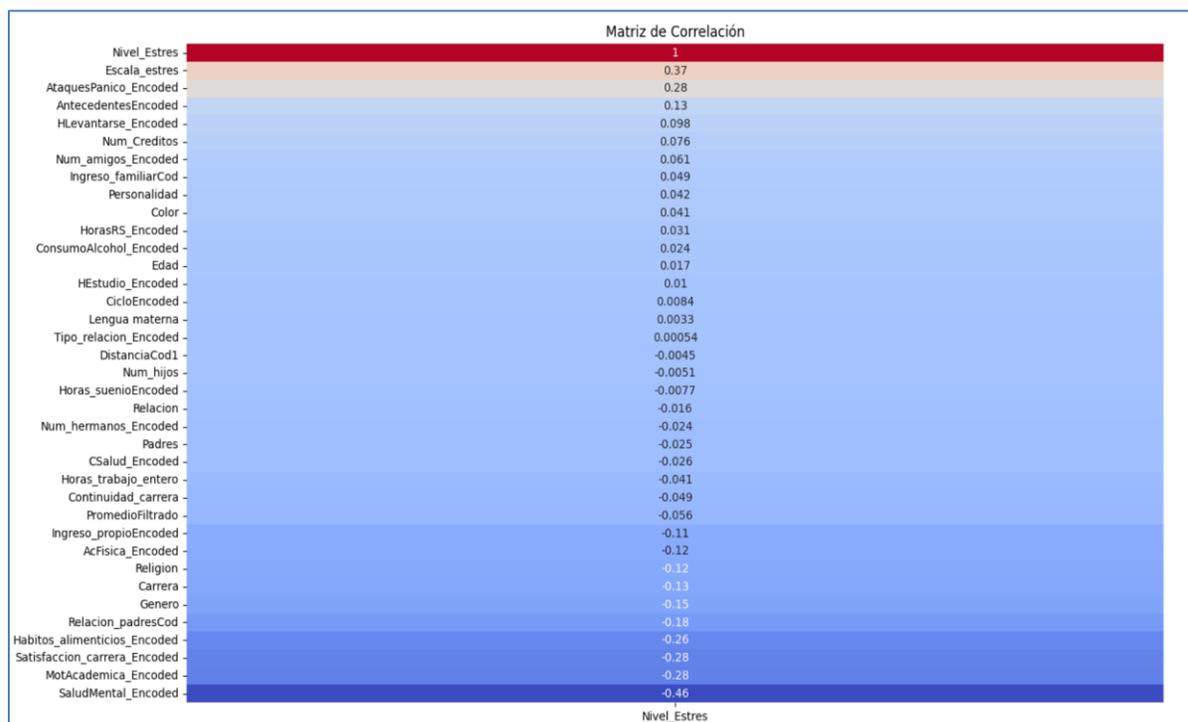
A continuación, se especifican los procedimientos que se siguieron para la construcción de los diferentes conjuntos de datos que se usaron en esta investigación y en especial para la data final con la que se obtuvo buenos resultados.

a) Exploración de datos

Como primera instancia, se realizó la exploración de datos para analizar la correlación que tenían los datos de entrada con la variable de salida. Para ello se realizó la visualización de la matriz de correlación con un mapa de calor como se muestra en la Figura 15.

Figura 15

Matriz de correlación

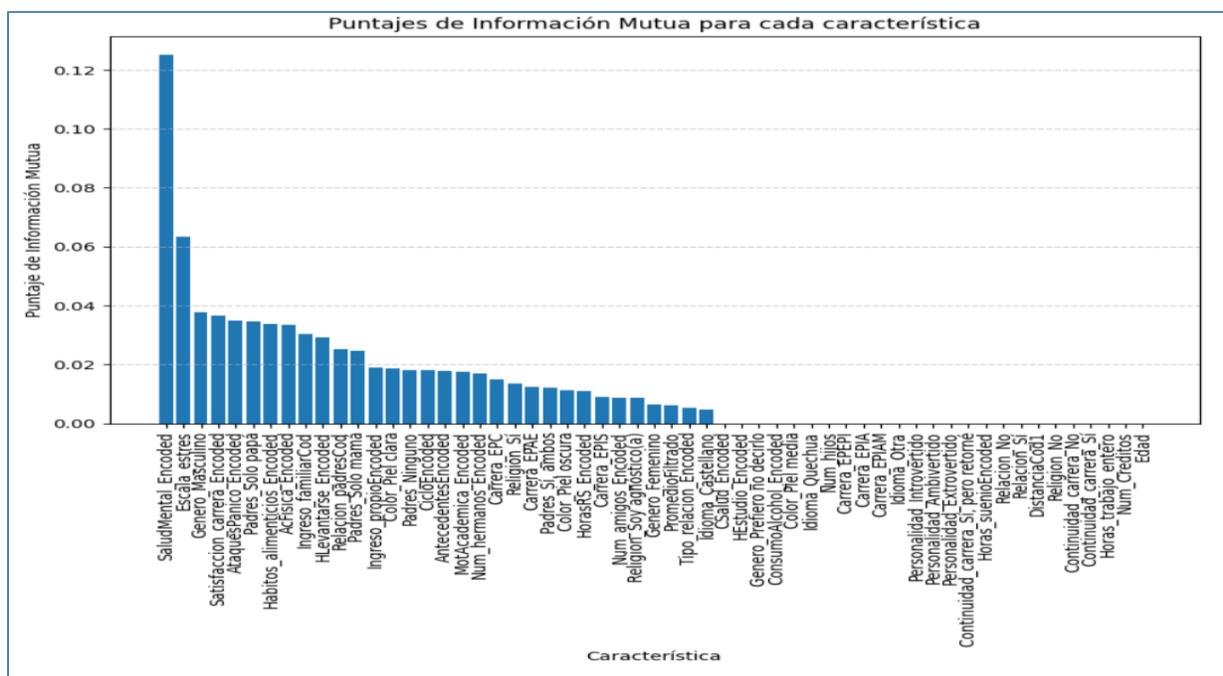


Se observa en la Figura 15 que las variables de entrada ‘Escala_estres’, ‘AtaquesPanico_Encoded’, ‘AntecedentesEncoded’, ‘HLevantarse_Encoded’ y así sucesivamente, son las variables que presentan mayor correlación con la variable de salida ‘Nivel_Estres’. Sin embargo, también se realizó la visualización de datos con la técnica de información mutua, esta técnica permite obtener un valor dependiente de la información compartida entre una variable de entrada y la variable objetivo.

En la Figura 16, se muestra un gráfico de barras en el que se representa la información mutua de cada una de las variables de entrada con la variable objetivo, una barra alta representa una fuerte dependencia o información compartida entre las dos variables, en este caso la variable ‘SaludMental_Encoded’ es la que presenta una fuerte dependencia con el nivel de estrés académico.

Figura 16

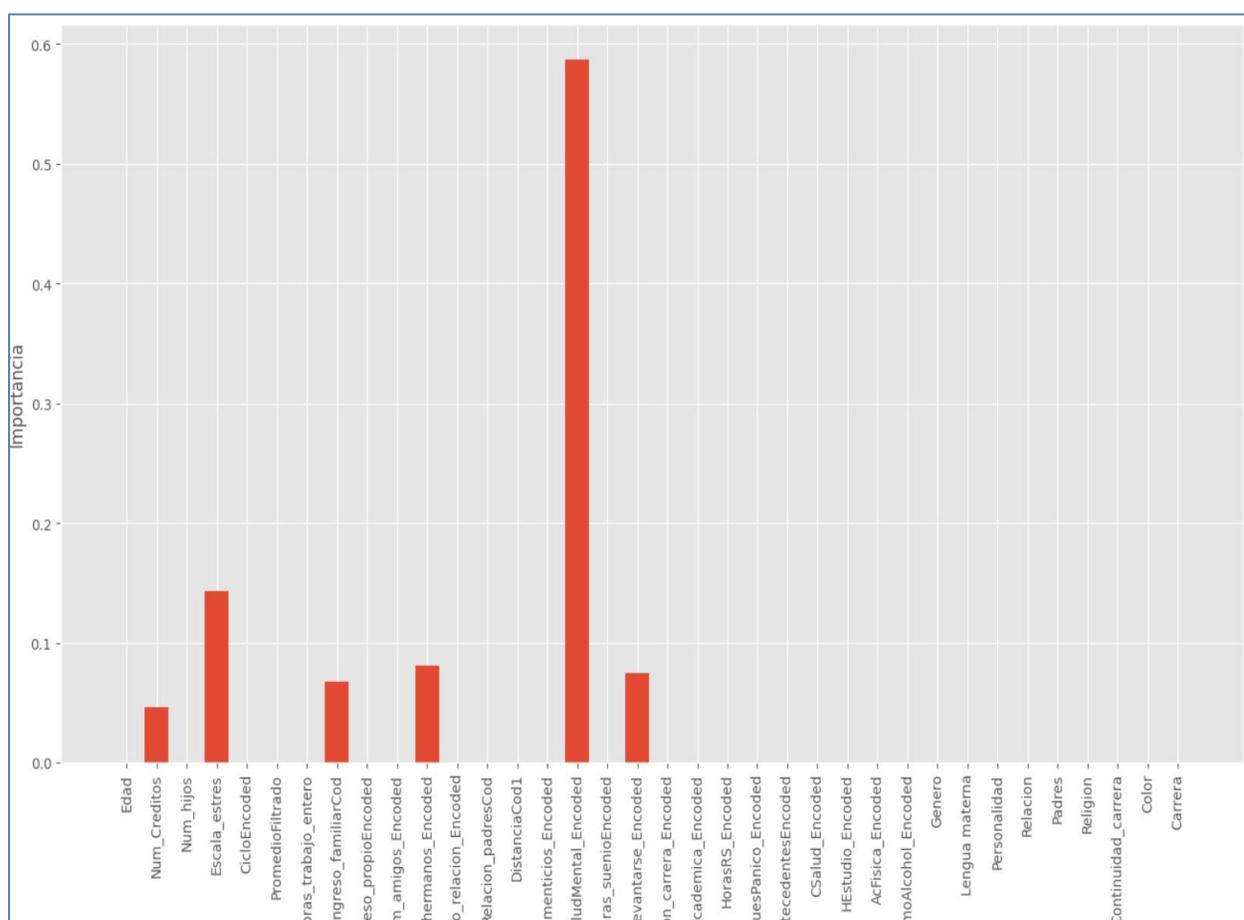
Puntuación de información mutua



Para evaluar que tan bien cada característica o variable de entrada contribuyó en la tarea de clasificación empleada con el modelo de árboles de decisión, se realizó un gráfico de barras como se muestra en la Figura 17. Ahí, se muestra la importancia que tuvieron las características en las decisiones tomadas por el modelo, por lo mismo se puede observar que ‘SaludMental_Encoded’ es la variable considerada con mayor importancia por el modelo de árboles de decisión, es decir esa variable mejora la capacidad del modelo para tomar decisiones informadas por lo que ayuda a realizar predicciones más precisas.

Figura 17

Importancia de características

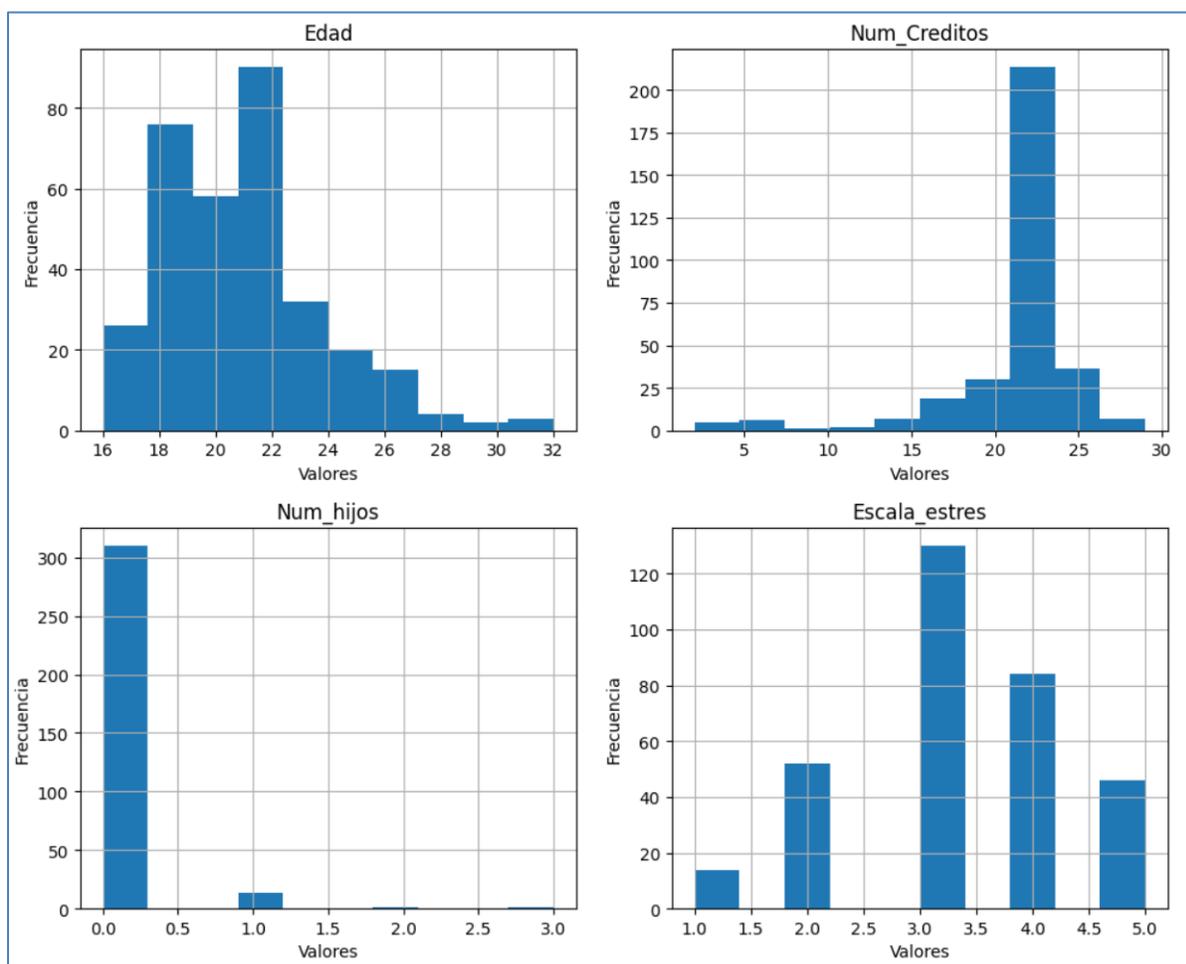


Por último, se realizó un gráfico de barras por cada característica o variable de entrada para visualizar la distribución de datos, fue de importancia para entender la dispersión de los datos por lo que ayudó en la decisión de selección de características observando la variabilidad y relevancia de cada uno.

En la Figura 18, se muestra 4 de los gráficos realizados como algo representativo, se puede observar completo en el notebook de reducción de la dimensionalidad que se encuentra en el Anexo 02.

Figura 18

Gráfico de barras de distribución



b) Selección de características

Después de haber realizado la exploración de datos, se tomaron en cuenta las características con mayor correlación e importancia para construir el nuevo conjunto de datos, además se observó la distribución de cada una de las variables de entrada para visualizar si es viable utilizarlos o no.

También se aplicó la técnica de PCA, o Análisis de Componentes Principales (Principal Component Analysis en inglés), donde se redujo la dimensionalidad del conjunto de datos a un número de componentes igual a 20. Al realizar las pruebas, se observó una mejora considerable en las métricas de evaluación de los diferentes modelos empleados. Sin embargo, en esta investigación se pretende reducir la dimensionalidad para realizar la recolección de datos de manera más sencilla y rápida. Con PCA no se puede lograr eso, ya que obtenemos un nuevo conjunto de componentes principales donde cada uno es una combinación lineal de todas las características originales, sin embargo, si se podría aplicar para mejorar el rendimiento de los modelos. Para visualizar a detalle, revisar el notebook PCA en el Anexo 2.

El conjunto de datos final y con el que se obtuvo buenos resultados, contiene 27 características o columnas, a diferencia del inicial que contenía 58 características. En la Figura 19, se muestra el conjunto de datos que se obtuvo al término de esta etapa.

Figura 19

Conjunto de datos final 'df_reducido'

Num_Creditos	Escala_estres	PromedioFiltrado	Horas_trabajo_entero	Ingreso_familiarCod	Ingreso_propioEncoded	Num_hermanos_Encoded	Tipo_relacion_Encoded	Relacion_padresCod	Distancia_Cod1	Habitos_alimenticios_Encoded	SaludMental_Encoded	HLevantarse_Encoded	Satisfaccion_carrera_Encoded	MotAcademica_Encoded	Ataques_Panico_Encoded	Antecedentes_Encoded	CSalud_Encoded	HEstudio_Encoded	AcFisica_Encoded	Genero	Lengua materna	Personalidad	Religion	Continuidad_carrera	Color	Nivel_Estres
24	4	16.29	0	2000	50	2	0	4	40	1	1	6	1	1	0	0	2	2	2	0	0	2	0	0	2	
22	2	14	10	0	10	7	0	4	30	2	2	6	2	0	0	0.5	0	4	1	1	0	2	2	0	1	2
20	4	15	0	0	900	4	0	0	40	3	2	10	2	1	0	0	4	30	1	0	2	2	0	0	2	
21	4	15	0	0	0	1	1	3	20	2	1	7	2	1	0	0.5	1	4	3	1	0	1	2	1	1	2
21	3	12	0	0	100	5	2	5	15	1	1	8	2	1	1	0	0	4	2	1	0	2	1	1	1	2
20	2	13.5	25	1000	400	6	1	4	60	3	2	6	2	0.5	0	0	15	4	1	2	2	2	0	1	2	
26	3	15	5	0	100	3	2	5	40	2	4	5.3	4	1	0	0	4	1	1	2	1	2	2	1	1	
23	1	13	24	0	1200	6	2	5	40	2	4	6.3	4	1	0	0.5	0	3	10	1	2	2	2	0	1	1
24	3	17	0	2000	920	2	2	4	10	2	1	6	4	1	0	0	1	20	7	1	0	0	1	0	1	2
20	3	15.5	0	1200	920	6	2	4	15	2	2	6.3	3	1	0	0	1	20	8	1	2	0	1	0	1	2
21	3	13	0	0	900	2	0	5	40	3	1	7	2	1	1	0	1	6	2	0	0	2	2	0	1	2
25	4	14	0	0	0	3	0	3	15	2	2	5.4	2	1	1	0	0	1	0	0	2	1	2	0	1	2
24	3	14	8	100	200	4	0	5	20	2	2	6.15	2	0.5	0	0	8	8	1	0	0	2	0	2	2	2
19	3	15	0	0	920	6	1	4	30	2	1	6	0	1	1	0	0	10	5	1	2	2	2	0	0	2
21	3	14	8	50	400	4	1	5	20	2	3	7	3	1	0	0	1	8	9	1	0	1	2	0	1	2
22	3	15	0	1300	920	6	0	5	5	2	3	6.3	3	1	0	0	1	5	7	1	2	0	2	0	1	1
21	3	14	0	0	0	1	0	4	40	2	2	7	3	1	0	0	4	4	2	0	2	2	0	1	1	
22	2	15	12	0	50	2	0	0	60	1	1	5	3	1	1	0	0	24	5	1	0	2	2	0	0	2
22	3	13	0	0	0	1	1	4	10	2	0	5	2	1	0	0	0	30	5	1	2	2	2	0	1	2
23	4	14	4	0	100	5	0	4	30	2	2	7	1	0.5	0	0	5	1	1	2	2	2	0	0	2	2
26	5	12	0	0	0	8	1	5	60	1	0	5	1	1	1	1	3	2	0	2	0	2	1	1	2	
20	3	15	15	1800	0	2	0	4	10	2	2	5.3	2	0.5	0	0.5	0	15	4	1	0	0	2	0	1	2
24	4	14.4	20	0	1000	9	0	4	10	2	1	6	2	0.5	0	0.5	0	10	12	1	2	2	0	0	0	2
21	3	12	0	0	0	8	2	5	30	3	1	6	2	1	0	0	0	10	5	0	2	2	2	0	0	1
22	2	13	8	500	150	3	2	5	60	2	2	5.3	2	1	0	0.5	0	15	4	1	0	1	2	2	1	2
23	4	15	0	600	0	4	0	5	20	3	3	5.3	3	0.5	0	0	7	4	1	2	2	2	0	1	1	
22	5	13	10	0	100	3	1	5	20	3	1	5	3	1	0	0	10	2	1	2	1	2	0	1	2	
23	3	14	10	0	920	2	3	3	15	3	2	7	2	0.5	0	0	1	20	1	0	1	2	0	1	1	2
5	3	11	48	0	500	6	2	4	15	2	2	6	2	1	1	0	1	18	24	0	2	0	2	0	1	1
22	2	15	20	1000	400	6	2	5	20	2	2	5	2	0	0	0.5	0	5	4	1	2	2	2	0	1	2
23	3	13	0	50	0	1	0	4	30	2	2	7.2	2	0.5	1	0.5	0	5	1	1	0	0	2	0	1	2
22	3	13	21	0	150	1	0	5	20	2	2	6.5	1	0.5	0	0	1	3	3	1	2	2	2	0	1	2
22	4	14	10	0	0	2	0	5	35	2	2	6.3	3	1	1	0	0	21	7	1	2	0	0	0	1	2
22	2	13	0	0	0	7	0	4	30	2	2	6	1	0	0	0.5	0	2	1	2	2	0	2	0	1	2

Etapa IV: Modelamiento

En esta etapa se realizó el proceso de modelamiento utilizando algoritmos de aprendizaje supervisado como son árboles de decisión, máquina de vectores de soporte, regresión logística, bosques aleatorios y redes neuronales.

Se realizó el entrenamiento de los datos con el 80% que se había dividido anteriormente para posteriormente evaluar los valores predictivos. Se puede contemplar con mayor detalle en el Anexo 02.

Etapa V: Evaluación

Se aplicaron medidas de rendimiento a los diferentes modelos de IA, tales como la exactitud, precisión y validación cruzada, este último es considerado como la manera más robusta y confiable de evaluar el rendimiento de un modelo, en lugar de dividir el conjunto de datos en una sola partición de entrenamiento y otra de prueba, divide el conjunto de datos en múltiples partes llamadas “folds” y realiza varias rondas de entrenamiento y evaluación.

Esta etapa fue importante para determinar cuál de los algoritmos utilizados fue adecuado y obtuvo mejores resultados para la predicción del nivel de estrés académico en estudiantes de la UNAJMA. Se puede visualizar de mejor manera en el Anexo 02.

Capítulo V: Resultados

En este capítulo se mostrarán los resultados que se obtuvieron después de realizar el entrenamiento y prueba del modelo de IA planteado con cada uno de los conjuntos de datos que se obtuvieron a partir de la etapa de preprocesamiento de datos.

El conjunto de datos con el que se trabajó y se obtuvieron los mejores resultados fue 'df_reducido.csv', el cual contiene 27 características o variables de entrada, incluyendo la variable objetivo 'Nivel_Estres' y 801 instancias, se logró realizar la reducción de la dimensionalidad, el cual se encuentra especificado en la etapa de preprocesamiento de la presente investigación.

Los resultados se mostrarán por cada algoritmo empleado en la elaboración del modelo, para esta investigación se trabajó con cinco algoritmos: Árbol de decisión, máquina de vectores de soporte, K vecinos más cercanos, redes neuronales y bosques aleatorios.

5.1. Modelo de Árbol de Decisión

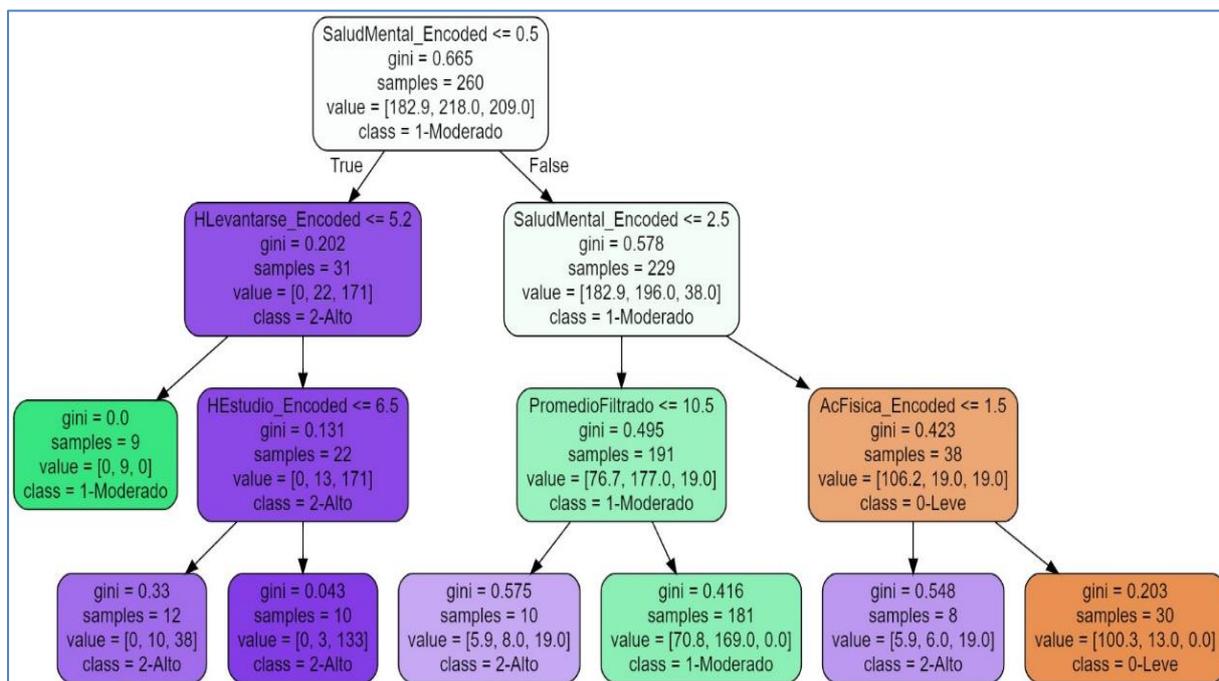
Se construyó un modelo de árbol de decisión con los parámetros de configuración:

- ❖ Criterion (Criterio). Está establecido en 'gini', lo que significa que el algoritmo utilizará el índice Gini como medida para evaluar la calidad de divisiones en los nodos del árbol. El índice de Gini mide la impureza de un conjunto de datos.
- ❖ Min Samples Split (Mínimo de muestras para dividir). Está configurado en 20, el árbol solo se dividirá en un nodo si hay al menos 20 muestras en ese nodo.
- ❖ Min Samples Leaf (Mínimo de muestras por hoja). Está configurado en 7, por lo que cada hoja del árbol debe contener por lo menos 7 muestras.
- ❖ Max Depth (Profundidad máxima). La profundidad máxima del árbol son 3 niveles, el cual se determinó hallando la exactitud en cada nivel del árbol y escogiendo el mejor puntaje.

Con ello se realizó el entrenamiento y prueba de rendimiento del mismo. En la Figura 20 se muestra la estructura del árbol de decisión y las reglas utilizadas para determinar el nivel de estrés académico de un estudiante.

Figura 20

Árbol de decisión para la segunda fase



Según la Figura 20, la mayoría de las reglas generadas son para la clasificación del estrés académico moderado y alto. Los cuadros en color morado representan la clase 2, estrés académico alto; los cuadros en color verde representan la clase 1, estrés académico moderado y los cuadros en color naranja representan la clase 0, estrés académico leve.

En la tabla 7, se detallan las reglas que fueron establecidas por este algoritmo para realizar la clasificación.

Tabla 7

Reglas establecidas por el modelo de árboles de decisión para predecir el nivel de estrés académico de un estudiante de la UNAJMA

Reglas establecidas por el modelo de árbol de decisión	Variable objetivo (Clasificación)
Salud mental ≤ 0.5 Hora de levantarse ≤ 5.2	Moderado
Salud mental ≤ 0.5 Hora de levantarse > 5.2 Horas semanales de estudio ≥ 6.5	Alto
Salud mental ≤ 2.5 Promedio del estudiante ≤ 10.5	Alto
Salud mental ≤ 2.5 Promedio del estudiante ≥ 10.5	Moderado
Salud mental ≥ 2.5 Horas de actividad física ≤ 1.5	Alto
Salud mental ≥ 2.5 Horas de actividad física ≥ 1.5	Leve

Observamos las reglas establecidas por el modelo de árbol de decisión. Según el algoritmo los estudiantes que tienen una consideración de su salud mental por debajo de 0.5 y se levantan antes de las 5:20 de la mañana, tienen un estrés moderado. Por otro lado, si tiene la misma consideración de su salud mental, pero se levantan después de las 5:20 de la mañana y estudian más de seis horas y media a la semana fuera del horario académico, tienen estrés académico alto.

Si consideran su salud mental por debajo de 2.5 y tienen un promedio desaprobatorio tienen un estrés académico alto, pero si su promedio es aprobatorio se considera como estrés académico moderado.

Por último, si su salud mental es considerada mayor a 2.5 y las horas de actividad física que realizan a la semana es menor a una hora y media, es clasificado como estrés académico alto. Sin embargo, si realizan más de una hora y media de actividad física a la semana, es clasificado como estrés académico leve.

En la Figura 21, se observa que se obtuvo un 83% de exactitud en la predicción de nivel de estrés académico de los estudiantes de la UNAJMA. En relación a la precisión, la clase 0 obtuvo un puntaje de 86%, la clase 1 un 78% y la clase 2 obtuvo 96%, con lo que se puede afirmar que en relación a la clase 2, el modelo realiza una buena predicción.

Figura 21

Métricas obtenidas en la fase 2 del modelo de árboles de decisión

	precision	recall	f1-score	support
0	0.86	0.57	0.69	54
1	0.78	0.94	0.85	109
2	0.96	0.86	0.91	51
accuracy			0.83	214
macro avg	0.87	0.79	0.82	214
weighted avg	0.84	0.83	0.83	214

Con respecto a la validación cruzada, se obtuvo un puntaje general de 80%, el cual es menor al que se obtuvo en la Figura 21, sin embargo, en el pliegue 3 se observa que se llegó a un 83% de exactitud, es el pliegue con el que se realizó mejor la predicción.

Tabla 8

Árboles de decisión - Resultados con validación cruzada

Métricas de evaluación	Porcentaje obtenido
Validación cruzada por pliegues	Pliegue 1 = 78%
	Pliegue 2 = 78%
	Pliegue 3 = 83%
	Pliegue 4 = 80%
	Pliegue 5 = 80%
Media de la validación cruzada	80%

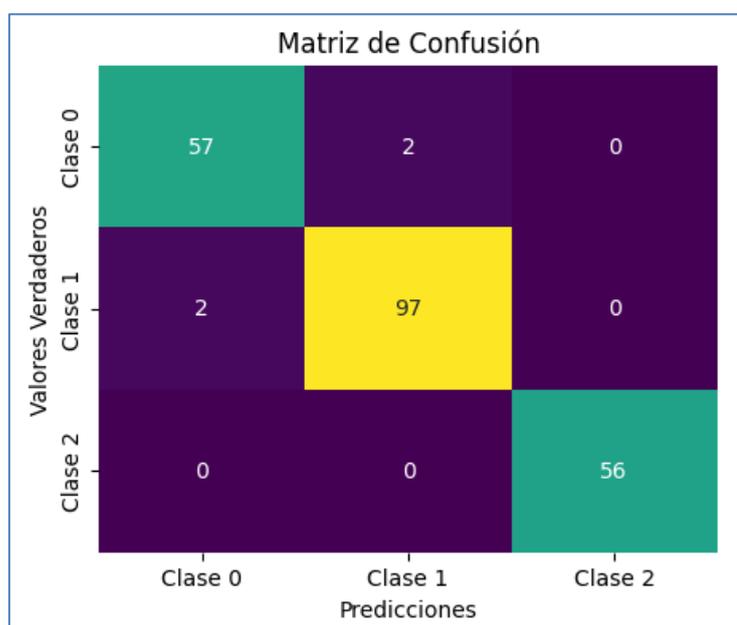
5.2. Modelo de Máquina de Vectores de Soporte

Se realizó el modelo con máquina de vectores de soporte con un kernel polinómico y un parámetro de regularización C establecido en 10.

A continuación, se muestra la matriz de confusión generada, la cual servirá para observar mejor la evaluación del rendimiento del modelo.

Figura 22

Matriz de confusión con MVS



Según la figura 22, los valores que fueron clasificados como verdaderos positivos que son un total de 210, son los que fueron clasificados correctamente como valores positivos para el modelo. Por otro lado, las instancias que fueron incorrectamente clasificadas como positivas para el modelo fueron 4, al igual que las instancias que fueron incorrectamente clasificadas como negativas.

En la Figura 23, observamos que la exactitud obtenida por el modelo es de 98% en la predicción del nivel de estrés académico. Es un valor bastante alto y óptimo para la para esta investigación.

Figura 23

Métricas obtenidas del modelo de MVS

	precision	recall	f1-score	support
0	0.97	0.97	0.97	59
1	0.98	0.98	0.98	99
2	1.00	1.00	1.00	56
accuracy			0.98	214
macro avg	0.98	0.98	0.98	214
weighted avg	0.98	0.98	0.98	214

En adición, al realizar la validación cruzada se obtuvo una media de 99% como se muestra en la tabla 9.

Tabla 9

MVS - Resultados con validación cruzada

Métricas de evaluación	Porcentaje obtenido
Validación cruzada por pliegues	Pliegue 1 = 97%
	Pliegue 2 = 100%
	Pliegue 3 = 99%
	Pliegue 4 = 99%
	Pliegue 5 = 100%
Media de la validación cruzada	99%

La diferencia entre el cálculo de las métricas de evaluación que considera solo 20% del conjunto de datos y la validación cruzada que trabaja con el total de datos; es del 1%, por lo que se puede asegurar que el modelo está aprendiendo bien.

5.3. Modelo de K-Vecinos más Cercanos

Este modelo fue configurado con el hiper parámetro 'n_neighbors=5', lo que significa que la clasificación se basará en las etiquetas de las 5 instancias más cercanas a un punto dado. La elección del parámetro 'random_state' influyó significativamente en el rendimiento del modelo, este se manejó con el valor 20.

Figura 24

Métricas obtenidas del modelo K-Vecinos más cercanos

	precision	recall	f1-score	support
0	0.68	1.00	0.81	50
1	1.00	0.66	0.80	104
2	0.85	1.00	0.92	60
accuracy			0.84	214
macro avg	0.84	0.89	0.84	214
weighted avg	0.88	0.84	0.83	214

Se obtuvo un 84% de exactitud en la predicción de nivel de estrés académico de los estudiantes de la UNAJMA. Es un valor menor en comparación a los obtenidos con los algoritmos previos, por lo que no se recomendaría su uso para esta investigación.

Tabla 10

KNN - Resultados con validación cruzada

Métricas de evaluación	Porcentaje obtenido
	Pliegue 1 = 80%
	Pliegue 2 = 80%
Validación cruzada por pliegues	Pliegue 3 = 75%
	Pliegue 4 = 79%
	Pliegue 5 = 80%
Media de la validación cruzada	79%

La media que se obtuvo de la validación cruzada según la tabla 10 es de 79%. Un porcentaje menor al que se obtuvo con la validación del 20% de la data.

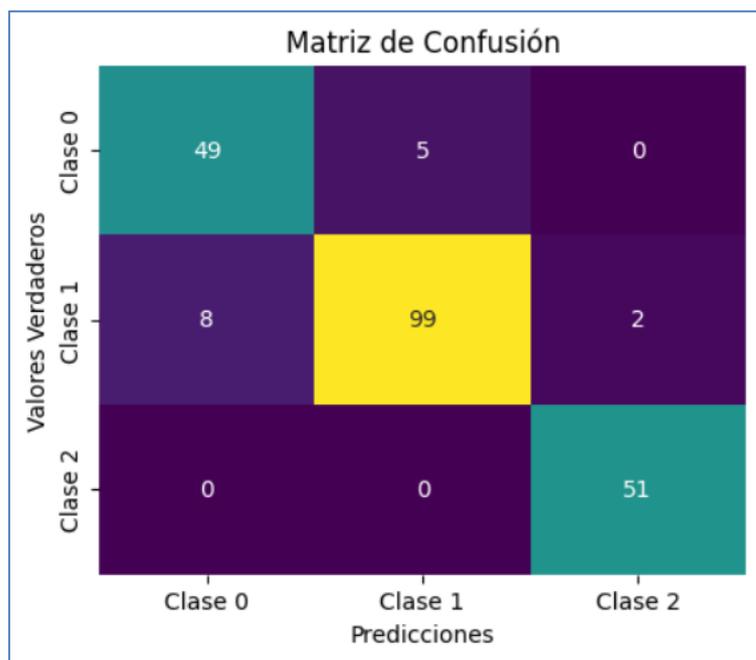
5.4. Modelo de Redes Neuronales

El modelo se configuró con tres capas ocultas, específicamente 20 neuronas en la primera, 5 en la segunda y 20 en la tercera. Se entrenó durante un máximo de 100 iteraciones con el parámetro 'random_state' y se usó una semilla aleatoria de 42 para garantizar la

reproductividad del entrenamiento. Los resultados que se obtuvieron se muestran en la matriz de confusión presentada en la Figura 25.

Figura 25

Matriz de confusión con redes neuronales



Los valores que fueron clasificados como verdaderos positivos que son un total de 199, son los que fueron clasificados correctamente como valores positivos para el modelo. Por otro lado, las instancias que fueron incorrectamente clasificadas como positivas para el modelo fueron 15, al igual que las instancias que fueron incorrectamente clasificadas como negativas.

En la Figura 26, se muestra que con el modelo empleado obtuvo un 93% de exactitud en la predicción de nivel de estrés académico de los estudiantes de la UNAJMA.

Figura 26

Métricas obtenidas del modelo de redes neuronales

	precision	recall	f1-score	support
0	0.86	0.91	0.88	54
1	0.95	0.91	0.93	109
2	0.96	1.00	0.98	51
accuracy			0.93	214
macro avg	0.92	0.94	0.93	214
weighted avg	0.93	0.93	0.93	214

Con respecto a la validación cruzada, se obtuvo un puntaje general de 94%. En los pliegues 4 y 5 se observa un mayor porcentaje del 96% a diferencia de los demás pliegues.

Tabla 11

Redes neuronales - Resultados con validación cruzada

Métricas de evaluación	Porcentaje obtenido
Validación cruzada por pliegues	Pliegue 1 = 88%
	Pliegue 2 = 94%
	Pliegue 3 = 94%
	Pliegue 4 = 96%
	Pliegue 5 = 96%
Media de la validación cruzada	94%

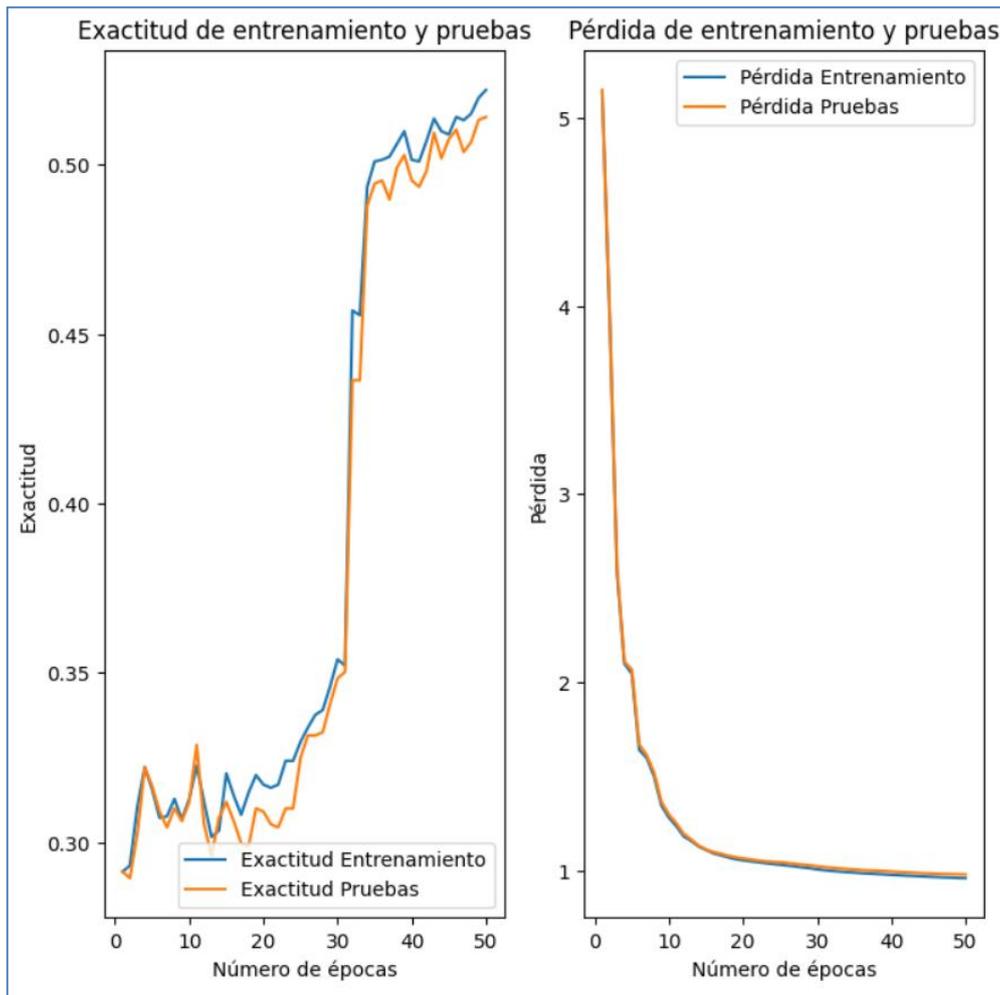
Observando la Figura 27, en lo que respecta a la curva de exactitud de entrenamiento y pruebas ambas curvas se mantienen casi unidas y van en aumento, lo que significa que el modelo ha ido mejorando en su capacidad para realizar predicciones correctas.

En lo que respecta a la gráfica de pérdida, de igual manera la curva de pérdida de entrenamiento y pruebas va disminuyendo, por lo que se puede afirmar que el modelo está mejorando en la minimización de la función de pérdida. Cuanto más bajo sea el valor de la pérdida, el rendimiento del modelo va mejorando.

En resumen, el modelo está aprendiendo bien y realiza predicciones más exactas generalizando bien a los nuevos datos.

Figura 27

Gráfica de exactitud y pérdida de entrenamiento y pruebas - RN



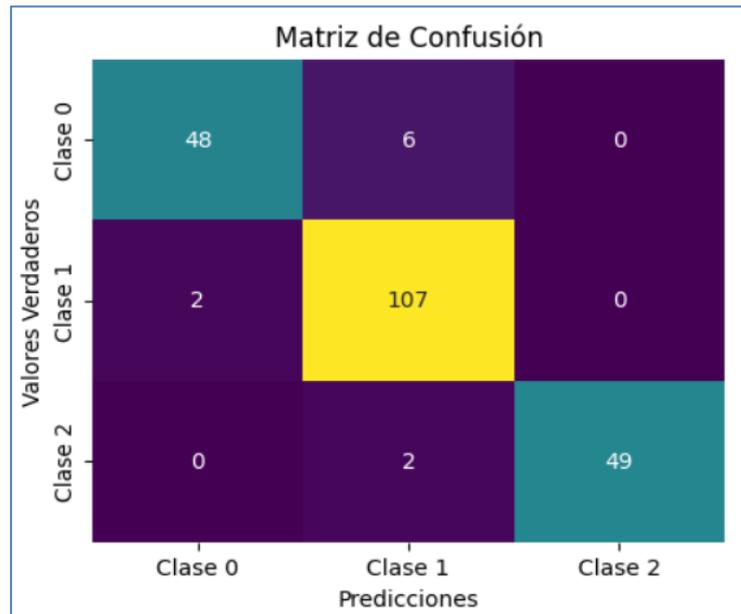
5.5. Modelo de Bosques Aleatorios

En la elaboración del modelo de bosques aleatorios, se trabajó con los siguientes hiperparámetros: N Estimators = 60, Criterion = 'Entropy', Max features = 'sqrt', Bootstrap = True, Max Samples = 1/3, Out-of-bag Score = True y Random State = 49.

En la Figura 28, se presenta la matriz de confusión con los resultados obtenidos después del entrenamiento y prueba.

Figura 28

Matriz de confusión con bosques aleatorios



Con estos resultados, observamos que el modelo fue bastante acertado en sus predicciones.

En la Figura 29, se muestra que con el modelo obtuvo un 95% de exactitud en la predicción de nivel de estrés académico de los estudiantes de la UNAJMA. En relación a la precisión, la clase 0 obtuvo un puntaje de 96%, la clase 1 un 93% y la clase 2 obtuvo 100%, son puntajes bastante alentadores.

Figura 29

Métricas obtenidas del modelo de bosques aleatorios

	precision	recall	f1-score	support
0	0.96	0.89	0.92	54
1	0.93	0.98	0.96	109
2	1.00	0.96	0.98	51
accuracy			0.95	214
macro avg	0.96	0.94	0.95	214
weighted avg	0.95	0.95	0.95	214

Con respecto a la validación cruzada, se obtuvo un puntaje general de 95%. En lo que respecta a los pliegues, los tres últimos obtuvieron un porcentaje mayor como se muestra en la tabla 12.

Tabla 12

Bosques aleatorios - Resultados con validación cruzada

Métricas de evaluación	Porcentaje obtenido
	Pliegue 1 = 89%
	Pliegue 2 = 94%
Validación cruzada por pliegues	Pliegue 3 = 97%
	Pliegue 4 = 97%
	Pliegue 5 = 97%
Media de la validación cruzada	95%

Al construir el modelo de bosques aleatorios, se realizó con un estimador de 60, por lo cual se generaron 60 árboles de decisión, estos se pueden visualizar en el Anexo 04.

Con la visualización de las figuras de los árboles de decisión pertenecientes al modelo de bosques aleatorios, se afirma que cada uno es diferente con diferentes reglas y estructura. Sin embargo, para visualizar de mejor manera se explican los colores empleados para cada clase.

Para la clase 0 – estrés académico leve, se encuentra de color anaranjado; para la clase 1 – estrés académico moderado, se encuentra de color verde y para la clase 2 – estrés académico alto, se encuentra de color morado.

En la tabla 13, se presenta un resumen de la evaluación de los diferentes modelos aplicados en esta investigación con los puntajes obtenidos en cuanto a la exactitud de la evaluación de predicción del nivel de estrés académico y también la validación cruzada.

Se considera como primera fase a los modelos que se realizaron en base al primer conjunto de datos obtenido y segunda fase a lo realizado con el conjunto de datos reducido.

Tabla 13*Evaluación de modelos*

Modelo	Nivel de exactitud consolidado			
	Exactitud primera fase	Validación cruzada primera fase	Exactitud segunda fase	Validación cruzada segunda fase
Árbol de decisión	73%	73%	83%	80%
MVS	82%	81%	98%	99%
KNN	76%	76%	84%	79%
Redes neuronales	82%	81%	93%	94%
Bosques aleatorios	74%	74%	95%	95%

Observando la tabla 13, se afirma que MVS es el mejor algoritmo para determinar el nivel de estrés académico en estudiantes de la UNAJMA, puesto que alcanza el 98% de exactitud con la data de prueba y con validación cruzada logra un 99% de exactitud.

Capítulo VI: Discusión

Tras examinar los resultados de este estudio, se afirma que el modelo de máquina de vectores de soporte entrenado, permite predecir el nivel de estrés académico en estudiantes de la Universidad Nacional José María Arguedas con una exactitud de 99%. En una investigación similar de la universidad de Juliaca, donde se realiza la predicción del nivel de estrés, con una muestra de solo 50 estudiantes y utilizando árboles de decisión, obtuvieron una precisión del 97%; sin embargo el método de recolección de datos que se realizó en dicha investigación fue a través de una encuesta con el método MSI, el cual no está completo por lo que se podría considerar algún sesgo, además la clasificación que se realiza es para cinco niveles de estrés académico: “Sin estrés”, “Leve”, “Medio”, “Alto” y “Grave”. En consecuencia, el resultado obtenido en la investigación de Vilca (2019) fue muy bueno, lo cual se logró superar con esta investigación.

En los resultados obtenidos de Elzeiny & Qaraq (2018) donde realizan la detección automática del estrés con diferentes enfoques como la fatiga al conducir, estrés en el trabajo y estudio; esto a través de la voz, detección de parpadeos, temperatura y comportamientos, obteniendo la data a través de sensores, tecleo en el computador, el uso de celulares, etc. Obtuvieron una exactitud de clasificación entre 80 a 95%, esto enfocado en señales fisiológicas; por otro lado, obtuvieron un resultado menor entre 70-80% con enfoques basados en voz, movimientos corporales y el uso de dispositivos móviles. Este último resultado es mucho menor al que se obtuvo en esta investigación que es un 99% de exactitud. Después, mencionan que encontraron más efectivos los algoritmos de bosques aleatorios, MVS y árboles de decisión; son los mismos algoritmos con los que en esta investigación también se obtuvieron excelentes resultados, haciendo enfoque en MVS y bosques aleatorios.

De una manera similar Umematsu et al. (2019), realizó el pronóstico del estrés en estudiantes en la vida diaria, sin embargo, utilizaron redes neuronales LSTM con datos de 142 participantes durante 30 días, trabajaron también con señales fisiológicas, uso de teléfonos celulares, ubicación y encuestas conductuales. La clasificación que realizaron fue solamente estrés alto y bajo, acompañaron el desempeño de LSTM con regresión logística y MVS con lo que obtuvieron un 82.4% de precisión, pero al utilizar solo datos objetivos (sin encuestas), LSTM mantuvo un 81.4% de precisión, mientras que regresión logística y MVS decayeron significativamente. Por otro lado, en esta investigación con el algoritmo de MVS, se obtuvieron mejores resultados, una exactitud de 99%.

Finalmente, la investigación que guarda mayor relación con este trabajo de investigación es el desarrollado por Ahuja & Banga (2019), quienes presentan un enfoque de detección de niveles de estrés mental en estudiantes universitarios utilizando algoritmos de aprendizaje automático. Trabajaron con un conjunto de datos de 206 estudiantes de una universidad de India, a los que se les aplicó la Escala de Estrés Percibido (PSS). Lo que indica una cantidad de datos incluso menor a lo recolectado en esta investigación. Utilizaron cuatro algoritmos de clasificación: Regresión lineal, Naive Bayes, bosques aleatorios y MVS; de los que el algoritmo con mayor desempeño fue MVS con una precisión de 85.71%, el segundo mejor algoritmo fue bosques aleatorios, con 83.33% de precisión. En cuanto a los resultados en esta investigación se obtuvieron porcentajes más altos y el algoritmo con mayor desempeño fue MVS, incluso se podría considerar a bosques aleatorios y redes neuronales entre los mejores algoritmos debido a que llegaron al 95% y 94% de exactitud respectivamente. La diferencia en las métricas de evaluación, también podría estar relacionado al enfoque; en este trabajo de investigación el enfoque está en la predicción del nivel de estrés académico, mientras que en la presentada por Ahuja & Banga (2019), está enfocada en la detección del nivel de estrés mental de los estudiantes.

Conclusiones

1. En esta investigación, se logró predecir el nivel de estrés académico de los estudiantes de la UNAJMA con un modelo de inteligencia artificial.
2. Se logró determinar el nivel de exactitud del modelo de inteligencia artificial en la predicción del nivel de estrés académico en estudiantes de la UNAJMA, obteniendo un 99% de exactitud, el cual es un excelente resultado.
3. Se logró determinar el algoritmo de IA adecuado para realizar la predicción del nivel de estrés académico en estudiantes de la UNAJMA en base a un conjunto de datos obtenido de la etapa de preprocesamiento de datos, el algoritmo en mención es máquina de vectores de soporte, el cual obtuvo una exactitud del 99%.
4. Se logró realizar la reducción de características o dimensionalidad con la finalidad de disminuir la información requerida para la predicción y mejorar el nivel de exactitud que se obtuvo de un modelo de IA usando MVS para la predicción del nivel de estrés académico en estudiantes de la UNAJMA, con un porcentaje de exactitud del 99%.

Recomendaciones

- ❖ Se recomienda agregar una mayor cantidad de datos, principalmente de estudiantes que tengan un nivel de estrés leve y alto, debido a que son las clases con menor cantidad de instancias. Si se obtiene una mayor cantidad de esos datos, el modelo tendría una mejor capacidad de generalización de datos.
- ❖ Mejorar la redacción de preguntas en la encuesta, realizarlo de manera más puntual y específica para una mejor comprensión de los estudiantes encuestados.
- ❖ Considerar de ser posible la intervención del Centro Comunitario de Salud Mental de la UNAJMA, para que puedan brindar mayor cantidad de datos con el diagnóstico del nivel de estrés académico determinado por especialistas.
- ❖ Se recomienda realizar las etapas de la metodología propuesta CRISP-DM en caso se quiera realizar mejoras, siempre realizando de la mejor manera el pre procesamiento de datos, categorización, normalización, exploración y análisis de datos, para obtener mejores resultados.

Bibliografía

- Ahuja, R., & Banga, A. (2019). Mental Stress Detection in University Students using Machine Learning Algorithms. *Science Direct*, 349-353.
- Alania, & al., e. (2021). Baremación del Inventario de Estrés Académico SISCO SV adaptado al contexto de COVID-19 en una población universitaria peruana. *SOCIALIUM (Revista Científica de Ciencias Sociales)*, 242-260.
- Amazon. (2023). *Amazon Web Services, Inc. o sus filiales-AWS*. Obtenido de <https://aws.amazon.com/es/what-is/python/>
- Avila, J. (2014). El estrés un problema de salud del mundo actual. *CON-CIENCIA*, 117-125.
- Barraza, A. M. (07 de Setiembre de 2007). *Research Gate*. Obtenido de https://www.researchgate.net/publication/28175062_El_Inventario_SISCO_del_Estrés_Academico
- Barraza, A. M. (2018). INVENTORY SISCO SV-21. Inventario SISTémico COgnoscitivista para el estudio del estrés académico. Segunda versión de 21 ítems. En *Books-©ECORFAN-México* (págs. 9-11). Durango: Books-©ECORFAN.
- Barrios, J. (2022). *Health Big Data*. Obtenido de <https://www.juanbarrios.com/que-es-un-modelo-de-aprendizaje-automatico-o-de-machine-learning/>
- Berrió, G. N., & Zea, R. M. (2011). Estrés Académico. *Revista de Psicología Universidad de Antioquia*.
- Bouza, C., & Santiago, A. (2012). *LA MINERÍA DE DATOS: ARBOLES DE DECISIÓN Y SU APLICACIÓN EN ESTUDIOS MÉDICOS*. La habana.
- Cañadas, R. (04 de Noviembre de 2021). *AbDatum*. Obtenido de <https://abdatum.com/tecnologia/redes-neuronales-artificiales>
- Contacto Directo Noticias. (2019). *Ola de suicidios en Andahuaylas*. Andahuaylas: Página de Facebook.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *SpringerLink*, 273-297.
- Díaz et al., M. (2007). *Una aproximación psicosocial al estrés escolar*. Cundinamarca: Educación y Educadores.
- EL PERUANO. (2022). Minsa alerta que el 52.2% de limeños sufre de estrés provocado por el covid-19. *El Peruano*.
- Elzeiny, S., & Qaraqe, M. (2018). *Machine Learning Approaches to Automatic Stress Detection: A Review*. College of Science and Engineering, Doha.
- ESIC Business & Marketing School. (2021). *Lenguajes de programación*. Obtenido de http://fcasua.contad.unam.mx/apuntes/interiores/docs/98/4/informatica_4.pdf

- FAYRIX. (19 de Enero de 2019). *FAYRIX*. Obtenido de https://fayrix.com/machine-learning-metrics_es
- Franco, V. T., Gonzales, R. C., & Souto, A. G. (2019). La percepción de estresores académicos según las orientaciones motivacionales a metas en estudiantes universitarios. *European Journal of Investigation in Health, Psychology and Education*, 189-200.
- Gogas, P., Papadimitriou, T., & Agrapetidou, A. (2018). Forecasting bank failures and stress testing: A machine learning approach. *International Journal of Forecasting*, 440-455.
- Goicochea, A. (11 de Agosto de 2009). *Tecnologías de la Información y Estrategia*. Obtenido de <https://anibalgoicochea.com/2009/08/11/crisp-dm-una-metodologia-para-proyectos-de-mineria-de-datos/>
- Gonzales, L. (30 de Noviembre de 2022). *APRENDE IA*. Obtenido de <https://aprendeia.com/algoritmo-regresion-logistica-machine-learning-teoria/>
- Google. (13 de Julio de 2023). *Google*. Obtenido de Colaboratory: <https://research.google.com/colaboratory/intl/es/faq.html#:~:text=Conceptos%20b%3%AIs%20icos&text=Colaboratory%2C%20o%20%22Colab%22%20para,an%3%AIlisis%20de%20datos%20y%20educaci%C3%B3n.>
- Hurwitz, J., & Kirsch, D. (2018). *Machine Learning for dummies*. Hoboken: John Wiley & Sons, Inc.
- IBERDROLA. (2022). *IBERDROLA*. Obtenido de <https://www.iberdrola.com/innovacion/machine-learning-aprendizaje-automatico>
- IBM. (12 de Julio de 2023). *IBM*. Obtenido de <https://www.ibm.com/topics/naive-bayes#:~:text=The%20Na%C3%AFve%20Bayes%20classifier%20is,a%20given%20class%20or%20category.>
- IBM Analytics. (2015). *2 Metodología Fundamental para la Ciencia de Datos*. New York: Somers, NY 10589.
- Jaques, N., Rudovic, O., Taylor, S., Sano, A., & Picard, R. (2017). Predicting Tomorrow's Mood, Health, and Stress Level using Personalized Multitask Learning and Domain Adaptation. *Media Lab*.
- Jorge, D. M. (2001). *Redes Neuronales: Conceptos Básicos y aplicaciones*. Rosario.
- Jupyter. (13 de Julio de 2023). *Jupyter*. Obtenido de <https://jupyter.org/>
- LaMotte, S. (2021). El 2020 fue un año récord para el estrés, especialmente para las madres. *CNN*.
- Lee, H., Grosse, R., & Ranganath, R. (2009). Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations. *Proceedings of the 26th Annual International Conference on Machine Learning*.

- Microsoft. (1 de Diciembre de 2022). *Microsoft*. Obtenido de learn.microsoft.com/es-es/windows/ai/windows-ml/what-is-a-machine-learning-model
- Ministerio de Educación (MINEDU). (2019). *Minedu y el Minsa trabajan con 21 universidades públicas en el cuidado de la salud mental*. Lima: Nota de prensa-MINEDU.
- Mitchell, T. (1997). *Machine Learning*. McGraw-Hill Science/Engineering/Math.
- Muñoz, C. R. (2018). *Metodología de la investigación*. Mexico: OXFORD.
- Nieto, N. T. (2018). *TIPOS DE INVESTIGACIÓN*. Santo Domingo: Universidad Santo Domingo de Guzmán.
- Ochoa, L. (2023). Expertos alertan sobre la salud mental de los universitarios: cerca del 40 % tienen depresión o ansiedad. *Medicina y Salud Pública (MSP)*.
- ORACLE. (14 de Enero de 2023). *OCI*. Obtenido de <https://www.oracle.com/pe/artificial-intelligence/what-is-ai/>
- Osorio, C. S., Giraldo, M. L., Jaramillo, V. h., & Piedrahita, E. C. (2021). Machine learning aplicado en la clasificación y predicción de la depresión: Una revisión sistemática. *REVISTA IBÉRICA DE SISTEMAS Y TECNOLOGÍAS DE INFORMACIÓN (RISTI)*, 363-375.
- Pérez, M. J. (s.f.). *Trata el estrés con PNL*
- Piryonesi, M., & El-Diraby, T. (2020). Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. *Journal of Transportation Engineering*
- Ramirez, A. (04 de Septiembre de 2021). *MEDIUM*. Obtenido de <https://medium.com/@aj.ramirez23/3-most-popular-data-science-metho-e61f6600b83f>
- Regueiro, A. (s.f.). *Conceptos básicos: ¿Qué es el estrés y cómo nos afecta?* Málaga.
- Russell, S. J., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach*.
- Shafiee, M. N., & Mutalib, S. (2020). Prediction of Mental Health Problems among Higher Education Student Using Machine Learning. *Modern Education and Computer Science*, 1-9.
- Trujillo, J. M., & Perez, M. M. (2021). *Evidencias psicométricas de la escala de Estrés percibido (PSS) en estudiantes universitarios de Huaraz*. Huaraz: Investiga UCV.
- Umematsu, T., Sano, A., Taylor, S., & Picard, R. (2019). *Improving Students' Daily Life Stress Forecasting using LSTM Neural Networks*. Cambridge.
- Veloso, F. (05 de Setiembre de 2019). *Feeding the Machine*. Obtenido de <https://www.feedingthemachine.ai/arboles-de-decision-en-regresion-machine-learning/>
- Vergara, C. (15 de Noviembre de 2022). *Actualidad en Psicología*. Obtenido de <https://www.actualidadenpsicologia.com/estres-y-educacion/>

Vilca, M. H. (2019). *Predicción del nivel de estrés en estudiantes universitarios utilizando técnicas de machine learning*. Puno, Juliaca.

Yepes, V. P. (15 de Setiembre de 2020). *UNIVERSITAT POLITECNICA DE VALENCIA*.
Obtenido de <https://victoryepes.blogs.upv.es/2020/09/15/el-aprendizaje-profundo-deep-learning-en-la-optimizacion-de-estructuras/>

Anexos

Anexo 01. Validación del instrumento de recolección de datos

Para realizar la recolección de datos, se aplicó un cuestionario a través de Google Forms en el que los estudiantes solo podían acceder con el correo institucional de la UNAJMA.

Validación del instrumento de recolección de datos

Yo, Carmen Yanac Guzmán, con número de licencia 19023, certifico que he llevado a cabo una revisión exhaustiva del formulario "Inventario SISCO del Estrés Académico" aplicado por el bachiller Geraldine Nila Flores Moscoso a través de un formulario de Google a los estudiantes de la UNAJMA. Dicha revisión tuvo como objetivo evaluar la validez en Perú, y la pertinencia del instrumento en el contexto de la medición del estrés académico.

En el presente documento, se adjunta una copia del inventario SISCO original y una copia del cuestionario aplicado.

Andahuaylas 01 de junio del 2023

1.- Durante el transcurso de este semestre ¿has tenido momentos de preocupación o nerviosismo (estrés)?

- Sí
 No

En caso de seleccionar la alternativa "no", el cuestionario se da por concluido, en caso de seleccionar la alternativa "sí", pasar a la pregunta número dos y continuar con el resto de las preguntas.

2.- Con la idea de obtener mayor precisión y utilizando una escala del 1 al 5 señala tu nivel de estrés, donde (1) es poco y (5) mucho.

1	2	3	4	5

3.-Dimensión estresores

Instrucciones: A continuación se presentan una serie de aspectos que, en mayor o menor medida, suelen estresar a algunos alumnos. Responde, señalando con una X, ¿con que frecuencia cada uno de esos aspectos te estresa? tomando en consideración la siguiente escala de valores:

Nunca	Casi nunca	Rara vez	Algunas veces	Casi siempre	Siempre
N	CN	RV	AV	CS	S

¿Con qué frecuencia te estresa:

Estresores	N	CN	RV	AV	CS	S
3.1.- La competencia con mis compañeros del grupo						
3.2.- La sobrecarga de tareas y trabajos escolares que tengo que realizar todos los días.						
3.3.- La personalidad y el carácter de los/as profesores/as que me imparten clases.						
3.4.- La forma de evaluación de mis profesores/as (a través de ensayos, trabajos de investigación, búsquedas en Internet, etc.)						
3.5.- El nivel de exigencia de mis profesores/as						
3.6.- El tipo de trabajo que me piden los profesores (consulta de temas, fichas de trabajo, ensayos, mapas conceptuales, etc.)						
3.7.- Que me toquen profesores/as muy teóricos/as.						
3.8.- Mi participación en clase (responder a preguntas, hacer comentarios, etc.)						
3.9.- Tener tiempo limitado para hacer el trabajo que me encargan los/as profesores/as						
3.10.- La realización de un examen						
3.11.- Exposición de un tema ante los compañeros de mi grupo.						
3.12.- La poca claridad que tengo sobre lo que quieren los/as profesores/as.						
3.13.- Que mis profesores/as estén mal preparados/as.						
3.14.- Asistir a clases aburridas o monótonas						
3.15.- No entender los temas que se abordan en la clase						

4.- Dimensión síntomas (reacciones)

Instrucciones: A continuación se presentan una serie de reacciones que, en mayor o menor medida, suelen presentarse en algunos alumnos cuando están estresados. Responde, señalando con una X, ¿con que frecuencia se te presentan cada una de estas reacciones cuando estás estresado? tomando en consideración la misma escala de valores del apartado anterior.

¿Con qué frecuencia se te presentan las siguientes reacciones cuando estás estresado:

Sintomas	N	CN	RV	AV	CS	S
4.1.- Trastornos en el sueño (insomnio o pesadillas)						
4.2.- Fatiga crónica (cansancio permanente)						
4.3.- Dolores de cabeza o migraña						
4.4.- Problemas de digestión, dolor abdominal o diarrea						
4.5.- Rascarse, morderse las uñas, frotarse, etc.						
4.6.- Somnolencia o mayor necesidad de dormir						
4.7.- Inquietud (incapacidad de relajarse y estar tranquilo)						
4.8.- Sentimientos de depresión y tristeza (decaído)						
4.9.- Ansiedad, angustia o desesperación.						
4.10.- Problemas de concentración						
4.11.- Sentimiento de agresividad o aumento de irritabilidad						
4.12.- Conflictos o tendencia a polemizar o discutir						
4.13.- Aislamiento de los demás						
4.14.- Desgano para realizar las labores escolares						
4.15.- Aumento o reducción del consumo de alimentos						

5.- Dimensión estrategias de afrontamiento

Instrucciones: A continuación se presentan una serie de acciones que, en mayor o menor medida, suelen utilizar algunos alumnos para enfrentar su estrés. Responde, encerrando en un círculo, ¿con qué frecuencia utilizas cada una de estas acciones para enfrentar tu estrés? tomando en consideración la misma escala de valores del apartado anterior.

¿Con qué frecuencia para enfrentar tu estrés te orientas a:

Estrategias	N	CN	RV	AV	CS	S
5.1.- Habilidad asertiva (defender nuestras preferencias, ideas o sentimientos sin dañar a otros)						
5.2.- Escuchar música o distraerme viendo televisión						
5.3.- Concentrarse en resolver la situación que me preocupa						
5.4.- Elogiar mi forma de actuar para enfrentar la situación que me preocupa (echarme porras)						
5.5.- La religiosidad (hacer oraciones o asistir a misa)						
5.6.- Búsqueda de información sobre la situación que me preocupa						
5.7.- Solicitar el apoyo de mi familia o de mis amigos						
5.8.- Ventilación y confidencias (verbalización o plática de la situación que preocupa).						
5.9.- Establecer soluciones concretas para resolver la situación que me preocupa						
5.10.- Analizar lo positivo y negativo de las soluciones pensadas para solucionar la situación que me preocupa						
5.11.- Mantener el control sobre mis emociones para que no me afecte lo que me estresa						
5.12.- Recordar situaciones similares ocurridas anteriormente y pensar en cómo las solucione						
5.13.- Salir a caminar o hacer algún deporte						
5.14.- Elaboración de un plan para enfrentar lo que me estresa y ejecución de sus tareas						
5.15.- Fijarse o tratar de obtener lo positivo de la situación que preocupa						



Cuestionario para evaluar el nivel de estrés académico en estudiantes universitarios - UNAJMA

Le agradecería tomar unos minutos de su tiempo para completar la encuesta presentada a continuación. Las respuestas que usted nos brinde serán tratadas de forma confidencial ya que serán utilizadas únicamente para el uso en el proyecto de investigación titulado: "Predicción del nivel de estrés académico en estudiantes de la Universidad Nacional José María Arguedas aplicando un modelo de inteligencia artificial".

La encuesta tiene una duración de 10 minutos aproximadamente.

1000220171@unajma.edu.pe [Cambiar de cuenta](#)



* Indica que la pregunta es obligatoria

Primera sección:

Correo *

Tu dirección de correo electrónico

Selecciona la escuela profesional a la que perteneces. *

Elige

1. Indica tu edad: *

Tu respuesta

2. Selecciona tu género: *

- Femenino
- Masculino
- Prefiero no decirlo

3. Selecciona tu lengua materna: *

- Quechua
- Castellano
- Otra

4. Indica el ciclo que estas cursando actualmente: *

Elige

5. Indica el número de créditos que estas llevando en el presente semestre: *

Solo responde con un número.

Tu respuesta

6. ¿Cuál es tu promedio de calificaciones actual? *

Solo responde con un número

Tu respuesta

7. Si trabajas indica, ¿Cuántas horas a la semana trabajas? *

Si no trabajas, responde 0.

Tu respuesta

8. ¿Cuál es tu ingreso económico familiar mensual? *

En caso no tengas un ingreso económico familiar, coloca 0.

Tu respuesta

9. ¿Cuál es tu ingreso económico propio mensual? *

En caso no tengas un ingreso económico propio, coloca 0.

Tu respuesta

10. ¿Cuántos amigos consideras que tienes en la universidad? *

Solo responde con un número

Tu respuesta

11. ¿Cómo te consideras según las siguientes alternativas? *

- Introverso(a) - (Es una persona reservada y tranquila que prefiere la soledad y la reflexión interna)
- Ambiverso(a) - (Es una persona que exhibe tanto características introversas como extroversas)
- Extroverso(a) - (Es una persona sociable y energética, disfruta de la compañía de otros)

12. Indica, ¿Cuántos hermanos tienes? *

Si no tienes hermanos, indica 0

Tu respuesta

13. Actualmente, ¿Te encuentras en una relación amorosa? *

- Sí
- No

14. Si en la pregunta anterior respondiste "Sí", indica ¿Cuál es el tipo de relación?

- Salientes, pero nada concreto
- Enamorados
- Esposos

15. ¿Tienes hijos? *

- Sí
- No

16. Si la respuesta en la anterior pregunta es "Sí", indica cuántos hijos tienes.
Solo responde con un número

Tu respuesta

17. ¿Tienes padres? *

- Sí, ambos
- Solo mamá
- Solo papá
- Ninguno

18. Si en la anterior pregunta marcaste alguna de las 3 primeras opciones,
responde: ¿Cuál es tu relación con tus padres o solo mamá o papá?

- Muy mala
- Mala
- Regular
- Buena
- Excelente

19. ¿Cuántos minutos tardas en trasladarte de tu domicilio a la universidad? *

Solo responde con un número

Tu respuesta

20. ¿Crees en Dios? *

- Sí
- No
- Soy agnóstico(a)

21. ¿Cómo consideras que son actualmente tus hábitos alimenticios? *

- Muy malos
- Malos
- Regular
- Buenos
- Excelentes

22. ¿Cuán satisfecho estas con tu salud mental? *

- Insatisfecho
- Algo satisfecho
- Satisfecho
- Muy satisfecho
- Extremadamente satisfecho

23. Responde, ¿Cuántas horas al día duermes? *

Solo responde con un número

Tu respuesta

24. Responde, ¿A qué hora te levantas por las mañanas usualmente? *

Ejemplo de respuesta: 6:15

Tu respuesta

25. ¿Cuál es el nivel de satisfacción que tienes con la carrera que estás cursando? *

- Insatisfecho
- Algo satisfecho
- Satisfecho
- Muy satisfecho
- Extremadamente satisfecho

26. ¿Alguna vez dejaste la carrera? *

- Sí
- No
- Sí, pero retorné

27. ¿Te sientes motivado en el ámbito académico actualmente? *

- Sí
- No
- Tal vez

28. ¿Cuál es tu color de piel? *

- Piel clara
- Piel media
- Piel oscura

29. ¿Cuántas horas a la semana inviertes en el uso de redes sociales? *

Solo responde con un número.

Tu respuesta

30. ¿Has experimentado ataques de pánico? *

- Sí
- No

31. ¿Tienes antecedentes familiares de alguna enfermedad mental? *

- Sí
- No
- No lo sé

32. Durante los últimos 12 meses, ¿Has visitado el Centro de Salud de la UNAJMA o algún otro establecimiento de salud en busca de ayuda psicológica? *

- Sí
- No

33. ¿Cuántas horas a la semana estudias fuera del horario académico de la universidad? *

Solo responde con un número.

Tu respuesta

34. ¿Cuántas horas a la semana inviertes en realizar alguna actividad física(ejercicio, caminar o practicar algún deporte)? *

Solo responde con un número.

Tu respuesta

35. ¿Con que frecuencia consumes bebidas alcohólicas? *

- Nunca
- Raramente
- A veces
- Frecuentemente
- Siempre

3.2.- La sobrecarga de tareas y trabajos escolares que tengo que realizar todos los días

3.3.- La personalidad y el carácter de los/as profesores/as que me imparten clases.

3.4.- La forma de evaluación de mis profesores/as (a través de ensayos, trabajos de investigación, búsquedas en Internet, etc.)

3.5.- El nivel de exigencia de mis profesores/as

3.6.- El tipo de trabajo que me piden los profesores (consulta de temas, fichas de trabajo, ensayos, mapas conceptuales, etc.)

- 3.7.- Que me toquen profesores/as muy teóricos/as.
- 3.8.- Mi participación en clase (responder a preguntas, hacer comentarios, etc.)
- 3.9.- Tener tiempo limitado para hacer el trabajo que me encargan los/as profesores/as
- 3.10.- La realización de un examen
- 3.11.- Exposición de un tema ante los compañeros de mi grupo
- 3.12.- La poca claridad que tengo sobre lo que quieren los/as profesores/as.
- 3.13.- Que mis profesores/as estén mal preparados/as.

5.6.- Búsqueda de información sobre la situación que me preocupa

5.7.- Solicitar el apoyo de mi familia o de mis amigos

5.8.- Ventilación y confidencias (verbalización o plática de la situación que preocupa).

5.9.- Establecer soluciones concretas para resolver la situación que me preocupa

5.10.- Analizar lo positivo y negativo de las soluciones pensadas para solucionar la situación que me preocupa

5.11.- Mantener el control sobre mis emociones para que no me afecte lo que me estresa

5.12.- Recordar situaciones similares ocurridas anteriormente y pensar en cómo las solucione

5.13.- Salir a caminar o hacer algún deporte

5.14.- Elaboración de un plan para enfrentar lo que me estresa y ejecución de sus tareas

5.15.- Fijarse o tratar de obtener lo positivo de la situación que preocupa

 *Carmen J. Yonac*
Lic. Carmen J. Yonac Guzman
PSICOLOGA
C. Ps. P. 19023

Anexo 02. Código fuente de la investigación

El conjunto de notebooks que se fueron desarrollando para la investigación, se pueden visualizar a detalle en el siguiente link:

https://drive.google.com/drive/folders/1t-BEQAs20Flumw_zRQIFxF-optWoQFvM?usp=sharing

Anexo 03. Conjuntos de datos o datasets

Para visualizar a detalle los conjuntos de datos obtenidos a partir de los diferentes procedimientos, ingresar al siguiente link:

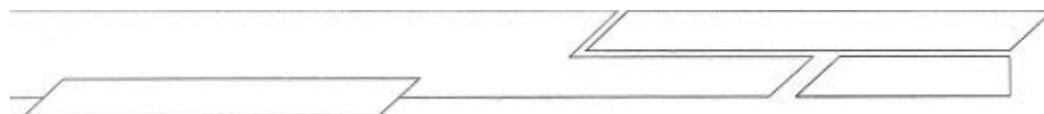
https://drive.google.com/drive/folders/1AnFG8Q-ResXDB_CvDCDqafxF8-FftIJW?usp=sharing

Anexo 04. Gráficos de bosques aleatorios

Para visualizar los gráficos generados por el modelo con bosques aleatorios tanto de la primera etapa, como de la segunda ingresar al siguiente link:

<https://drive.google.com/drive/folders/13mQXIALkAfo5UsqaqrJADpJPjMUPmmme?usp=sharing>

Anexo 05. Autorización para encuestar en la Escuela Profesional de Ingeniería de Sistemas – EPIS



Universidad Nacional José María Arguedas

Identidad y Excelencia para el Trabajo Productivo y el Desarrollo
Escuela Profesional de Ingeniería de Sistemas



Andahuaylas, 05 de junio de 2023

CARTA N° 001-2023-UNAJMA-VP/ACAD-D/EPIS/GNFM

Señor (a):
M.Sc. Neptalí Menejes Palomino
Director de la Escuela Profesional de Ingeniería de Sistemas
Presente. –

UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS	
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS	
05 JUN. 2023	
N° REG. 354	N° FOLIOS 01
HORA 12:00	FRMA.
VERIFICACION DEL DOCUMENTO NO ES ORIGINAL CONFORME	

Asunto : Solicito autorización para aplicar un cuestionario a los estudiantes de la Escuela Profesional que dirige.

De mi mayor consideración:

Mediante el presente, tengo el honor de dirigirme a usted para saludarle cordialmente y a la vez, **solicitar autorización para la aplicación de un cuestionario en los estudiantes de la Escuela Profesional de Ingeniería de Sistemas y de ser necesario ingresar a las aulas;** con el fin de aplicar un cuestionario a través de Google Forms el cual tiene una duración aproximada de 10 minutos y servirá para el desarrollo del proyecto de investigación titulado: "Predicción del nivel de estrés académico en estudiantes de la Universidad Nacional José María Arguedas aplicando un modelo de Inteligencia Artificial".

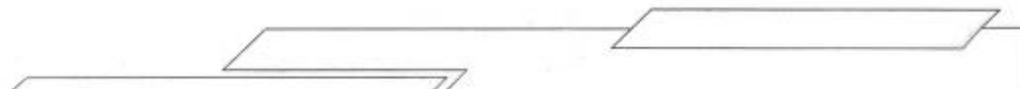
Se adjunta el link del formulario;

Link: <https://forms.gle/gwtVEJ4sElvMtwyL8>

Sin otro en particular y esperando una respuesta favorable, es propicia la oportunidad para manifestarle las muestras de mi mayor consideración y estima personal.

Atentamente,

Bach. Geraldine Nila Flores Moscoso
70420760



Anexo 06. Autorización para encuestar en la Escuela Profesional de Contabilidad – EPC



Universidad Nacional José María Arguedas
Identidad y Excelencia para el Trabajo Productivo y el Desarrollo
Escuela Profesional de Ingeniería de Sistemas



Andahuaylas, 27 de junio de 2023

CARTA N° 002-2023-UNAJMA-VP/ACAD-D/EPIS/GNFM

Señor (a):
Mgtr. Maruja Hermelinda Valencia Llamoca
Director de la Escuela Profesional de Contabilidad
Presente. –

Asunto : Solicito autorización para ingresar a las aulas de la Escuela Profesional que dirige para aplicar un cuestionario.

De mi mayor consideración:

Mediante el presente, tengo el honor de dirigirme a usted para saludarle cordialmente y a la vez, **solicitar autorización para ingresar a las aulas de la Escuela Profesional de Contabilidad**; con el fin de aplicar un cuestionario a través de Google Forms el cual tiene una duración aproximada de 10 minutos y servirá para el desarrollo del proyecto de investigación titulado: "Predicción del nivel de estrés académico en estudiantes de la Universidad Nacional José María Arguedas aplicando un modelo de Inteligencia Artificial".

Se adjunta el link del formulario:
Link: <https://forms.gle/gwtVEJ4sELvMtwyLB>

Sin otro en particular y esperando una respuesta favorable, es propicia la oportunidad para manifiestarle las muestras de mi mayor consideración y estima personal.

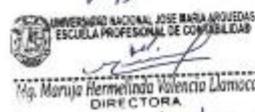


Atentamente,



Bach. Geraldine Nila Flores Moscoso
70420760

vob^o



Mg. Maruja Hermelinda Valencia Llamoca
DIRECTORA

Autorizado



**UNIVERSIDAD NACIONAL
JOSÉ MARÍA ARGUEDAS**

Licenciada por la Sunedu

FORMATO DE AUTORIZACIÓN DE PUBLICACIÓN EN EL REPOSITORIO INSTITUCIONAL DE LA UNAJMA

Formulario de autorización para la publicación electrónica en la página web del Repositorio Institucional de la Universidad Nacional de José María Arguedas, de conformidad con la Ley N° 30035 Ley que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto y Resolución de Comisión Organizadora N° 0203-2021 que aprueba el Reglamento que regula el Repositorio Institucional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto de la Universidad Nacional de José María Arguedas.

1. DATOS DEL AUTOR:

Nombres y Apellidos: Geraldine Nila Flores Moscoso

DNI / Carnet de Extranjería: 70420760

Pasaporte N°: 119155794

Correo Institucional: 1000220171@unajma.edu.pe

Teléfono/Celular: 929 794 878

Correo Electrónico: geraldinenila@gmail.com

2. GRADO, TÍTULO O ESPECIALIZACIÓN (Marque con una X)

Bachiller	Título Profesional	Magister	Doctor	Segunda Especialidad
X				

3. TIPO DE PUBLICACIÓN (Marque con una X)

Tesis	Trabajo Académico	Trabajo de investigación	Trabajo de Suficiencia profesional
X			



**UNIVERSIDAD NACIONAL
JOSÉ MARÍA ARGUEDAS**

Licenciada por la Sunedu

4. DATOS DE LA PUBLICACION:

Título: PREDICCIÓN DEL NIVEL DE ESTRÉS ACADÉMICO EN ESTUDIANTES DE LA UNIVERSIDAD NACIONAL JOSÉ MARÍA ARGUEDAS APLICANDO UN MODELO DE INTELIGENCIA ARTIFICIAL.

Fecha de Publicación:

5. DATOS DEL ASESOR:

Nombres y Apellidos: Iván Soria Solís

DNI/Pasaporte/Cédula: 40725333

ORCID: 0000-0003-4202-9402

.....
Firma

70420760

Geraldine Nila Flores Moscoso



**UNIVERSIDAD NACIONAL
JOSÉ MARÍA ARGUEDAS**

Licenciada por la Sunedu

LICENCIAS

Licencia Estándar

Bajo los siguientes términos, autorizo el depósito de mi trabajo de investigación en el Repositorio Institucional Digital de la UNAJMA.

Con la autorización de depósito de mi trabajo de investigación otorgo a la Universidad Nacional de José María Arguedas, una Licencia No Exclusiva para reproducir, distribuir, comunicar al público, transformar (únicamente mediante su traducción a otros idiomas) y poner a disposición del público mi trabajo de investigación (incluido el resumen), en formato físico o digital en cualquier medio, conocido o por conocerse, a través de los diversos servicios provistos por la Universidad, creados o por crearse, tales como el Repositorio Institucional Digital de la UNAJMA, Colección de Tesis, entre otros, en el Perú y en el extranjero, por el tiempo y veces que considere necesarias y libre de remuneraciones.

En virtud de dicha Licencia, la Universidad Nacional de José María Arguedas, podrá reproducir mi trabajo de investigación en cualquier tipo de soporte y en más de un ejemplar, sin modificar su contenido, solo con los propósitos de seguridad, respaldo y preservación.

Declaro que el trabajo de investigación es una creación de mi autoría y exclusiva titularidad, o coautoría con titularidad compartida, y me encuentro facultado a conceder la presente licencia y, asimismo, garantizo que dicho trabajo de investigación no infringe derechos de autor de terceras personas.

La Universidad Nacional de José María Arguedas consignará el nombre del/los autores/es del trabajo de investigación, y no le hará ninguna modificación más de la permitida en la presente licencia.

Autorizo el depósito (marque con una X).

SI autorizo que se deposite inmediatamente.

No autorizo.

Fecha: 17 de junio del 2024

Firma

Documento de identidad: 70420760

Nombres y apellidos: Geraldine Nila Flores Moscoso

