



UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA ESTADÍSTICA E
INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E
INFORMÁTICA



**MODELO DE APRENDIZAJE AUTOMÁTICO BASADO EN EL
CLASIFICADOR NAIVE DE BAYES PARA LA DETECCIÓN DEL
NIVEL DE DEPRESIÓN EN ESTUDIANTES DE LA UNA-PUNO
2022**

TESIS

PRESENTADA POR:

Bach. EVELYN ELIANA COAQUIRA FLORES

**PARA OPTAR AL TÍTULO PROFESIONAL DE:
INGENIERO ESTADÍSTICO E INFORMÁTICO**

PUNO – PERÚ

2023



NOMBRE DEL TRABAJO

MODELO DE APRENDIZAJE AUTOMÁTICO BASADO EN EL CLASIFICADOR NAIVE DE BAYES PARA LA DETECCIÓN DEL NIVEL DE DEPRESIÓN EN ESTUDIANTES DE LA UNA-PUNO 2022

AUTOR

EVELYN ELIANA COAQUIRA FLORES

RECUESTO DE PALABRAS

34594 Words

RECUESTO DE CARACTERES

193335 Characters

RECUESTO DE PÁGINAS

159 Pages

TAMAÑO DEL ARCHIVO

2.6MB

FECHA DE ENTREGA

Dec 11, 2023 12:33 PM GMT-5

FECHA DEL INFORME

Dec 11, 2023 12:35 PM GMT-5

● **14% de similitud general**

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

- 11% Base de datos de Internet
- Base de datos de Crossref
- 9% Base de datos de trabajos entregados
- 3% 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 10 palabras)



Firmado digitalmente por TITO LUNA
Josefredo PUNO 202305090170 aut
Motivo: Soy el autor del documento
Fecha: 11.12.2023 12:33:51 -05:00



Firmado digitalmente por COYLA
CAME Leonora PUNO 202305090170 aut
Motivo: Soy el autor del documento
Fecha: 11.12.2023 12:35:52 -05:00



DEDICATORIA

*A aquellos que son la luz en mi andar,
a los que en la tormenta fueron mi puerto.
Esta tesis es más que palabras en el mar,
es la danza del conocimiento, mi más sincero concierto.*

*A mi familia, raíces fuertes que sostienen,
en su amor encontré fuerza y refugio.
A cada risa compartida y lágrima que se llevan,
dedico este logro, mi tesoro máspreciado.*

*A mis seres queridos de risas y desvelos,
juntos exploramos los confines del saber.
A sus manos extendidas en momentos bellos,
mi gratitud eterna, como un río sin entender.*

Evelyn Eliana Coaquira Flores



AGRADECIMIENTOS

*En la senda de la pesquisa, con anhelo y pasión,
agradezco a mi Señor por guiarme con devoción.
Dios, faro eterno, luz que me ilumina,
gracias por permitir que mi investigación culmine.*

*A mi madre, pilar inquebrantable de amor,
tu aliento me sostuvo, fuente de fervor.
En cada paso, en los días de incertidumbre,
Tu apoyo constante, mi fortaleza cumbre.*

*Asesor y coasesor, maestros en el saber,
A ustedes debo el conocimiento, el crecer.
Enseñanzas como ríos que fluyen con claridad,
Gracias por guiar mi barca, con humildad.*

*Mi gratitud se extiende al jurado que evaluó,
Con sabiduría, con mirada que todo lo ve.
Sus luces, guías sabias en este viaje,
Agradezco la atención, el consejo, el encaje.*

*A mis seres queridos, faros en la tormenta y más,
En las buenas y malas, en risas y paz.
Gracias por ser anclas en mi mar agitado,
Su amor, mi puerto seguro, mi refugio anhelado*

.

Evelyn Eliana Coaquira Flores



ÍNDICE DE GENERAL

	Pág.
DEDICATORIA	
AGRADECIMIENTOS	
ÍNDICE DE GENERAL	
ÍNDICE DE TABLAS	
ÍNDICE DE FIGURAS	
ÍNDICE DE ANEXOS	
ACRÓNIMOS	
RESUMEN	16
ABSTRACT.....	17
CAPÍTULO I	
INTRODUCCIÓN	
1.1 DESCRIPCIÓN DEL PROBLEMA	19
1.2 FORMULACIÓN DEL PROBLEMA	20
1.2.1 Problema General.....	20
1.2.2 Problemas Específicos.....	21
1.3 OBJETIVOS DE LA INVESTIGACIÓN.....	21
1.3.1 Objetivo general	21
1.3.2 Objetivos específicos.....	21
1.4 HIPÓTESIS DE LA INVESTIGACIÓN	22
1.4.1 Hipótesis general	22
1.4.2 Hipótesis específicas	22
1.5 JUSTIFICACIÓN DE LA INVESTIGACIÓN	22
1.6 LIMITACIONES DE LA INVESTIGACIÓN	24



CAPÍTULO II

REVISIÓN DE LA LITERATURA

2.1	MARCO REFERENCIAL	25
2.1.1	Ámbito Internacional.....	25
2.1.2	Ámbito Nacional	37
2.1.3	Ámbito regional.....	42
2.2	MARCO TEÓRICO	47
2.2.1	Machine Learning (Aprendizaje Automático)	47
2.2.1.1	Tipos de Análisis de Aprendizaje Automático	47
2.2.1.2	Proceso de aprendizaje automático	48
2.2.2	Minería de datos	50
2.2.2.1	Características de la minería de datos	51
2.2.3	Algoritmos y técnicas de explotación de datos	52
2.2.4	Clasificador (Naive de Bayes).....	53
2.2.4.1	Para atributos discretos	56
2.2.5	Indicadores para evaluar el rendimiento del modelo	58
2.3	MARCO CONCEPTUAL	62
2.3.1	Depresión	62
2.3.1.1	Fisiología de la Depresión.....	64
2.3.1.2	Evaluación de la depresión	64
2.3.2	Factores de riesgo.....	64
2.3.3	Manifestaciones clínicas	65
2.3.4	Diagnostico.....	66
2.3.5	Escala de valoración de Hamilton.....	67
2.3.6	Síntomas de la depresión.....	70



2.3.6.1	Humor Depresivo:.....	70
2.3.6.2	Sentido de Culpabilidad.....	70
2.3.6.3	Insomnio Nocturno e Insomnio Matinal.....	70
2.3.6.4	Comunicación Fluida.....	71
2.3.6.5	Manía Corporal.....	71
2.3.6.6	Preocupación.....	71
2.3.6.7	Síntomas de Alguna Enfermedad.....	72
2.3.6.8	Pérdida de Apetito, Peso y Fatiga:.....	72
2.3.7	Impacto de la depresión en la salud publica.....	72
2.4	OPERACIONALIZACIÓN DE VARIABLES.....	73

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1	LUGAR DE ESTUDIO.....	74
3.2	METODO DE INVESTIGACION.....	75
3.3	DISEÑO DE LA INVESTIGACIÓN.....	75
3.4	TIPO DE INVESTIGACIÓN.....	75
3.5	NIVEL DE LA INVESTIGACIÓN.....	75
3.6	ENFOQUE DE LA INVESTIGACIÓN.....	76
3.7	POBLACIÓN Y MUESTRA.....	76
3.7.1	Población.....	76
3.7.2	Muestra.....	76
3.8	TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS.....	77
3.8.1	Técnica.....	77
3.8.2	Instrumento.....	77
3.9	PROCEDIMIENTO DE LA INVESTIGACIÓN.....	78



3.10 PRUEBAS PARA LA VERIFICACIÓN DE LAS HIPÓTESIS 80

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

**4.1 ANÁLISIS DE LAS CARACTERÍSTICAS ASOCIADAS A LA
DEPRESIÓN DE LOS ESTUDIANTES..... 82**

4.1.1 Características demográficas..... 82

4.1.1.1 Género..... 82

4.1.1.2 Lugar de residencia..... 84

4.1.1.3 Tipo de vivienda..... 85

4.1.2 Características académicas..... 88

4.1.2.1 Área de estudio..... 88

4.1.2.2 Horas de estudio..... 90

4.1.3 Características de conectividad..... 92

4.1.3.1 Servicio de internet..... 92

4.1.3.2 Tipo de conexión..... 93

4.1.4 Características de estilos de vida y salud..... 95

4.1.4.1 Practica de deporte..... 95

4.1.4.2 Consumo de sustancias..... 97

4.1.4.3 Afectado por el Covid-19..... 99

4.1.5 Correspondencia múltiple de las características asociadas a la depresión.
..... 100

**4.2 NIVEL DE DEPRESIÓN DE LOS ESTUDIANTES CON LA ESCALA DE
VALORACIÓN DE HAMILTON (HDRS)..... 104**

4.3 MODELO DE APRENDIZAJE AUTOMÁTICO BAYESIANO 107

4.3.1 Distribución de probabilidad condicional según el nivel de depresión. 108



4.3.2	Probabilidades a priori según los niveles de depresión.....	110
4.3.3	Probabilidad total para el modelo Naive Bayes	111
4.3.4	Probabilidades a posteriori según el modelo Naive Bayes	112
4.4	MÉTRICAS DE EVALUACIÓN DEL MODELO.....	112
4.5	CONTRASTACIÓN DE HIPÓTESIS	116
4.5.1	Contrastación de la Hipótesis General	116
4.5.2	Contrastación de la Hipótesis Específica 1	117
4.5.3	Contrastación de la Hipótesis Específica 2	118
4.5.4	Contrastación de la Hipótesis Específica 3	119
V.	DISCUSIÓN	121
VI.	CONCLUSIONES	123
VII.	RECOMENDACIONES	125
VIII.	REFERENCIAS BIBLIOGRÁFICAS	127



ÍNDICE DE TABLAS

	Pág.
Tabla 1: Interpretación del coeficiente Kappa	59
Tabla 2: Escala de valoración de Hamilton.....	67
Tabla 3: Puntuación de la Escala de Hamilton.....	69
Tabla 4: Cuadro de operacionalización de variables.....	73
Tabla 5: Frecuencia del nivel de depresión con el género en los estudiantes de la UNA-Puno 2022	82
Tabla 6: Frecuencia del nivel de depresión y lugar de residencia en los estudiantes de la UNA-Puno 2022	84
Tabla 7: Frecuencia del nivel de depresión y el tipo de vivienda en estudiantes de la UNA-Puno 2022	86
Tabla 8: Frecuencia del nivel de depresión y el área de estudio en los estudiantes de la UNA-Puno 2022	88
Tabla 9: Frecuencia del nivel de depresión y las horas de estudio en estudiantes de la UNA-Puno 2022	90
Tabla 10: Frecuencia del nivel de depresión y el servicio de internet en estudiantes de la UNA-Puno 2022	92
Tabla 11: Frecuencia del nivel de depresión y el tipo de conexión en estudiantes de la UNA-Puno 2022	93
Tabla 12: Frecuencia del nivel de depresión con la práctica de deporte en estudiantes de la UNA-Puno 2022	95
Tabla 13: Frecuencia del nivel de depresión y el consumo de sustancias nocivas en los estudiantes de la UNA-Puno 2022.....	97



Tabla 14: Frecuencia del nivel de depresión y ser afectado por el covid-19 en estudiantes de la UNA-Puno 2022.....	99
Tabla 15: Frecuencia de los niveles de depresión en los estudiantes de la UNA-Puno 2022.....	105
Tabla 16: Resumen de las características asociadas al nivel de depresión	107
Tabla 17: Distribución de probabilidad condicional según el nivel de depresión	108
Tabla 18: Matriz de confusión del nivel de depresión en estudiantes de la UNA-Puno 2022.....	112
Tabla 19: Métricas de evaluación del modelo basado en Naive Bayes	114
Tabla 20: Matriz de correspondencias múltiples de las características asociadas a la depresión.....	136



ÍNDICE DE FIGURAS

	Pág.
Figura 1: Lugar de estudio de la investigación	74
Figura 2: Procedimiento de la investigación.....	79
Figura 3: Asociación entre la depresión y el género en los estudiantes de la UNA- Puno 2022	83
Figura 4: Asociación entre depresión y lugar de residencia en estudiantes de la UNA- Puno 2022	85
Figura 5: Asociación entre depresión y tipo de vivienda en estudiantes de la UNA- Puno 2022	87
Figura 6: Asociación entre depresión y área de estudio en estudiantes de la UNA-Puno 2022	89
Figura 7: Asociación entre la depresión y horas de estudio en los estudiantes de la UNA-Puno 2022	91
Figura 8: Asociación entre depresión e internet en estudiantes de la UNA-Puno 2022	93
Figura 9: Asociación entre depresión y tipo de conexión en estudiantes de la UNA- Puno 2022	95
Figura 10: Asociación entre la depresión y la práctica de deporte en estudiantes de la UNA-Puno 2022	96
Figura 11: Asociación de la depresión con el consumo de sustancias nocivas en los estudiantes de la UNA-Puno 2022.....	98
Figura 12: Asociación entre la depresión y los estudiantes de la UNA-Puno afectados por Covid-19.....	100



Figura 13: Correspondencia múltiple de las características asociadas a la depresión	103
Figura 14: Nivel de depresión en estudiantes de la UNA-Puno 2022.....	105
Figura 15: Métricas de evaluación del modelo basado en Naive Bayes	114



ÍNDICE DE ANEXOS

	Pág.
ANEXO 1: Matriz de correspondencias múltiples de las características asociadas a la depresión.....	136
ANEXO 2: Código en RStudio para el preprocesamiento de datos.....	139
ANEXO 3: Código en RStudio para el modelo de aprendizaje automático Naive de Bayes	142
ANEXO 4: Instrumento de recolección de datos	143
ANEXO 5: Matriz de consistencia.....	145



ACRÓNIMOS

OMS:	Organización Mundial de la Salud
HDRS:	Escala de valoración de Hamilton
MINSA:	Ministerio de Salud
UNA:	Universidad Nacional del Altiplano
DSM-5:	Manual Diagnóstico y Estadístico de los trastornos mentales.
CSV:	Valores separados por comas



RESUMEN

La depresión es un trastorno mental común que afecta a muchas personas en todo el mundo, su detección temprana es crucial para brindar el apoyo y tratamiento adecuado por lo que la investigación tiene como objetivo determinar un modelo de Aprendizaje Automático basado en el Clasificador Naive de Bayes para la detección del nivel de depresión en estudiantes de la UNA-Puno. Se recopiló una muestra de 519 estudiantes universitarios; se realizó la limpieza de datos, un análisis de datos, también se efectuó el preprocesamiento de la variable endógena utilizando el cuestionario HDRS y posteriormente se desarrolló el clasificador Naive Bayes utilizando aprendizaje automático con el 70% de los datos como entrenamiento y 30% para validación, tomando en cuenta características del estudiante asociados a la depresión, finalmente se evaluaron las métricas de evaluación. El modelo alcanzó una precisión global del 83.33% con un intervalo de confianza del 96% que oscila entre (0.7833 y 0.8834) con un Kappa del 74.52% lo que muestra concordancia mayor entre las predicciones del modelo y los valores reales, la precisión equilibrada que oscila entre el 68.93%, 90.69% indica que el modelo es altamente eficaz para la detección del nivel de depresión en estudiantes. Estos resultados subrayan la eficacia del modelo en la detección precoz y el tratamiento de la depresión beneficiando especialmente a sectores vulnerables, contribuyendo a la mejora de la salud mental en la población estudiantil.

Palabras Clave: Depresión, Aprendizaje Automático, Escala HDRS, Naive de Bayes.



ABSTRACT

Depression is a common mental disorder that affects many people around the world, its early detection is crucial to provide adequate support and treatment, so the research aims to determine a Machine Learning model based on the Bayes Naive Classifier for the detection of the level of depression in students of the UNA-Puno. A sample of 519 university students was collected, the data cleaning was carried out, data analysis, preprocessing of the endogenous variable using the HDRS questionnaire, then the Naive Bayes classifier was developed using machine learning with 70% training and 30% for validation, taking into account student characteristics associated with depression, Finally, the evaluation metrics were evaluated. The model achieved an overall accuracy of 83.33% with a 96% confidence interval ranging from (0.7833 to 0.8834) with a Kappa of 74.52% which shows greater agreement between the predictions of the model and the real values, the balanced accuracy that ranges between 68.93%, 90.69% indicates that the model is highly effective for the detection of the level of depression in students. These results underline the effectiveness of the model in the early detection and treatment of depression, especially benefiting vulnerable sectors, contributing to the improvement of mental health in the student population.

Keywords: Depression, Machine Learning, HDRS Scale, Naïve Bayes



CAPÍTULO I

INTRODUCCIÓN

La depresión es un trastorno mental prevalente que afecta a millones de personas en todo el mundo dado su importante impacto en el estilo de vida de una persona y en su capacidad para funcionar con normalidad, la identificación temprana del riesgo de depresión es crucial para intervenir a tiempo y proporcionar el apoyo adecuado (Moon et al., 2021) Recientemente, el aprendizaje automático ha surgido como un enfoque prometedor para predecir el riesgo de depresión, lo que permite la detección temprana y la adopción de medidas preventivas. El aprendizaje automático, también conocido como inteligencia artificial, permite a los ordenadores aprender de los datos y hacer predicciones sin intervención humana. Entre los algoritmos más utilizados destaca el clasificador Naive Bayes, que aprovecha el teorema de Bayes y asume la independencia condicional de las características, lo que lo convierte en un modelo sencillo pero eficaz para la clasificación de datos. La presente investigación tiene como objetivo explorar las capacidades del clasificador Naive Bayes como modelo de aprendizaje automático para prever el riesgo de depresión, el objetivo es desarrollar un modelo capaz de reconocer patrones subyacentes y calcular la probabilidad de que un individuo esté en riesgo de desarrollar depresión en el futuro. Utilizaremos conjuntos de datos clínicos y psicológicos de individuos con y sin síntomas depresivos, abarcando información demográfica, historial médico, evaluaciones de síntomas, factores de estrés y otras variables pertinentes. Al entrenar al clasificador Naive Bayes con estos datos, anticipamos que aprenderá a discernir patrones característicos asociados con el riesgo de depresión y proporcionará predicciones precisas. La capacidad de predecir el riesgo de depresión mediante aprendizaje automático, específicamente con el clasificador Naive Bayes, posee prometedoras implicaciones clínicas y de salud pública al permitir la detección temprana,



intervenciones oportunas y brindar apoyo psicológico (Sau & Bhakta, 2019). Un modelo predictivo preciso puede identificar a personas que se beneficiarían de intervenciones preventivas y programas de apoyo temprano, contribuyendo así a reducir la carga de la enfermedad y mejorar la calidad de vida de los afectados.

1.1 DESCRIPCIÓN DEL PROBLEMA

A nivel mundial, la problemática de la depresión en la población estudiantil se ha convertido en un fenómeno de preocupación creciente. Investigaciones y estudios epidemiológicos indican un aumento significativo en los trastornos mentales entre los estudiantes universitarios, atribuido a factores académicos, sociales y económicos. Este fenómeno impacta no solo el rendimiento académico sino también la calidad de vida y el bienestar emocional de los jóvenes en un contexto global, según la Organización Mundial de la Salud (OMS) que señala la creciente carga de los trastornos mentales, incluida la depresión, en la población global, más de 264 millones de personas en todo el mundo sufren de depresión, convirtiéndola en una de las principales causas de discapacidad ya que afecta a todas las edades, incluida la población estudiantil, donde se observa un aumento significativo de los trastornos mentales.

Al desplazarnos hacia el ámbito nacional en el contexto de Perú en el Ministerio de Salud MINSA del país informa que la depresión es una de las principales causas de consulta en los servicios de salud mental, se observa una tendencia similar, la presión académica, las expectativas sociales y la transición a la vida universitaria pueden contribuir al aumento de los niveles de estrés y depresión en los estudiantes. La conciencia de esta problemática a nivel nacional se refleja en la necesidad de abordar estos problemas de salud mental en el ámbito educativo para garantizar el desarrollo integral de los estudiantes y el éxito académico.



En el contexto regional, específicamente en la región de Puno, se pueden identificar factores adicionales que podrían influir en la salud mental de los estudiantes universitarios. Las características culturales, las condiciones socioeconómicas y las particularidades de la vida en la región pueden tener un impacto significativo en la manifestación y percepción de la depresión entre los estudiantes.

Finalmente, al centrarnos en la Universidad Nacional del Altiplano (UNA-Puno), la necesidad de comprender y abordar la depresión se vuelve aún más urgente y específica, ya que los desafíos académicos, las dinámicas sociales dentro de la universidad y las características propias de la población estudiantil de la UNA-Puno pueden influir de manera única en la prevalencia y gravedad de la depresión. Por lo tanto, la implementación de un modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión en estudiantes se presenta como una solución adaptada a las necesidades específicas de esta comunidad académica. Este enfoque permitirá no solo entender mejor el fenómeno en la universidad sino también implementar intervenciones y recursos de apoyo de manera más efectiva y personalizada en otras universidades.

1.2 FORMULACIÓN DEL PROBLEMA

1.2.1 Problema General

¿Cómo el modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes permitirá detectar el nivel de depresión en estudiantes de la UNA-Puno 2022?



1.2.2 Problemas Específicos

- ¿Cuáles son las características asociadas a la depresión en estudiantes universitarios de la UNA-Puno 2022?
- ¿Cuál es el nivel más alto de depresión en los estudiantes de la UNA-Puno mediante la Escala de valoración de Hamilton HDRS?
- ¿Cuál es la eficacia del modelo de Aprendizaje Automático con el clasificador Naive de Bayes en la detección del nivel de depresión en estudiantes universitarios de la UNA-Puno?

1.3 OBJETIVOS DE LA INVESTIGACIÓN

1.3.1 Objetivo general

Determinar un modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión en estudiantes de la UNA Puno 2022.

1.3.2 Objetivos específicos

- Analizar las características asociadas a la depresión en estudiantes universitarios de la UNA-Puno 2022
- Determinar el nivel más alto de depresión en los estudiantes de la UNA-Puno mediante la Escala de valoración de Hamilton HDRS
- Validar la eficacia del modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión en estudiantes de la UNA-Puno 2022.



1.4 HIPÓTESIS DE LA INVESTIGACIÓN

1.4.1 Hipótesis general

El modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes permitirá detectar de manera óptima el nivel de depresión en estudiantes de la UNA-Puno 2022.

1.4.2 Hipótesis específicas

- Las características demográficas, académicas, conectividad, estilos de vida y salud se asocian significativamente a la depresión en estudiantes universitarios de la UNA-Puno.
- El nivel más alto de depresión en los estudiantes de la UNA-Puno mediante la Escala de valoración de Hamilton HDRS es una depresión severa
- El modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes es eficaz para la detección del nivel de depresión en estudiantes de la UNA-Puno 2022

1.5 JUSTIFICACIÓN DE LA INVESTIGACIÓN

La realización de este estudio en la población universitaria surge ante la necesidad de comprender y abordar los trastornos mentales, estados de tristeza y alteraciones anímicas que afectan a los estudiantes de la Universidad Nacional del Altiplano, si bien la depresión es una realidad presente en este contexto, sus causas son múltiples y complejas, y no se limitan a un único factor desencadenante. Los jóvenes universitarios enfrentan desafíos particulares que pueden contribuir a la aparición de la depresión, entre estos desafíos se incluyen la presión académica, problemas socioeconómicos, dificultades familiares, y la transición hacia la vida universitaria, entre otros, estos factores,



combinados de manera única en cada individuo, pueden influir de manera significativa en la salud mental de los estudiantes. El objetivo de este estudio es identificar y comprender las diversas variables asociadas a la depresión en estudiantes universitarios de la UNA-Puno, se busca analizar características específicas, evaluar el nivel de depresión utilizando la Escala de valoración de Hamilton (HDRS) y validar un modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión en esta población. Al comprender estas variables, se pretende no solo ofrecer un diagnóstico preciso del estado de depresión, sino también sentar las bases para el desarrollo de estrategias preventivas y de intervención personalizadas, el enfoque de este estudio es contribuir al bienestar psicológico y emocional de los estudiantes universitarios, promoviendo un entorno que fomente la salud mental y la resiliencia, independientemente de las circunstancias externas. Adicionalmente, es crucial destacar que los estudiantes universitarios no solo representan el presente de nuestra sociedad, sino también su futuro, su formación académica y desarrollo personal tienen un impacto directo en el rumbo que tomará la región y el país en los años venideros. Por lo tanto, comprender y abordar las complejidades asociadas a la depresión en esta etapa de sus vidas es esencial para garantizar un futuro más saludable y equilibrado, la salud mental de los jóvenes universitarios no solo afecta su rendimiento académico, sino que también influye en su capacidad para contribuir de manera positiva a la sociedad. Promover un entorno universitario que fomente la salud mental y proporcione recursos adecuados para abordar la depresión es una inversión en el potencial y el bienestar no solo de los estudiantes, sino de la comunidad en su conjunto. Este estudio, al centrarse en la identificación y comprensión de las características demográficas, académicas, de conectividad, y de estilos de vida y salud, busca no solo aliviar el sufrimiento individual, sino también sentar las bases para una comunidad universitaria más fuerte, resiliente y



preparada para enfrentar los desafíos futuros. Las características de los estudiantes universitarios se rigen como pilares fundamentales para la construcción de un futuro promisorio y sostenible para la región y el país.

1.6 LIMITACIONES DE LA INVESTIGACIÓN

La restricción temporal impidió la expansión del estudio a más universidades del sur del Perú, limitando la representatividad geográfica del resultado, además la exploración de técnicas de aprendizaje automático se vio acotada, ya que se enfocó principalmente en el clasificador Naive de Bayes ya que la inclusión de más técnicas podría haber enriquecido la comprensión del fenómeno. Aunque estas limitaciones son leves, es importante tenerlas en cuenta al interpretar los resultados y pueden abrir la puerta a investigaciones futuras que aborden estas áreas específicas.



CAPÍTULO II

REVISIÓN DE LA LITERATURA

2.1 MARCO REFERENCIAL

Estos estudios de investigación proporcionan información valiosa sobre los factores que influyen en la depresión y muestran el potencial del aprendizaje automático en la predicción y gestión de los problemas relacionados con la depresión, los diversos enfoques empleados en estas investigaciones contribuyen a una comprensión global de los factores de riesgo de la depresión y abren vías para la detección precoz y la intervención para mejorar los resultados de salud mental

2.1.1 *Ámbito Internacional*

Moon et al. (2021) en su investigación titulada *“Enfoque de aprendizaje automático para predecir la depresión en los sectores laborales de Bangladesh”* con el objetivo principal de este estudio fue prever la presencia de la depresión en el sector laboral de Bangladesh, con un enfoque específico en identificar el género que experimenta una mayor tristeza y explorar las disparidades en la satisfacción laboral. La metodología empleada en este estudio fue exhaustiva y sistemática. Se aplicaron diversas técnicas, como análisis factorial, clasificador de bosques aleatorios, regresión de bosques aleatorios, clasificador bayesiano ingenuo y clasificador de vecinos K. Los resultados revelaron una prevalencia significativa de síntomas de depresión en el sector, notables disparidades en la satisfacción laboral entre géneros y correlaciones significativas entre factores como la edad, los ingresos mensuales, la ocupación, el estado civil y la satisfacción laboral y los niveles de depresión. El estudio demostró cómo los algoritmos de aprendizaje



automático pueden ser útiles para predecir problemas psicológicos en la vida moderna, se estudió la predicción de la depresión en los sectores laborales de Bangladesh utilizando un enfoque de aprendizaje automático, y se recogieron datos de hombres y mujeres

Vergaray et al. (2023), presentó una investigación denominada *“Predicción de la depresión en estudiantes universitarios mediante técnicas de ensemble de apilamiento sobre el método de sobremuestreo”* cuyo objetivo es desarrollar un método y tres modelos combinados basados en Stacking con el fin de anticipar la manifestación de la depresión en los estudiantes de Ingeniería Informática y Sistemas en una universidad pública en Perú. En la fase metodológica, se formó un conjunto de datos con 284 estudiantes universitarios, seguido de una limpieza y preprocesamiento detallado con Python. Se aplicó un método de sobremuestreo para equilibrar los datos y se utilizó la validación cruzada para la partición de los mismos. Tres modelos combinados fueron propuestos y evaluados. Entre los modelos propuestos, Ensemble Stacking 1 y Stacking 2 destacaron con las puntuaciones más altas de precisión y curva ROC, alcanzando valores de 94,69% y 100,00% en las categorías micro y macro. Stacking 1 demostró el mejor desempeño en sensibilidad, precisión y puntuación F1, con valores respectivos de 94,22%, 94,09%, y 94,12%. Esta investigación resalta la implementación exitosa de la técnica de Ensemble Stacking para identificar de manera precoz la existencia de depresión en estudiantes universitarios pertenecientes a una institución pública en Perú. La implementación de este enfoque combinado evidenció mejoras sustanciales en la eficacia de la predicción de la depresión en comparación con el uso independiente de algoritmos.



Gutierrez et al. (2020) en su investigación titulada “*Factores psicosociales asociados a la depresión en gestantes atendidas en un centro materno infantil peruano, 2018*”, el propósito fue identificar los factores psicosociales vinculados a la depresión durante el embarazo. El estudio adoptó un enfoque de casos y controles, involucrando a 95 gestantes atendidas en un centro de atención primaria, donde la depresión se evaluó mediante el Cuestionario de la Salud del Paciente (PHQ-9). Para el análisis bivariante, se aplicó la prueba Chi-cuadrado de Pearson, mientras que el análisis multivariante se basó en la regresión logística con un nivel de confianza del 95%. La muestra comprendió 45 casos y 45 controles, sumando un total mínimo de 90 embarazadas. La validez del PHQ-9 y el formulario de recopilación de datos fue respaldada por tres expertos: dos psicólogos y un obstetra especializados en salud mental durante el embarazo, estos profesionales revisaron y aportaron observaciones antes de la ejecución del estudio. Considerando una probabilidad de aborto del 5%, se estableció una muestra total de 95 mujeres embarazadas. Mediante un análisis probabilístico aleatorio directo, se identificaron las variables de riesgo asociadas a la depresión durante el embarazo.

Aveiro-Róbalo et al.(2021) en su proyecto “*Ansiedad, depresión y estrés por COVID-19 en profesionales de la salud de Latinoamérica: Características y factores asociados*”, cuyo objetivo fue identificar los factores vinculados a la percepción de repercusiones en la salud mental de profesionales de la salud en Latinoamérica frente al COVID-19, mediante un análisis multivariado. Se observó que, a medida que aumentaba la edad, la percepción de repercusiones en la esfera mental disminuía (RPa: 0,98; IC95%: 0,97-0,99; valor $p=0,012$). Además, aquellos con una mayor percepción de impacto en la salud mental también



presentaron niveles más bajos de ansiedad (RPa: 1,84; IC95%: 1,14-2,98; valor $p=0,013$) y de estrés postraumático (RPa: 2,28; IC95%: 1,61-3,22; valor $p<0,001$), ajustando por la depresión y el estrés. Esto sugiere que los individuos más jóvenes tienden a experimentar una mayor percepción de repercusiones en la esfera mental. Además, aquellos con mayores percepciones de impacto en la esfera mental fueron aquellos que experimentaron ansiedad en niveles bajos y estrés postraumático. Asimismo, se identificaron percepciones significativas y descriptivas en relación con la esfera mental en el personal de salud, destacando sus miedos, preocupaciones, agotamiento y percepciones de maltrato en diversas situaciones, en el contexto de la atención durante la pandemia de la COVID-19.

Julio et al. (2019) en su proyecto "*Factores asociados a la depresión en personas mayores de 75 años de edad en un área urbana*" se abordó el objetivo de determinar la presencia de la depresión y factores asociados en esta población. La metodología adoptada consistió en un enfoque observacional, transversal y analítico, que incluyó la administración de un cuestionario a una muestra seleccionada de 400 adultos mayores comunitarios, mediante muestreo intencional, durante el año 2015. Los resultados revelaron una prevalencia general de síntomas depresivos del 30.5%, siendo más pronunciada en mujeres (40.1%). Los adultos mayores que vivían solos exhibieron una mayor prevalencia de síntomas depresivos (26.6%) en comparación con aquellos que compartían su residencia (16.1%) ($p = 0.011$). La buena autopercepción de la salud se asoció con una menor presencia de síntomas depresivos. Asimismo, los individuos no frágiles mostraron tasas más bajas de síntomas depresivos (12.1%) en comparación con los frágiles (30.8%) y los prefrágiles (57%) ($p < 0.05$). Además, los aspectos negativos en las relaciones sociales presentaron una puntuación media inferior



(0.003) en individuos sin síntomas depresivos ($p = 0.006$). En conclusión, se estableció una conexión entre la presencia de síntomas depresivos y variables como el género, la autopercepción de la salud, las condiciones de convivencia y la fragilidad en adultos mayores, resaltando la importancia de comprender estas asociaciones para informar estrategias de intervención dirigidas y mejorar el bienestar mental en esta población.

Macedo-Poma et al. (2019) en su proyecto "*Factores asociados a síntomas depresivos en madres con hijos hospitalizados en unidades de pediatría y neonatología en Perú: estudio de casos y controles*" el propósito de este estudio consistió en identificar y determinar los factores de riesgo asociados al desarrollo de depresión postparto en madres cuyos hijos han sido hospitalizados en unidades de pediatría y neonatología en la ciudad de Huancayo. Se llevó a cabo una investigación de tipo observacional y analítico, utilizando un diseño de casos y controles que incluyó las fichas de recolección de datos de madres cuyos hijos estuvieron hospitalizados en dichas unidades durante el año 2017. Al realizar un análisis multivariado, se observó que aquellas madres que se encontraban desempleadas ($p < 0,001$), las que eran solteras (valor $p < 0,001$) y las que experimentaron un embarazo no planificado ($p = 0,003$) presentaban una mayor probabilidad de desarrollar depresión postparto. Donde se encontró diversos factores relacionados con la depresión postparto: el estado civil, la relación con la pareja, la condición laboral y la no planificación del embarazo.

Nuankaew et al. (2022) en su proyecto "*Un modelo predictivo para el riesgo de depresión en jóvenes tailandeses durante COVID-19*" cuyo objetivo fue estudiar la situación de riesgo de depresión de los jóvenes en Tailandia y (2) desarrollar un modelo para predecir la depresión entre los jóvenes de Tailandia.



Los datos utilizados en la investigación fueron 1413 muestras de 9 facultades de la Universidad Rajabhat Maha Sarakham y la Escuela Phadungnaree en el distrito de Mueang de la provincia de Maha Sarakham, Tailandia, los resultados mostraron que la mayoría de los encuestados no tenían problemas de depresión con 1.059 muestras (74,95%) como se detalla en la Tabla IV. Sin embargo, todavía hay tres grupos de riesgo que necesitan ser monitoreados: nivel leve con 260 muestras (18,40%), nivel moderado con 78 muestras (5,52%) y nivel severo con 16 muestras (1,13%). Los datos utilizados en la investigación fueron 1,413 muestras

Zulfiker et al. (2021) en su proyecto *“Un análisis en profundidad de los enfoques de aprendizaje automático para predecir la depresión”* cuyo objetivo es identificar si una persona está deprimida o no, determinar los factores clave que inducen la depresión y determinar el mejor enfoque de aprendizaje automático para identificar a las personas deprimidas contiene seis clasificadores de aprendizaje automático diferentes utilizando información sociodemográfica y psicosocial para detectar si una persona está deprimida o no. Además, tres métodos diferentes de selección de características, como Seleccionar K-Best Features (SelectKBest), Redundancia mínima y relevancia máxima (mRMR) y el algoritmo de selección de características de Boruta se han utilizado para extraer las características más relevantes del conjunto de datos con el clasificador AdaBoost con la técnica de selección de características SelectKBest ha superado a todos los demás enfoques con una precisión del 92,56 %. Además, se han calculado otras métricas de evaluación, a saber, sensibilidad, especificidad, precisión, puntuación F1 y la medida de la superficie bajo la curva (AUC) de diversos modelos con el fin de determinar el modelo más eficiente para predecir



la depresión. Para lograr una mayor precisión en la predicción de la depresión, se ha utilizado la técnica de sobremuestreo de minorías sintéticas (SMOTE) que reduce el desequilibrio de clase de los datos de entrenamiento

Khubchandani et al. (2021) en su proyecto *“Depresión y ansiedad posteriores al confinamiento en EE. UU. durante la pandemia de COVID-19”* tiene como objetivo realizar una evaluación nacional integral y sistemática de la prevalencia de la depresión y la ansiedad en la población adulta de EE. UU llegando a la prevalencia de depresión (39 %), ansiedad (42 %) y malestar psicológico (39 %) se calculó a partir de la escala PHQ-4. En análisis de regresión múltiple, la carga de depresión, ansiedad y angustia psicológica (evaluada por la escala PHQ-4) se predijo significativamente en función de la raza, el origen étnico, la edad, tener hijos en el hogar, el empleo como trabajador de la salud, el ingreso familiar anual y el área de residencia. Los hombres tenían más probabilidades de tener depresión y las mujeres tenían más probabilidades de tener síntomas de ansiedad llegando a la conclusión de una fuerte persistencia de depresión y ansiedad, en los EE. UU. se recomiendan enfoques interdisciplinarios y multisectoriales junto con intervenciones basadas en la población para mejorar la salud mental

De Andrade et al. (2022) en su proyecto *“Experiencias adversas en infancia, características sociodemográficas e síntomas de depresión en adolescentes de un municipio do Rio de Janeiro, Brasil”* cuyo objetivo es identificar patrones de experiencias infantiles adversas entre estudiantes adolescentes en una localidad de Río de Janeiro, Brasil, considerando factores sociodemográficos (género, tono de piel y estatus socioeconómico) y la presencia de síntomas depresivos. El análisis comprendió evaluaciones bivariados y análisis



de correspondencias múltiples (ACM) de adversidades infantiles, variables sociodemográficas (género, tono de piel y situación socioeconómica) y síntomas depresivos. Los resultados revelan la agrupación en ocho categorías: niñas y adolescentes de estratos sociales bajos con experiencias infantiles adversas vinculadas al entorno; varones de estratos sociales más altos sin adversidades en la infancia; adolescentes con síntomas depresivos y experiencias infantiles adversas centradas en ellos/familia; adolescentes de piel blanca sin síntomas depresivos y sin experiencias infantiles adversas; adolescentes negros/mestizos/amarillos/indígenas con experiencias adversas en familia y comunidad; adolescentes que perdieron a sus padres por muerte y experimentan carencias alimenticias; adolescentes que sufrieron violencia psicológica; y adolescentes con experiencias sexuales con sus padres.

Martínez-García et al. (2022) en su proyecto *“Prevalencia de depresión, ansiedad y factores asociados en médicos residentes de centros hospitalarios durante la pandemia de COVID-19”* cuyo objetivo fue evaluar la prevalencia de ansiedad y depresión en médicos residentes de instituciones hospitalarias en México durante el periodo de la pandemia de COVID-19. Se llevó a cabo un estudio observacional transversal en el lapso de junio a diciembre de 2020. La muestra estuvo compuesta por 247 médicos residentes, cada uno de los cuales fue entrevistado individualmente mediante un cuestionario del CEMEVI . En el análisis de la muestra, se observó que la ansiedad se manifestó en un 17% de las ocasiones, mientras que la desesperación fue experimentada en un 45% de los casos. El rango de edad más común fue de 26 a 31 años, con la participación de 124 médicas y 123 médicos. Se identificó que las disciplinas de Anestesiología, Medicina Interna y Pediatría presentaron las tasas más altas de depresión y



ansiedad, respectivamente. Asimismo, se encontró que el contacto con COVID-19 (OR = 2,49, $p = 0,035$) y la ansiedad de los padres (OR = 2,6, $p = 0,009$) se destacaron como factores de riesgo para la depresión. La razón de momios (RM) observada fue de 2,3, con un valor p correspondiente de 0,031. Se estableció una relación significativa entre la ansiedad y los antecedentes, lo que lleva a la conclusión de que tanto la depresión como la ansiedad son factores contribuyentes a las alteraciones de la estabilidad emocional en los médicos residentes de los centros hospitalarios

Espinoza et al. (2022) con el proyecto *“Prevalencia y Factores Asociados con Depresión en Personal de Salud Durante la Pandemia de SARS-CoV-2 en el Departamento de Piura, Perú”*, con el propósito de determinar la prevalencia e identificar los factores vinculados a la depresión en el personal de salud en el contexto de la pandemia causada por el SARS-CoV-2, se llevó a cabo un estudio transversal analítico entre mayo y septiembre de 2020 en instituciones de salud. La recolección de datos se realizó mediante una encuesta aplicada a una muestra de 136 profesionales sanitarios. La variable dependiente, la depresión, fue cuantificada utilizando la escala autoaplicada de Zung. Los resultados revelaron que aproximadamente 1 de cada 10 profesionales y técnicos sanitarios experimentaba depresión durante la epidemia de COVID-19. Además, se observó una correlación negativa entre la depresión y el historial de pérdidas de familiares o amigos relacionadas con la COVID-19, mientras que el uso de equipo de protección personal (EPP) se identificó como un factor protector contra la depresión.

Rabinowitz et al. (2022) en su proyecto *“Comprobaciones de coherencia para mejorar la medición con la Hamilton Rating Scale for Depression (HAM-*



D)” tiene como objetivo aplicar comprobaciones de coherencia lógica y estadística a la Escala de Calificación de Hamilton para la Depresión (HAM-D17) con el fin de mejorar la precisión en la medición de los síntomas de trastornos del estado de ánimo, particularmente en el contexto de ensayos clínicos, se formó un grupo de trabajo de expertos bajo la Sociedad Internacional de Metodología y Ensayos Clínicos del SNC, este grupo desarrolló indicadores de consistencia/inconsistencia para la HAM-D17. Estos indicadores se aplicaron a 95,468 evaluaciones de HAM-D provenientes de 32 ensayos de medicamentos antidepresivos y a datos simulados por Monte Carlo para evaluar indicadores en condiciones de inconsistencia conocida. Los resultados demostraron una mejora significativa en las puntuaciones de depresión y discapacidad, el protocolo de EMTr intensiva se consideró seguro y factible, ofreciendo una alternativa prometedora con mayor eficacia y menor tiempo de exposición en comparación con los protocolos de EMTr convencionales, para mejorar la precisión de la medición de la Hamilton Depression Rating Scale HAM D se propuso un enfoque de control de consistencia en otro estudio

Santana et al. (2022) en la investigación *“Reconocimiento del efecto de los ejercicios vocales mediante Fuzzy Triangular Naive Bayes, un clasificador de aprendizaje automático: Un análisis preliminar”* con el objetivo de analizar la efectividad del clasificador de Naive Bayes en detectar cambios en parámetros vocales antes y después de ejercicios en mujeres con y sin disfonía. Este estudio evaluó el rendimiento del clasificador Fuzzy Triangular Naive Bayes (FTriangNB) en el análisis de patrones vocales antes y después de ejercicios vocales en mujeres con y sin disfonía. Después de la aplicación de los ejercicios, el FTriangNB demostró una precisión del 87,5% y un coeficiente Kappa del



81,3%, mientras que antes de los ejercicios mostró una precisión baja del 33,3% y un coeficiente Kappa negativo. Los resultados muestran que el FTriangNB es eficaz en reconocer los efectos de los ejercicios vocales, destacando la necesidad de más estudios con muestras más grandes y otros clasificadores de aprendizaje automático en Ciencias de la Voz. concluyendo una precisión en el reconocimiento del efecto de los ejercicios vocales. Para este propósito, se deben realizar más estudios exploratorios con muestras más grandes utilizando FTriangNB, así como otros clasificadores de aprendizaje automático, en Voice Science para permitir inferencias.

Sau & Bhakta (2019), Realizó una investigación que lleva por nombre *“Detección de la ansiedad y la depresión entre los marinos mediante tecnología de aprendizaje automático”*, con la finalidad de realizar una comparación del rendimiento entre diversos algoritmos de aprendizaje automático para identificar ansiedad y depresión en la población marítima, se entrevistó a un total de 470 marinos en Haldia Dock Complex, India, tras la adquisición de los permisos y la autorización ética necesarios. Se procedió a recopilar una variedad de datos sociodemográficos, ocupacionales y de salud. Se realizaron evaluaciones de cinco clasificadores de aprendizaje automático: Catboost, Logistic Regresión, Naïve Bayes, Random Forest y Support Vector Machine, utilizando el lenguaje de programación Python. Como consecuencia, se encontró que Catboost parecía ser el más adecuado para este propósito, logrando una exactitud y precisión del 82,6 % y 84,1% respectivamente, con el uso de esta tecnología, es posible sustituir el procedimiento de detección de ansiedad y depresión, que normalmente implica una considerable inversión de tiempo, por una técnica automatizada basada en computadora que ofrece una precisión aceptable.



Cheffi et al. (2022), presentaron su investigación denominada “*Validación de la Escala de Evaluación de la Depresión de Hamilton (HDRS) en el dialecto tunecino*”, cuyo objetivo es investigar la confiabilidad y validez de la HDRS entre Estudiantes tunecinos que han sido hospitalizados por un intento de suicidio. El estudio adoptó un diseño transversal en el servicio de urgencias, seleccionando Estudiantes hospitalizados por intento de suicidio. Se examinaron propiedades psicométricas, incluyendo validez de constructo, validez interna y externa. El análisis mostró alta correlación entre los ítems de HDRS y la puntuación total, con una varianza explicada del 64,4%. La versión tunecina demostró ser un instrumento efectivo para detectar la depresión en personas con intento de suicidio, con buena sensibilidad (80,8%) y especificidad (91,1%). Basándose en los resultados, se concluye que la versión tunecina de la HDRS es un instrumento válido y aceptable para detectar la depresión en individuos que han intentado suicidarse. Su capacidad para proporcionar información precisa y sensible destaca su utilidad clínica en el ámbito de las emergencias.

Liu et al. (2022) en la investigación “*Identificación de la diferencia en la percepción del tiempo entre el trastorno depresivo mayor y la depresión bipolar mediante una tarea de bisección temporal*” con el objetivo de explorar la diferencia en la percepción del tiempo entre MDD y BD utilizando una tarea de bisección temporal. En la evaluación de la percepción temporal, 30 Estudiantes con TDM, 30 con BD y 30 controles sanos participaron en una tarea de bisección. Se aplicó un análisis estadístico y se calcularon medidas como BP, DL y WR. Se halló una interacción significativa en el tipo de sujeto y la duración, indicando diferencias en las respuestas temporales de TDM y BD respecto a HC. Las variaciones en BP y WR entre los grupos resaltan la percepción temporal única de



TDM y BD. En conclusión, el TDM y BD exhibieron una percepción temporal distorsionada, especialmente en periodos cortos. TDM experimentó una sensación de tiempo más lenta que BD. Ambos grupos tuvieron menor sensibilidad temporal en comparación con HC, sin diferencias significativas entre TDM y BD.

Obagbuwa et al. (2023) en el estudio *“Modelos supervisados de aprendizaje automático para el análisis de sentimientos de depresión”* con el objetivo de demostrar cómo los algoritmos de aprendizaje automático pueden ser útiles para predecir problemas psicológicos en la vida moderna, se estudió la predicción de la depresión en los sectores laborales de Bangladesh utilizando un enfoque de aprendizaje automático, y se recogieron datos de hombres y mujeres. Los conjuntos de datos utilizados en esta investigación se obtuvieron de publicaciones en Twitter. Los resultados de esta investigación subrayan el potencial de emplear análisis de sentimientos y modelos de aprendizaje automático para detectar la melancolía en los usuarios de redes sociales en una etapa temprana. La implementación práctica e implementación de los modelos SVM y Regresión Logística en contextos del mundo real está respaldada por sus respectivos tiempos de ejecución, donde la Regresión Logística demuestra una mayor eficiencia. Los modelos SVM y de regresión logística en contextos del mundo real están respaldados por sus respectivos tiempos de ejecución, y la regresión logística demuestra una mayor eficiencia.

2.1.2 Ámbito Nacional

Orué Medina (2018) en la investigación *“Detección de depresión a través de análisis textual utilizando aprendizaje automático, 2017”* con el objetivo de esta investigación consiste en mejorar la eficiencia en la atención oportuna a



posibles Estudiantes con depresión, identificar características particulares en individuos considerados de alto riesgo de desarrollar depresión, y determinar los niveles de depresión en posibles Estudiantes mediante el análisis de texto utilizando técnicas de aprendizaje automático. nivel mundial, el "Inventario de Beck II" y el "Test de Frases Incompletas Sacks", con el objetivo de evaluar la presencia de síntomas depresivos y detectar en aquellos individuos que presenten el trastorno mental mencionado. Es importante destacar que el enfoque de investigación utilizado en este estudio es de naturaleza aplicada y exploratoria. Cabe vale la pena destacar que , para el procesamiento y tratamiento de los datos, se utilizó la medida numérica de análisis textual TF-IDF , junto con las técnicas de aprendizaje automático de Support Vector Machine, y el servicio Microsoft Azure Machine Learning con R. para el procesamiento y tratamiento de datos se utilizó la medida numérica de análisis textual TF-IDF , junto con las técnicas de aprendizaje automático de Support Vector Machine, y el servicio Microsoft Azure Machine Learning con R. La creación de una aplicación móvil para la detección de la depresión presenta una precisión del 99.2%.

Lucero (2022) en la investigación *“Comparación de técnicas de clasificación de aprendizaje de máquina en el diagnóstico del trastorno depresivo leve”* con el objetivo de comparar técnicas de clasificación de aprendizaje de máquina en el diagnóstico de trastorno depresivo leve. Estos estudios indican que el documento técnico "Auto Escala de Zung" del Ministerio de Salud es eficaz para identificar la depresión, con una especificidad del 63% y una sensibilidad del 97%, logrando un acierto del 82% en la discriminación de la depresión. En esta investigación, se desarrolló un método que comienza con la creación de un conjunto de datos basado en las variables de entrada y salida, así como el nivel de



prioridad según el cuestionario de Zung. Posteriormente, se seleccionaron técnicas de aprendizaje de máquina, como Naive Bayes, árboles de decisión, redes neuronales y máquinas de vectores de soporte, para mejorar la precisión en el diagnóstico de la depresión. Estas técnicas se implementaron y compararon utilizando la plataforma de Google Colaboratory y el lenguaje de programación Python. Según el método propuesto, desarrollado y evaluado, se concluye que las redes neuronales alcanzan una precisión del 100% en el diagnóstico de la depresión.

Peralta (2021) en la investigación *“Depresión, ansiedad y estrés en estudiantes de educación superior en tiempos del COVID-19”* con el objetivo de determinar los niveles de depresión, ansiedad y estrés en estudiantes de Educación Superior durante el periodo de confinamiento debido al Covid-19. Empleando un enfoque cuantitativo y descriptivo con un diseño de investigación no experimental de corte transversal, se seleccionó una muestra de 522 estudiantes de ambos géneros. Se aplicó un cuestionario sociodemográfico y la Escala de Depresión, Ansiedad y Estrés (DASS-21) para la recopilación de datos. Los resultados indicaron que el 53.9% de los estudiantes presentan niveles de depresión que van desde severos hasta extremadamente severos, el 64.5% exhiben niveles de ansiedad desde leves hasta extremadamente severos, y el 30.1% experimentan niveles de estrés en el rango de leve a extremadamente severo. Asimismo, se identificaron diferencias significativas en los niveles de ansiedad según el género, siendo las mujeres quienes muestran niveles más elevados de ansiedad.

Velez (2023) en la investigación *“Estrés, ansiedad y depresión en relación con el rendimiento académico durante la pandemia Covid-19 en los estudiantes de la escuela de Odontología de la Universidad Nacional Jorge Basadre*



Grohmann – 2021”, el propósito de esta investigación es examinar la conexión entre los niveles de estrés, ansiedad y depresión y el desempeño académico de los estudiantes de la escuela de Odontología de la Universidad Nacional Jorge Basadre Grohmann durante la pandemia de COVID-19 en 2021. Utilizando un enfoque correlacional y un diseño transversal, la encuesta virtual aplicada a 155 estudiantes empleó la Escala de Depresión, Ansiedad y Estrés (DASS-21). Aunque se encontraron niveles significativos de estrés, ansiedad y depresión, los resultados no demostraron una relación significativa entre estos factores y el rendimiento académico de los estudiantes. En general, la mayoría de los estudiantes mostraron un rendimiento académico medio. Por lo tanto la complejidad de la interacción entre la salud mental y el rendimiento académico durante la pandemia del COVID-19.

Pinto (2019) en la investigación *“Ansiedad y depresión en los estudiantes de la Especialidad de Físico - matemática de la Facultad de Ciencias de la educación de la Universidad Nacional de San Agustín de Arequipa – 2018”* cuyo objetivo fue determinar los niveles de ansiedad y depresión en estudiantes de la especialidad de Físico-Matemática. La metodología cuantitativa y descriptivo-exploratoria empleada se basó en encuestas utilizando el Cuestionario de ansiedad estado-rasgo STAI y el Inventario de Depresión de Beck-II (BDI-II) aplicados a 130 estudiantes. Los resultados revelaron que, en general, los niveles de ansiedad y depresión se encontraron en un rango moderado, con el 69% y el 43%, respectivamente. Se evidenció que los estudiantes de tercer año presentaban una inclinación mayor hacia la ansiedad, mientras que los de primer año exhibían una mayor propensión hacia la depresión. Estos resultados ofrecen una perspectiva



detallada sobre el bienestar psicológico de los estudiantes de Físico-Matemática en dicha facultad.

Lacuta (2019), en su investigación *“Prevalencia de depresión y factores asociados en internos de Medicina de la UNMSM-2016”*, con el objetivo de establecer la prevalencia de la depresión y sus factores asociados en estudiantes de Medicina de la UNMSM durante el año 2016, se llevó a cabo una investigación de tipo descriptivo y observacional de diseño transversal que analizó a 106 participantes en diciembre de 2016. Se emplearon el test de Zung y la prueba de Hamilton con el fin de evaluar la depresión, recopilando datos acerca de la edad, género, estado civil, pasatiempos y carga laboral. Se realizó un análisis factorial y se aplicó la prueba de alfa de Crombach para evaluar la consistencia interna. El análisis de correlación se llevó a cabo para cada test de depresión, utilizando el análisis de clúster para determinar el test más adecuado para medir la depresión y su prevalencia. El Test de Hamilton resultó ser el más predictivo, con una tasa de depresión del 14,2%. No se encontró una asociación significativa entre la depresión y las variables estudiadas, concluyendo que la prevalencia de la depresión en internos de Medicina es similar a la de la población general. Además, se determinó que factores como el Género, la edad, el estado civil, las actividades recreativas y la carga de trabajo no están asociados con una mayor proporción de casos de trastorno depresivo.

Quispe (2021), en la investigación *“Factores de riesgo asociados a depresión en estudiantes de enfermería de una universidad pública de Lima, 2021”*, el propósito de esta investigación es reconocer los elementos de riesgo asociados a la depresión en alumnos de enfermería en la Universidad Nacional Mayor de San Marcos. Se eligió una muestra de 73 estudiantes de segundo a



quinto año que presentaban algún nivel de depresión, utilizando un método de muestreo no probabilístico. Se aplicaron la "Escala de Depresión Autoaplicada" de Zung y un "Cuestionario de Factores de Riesgo Asociados a Depresión" diseñado por los investigadores a través de Google Forms. Los resultados indican que la mayoría de los estudiantes presentan depresión leve (84%) y moderada (16%), con la presencia de factores de riesgo asociados en el 52% de los casos, destacando los factores sociales (70%) y académicos (68%). El análisis de Spearman muestra una correlación moderada ($\rho: 0.503$) entre los factores de riesgo y la depresión. En resumen, los factores de riesgo de índole personal, social, familiar y académica están moderadamente vinculados con la depresión en estudiantes de enfermería, siendo la depresión leve la más prevalente. Se subraya la influencia de variables como el género, la percepción de salud, el aislamiento social debido a la COVID-19, las relaciones interpersonales, la situación económica familiar, la interferencia familiar, la percepción del programa académico, el respaldo docente, el interés en enfermería y la sobrecarga académica. Además, se evidencia que los factores personales y los relacionados con el ámbito académico mantienen una conexión moderada con la depresión en relación con los factores personales y académicos, mientras que la vinculación con los factores sociales y familiares es más débil.

2.1.3 Ámbito regional

Maquera (2020) en la investigación "*Depresión y estrés académico en estudiantes de Enfermería de la Universidad Nacional del Altiplano, Puno – 2019*" tuvo como objetivo principal explorar la conexión entre la depresión y el estrés académico en estudiantes del séptimo ciclo de la Facultad de Enfermería de la Universidad Nacional del Altiplano - Puno. Adoptando un enfoque



correlacional con un diseño no experimental y de corte transversal, se encuestó a una muestra de 54 estudiantes mediante un muestreo no probabilístico. Se utilizaron la escala de Zung y el inventario SISCO como herramientas validadas y confiables para evaluar la depresión y el estrés académico, respectivamente. El análisis de datos se realizó mediante el software estadístico SPSS versión 23.0, y la prueba Chi cuadrada se empleó para verificar la hipótesis. Los resultados indicaron que el 57.4% de los estudiantes presenta un nivel normal de depresión, mientras que el 81.5% experimenta un nivel moderado de estrés académico. La conclusión principal es que existe una relación significativa entre la depresión y el estrés académico en este grupo de estudiantes, respaldada por la prueba X2 con un valor de 30.702 y un valor de $p=0.000$, siendo inferior a $p=0.05$.

Quispe Pari (2021) en la investigación *“Factores de riesgo en la salud mental de los estudiantes de la Facultad de Ingeniería Estadística e Informática de la UNA Puno ante la pandemia de COVID-19”* cuyo objetivo es identificar los factores de riesgo que afectan la salud mental de los estudiantes de la Facultad de Ingeniería Estadística e Informática de la UNA Puno durante la pandemia de COVID-19. Utilizando un enfoque cuantitativo no experimental, se encuestó a 127 estudiantes de diversos semestres durante el año académico 2020-II mediante un muestreo aleatorio simple. Se utilizó un cuestionario evaluado con un coeficiente de Alfa de Crombach de 0.779 para garantizar su consistencia interna. El análisis factorial reveló que factores tales como el género, problemas relacionados con drogas, ansiedad, violencia física, depresión, problemas con el consumo de alcohol, dificultades para conciliar el sueño, violencia psicológica, la exposición negativa a noticias vinculadas al virus y el ingreso económico incidieron de manera significativa en la salud mental de los estudiantes. Se llevó



a cabo un análisis descriptivo exhaustivo de estos factores, que incluyó aspectos numéricos y porcentuales, y se identificaron factores socioeconómicos como la actividad física, ingreso económico, acceso a internet, alimentación, lugar de residencia y tipo de vivienda familiar como elementos destacados que afectan la salud mental.

Gallegos (2023) en su investigación *“Asociación del nivel de ansiedad, depresión y calidad de sueño con el síndrome de Burnout en internos del área de salud en el HCMM Juliaca-2022”*, cuyo propósito fue analizar la relación entre la ansiedad, la depresión y la calidad del sueño con el Síndrome de Burnout (SBO) en internos del área de salud que desempeñan sus funciones en el Hospital Carlos Monge Medrano de Juliaca en 2022. La metodología adoptada fue descriptiva, observacional, transversal y de naturaleza asociativa, utilizando una muestra conformada por internos de medicina, enfermería y obstetricia. Se utilizaron instrumentos como la Ficha de recolección de datos, la escala de ansiedad de Hamilton, el inventario de depresión de Beck-II, la escala de calidad de sueño de Pittsburgh y el Inventario de Maslach Burnout (MBI). Los resultados revelaron que el 19.4% de los 93 internos presentaron SBO, siendo el agotamiento emocional la dimensión más afectada (43%). Se identificaron asociaciones significativas entre niveles de depresión, ansiedad y la calidad de sueño inadecuada con el SBO. Se concluye, se estableció una conexión estadísticamente significativa entre los factores mencionados y el SBO en los internos del HCMM.

Holgado (2022) en la investigación *“Modelamiento de la satisfacción laboral de docentes de educación básica mediante técnicas Machine Learning”* cuyo objetivo es establecer un modelo predictivo para la satisfacción laboral de docentes de educación básica, utilizando técnicas de aprendizaje automático. Se



empleó un conjunto original de datos que consta de 15,087 instancias y 942 atributos, extraídos de la Encuesta Nacional a Docentes de instituciones educativas de educación básica regular (ENDO-2018) del Ministerio de Educación del Perú. Las técnicas de selección de características abarcaron el filtro ANOVA F-test y el filtro Chi-Cuadrado. En la etapa de modelado, se aplicaron algoritmos como la Regresión logística, Gradient Boosting, Random Forest, XGBoost y Decision Trees-CART. El algoritmo Random Forest obtuvo una precisión del 73%, una sensibilidad del 74.8%, un AUC de 0.82, con resultados destacados en la matriz de confusión: 163 falsos negativos y 484 verdaderos positivos. Se identificaron predictores claves, siendo los ingresos económicos, la satisfacción con la vida, la autoestima, la actividad pedagógica, la relación con el director(a), la percepción de las condiciones de vida, la satisfacción con relaciones familiares, problemas de salud relacionados con la depresión y la satisfacción en la relación con colegas los más significativos en la determinación de la satisfacción laboral de los docentes.

Mamani (2023) en la investigación *“La depresión, ansiedad y estrés asociado al rendimiento académico en escolares de nivel secundario en el contexto de la COVID - 19 en I.E.S. San Juan Bosco, Puno – 2021”* cuyo objetivo de este estudio fue examinar la influencia de las estrategias de afrontamiento en la relación entre el estrés académico y el bienestar psicológico en estudiantes universitarios. La investigación se llevó a cabo mediante un diseño correlacional y se aplicó a una muestra de 300 estudiantes de diversas disciplinas. Se utilizaron cuestionarios para medir el estrés académico, las estrategias de afrontamiento y el bienestar psicológico. Los resultados revelaron una correlación negativa significativa entre el estrés académico y el bienestar psicológico. Además, se



encontró que las estrategias de afrontamiento mediaban esta relación, ya que los estudiantes que empleaban estrategias efectivas experimentaban menos estrés académico y, a su vez, reportaban un mayor bienestar psicológico. En conclusión, este estudio destaca la importancia de las estrategias de afrontamiento en el manejo del estrés académico y su impacto en el bienestar psicológico de los estudiantes universitarios.

Cosi (2022) en la investigación *“Las clases virtuales y su relación con la salud mental en estudiantes de la IES César Vallejo de la provincia de Huancané en el contexto de la cuarentena por Covid-19, 2020”* el objetivo de este estudio fue establecer la relación entre las clases virtuales y la salud mental de los estudiantes de la institución educativa secundaria César Vallejo durante la cuarentena por Covid-19 en 2020. Se empleó un enfoque hipotético deductivo con un diseño no experimental, utilizando una encuesta y la escala de Likert como instrumento aplicado a 192 estudiantes. Los resultados, respaldados por un valor de correlación de Spearman de 0.674 y una significancia de 0.000, indicaron una correlación positiva significativa entre las clases virtuales y la salud mental. Los resultados detallaron niveles diversos en áreas como el aprendizaje, el uso de recursos, síntomas somáticos, ansiedad, insomnio, aprendizaje colaborativo, satisfacción en actividades, logro de competencias y niveles de depresión, destacando que la mayoría de los estudiantes experimentaron niveles bajos o leves de depresión durante la cuarentena.



2.2 MARCO TEÓRICO

2.2.1 Machine Learning (Aprendizaje Automático)

La adopción del aprendizaje automático ha experimentado un aumento notable en su preferencia por parte de las empresas y organizaciones de diversos sectores para mejorar la eficiencia y la productividad. Utilice Internet para investigar casos de uso de aprendizaje automático. Identifique cinco casos de uso que abarcan diferentes sectores o aplicaciones y utilice el espacio a continuación para analizar sus conclusiones.

El aprendizaje automático abarca diferentes algoritmos o modelos con una amplia aplicabilidad, mientras que otros son adecuados para aplicaciones específicas. El aprendizaje automático se divide en tres enfoques principales de modelos de aprendizaje supervisados, no supervisados y de refuerzo. Cada modelo difiere en el entrenamiento; cada uno tiene sus fortalezas y enfrenta diferentes tareas o problemas. Al elegir un modelo de aprendizaje automático para implementar, una organización debe comprender los datos disponibles y el problema a resolver.

2.2.1.1 Tipos de Análisis de Aprendizaje Automático

- **No supervisado**

Los algoritmos de aprendizaje automático no supervisados prescinden de la intervención de expertos humanos, ya que son capaces de descubrir patrones en los datos de manera autónoma. Se ocupa principalmente de datos sin etiquetar. El modelo tiene que funcionar por sí solo para encontrar patrones e información. Ciertos casos de situaciones

abordadas mediante enfoques no supervisados incluyen la formación de clústeres y la identificación de asociaciones.(Efecode, n.d.)

- **Supervisado**

Los algoritmos de aprendizaje automáticos supervisados son los más utilizados para el análisis predictivo. El aprendizaje automático supervisado requiere la interacción humana para etiquetar los datos leídos para un aprendizaje supervisado preciso. En el aprendizaje supervisado, el modelo se enseña con ejemplos mediante el uso de conjuntos de datos de entrada y salida procesados por expertos humanos, generalmente científicos de datos. El modelo aprende las relaciones entre los datos de entrada

- **Refuerzo**

El aprendizaje por refuerzo enseña a la máquina a través de la prueba y el error mediante la retroalimentación de sus acciones y experiencias, también conocida como aprendizaje de errores. Implica asignar valores positivos a los resultados deseados y valores negativos a los efectos no deseados. El resultado son soluciones optimas; el sistema aprende a evitar resultados adversos y buscar lo positivo. Las aplicaciones prácticas del aprendizaje por refuerzo incluyen la creación de inteligencia de ratificación para juegos de video, robótica y automatización industrial.

2.2.1.2 Proceso de aprendizaje automático

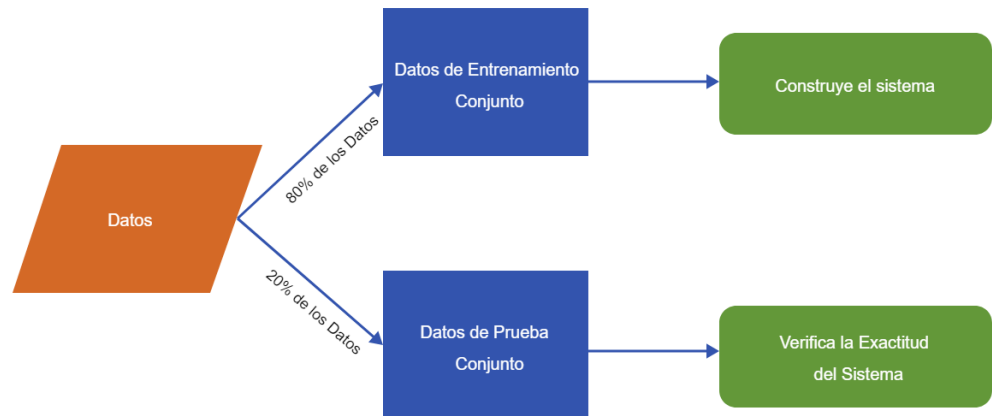
La dinámica de aprendizaje automático, en el área de inteligencia artificial (Iued, n.d.), constituye un ciclo dinámico y evolutivo que busca



dotar a los sistemas informáticos de la capacidad para mejorar su rendimiento mediante la experiencia adquirida. Este proceso comprende diversas etapas cruciales que permiten que un modelo o algoritmo se adapte y evolucione frente a datos nuevos las cuales son:

- Paso 1 - Preparar los datos con procedimientos de limpieza de datos.
- Paso 2 - Crear un grupo de datos de aprendizaje destinado para la capacitación del modelo y probar un conjunto de datos para evaluar el desempeño.
- Paso 3 - El algoritmo se elige en función del problema que se debe resolver.
- Paso 4 - Se evalúa el algoritmo en los datos de aprendizaje.
- Paso 5 - La solución se prueba con los datos de prueba.
- Paso 6 - Se implementa el modelo.

El reconocimiento de patrones requiere que el sistema “aprenda” para obtener resultados precisos. El aprendizaje es la fase más crítica para determinar qué tan bien se desempeña el sistema con los datos proporcionados. Para el proceso de aprendizaje, todos los datos se dividen en dos conjuntos. El primer conjunto es el conjunto de entrenamiento utilizado en la construcción y capacitación del modelo. El segundo es el conjunto de prueba, que se utiliza para probar el modelo y verificar que produce la salida correcta. En general, el conjunto de entrenamiento consta del 80% de los datos y el conjunto de pruebas consta del otro 20%.



2.2.2 Minería de datos

La minería de datos se presenta como una herramienta esencial que posibilita la adquisición de información a partir de conjuntos de datos almacenados, sometiéndolos a un proceso complejo y no trivial con el fin de transformarlos en información objetiva y utilizable, contribuyendo así a la toma de decisiones empresariales más fundamentadas. En concordancia con la definición propuesta por (Fayyad, 2000), la minería de datos se configura como un procedimiento de reconocimiento de patrones significativos, previamente desconocidos, pero potencialmente valiosos y comprensibles que yacen ocultos en la vastedad de los datos.

La Minería de Datos, según la conceptualización adoptada, se concibe como la extracción profunda de información implícita, la cual es de naturaleza novedosa y posee un potencial uso práctico, todo ello derivado de los datos existentes. Dentro del marco actual de la era de informática, la minería de datos emerge como herramienta fundamental, tal como se destaca en la perspectiva de (D.M., 2015), permitiendo el análisis y la explotación eficiente de los datos en pos de los objetivos estratégicos de cualquier organización.



2.2.2.1 Características de la minería de datos

Las características inherentes al empleo de la minería de datos son diversas y fundamentales para su aplicación eficaz:

- La minería de datos se destaca por su capacidad para explorar información alojada en las profundidades extensas de grandes bases de datos. Estos depósitos a menudo contienen información acumulada a lo largo de varios años, ofreciendo una riqueza de datos históricos.
- El entorno en el que opera la minería de datos generalmente adopta una estructura cliente-servidor, proporcionando una plataforma eficiente para la interacción y el procesamiento de datos entre sistemas.
- Las herramientas de minería de datos desempeñan un papel crucial al obtener el "mineral" de información que yace resguardado en documentos corporativos o registros públicos archivados. Este proceso permite desenterrar conocimientos valiosos que pueden haber pasado desapercibidos.
- La acción de hurgar y sacudir en los datos conlleva la posibilidad de descubrir resultados valiosos e inesperados. Este enfoque exploratorio puede revelar patrones y relaciones no evidentes en una primera instancia, brindando perspectivas valiosas.
- Las herramientas de minería de datos se caracterizan por su capacidad para combinarse de manera fluida, permitiendo un



análisis y procesamiento rápido de los datos. Esta versatilidad facilita la integración de diversas técnicas y enfoques para obtener resultados más completos.

- La minería de datos proporciona cinco clases fundamentales de información: conexiones, sucesiones, categorizaciones, agrupaciones y previsiones. Estos tipos de información abarcan desde relaciones entre variables hasta la predicción de tendencias futuras, proporcionando una visión integral de los datos analizados.

Antes de sumergirnos en los algoritmos y técnicas de exploración de datos, es esencial comprender las distintas fases de un proceso de Data Mining, sin profundizar completamente en cada una de ellas. En la fase inicial, se realiza la selección de datos, la cual puede estar alojada en un Data Warehouse. A continuación, se lleva a cabo el procesamiento de la información, una etapa que podríamos describir como una limpieza de datos, donde se eliminan aquellos que no son necesarios, según lo señalado por Fayad (2006).

2.2.3 Algoritmos y técnicas de explotación de datos

La fase culminante implica la aplicación de diversos modelos o técnicas de explotación para obtener información valiosa,

- a) **Predicción:** Enfocado en anticipar eventos futuros o valores basándose en patrones históricos de datos.
- b) **Asociación:** Detecta asociaciones entre variables y es ampliamente utilizado en sectores como supermercados para analizar patrones de compra.



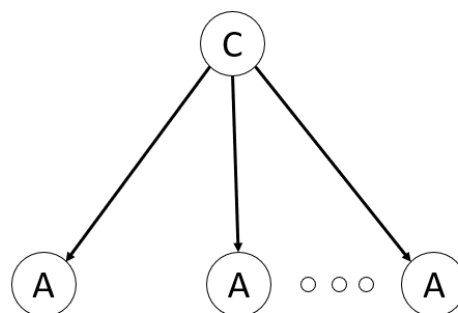
- c) **Clustering:** Busca identificar elementos afines en un conjunto de datos, agrupándolos según similitudes.
- d) **Árboles de Decisiones:** Herramienta de clasificación que utiliza un enfoque jerárquico para organizar las decisiones basándose en características específicas
- e) **Series Temporales:** Algoritmo diseñado específicamente para prever valores de magnitudes en función del tiempo, vital en análisis temporales.
- f) **Algoritmo Naive Bayes:** Comúnmente utilizado como punto inicial en el análisis exploratorio de datos, busca correlaciones entre atributos para realizar clasificaciones.
- g) **Redes Neuronales:** Resuelve problemas de clasificación y regresión, compitiendo con los árboles de decisiones y ofreciendo una alternativa en el análisis de datos complejos.

2.2.4 Clasificador (Naive de Bayes)

En el ámbito del aprendizaje de máquina, los clasificadores de Bayes constituyen una familia de clasificadores probabilísticos fundamentados en la implementación del teorema de Bayes, haciendo supuestos de independencia entre los atributos (Barredad Barreda & Katherine Michelle, 2022). El Naive Bayes ha sido objeto de estudio extenso desde la década de 1950. Su introducción inicial en la década de 1960, con un nombre diferente, se vincula a la recuperación de textos y, a lo largo del tiempo, ha perdurado como un método popular y una línea de base para la categorización de textos, especialmente al emplear frecuencias de palabras como características.

Con un preprocesamiento adecuado, el Naive Bayes demuestra ser altamente competitivo en la categorización de textos, incluso frente a métodos más avanzados como los clasificadores de vectores de soporte. Asimismo, su aplicabilidad se extiende al ámbito del diagnóstico automático médico, como se evidencia en las investigaciones de Mitchell (2016). Este enfoque probabilístico sigue siendo relevante y eficaz en diversas aplicaciones, destacándose por su simplicidad conceptual y su capacidad para proporcionar resultados sólidos en la clasificación de datos en distintos dominios.

El clasificador Naive Bayes clasifica un objeto u observación en la clase con la mayor probabilidad, asumiendo que todas las variables o atributos son independientes condicionalmente dada la clase. Dadas las variables o atributos A_1, A_2, \dots, A_n que predicen el valor de la clase C , la hipótesis de independencia adoptada por el clasificador Naive Bayes conduce a un modelo de red bayesiana (Alarcón J.C., 2015). En él, se encuentra un solo nodo raíz, que representa la clase, y todos los atributos actúan como nodos hoja con la variable clase como único progenitor.



Estructura del clasificador Naive Bayes

El clasificador Naive Bayes utiliza la ecuación (1) para clasificar una observación o conjunto de atributos. Esta ecuación se basa en la suposición de que

los atributos son independientes entre sí, dado el valor de la clase. En consecuencia, la probabilidad de que una observación o conjunto de atributos $\{a_1, a_2, \dots, a_n\}$ pertenezca a la clase "c" es:

Ecuación (2).

$$P(C = c/a_1, a_2, \dots, a_n) = \frac{P(C = c)P(a_1/C = c)P(a_2/C = c) \dots P(C = c)P(a_n/C = c)}{P(a_1, a_2, \dots, a_n)}$$

Factorizando:

$$P(C = c/a_1, a_2, \dots, a_n) = \frac{P(C = c) \prod_{i=1}^n P(a_i / C = c)}{P(a_1, a_2, \dots, a_n)} \quad (3)$$

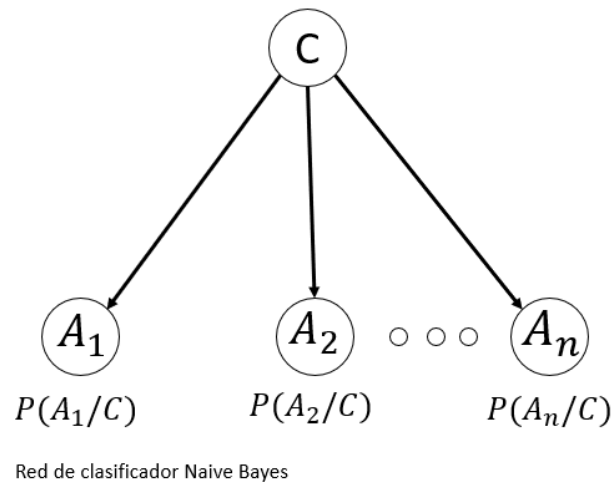
Con $c = c_1, c_2, \dots, c_m$

esta ecuación será empleada para calcular las probabilidades en la tarea durante el proceso de clasificación.

Especificación de los parámetros de la red del clasificador Naive Bayes

La definición de valores de los parámetros de la red Naive Bayes consiste en estimar los parámetros de las distribuciones de probabilidad que se definen en la red.

Una vez definida la red Naive Bayes, se deben estimar los parámetros de la distribución de probabilidad de la variable clase $P(C)$ y de las distribuciones condicionales de los atributos dada la clase $P\left(\frac{A_1}{C}\right)$. Estas distribuciones se representan en la siguiente figura:



La manera de estimar los parámetros dependerá de si las variables o atributos A_i son de naturaleza discreta o continua.

2.2.4.1 Para atributos discretos

El clasificador Naive Bayes es un algoritmo probabilístico empleado habitualmente en tareas de clasificación, que opera bajo el supuesto de independencia entre predictores o características (Del et al., 2023), o que implica que la presencia de una característica concreta no afecta a las demás (González-Ortega et al., 2022) La perspectiva bayesiana, presentada según Thomas Bayes, tiene especial relevancia en la predicción de la depresión en el campo médico, este enfoque se fundamenta en conceptos de probabilidades y asume la falta de dependencia estadística entre diferentes características, por lo que considera que las características específicas de una clase pueden existir de forma autónoma independientemente de otras características.



La evaluación de la probabilidad de distribución para la clase $P(C)$ y distribución condicional $P(A_i / C)$ se fundamenta en el número de veces que se producen los sucesos en cuestión que se determinan en el conjunto de datos. Entonces el valor estimado de probabilidad es:

$$P(a_i / C = c) = \frac{n(a_i, C = c)}{n(C = c)}$$

Donde,

- $n(a_i, C = c)$: Representa la cantidad de instancias en el conjunto de datos en las que la variable A_i tiene el valor a_i y su variable padre (Clase) C toma el valor c .
- $n(C = c)$: Indica la cantidad de instancias en el conjunto de datos en las que la clase C tiene el valor c .

Aunque esta estimación inicialmente parecer una aproximación razonable, en problemas del mundo real es común encontrarse con la limitación de que los casos en el conjunto de datos no abarcan todas las combinaciones posibles de valores entre la variable clase y las variables predictoras, por lo que esta limitación puede conducir a una estimación de parámetros con un exceso de ceros.

Se han desarrollado varios métodos para abordar estas problemáticas, y uno de ellos es el estimador basado en la Ley de Sucesión de Laplace [15]. En este enfoque, en lugar de estimar directamente la probabilidad como,

$$\frac{\text{Casos favorables}}{\text{Casos Totales}},$$

emplea como estimación el valor obtenido al dividir



$$\frac{n(a_i, C = c) + 1}{n(C = c) + r_{A_i}}$$

En otras palabras, se obtiene dividiendo la suma de los casos favorables y uno adicional entre la suma de todos los casos y el número de valores potenciales del atributo A_i denotado r_{A_i} o equivalente,

$$\frac{N_{ik} + 1}{N_i + r_i}$$

Donde,

N_{ik} : Número de casos del conjunto de datos donde la variable A_i opta por el k -ésimo valor y la clase C su valor c .

N_i : Cantidad de casos del conjunto de datos donde la variable A_i toma las clases y la clase C toma su valor c .

r_i : Número de posibles valores de la variable A_i .

2.2.5 Indicadores para evaluar el rendimiento del modelo

Siguiendo el planteamiento adoptado en este estudio, la precisión del modelo de aprendizaje automático se evaluó utilizando una matriz de confusión N-N, donde N representa el número de etiquetas de clase, esta matriz sirve para determinar el verdadero rendimiento de diversas técnicas de aprendizaje automático. Con base en esta matriz, es posible determinar la precisión del modelo, la cual indica la habilidad del algoritmo para prever casos con precisión. La exactitud se determina promediando los valores diagonales.

Kappa

Es una técnica estadística multivariada utilizada en pruebas de precisión, particularmente en el contexto de modelos de clasificación, para determinar si una matriz de error es significativamente diferente. Se basa en la comparación entre el acuerdo real en la matriz de error y el acuerdo esperado al azar, indicado por las filas y columnas marginales.

Tabla 1:

Interpretación del coeficiente Kappa

Estadística kappa	Fuerza de concordancia
< 0.00	Mala
0.00 - 0.20	Pobre
0.21 - 0.40	Débil
0.41 - 0.60	Aceptable
0.61 - 0.80	Bueno
0.81 - 1	Excelente

Nota: J. Cerda, L. Villarroel (2008)

Teniendo un modelo de muestreo multinomial, la estimación máxima de verosimilitud para Kappa se calcula mediante

$$\hat{K} = \frac{P_O - P_C}{1 - P_C}$$

Con fines de cálculos computacionales:

$$\hat{K} = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} n_{+i}} P_C$$

Accuracy



$$CA = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precisión: Es la métrica que representa el porcentaje total de valores correctamente clasificados, tanto positivos como negativos

$$Precision = \frac{TP}{TP + FP}$$

- Especificidad (Specificity): También conocida como tasa de verdaderos negativos, es la proporción de casos negativos correctamente identificados por el modelo. Es la capacidad del modelo para detectar los casos negativos de una condición o enfermedad.

$$Especificidad = \frac{TN}{TN + FP}$$

- Sensibilidad (Sensitivity): También conocida como tasa de verdaderos positivos, es la proporción de casos positivos correctamente identificados por el modelo. Es la capacidad del modelo para detectar los casos positivos de una condición o enfermedad.

$$Sensibilidad = \frac{TP}{TP + FN}$$

- Pos Pred Value (Positive Predictive Value): También conocido como valor predictivo positivo, es la proporción de casos positivos correctamente identificados en relación con todos los casos clasificados como positivos por el modelo. Es la probabilidad de que un caso clasificado como positivo sea realmente positivo

$$PPV = \frac{sens * prev}{(sens * prev) + ((1 - espe) * (1 - prev))}$$



- Neg Pred Value (Negative Predictive Value): También conocido como valor predictivo negativo, es la proporción de casos negativos correctamente identificados en relación con todos los casos clasificados como negativos por el modelo. Es la probabilidad de que un caso clasificado como negativo sea realmente negativo.

$$NPV = \frac{espe * (1 - prev)}{((1 - sens) * prev) + ((espe) * (1 - prev))}$$

- Tasa de detección: También conocido como tasa de detección, es la proporción de casos correctamente identificados por el modelo en relación con todos los casos reales en la población. Es una medida de la capacidad del modelo para detectar la presencia de una condición o enfermedad en la población.

$$Tasa\ de\ Detección = \frac{TP}{TP + FP + FN + TN}$$

- Prevalencia de detección: Representa la fracción de casos que el modelo ha identificado correctamente en comparación con la totalidad de los casos clasificados por el modelo como positivos. Representa la frecuencia o prevalencia de los casos detectados por el modelo.

$$Prevalencia\ de\ Detección = \frac{TP + FP}{TP + FP + FN + TN}$$

- Precisión de Equilibrio (Balanced Accuracy): También conocida como precisión equilibrada, es la medida promedio de la sensibilidad y la especificidad del modelo. Proporciona una medida general del rendimiento del modelo al considerar tanto los verdaderos positivos como los verdaderos negativos. Es útil cuando las clases están desequilibradas en el conjunto de datos.



$$\textit{Precisi3n de equilibrio} = \frac{(\textit{Sensibilidad} + \textit{especificidad})}{2}$$

Disponer de m3tricas de evaluaci3n s3lidas capacita a los profesionales sanitarios para tomar decisiones informadas y precisas en relaci3n con el diagn3stico y el tratamiento de la depresi3n en los Estudiantes, un modelo con m3tricas de alta sensibilidad y precisi3n puede identificar eficazmente a los individuos en riesgo de depresi3n y asignar con prontitud los recursos y la atenci3n adecuados.

2.3 MARCO CONCEPTUAL

2.3.1 Depresi3n

La depresi3n es un trastorno mental que suele repetirse y provoca una perturbaci3n en el estado emocional, manifest3ndose principalmente como tristeza y, en ocasiones, ansiedad. Esta afecci3n puede llevar a la persona a experimentar una serie de s3ntomas adicionales, como apat3a, sensaci3n de vac3o, retiro social, disminuci3n del apetito y problemas para dormir. Tambi3n puede causar agitaci3n, lentitud en el pensamiento y movimiento, sentimientos de culpabilidad e inutilidad, y pensamientos relacionados con la muerte o suicidio, as3 como s3ntomas f3sicos variados (Garc3a, n.d.).

La depresi3n se caracteriza por una reducci3n del bienestar emocional y una notable falta de inter3s o capacidad para disfrutar de las actividades diarias. Esta condici3n lleva consigo s3ntomas mentales como la tristeza y problemas de concentraci3n y memoria, as3 como manifestaciones f3sicas como cambios en el apetito y disminuci3n del deseo sexual. Adem3s de ver la depresi3n desde la perspectiva de una p3rdida emocional, es importante entenderla en relaci3n con la



fragilidad interna, la susceptibilidad y los sentimientos de desprotección e inferioridad del individuo (SH, n.d.)

Los trastornos conocidos como depresivos engloban variedades clínicas que van desde problemas leves hasta graves, como el episodio de depresión mayor y la distimia. Según el DSM-5, estos se clasifican como trastornos del estado de ánimo. Estos trastornos son síndromes que presentan comportamientos atípicos de importancia clínica, que se apartan de las reacciones estándar. Su rasgo distintivo es una perturbación en el estado de ánimo que conlleva malestar y dificultades funcionales, aumentando el riesgo de autolesión, dolor o pérdida de libertad (Association, n.d.)

Así, estos trastornos afectan principalmente el estado de ánimo y las emociones de quienes los padecen, provocando alteraciones en áreas como el apetito, el nivel de energía, el sueño, la capacidad de pensar, sentimientos de culpabilidad intensos e incluso pensamientos suicidas. A pesar de los avances en la investigación, todavía existen incógnitas sobre las causas exactas de la depresión. Algunos estudios apuntan a causas individuales, como predisposiciones genéticas, traumas emocionales o problemas vasculares cerebrales. Por otro lado, se reconoce el papel crucial de factores psicosociales, como eventos vitales significativos o situaciones económicas y políticas adversas. También es posible que haya una combinación de varios factores que interactúan de manera compleja, lo que sugiere que las soluciones exclusivamente individuales podrían no ser suficientes (Torre A. Vasco M. DE A, n.d.)



2.3.1.1 Fisiología de la Depresión.

Las alteraciones en neurotransmisores, citoquinas y hormonas parecen tener un papel crucial en el inicio y desarrollo de la enfermedad. La depresión se asocia a una reducción en la actividad sináptica, lo que significa un descenso en la neurotransmisión. Durante un episodio depresivo, los neurotransmisores que se ven afectados incluyen:

- Serotonina
- Acetilcolina
- Catecolaminas: Dopamina, Noradrenalina y Adrenalina.

2.3.1.2 Evaluación de la depresión

El diagnóstico de un trastorno depresivo a menudo surge de señales observacionales no muy definidas, tales como descuido en la presentación personal, lentitud en los movimientos, tono de voz apagado, expresión facial de tristeza, tendencia al llanto o llanto sin motivo aparente, atención reducida, expresión de pensamientos negativos (como sentimientos de culpa, preocupaciones de salud infundadas, sensación de fracaso) junto con problemas de sueño y molestias físicas no concretas. La clave para diferenciar estos síntomas patológicos de comportamientos habituales radica en la duración constante de los mismos, su intensidad y el nivel de impacto en la funcionalidad y vida social del individuo (Sanidad, n.d.)

2.3.2 Factores de riesgo

La salud mental y física mantienen una conexión intrínseca, cuya interacción puede ejercer un impacto sustancial en la salud global. Los trastornos



mentales no solo representan un desafío en sí mismos, sino que también pueden amplificar el riesgo de enfermedades físicas y psicológicas. Factores determinantes, tales como la desestructuración familiar, la pobreza extrema y la incidencia de embarazo adolescente, entre otros, pueden contribuir de manera significativa a la génesis de los trastornos mentales. Un reciente estudio ha arrojado luz sobre la potente influencia de la desestructuración familiar en la instauración de enfermedades psiquiátricas, destacando la compleja red de factores socioambientales que moldean la salud mental (40).

La íntima conexión entre la salud mental y física se manifiesta no solo en la predisposición a enfermedades concomitantes, sino también en la calidad de vida general de los individuos. La interrelación entre estos aspectos puede ser explorada desde perspectivas bio-psico-sociales, considerando no solo los componentes biológicos, sino también los factores psicológicos y sociales que influyen en la salud. La literatura científica abunda en evidencia que respalda esta relación compleja, subrayando la importancia de una evaluación holística para abordar de manera efectiva las condiciones de salud tanto físicas como mentales.

2.3.3 Manifestaciones clínicas

La depresión, un trastorno que impacta el estado de ánimo, se caracteriza por una serie de síntomas que abarcan desde la tristeza y la apatía hasta alteraciones en el sueño, apetito y peso, acompañados de fatiga, dificultad de concentración, sentimientos de culpa y pensamientos de muerte. Alarmanamente, cerca de un cuarto de la población general experimenta un episodio depresivo a lo largo de su vida. Sin embargo, la depresión a menudo se diagnostica y trata de manera incorrecta, lo que subraya la necesidad de una evaluación cuidadosa por



parte de los profesionales de la salud. La valoración del riesgo de suicidio es esencial, y en caso necesario, la derivación a un especialista se convierte en un paso crítico. La presencia de ansiedad o agitación aumenta significativamente el riesgo de suicidio.

Por otro lado, el trastorno depresivo persistente se caracteriza por síntomas depresivos constantes y crónicos que perduran al menos durante dos años. Este trastorno, a veces confundido con la depresión mayor, puede manifestarse como una "depresión doble". Su prevalencia es más pronunciada en mujeres y tiende a aumentar con la edad. Con raíces en factores genéticos y psicosociales, eventos estresantes intensos pueden actuar como desencadenantes. La duración de un episodio puede variar ampliamente, desde meses hasta años, y la recurrencia es común, siendo el número de episodios previos un factor de riesgo significativo. En casos extremos, la sintomatología psicótica puede manifestarse en una minoría de Estudiantes. Además, el trastorno afectivo estacional, más frecuente en mujeres, muestra una posible mejoría con la exposición a la luz, resaltando la importancia de considerar factores ambientales en la comprensión y abordaje de los trastornos depresivos.

2.3.4 Diagnostico

El diagnóstico de la depresión implica la evaluación cuidadosa de los síntomas y la consideración de criterios establecidos por manuales como el DSM-5. Recursos como "Psicopatología: Comprendiendo la Conducta Anormal" ofrecen una perspectiva detallada sobre los procesos de diagnóstico, subrayando la importancia de evaluar la duración, intensidad y repercusiones de los síntomas para una identificación precisa.

2.3.5 Escala de valoración de Hamilton

La escala en cuestión es una herramienta heteroaplicada, diseñada específicamente para la evaluación cuantitativa de la gravedad de los síntomas en Estudiantes previamente diagnosticados con depresión, con el propósito de monitorizar de manera precisa los cambios en el estado del paciente deprimido. Su aplicación se basa en la información recopilada durante la entrevista clínica, complementada con datos provenientes de fuentes secundarias.

Aunque la versión original de la escala constaba de 21 ítems, una versión reducida de 17 ítems fue posteriormente desarrollada y es recomendada por el Instituto Nacional de Salud Mental de los Estados Unidos. La validación de la versión en español de esta escala fue llevada a cabo en 1986 por Ramos-Brieva. Diversas evaluaciones han corroborado la validez discriminante, fiabilidad y sensibilidad al cambio tanto en poblaciones hospitalizadas como ambulatorias.

Tabla 2:

Escala de valoración de Hamilton

ITEM	INDICADOR
Humor depresivo (tristeza desesperación, autodesprecio)	0. Ausente
	1. Estos sentimientos solo se revelan cuando son preguntados.
	2. Estos sentimientos se dan a conocer de manera verbal y espontáneamente.
	3. Sensaciones no comunicadas oralmente (expresión facial, postura, voz o tendencia al llanto)
	4. Manifiesta estas sensaciones en su comunicación verbal y no verbal de forma espontánea.
Sentido de culpabilidad	0. Ausente
	1. Se recrimina si mismo, tiene la sensación de haber causado daño
	2. Idea de culpa o reflexión sobre errores pasados o acciones malas
	3. Ideas deliberantes de culpabilidad
	4. Oye voces acusatorias o de denuncia o alucinaciones amenazantes visuales y/o auditivas



Tendencia a morirse	<hr/> <ul style="list-style-type: none">0. Ausente1. Tiene la impresión de que no vale la pena vivir2. Desea morir o tiene imaginaciones que elevan posibilidad querer morirse.3. Ideas o actitudes suicidas4. Intentos o tentativa de suicidio <hr/>
Insomnio nocturno	<hr/> <ul style="list-style-type: none">0. Ausente1. Dificultad esporádica por conciliar el sueño2. Desvelo nocturno (se levanta de la cama sin justificación de orinar, tomar o dar medicación) <hr/>
Insomnio matinal	<hr/> <ul style="list-style-type: none">0. Ausente1. Despertar a primeras horas de la madrugada pero vuelve a dormirse2. Incapacidad de volver a dormir si se levanta de la cama <hr/>
Actividad académica en la universidad	<hr/> <ul style="list-style-type: none">0. Ninguna dificultad1. Ideas y pensamientos de sensación de incapacidad, Fatiga o debilidad relacionadas con su trabajo, actividad o aficiones.2. Desinterés de las actividades académicas.3. Disminución del tiempo de actividad o descenso en la productividad4. Ha dejado su trabajar por la enfermedad <hr/>
Comunicación fluida en las actividades diarias	<hr/> <ul style="list-style-type: none">0. Palabra y pensamiento normales1. Retraso ligero durante el diálogo.2. Retraso evidente durante el diálogo.3. Comunicación difícil4. Torpeza absoluta. <hr/>
Manía corporal	<hr/> <ul style="list-style-type: none">0. Ninguno.1. Juega con sus dedos, cabellos, etc.2. Inquieto con las manos, se muerde las uñas, los labios, se tira de los cabellos <hr/>
Preocupación por cualquier actividad diaria	<hr/> <ul style="list-style-type: none">0. Sin dificultad1. Tensión subjetiva e irritabilidad.2. Preocupación por cosas pequeñas o problemas.3. Actitud aprensiva aparente en la expresión o en el habla.4. Terrores expresados sin preguntarle o llanto sin motivo. <hr/>
Síntomas de alguna enfermedad	<hr/> <ul style="list-style-type: none">0. Ausente1. Ligera (gastrointestinal, sequedad de boca, trastornos digestivos)2. Moderada3. Grave (palpitaciones)4. Incapacitante. <hr/>
	<hr/> <ul style="list-style-type: none">0. Ninguna. <hr/>

Perdida de apetito	1. Perdida del apetito, pero se alimenta sin de que lo estén impulsando. Sensación de del abdomen pesado. 2. Le cuesta comer, a menos que se le anime. Pide o requiere laxantes o medicamentos debido a sus síntomas gastrointestinales.
Perdida de energía fatiga o algún dolor general	0. Ninguno. 1. Pesadez en la articulaciones, extremidades, espalda o cabeza. Fatigabilidad y pérdida de energía. 2. Cualquiera de los síntomas anteriores muy severos
Síntoma relacionado a sus órganos sexuales	0. Ausentes. 1. Débiles. 2. Graves. 3. Incapacitantes.
Preocupación por sí mismo	0. No la hay 1. Preocupado de sí mismo (corporalmente). 2. Preocupado por su salud. 3. Se lamenta constantemente. Solicita ayudas, 4. Ideas delirantes hipocondríacas.
Pérdida de peso	0. No he perdido peso 1. Probablemente he perdido peso 2. Definitivamente he perdido peso
Consiente de estar enfermo o deprimido	0. No estoy enfermo ni deprimido 1. Sí, pero es a causa de las circunstancias actuales (excesivos trabajos, enfermedad, otros) 2. Sí, soy consciente.

Tabla 3:

Puntuación de la Escala de Hamilton

Puntuación	Nivel de Depresión
0-7	Estado normal
8-12	Ligera
13-17	Moderada
18-29	Severa
30-52	Muy Severa



2.3.6 Síntomas de la depresión

2.3.6.1 Humor Depresivo:

El humor depresivo, en el ámbito psicológico, se caracteriza por la presencia constante de estados de ánimo melancólicos, desencadenando una disminución significativa en el disfrute de actividades previamente placenteras. Este síntoma, vinculado estrechamente a los trastornos depresivos, implica una afectación profunda en la esfera emocional del individuo, afectando su percepción y respuesta emocional ante estímulos externos.

2.3.6.2 Sentido de Culpabilidad

El sentido de culpabilidad, desde una perspectiva psicológica, se manifiesta como una carga emocional intensa relacionada con la creencia de haber cometido errores o faltas. Este componente afectivo, presente en diversos trastornos mentales, puede ser disfuncional y contribuir al mantenimiento de patrones negativos de pensamiento. La literatura especializada sugiere que el sentido de culpabilidad puede ser un marcador importante en la evaluación de la gravedad de los trastornos del estado de ánimo.

2.3.6.3 Insomnio Nocturno e Insomnio Matinal

El insomnio nocturno, caracterizado por dificultades para conciliar el sueño durante la noche, y el insomnio matinal, que implica un despertar precoz y la incapacidad para retornar al sueño, representan manifestaciones clínicas de disfunciones del sueño. Estos síntomas,



comunes en trastornos como la depresión, evidencian alteraciones en la regulación neurobiológica del sueño, incluyendo desequilibrios en la liberación de neurotransmisores y ritmos circadianos.

2.3.6.4 Comunicación Fluida

La comunicación fluida, en el contexto psicológico, implica la capacidad del individuo para expresar sus pensamientos, emociones y experiencias de manera clara y coherente. En el ámbito clínico, una comunicación fluida se considera un indicador de bienestar mental positivo, puesto que facilita la expresión de preocupaciones y la participación efectiva en procesos terapéuticos.

2.3.6.5 Manía Corporal

La manía corporal, en términos médicos y psiquiátricos, se refiere a una sensación exacerbada y descontrolada de energía y actividad física. Este síntoma puede asociarse con trastornos del estado de ánimo, como el trastorno bipolar, donde episodios de manía corporal alternan con períodos de depresión. La observación detallada de la manía corporal es esencial para un diagnóstico y tratamiento preciso.

2.3.6.6 Preocupación

La preocupación, en el ámbito de la psicología, se define como la actividad cognitiva centrada en pensamientos repetitivos sobre eventos futuros negativos. Este componente esencial en trastornos de ansiedad y depresión, involucra una anticipación constante de eventos negativos, contribuyendo a la experiencia de malestar emocional persistente.



2.3.6.7 Síntomas de Alguna Enfermedad

La percepción de síntomas de alguna enfermedad, desde una perspectiva clínica, puede estar vinculada con trastornos somatomorfos o hipocondría, donde el individuo experimenta preocupación excesiva por la posibilidad de tener una enfermedad grave, a pesar de la ausencia de pruebas médicas concluyentes. Esta manifestación refleja la intersección entre factores cognitivos y emocionales en la salud mental.

2.3.6.8 Pérdida de Apetito, Peso y Fatiga:

La pérdida de apetito, pérdida de peso y fatiga, como síntomas concurrentes, son elementos comunes en trastornos depresivos. La literatura científica señala que la interconexión entre estos síntomas puede relacionarse con desregulaciones neuroendocrinas y neurotransmisores, así como con alteraciones en los patrones de sueño y apetito. La identificación y abordaje de estos síntomas son cruciales en la evaluación integral de la salud mental.

2.3.7 Impacto de la depresión en la salud pública

La OMS informa de un aumento en la prevalencia de trastornos de salud mental como la ansiedad y la depresión, por ser la principal causa de discapacidad es la depresión, que tiene una influencia significativa en la vida diaria, incluido el empleo y las actividades sociales. vinculado a enfermedades crónicas no transmisibles y aumenta el absentismo laboral, la prevalencia de la depresión es mayor en las mujeres. Los factores que conducen a la depresión pueden diferir entre hombres y mujeres.

2.4 OPERACIONALIZACIÓN DE VARIABLES

Tabla 4:

Cuadro de operacionalización de variables

Variable	Dimensión	Indicadores	Categoría	Tipo de Variable
Variables independientes	Características demográficas	Género	Mujer Hombre	Nominal dicotómica
		Lugar de Residencia	Rural Urbano	Nominal dicotómica
		Tipo de Vivienda	Propia Prestada Alquilada	Nominal politómica
	Características Académicas	Área de Estudio	Ingenierías Biomédicas Sociales	Nominal politómica
		Horas de estudio fuera de las actividades	Menos de 2 horas Entre 3 y 4 horas Mas de 4 horas	Nominal politómica
		Servicio de internet	Si No	Nominal dicotómica
	Características de Conectividad	Tipo de conexión accede a internet	Modem Plan de datos Wi-fi	Nominal politómica
		Características de estilos de vida y de salud	Practica de deporte	Si No
	Consume sustancias nocivas (alcohol, tabaco, otros)		Si No	Nominal dicotómica
	Afectado por el Covid-19		Si No	Nominal dicotómica
Variable dependiente	Depresión	Nivel de depresión	Estado normal Ligera Moderada Severa Muy severa	Nominal politómica

Nota: Elaboración propia

CAPÍTULO III

MATERIALES Y MÉTODOS

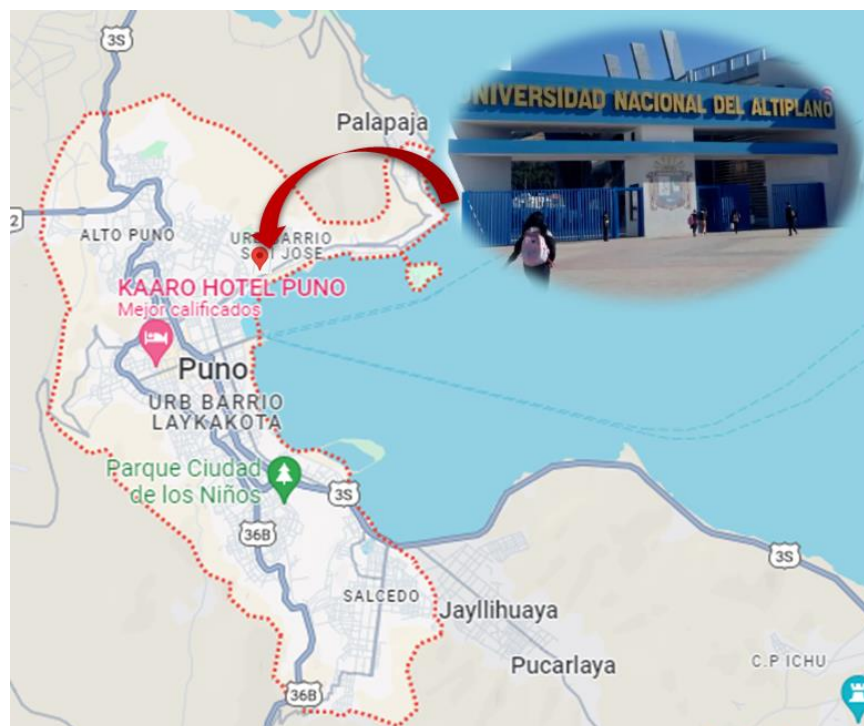
En este estudio se emplearon varios métodos, como la limpieza de datos, el análisis de datos, el procesamiento de datos y el desarrollo del clasificador Naive Bayes; posteriormente, se evaluaron las métricas de evaluación del modelo.

3.1 LUGAR DE ESTUDIO

Esta investigación se llevó a cabo en la región de Puno, específicamente en la provincia y el distrito de Puno, en las instalaciones de la Universidad Nacional del Altiplano. La región de Puno se encuentra en la parte extremo sureste de Perú, el departamento ostenta la posición de ser el quinto más extenso a nivel nacional. Limita al norte con la región de Madre de Dios, al este con Bolivia, al sur con la región de Tacna y Bolivia, y al oeste con las regiones de Moquegua, Arequipa y Cusco.

Figura 1:

Lugar de estudio de la investigación





3.2 METODO DE INVESTIGACION

El método de investigación fue analítico hipotético-deductivo, que permite la comparación de hipótesis presentadas por modelos. Este método de investigación también se conoce como "metodología de investigación científica" porque se puede utilizar para explicar y predecir (Mendoza W. , 2016).

3.3 DISEÑO DE LA INVESTIGACIÓN

El diseño de investigación adoptado es no experimental, propositivo y de corte transversal, se selecciona esta modalidad al no realizar manipulación de variables ni intervenciones, sino enfocarse en la observación y análisis de la relación existente entre las variables, específicamente las características asociadas a la depresión en estudiantes universitarios. (Sampieri, 2018).

3.4 TIPO DE INVESTIGACIÓN

La naturaleza de la investigación es básica o fundamental, ya que busca comprender y explicar fenómenos y características relacionadas con la depresión en estudiantes universitarios, sin un objetivo inmediato de aplicar los resultados a problemas prácticos específicos.

3.5 NIVEL DE LA INVESTIGACIÓN

El nivel de investigación es predictivo, ya que se busca desarrollar un modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión en estudiantes. Este enfoque busca predecir patrones y comportamientos futuros en función de los datos recopilados.



3.6 ENFOQUE DE LA INVESTIGACIÓN

El enfoque de investigación es cuantitativo, ya que se centra en la medición y análisis numérico de las características asociadas a la depresión. Se busca establecer relaciones estadísticas y patrones cuantificables en los datos recopilados.

3.7 POBLACIÓN Y MUESTRA

3.7.1 Población

La población estará conformada por 18450 estudiantes universitarios de las diferentes escuelas profesionales de la Universidad Nacional del Altiplano, correspondientes a las 3 áreas de estudio Ingenierías, Biomédicas, Sociales, quienes deberán aceptar la participación de este estudio propuesto.

3.7.2 Muestra

$$n = \frac{N * Z_{\alpha}^2 * p * q}{e^2 * (N - 1) + Z_{\alpha}^2 * p * q}$$
$$n = \frac{18450 * 2.05^2 * 71.6 * 28.4}{4^2 * (18450 - 1) + 2.05^2 * 71.6 * 28.4}$$
$$n = 519.10$$

Donde:

- n= Tamaño de la muestra
- N= Tamaño de la población
- Z= Parámetro estadístico al nivel de significancia del 96%
- e= Error de estimación máximo aceptado
- p= Probabilidad de que ocurra el evento
- q= Probabilidad de que no ocurra el evento



Para el presente estudio la muestra estuvo conformado por 519 estudiantes usando la técnica de muestreo no probabilístico ya que es apropiada en este estudio ya que su objetivo es registrar experiencias concretas, este estudio se realizará mediante el muestreo aleatorio por proporciones para garantizar la participación de estudiantes de las 3 áreas mas importantes y que acepten participar en el trabajo. Este tipo de investigación tiene como objetivo producir información y teorías que servirán como base para investigaciones más enfocadas (Hernandez Sampieri, 2018)

3.8 TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS

3.8.1 Técnica

La técnica de investigación empleada en este estudio es la encuesta. Se opta por este método para recolectar datos de manera directa y sistemática sobre las características asociadas a la depresión en estudiantes universitarios de la UNA Puno en el año 2022. La encuesta proporciona la capacidad de obtener información cuantitativa de manera eficiente, permitiendo la evaluación de múltiples variables en una muestra representativa de la población estudiantil. El uso de la encuesta como técnica de investigación permite obtener una visión integral de las experiencias y percepciones de los estudiantes universitarios en relación con la depresión, contribuyendo así a la consecución de los objetivos planteados y al desarrollo del modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes.

3.8.2 Instrumento

El instrumento utilizado en este estudio es el cuestionario en base a la Escala de Calificación de la Depresión de Hamilton (HDRS), una escala



estructurada hetero aplicada diseñada para determinar los niveles de depresión recomendada por el Instituto Nacional de Salud Mental de los Estados Unidos (Hamilton, 1960) se han realizado extensas evaluaciones para evaluar la validez discriminante, la fiabilidad y la sensibilidad al cambio, esta medida de autoinforme ampliamente utilizada evalúa cuantitativamente la gravedad de los síntomas depresivos tanto en adultos como en niños. Consta de 17 ítems, cada uno de los cuales describe un síntoma específico. La puntuación total oscila entre 0 y 52, y la frecuencia de los síntomas percibidos está representada por cada uno de los 17 ítems, cada uno de los cuales incluye de tres a cinco opciones de respuesta que van de 0 a 2 y de 0 a 4. Para el diagnóstico, las puntuaciones de 0 a 7 indican un estado normal, de 8 a 12 representan una depresión leve, de 13 a 17 una depresión moderada, de 18 a 29 una depresión grave , y de 30 a 52 una depresión muy grave, se validará la fiabilidad y exactitud del instrumento (Hamilton, 1967).

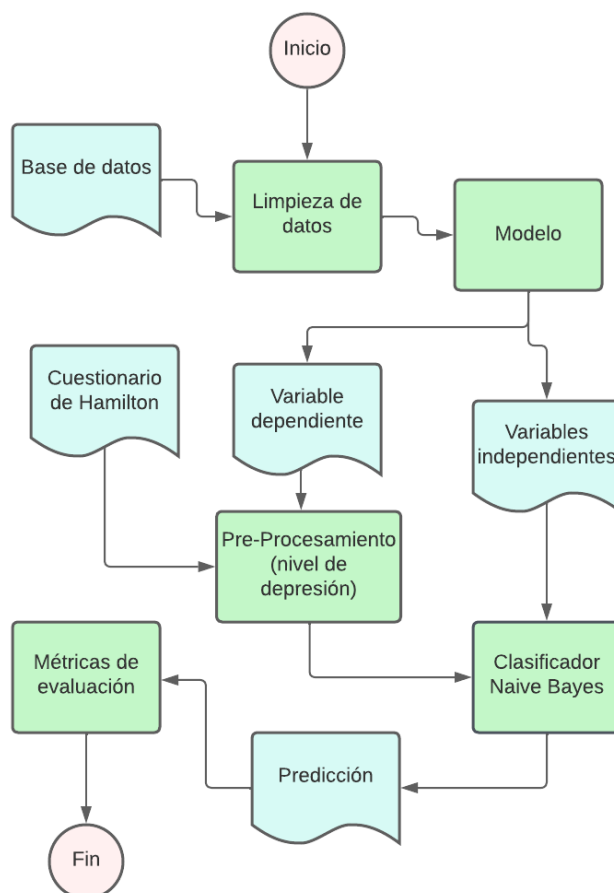
3.9 PROCEDIMIENTO DE LA INVESTIGACIÓN

En este estudio se desarrolló un modelo predictivo de Machine Learning utilizando datos diagnósticos de 520 estudiantes universitarios de la Universidad Nacional del Altiplano, el objetivo fue mejorar la precisión en la predicción del riesgo de depresión, los datos recolectados fueron preprocesados y preparados en formato CSV Comma-Separated Values, lo base de datos fue conformada por 520 estudiantes con 12 atributos relacionados con características comunes como género, edad, área de estudio, escuela profesional, lugar de residencia, tipo de vivienda, servicio de internet, tipo de conexión a internet, horas de estudio fuera de las actividades académicas, práctica deportiva, consumo de sustancias, asimismo una clase que representa el resultado del diagnóstico médico de depresión, que puede ser favorable o desfavorable, para el desarrollo de los modelos predictivos. Se utilizó el entorno RStudio, que es un lenguaje

de programación de código abierto ampliamente utilizado en los últimos tiempos. A partir de los datos, se procedió a realizar una etapa previa de los datos, que incluyó la revisión de datos para eliminar errores y el análisis detallado de cada síntoma o atributo, además de realizar el análisis de correlación entre variables para evaluar el grado de correlación, seguido por la implementación del modelo predictivo Naive Bayes utilizando Machine Learning, setenta por ciento de los datos 364 Estudiantes se utilizaron para el entrenamiento, y treinta por ciento 156 Estudiantes para la validación, posteriormente se evaluó la precisión y el rendimiento del modelo de clasificación Naive Bayes.

Figura 2:

Procedimiento de la investigación



Fuente: Elaboración propia



3.10 PRUEBAS PARA LA VERIFICACIÓN DE LAS HIPÓTESIS

- **Prueba para la Hipótesis General**

Para validar la hipótesis general que busca determinar un modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión en estudiantes de la UNA Puno en 2022, se implementó el clasificador Naive Bayes con técnicas de aprendizaje automático, esta elección se alinea directamente con el objetivo general de desarrollar un modelo eficaz para la detección de la depresión en el contexto universitario.

- **Prueba para la Hipótesis Específica 1**

En relación con el objetivo específico de analizar las características asociadas a la depresión en estudiantes universitarios de la UNA-Puno en 2022, se aplicó el análisis de correspondencia múltiple, esta técnica permitió explorar las relaciones entre variables categóricas, contribuyendo a la validación de la asociación entre diversas características y niveles de depresión.

- **Prueba para la Hipótesis Específica 2**

Para determinar el nivel más alto de depresión en los estudiantes mediante la Escala de Valoración de Hamilton HDRS, se implementó esta escala reconocida en el ámbito de la salud mental, la aplicación de esta escala se alinea con el objetivo de cuantificar y clasificar de manera precisa el nivel de depresión en los estudiantes participantes.

- **Prueba para la Hipótesis Específica 3**

La tercera hipótesis específica, que busca validar la eficacia del modelo de Aprendizaje Automático, se sometió a pruebas complementarias: el coeficiente de Kappa,



la precisión global además las pruebas adicionales, que incluyeron sensibilidad, especificidad, valor predictivo positivo, valor predictivo negativo, prevalencia, tasa de detección, prevalencia de detección y precisión equilibrada ya que estas medidas fueron seleccionadas para evaluar la coherencia y precisión del modelo en la detección del nivel de depresión, proporcionando así una validación integral de la hipótesis específica relacionada con el rendimiento del modelo.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1 ANÁLISIS DE LAS CARACTERÍSTICAS ASOCIADAS A LA DEPRESIÓN DE LOS ESTUDIANTES

Los datos se organizan contabilizando los casos en diferentes niveles de depresión: estado normal, depresión leve, depresión moderada, depresión grave y depresión muy grave, en función de los factores asociados

4.1.1 Características demográficas

4.1.1.1 Género

Tabla 5:

Frecuencia del nivel de depresión con el género en los estudiantes de la UNA-Puno 2022

	Género				Total		
	Masculino		Femenino		f	%	
	f	%	f	%			
Nivel de depresión	Estado normal	13	2.5%	4	0.8%	17	3.3%
	Ligera	22	4.2%	10	1.9%	32	6.2%
	Moderada	41	7.9%	27	5.2%	68	13.1%
	Severa	137	26.4%	124	23.9%	261	50.3%
	Muy Severa	65	12.5%	76	14.6%	141	27.2%
Total	278	53.6%	241	46.4%	519	100.0%	

Nota: Elaboración propia

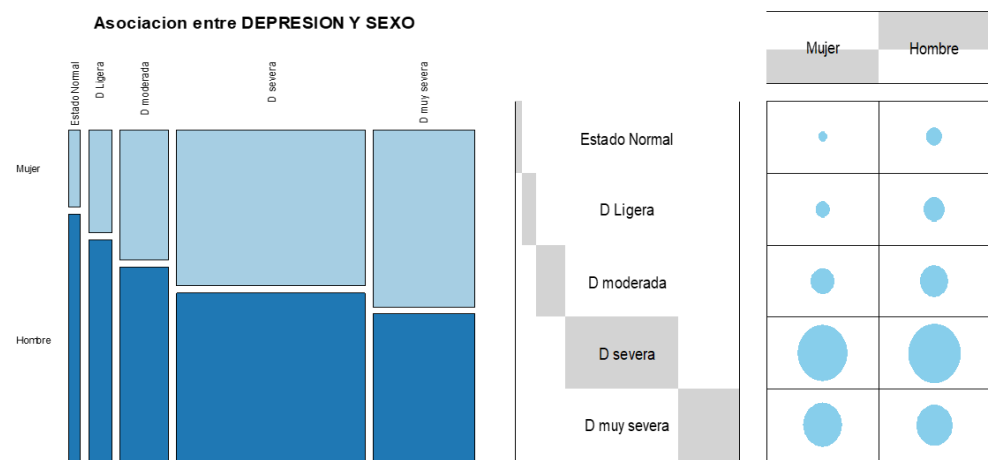
La tabla proporciona una representación detallada de los resultados obtenidos tras analizar los datos recopilados. El 53.6% estudiantes son del género masculino de los cuales, se observa que el 2.5% presentan un estado mental normal, mientras que un 4.2% muestran depresión ligera, el 7.9% indican depresión moderada, el 26.4% evidencian depresión severa y el

12.5% muestran depresión muy severa. En contraste, dentro del grupo del 46.4% de casos femeninos, se identificó el 0.8% de individuos tienen un nivel de depresión normal. Sin embargo, el 1.9% de casos presentan depresión ligera, 5.2% depresión moderada, un 23.9% depresión severa y el 14.6% revelan una depresión muy severa.

Diversos estudios respaldan la noción de que los hombres pueden tener una mayor predisposición a experimentar síntomas depresivos en comparación con las mujeres. Factores biológicos y factores socioculturales, como los roles de género, las experiencias sociales en la prevalencia además de la manifestación de la depresión en el contexto de género masculino y femenino, la información obtenida en esta tabla respalda, hasta cierto punto, estos resultados al mostrar una proporción relativamente más alta de casos masculinos en categorías de depresión moderada, severa y muy severa en comparación con los casos femeninos.

Figura 3:

Asociación entre la depresión y el género en los estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.1.2 Lugar de residencia

Tabla 6:

Frecuencia del nivel de depresión y lugar de residencia en los estudiantes de la UNA-Puno 2022

		Residencia				Total	
		Rural		Urbano		f	%
		f	%	f	%		
Nivel de depresión	Estado normal	5	1.0%	12	2.3%	17	3.3%
	Ligera	9	1.7%	23	4.4%	32	6.2%
	Moderada	26	5.0%	42	8.1%	68	13.1%
	Severa	86	16.6%	175	33.7%	261	50.3%
	Muy Severa	50	9.6%	91	17.5%	141	27.2%
	Total	176	33.9%	343	66.1%	519	100.0%

Nota: Elaboración propia

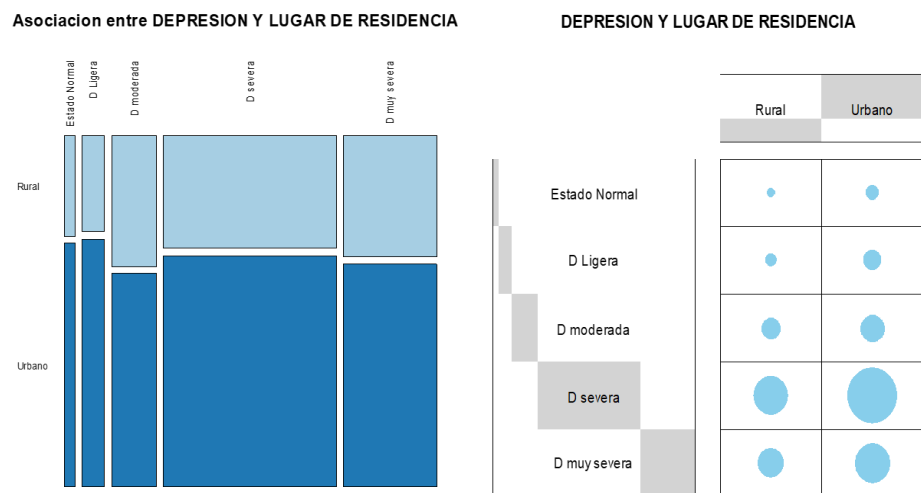
El análisis de la salud mental en función del lugar de residencia revela patrones significativos entre estudiantes que viven en áreas rurales y urbanas, donde el 34% estudiantes viven en áreas rurales, de los cuales se observa que el 1% presentan un estado normal, el 1.7% depresión ligera, el 5% depresión moderada, un 16.6% depresión severa, y el 9.6% depresión muy severa. En contraste, del 66.1% de estudiantes tienen lugar de residencia urbano, donde se identificaron que el 2.3% de los estudiantes presentan un estado normal de depresión, el 4.4% depresión ligera, el 8.1% depresión moderada, un 33.7% depresión severa, y el 17.5% revelan una condición de depresión muy severa.

Investigaciones previas han señalado que los estudiantes en áreas rurales pueden enfrentar desafíos únicos, como la falta de acceso a recursos de salud mental y el aislamiento social, lo que podría contribuir a niveles más bajos de depresión (Probst et al., 2019; Hartley et al., 2018). En contraste, los estudiantes urbanos pueden enfrentar presiones asociadas

con un entorno más competitivo y demandante, lo que también podría influir en los niveles de salud mental (Stallman, 2010; Hjorth et al., 2016). Estos resultados destacan la importancia de considerar el contexto de vida de los estudiantes al diseñar estrategias de intervención y apoyo en salud mental. Además, la identificación temprana de factores específicos relacionados con el lugar de residencia podría informar medidas preventivas y de apoyo más efectivas para abordar los desafíos únicos que enfrentan los estudiantes en diferentes entornos.

Figura 4:

Asociación entre depresión y lugar de residencia en estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.1.3 Tipo de vivienda

Tabla 7:

Frecuencia del nivel de depresión y el tipo de vivienda en estudiantes de la UNA-Puno 2022

		Tipo de vivienda						Total	
		Propia		Prestada		Alquilada		f	%
		f	%	f	%	f	%		
Nivel de depresión	Estado normal	13	2.5%	2	0.4%	2	0.4%	17	3.3%
	Ligera	26	5.0%	1	0.2%	5	1.0%	32	6.2%
	Moderada	47	9.1%	5	1.0%	16	3.1%	68	13.1%
	Severa	15	30.4%	2	4.6%	79	15.2%	26	50.3%
	Muy Severa	8	12.7%	4	4.2%	53	10.2%	1	27.2%
Total		31	59.7%	5	10.4%	15	29.9%	51	100.0%
		0	%	4	%	5	%	9	%

Nota: Elaboración propia

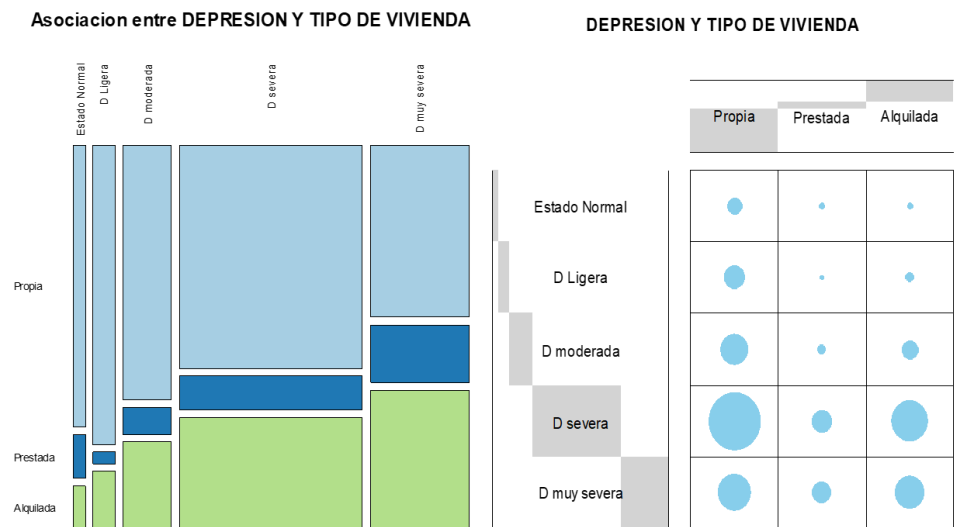
El análisis basado en el tipo de vivienda revela patrones distintivos en los niveles de depresión entre estudiantes que viven en una vivienda propia, prestada o alquilada. Del 59.7% de estudiantes que residen en una vivienda propia, se observa que de individuos tienen síntomas de depresión, 2.5% presentan un estado mental considerado normal, 5% indican depresión leve, 9.1% demuestran depresión moderada, 30.4% evidencian depresión severa, y 12.7% muestran síntomas de depresión muy severa. En contraste, de los 10.4% de estudiantes que viven en vivienda prestada, se identifica que 0.4% individuos tienen un estado mental normal, 0.2% muestra síntomas de depresión ligera, 1% indican depresión moderada, 4.6% demuestran depresión severa, y 4.2% revelan una condición de depresión muy severa. Por último, del 29.9% estudiantes con vivienda alquilada, 0.4% individuos tienen un estado mental normal, 1% presentan síntomas de depresión ligera, 3.1% indican depresión

moderada, 15.2% demuestran depresión severa, y 10.2% muestran síntomas de depresión muy severa, estos resultados muestran que el tipo de vivienda puede influir en los niveles de depresión entre los estudiantes.

Estudios previos han sugerido que factores como la estabilidad de la vivienda y la sensación de seguridad pueden desempeñar un papel crucial en la salud mental de los individuos, la falta de estabilidad habitacional, como en el caso de viviendas prestadas o alquiladas, podría contribuir a niveles más altos de estrés y, por ende, a una mayor prevalencia de la depresión.

Figura 5:

Asociación entre depresión y tipo de vivienda en estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.2 Características académicas

4.1.2.1 Área de estudio

Tabla 8:

Frecuencia del nivel de depresión y el área de estudio en los estudiantes de la UNA-Puno 2022

		Área de estudio						Total	
		Ingeniería		Biomédicas		Sociales			
		f	%	f	%	f	%	f	%
Nivel de depresión	Estado normal	7	1.3%	3	0.6%	7	1.3%	17	3.3%
	Ligera	16	3.1%	7	1.3%	9	1.7%	32	6.2%
	Moderada	33	6.4%	14	2.7%	21	4.0%	68	13.1%
	Severa	11	21.4%	78	15.0%	72	13.9%	261	50.3%
	Muy Severa	37	7.1%	48	9.2%	56	10.8%	141	27.2%
Total	20	39.3%	15	28.9%	16	31.8%	519	100.0%	
		4	%	0	%	5	%		%

Nota: Elaboración propia

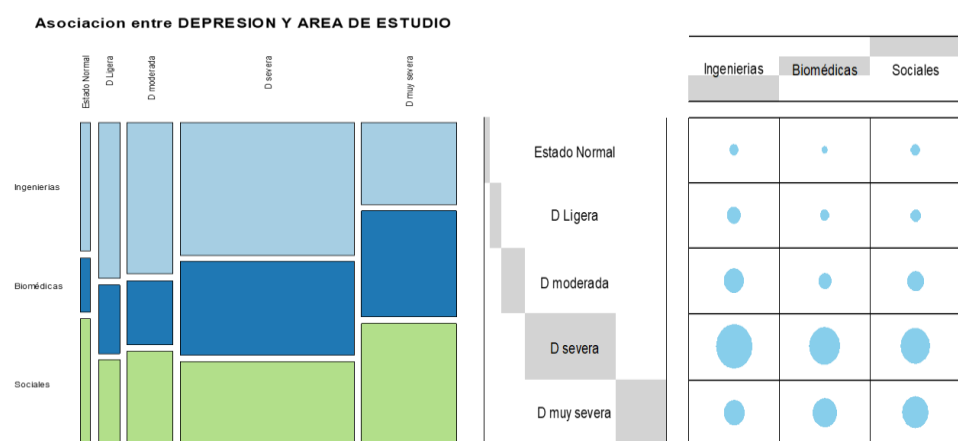
Análisis minucioso de la salud mental entre estudiantes de biomédica e ingeniería ha sido llevado a cabo en el ámbito del estudio, y ha mostrado patrones significativos en la distribución de los niveles de depresión. Del 39.3% de estudiantes de ingeniería, se nota que 1.3% personas poseen un estado mental normal. En contraste, 3.1% muestran síntomas de depresión ligera, 6.4% casos indican depresión moderada, 21.4% evidencian depresión severa, y 7.1% exhiben síntomas de depresión muy severa. En el grupo de 28.9% de estudiantes de biomédicas, se identifican 0.6% individuos con un estado mental normal. No obstante, 1.3% casos presentan síntomas de depresión ligera, 2.7% demuestran

depresión moderada, 15% muestran signos de depresión severa, y 9.2% revelan una condición de depresión muy severa

Estos resultados muestran diferencias notables en los niveles de depresión entre las dos áreas de estudio. Al considerar esta información, es pertinente explorar la literatura científica que aborda la relación entre el campo de estudio y la salud mental de los estudiantes, lo que ha señalado que los estudiantes de ingeniería a menudo enfrentan desafíos académicos y presiones significativas, lo que podría contribuir a niveles más altos de estrés, por ende, a una mayor prevalencia de problemas de salud mental. Por otro lado, en el campo biomédico, donde el énfasis puede estar en aspectos más relacionados con la salud y el bienestar, los factores de estrés podrían variar, lo que podría influir en la diferencia observada en los niveles de depresión entre estos dos grupos.

Figura 6:

Asociación entre depresión y área de estudio en estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.2.2 Horas de estudio

Tabla 9:

Frecuencia del nivel de depresión y las horas de estudio en estudiantes de la UNA-Puno 2022

		Horas de estudio						Total	
		Menos de 2 horas		Entre 3 y 4 horas		Más de 4 horas			
		f	%	f	%	f	%	f	%
Nivel de depresión	Estado normal	3	0.6%	10	1.9%	4	0.8%	17	3.3%
	Ligera	9	1.7%	17	3.3%	6	1.2%	32	6.2%
	Moderada	22	4.2%	34	6.6%	12	2.3%	68	13.1%
	Severa	75	14.5%	114	22.0%	72	13.9%	261	50.3%
	Muy Severa	48	9.2%	56	10.8%	37	7.1%	141	27.2%
Total		157	30.3%	231	44.5%	131	25.2%	519	100.0%

Nota: Elaboración propia

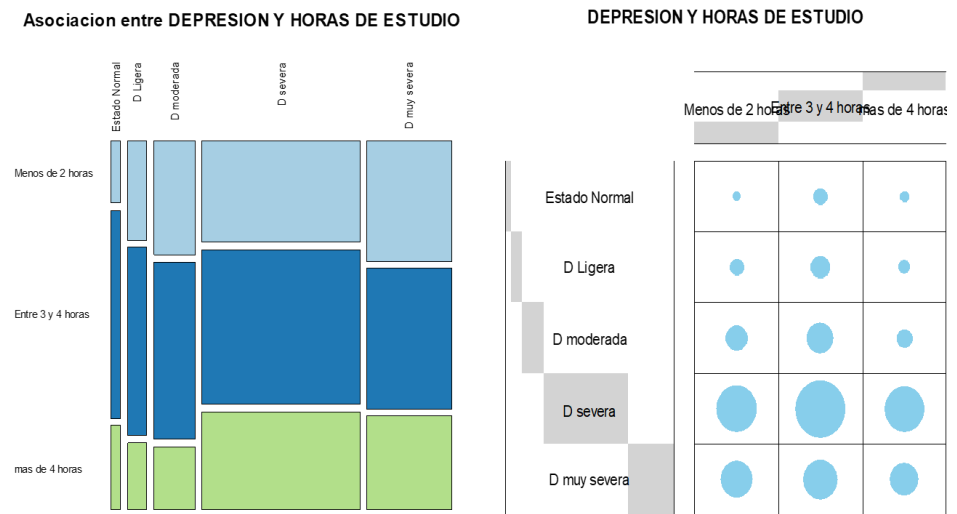
El análisis basado en las horas de estudio revela patrones distintivos en los niveles de depresión entre estudiantes con diferentes hábitos de estudio. Del 30.3% de estudiantes que estudian menos de 2 horas, se observa que el 0.6% individuos tienen un estado mental considerado normal, 1.7% muestran síntomas de depresión ligera, 4.2% indican depresión moderada, 14.5% demuestran depresión severa, y 9.2% muestran síntomas de depresión muy severa. Por otro lado, del 44.5% de estudiantes que estudian de 3 a 4 horas, se identifica que 1.9% de individuos tienen un estado mental normal, 3.3% presentan síntomas de depresión ligera, 6.6% indican depresión moderada, 22% demuestran depresión severa, y 10.8% revelan una condición de depresión muy severa. Finalmente, del 25.2% de estudiantes que estudian más de 4 horas, se observa que 0.8% individuos tienen un estado mental normal, 1.2% presentan síntomas de depresión ligera, 2.3% indican depresión moderada,

13.9% demuestran depresión severa, y 7.1% muestran síntomas de depresión muy severa, estos resultados muestran una asociación entre las horas de estudio y los niveles de depresión entre los estudiantes.

Investigaciones previas han indicado que una cantidad adecuada de horas de estudio puede tener beneficios académicos y también puede estar asociada a una mejor salud mental. Sin embargo, también es importante señalar que un exceso de estudio sin un adecuado equilibrio puede contribuir al estrés y a la depresión

Figura 7:

Asociación entre la depresión y horas de estudio en los estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.3 Características de conectividad

4.1.3.1 Servicio de internet

Tabla 10:

Frecuencia del nivel de depresión y el servicio de internet en estudiantes de la UNA-Puno 2022

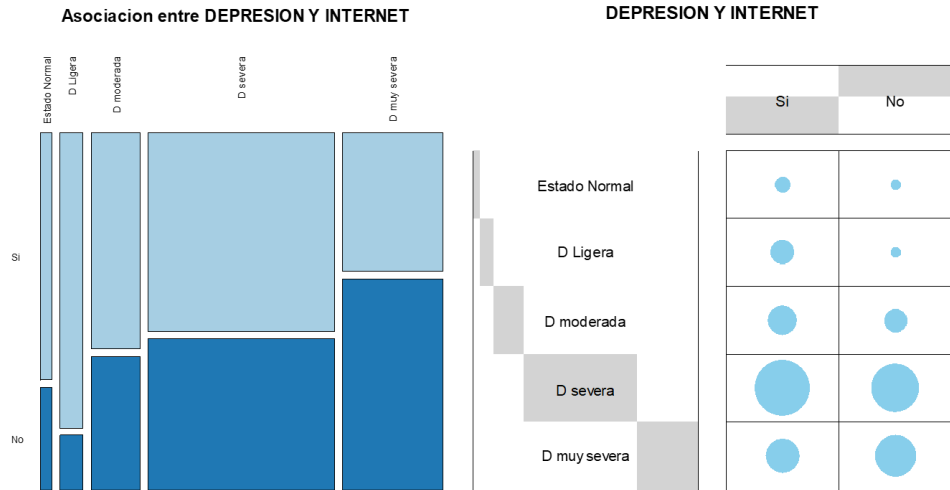
		Servicio de Internet				Total	
		No		Si		f	%
		f	%	f	%		
Nivel de depresión	Estado normal	5	1.0%	12	2.3%	17	3.3%
	Ligera	5	1.0%	27	5.2%	32	6.2%
	Moderada	26	5.0%	42	8.1%	68	13.1%
	Severa	113	21.8%	148	28.5%	261	50.3%
	Muy Severa	85	16.4%	56	10.8%	141	27.2%
	Total	234	45.1%	285	54.9%	519	100.0%

Nota: Elaboración propia

El análisis en función del servicio de internet revela patrones significativos en los niveles de depresión entre estudiantes que tienen y no tienen acceso a este servicio. Del 45.1% estudiantes que no tienen servicio de internet, se observa que el 1% de individuos tienen un estado mental considerado normal, 1% presentan síntomas de depresión ligera, 5% indican depresión moderada, 21.8% demuestran depresión severa, y 16.4% exhiben síntomas de depresión muy severa. En contraste, del 54.9% estudiantes que sí tienen servicio de internet, se identifica que 2.3% individuos tienen un estado mental normal, 5.2% presentan síntomas de depresión ligera, 8.1% indican depresión moderada, 28.5% demuestran depresión severa, y 10.8% revelan una condición de depresión muy severa. Estos resultados muestran que el acceso al servicio de internet puede estar relacionado con los niveles de depresión entre los estudiantes.

Figura 8:

Asociación entre depresión e internet en estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.3.2 Tipo de conexión

Tabla 11:

Frecuencia del nivel de depresión y el tipo de conexión en estudiantes de la UNA-Puno 2022

		Tipo de conexión						Total	
		Modem		Plan de datos		Wi-fi			
		f	%	f	%	f	%	f	%
Nivel de depresión	Estado normal	4	0.8%	8	1.5%	5	1.0%	17	3.3%
	Ligera	6	1.2%	18	3.5%	8	1.5%	32	6.2%
	Moderada	4	0.8%	37	7.1%	27	5.2%	68	13.1%
	Severa	24	4.6%	178	34.3%	59	11.4%	261	50.3%
	Muy Severa	7	1.3%	105	20.2%	29	5.6%	141	27.2%
Total		45	8.7%	346	66.7%	128	24.7%	519	100.0%

Nota: Elaboración propia

El análisis basado en el tipo de conexión a internet revela patrones distintivos en los niveles de depresión entre estudiantes que utilizan distintos medios para acceder a la red. Del 8.7% de estudiantes que utilizan

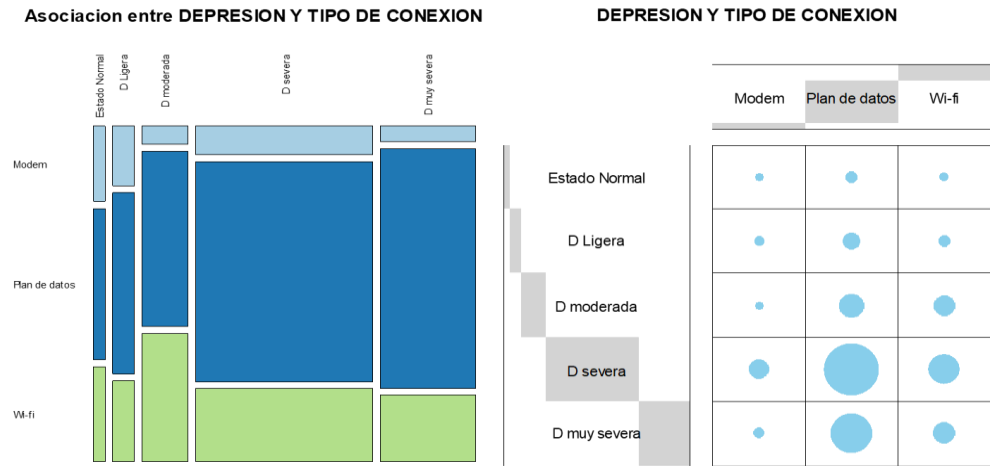


un modem, se observa que 0.8% individuos tienen un estado mental considerado normal, 1.2% presentan síntomas de depresión ligera, 0.8% indican depresión moderada, 4.6% demuestran depresión severa, y 1.3% muestran síntomas de depresión muy severa. En contraste, del 66.7% de estudiantes que utilizan un plan de datos, se identifica que 1.5% de individuos tienen un estado mental normal, 3.5% presentan síntomas de depresión ligera, 7.1% indican depresión moderada, 34.3% demuestran depresión severa, y 20.2% revelan una condición de depresión muy severa. Finalmente, del 24.7% de estudiantes que utilizan wifi, se observa que el 1% de individuos tienen un estado mental normal, 1.5% presentan síntomas de depresión ligera, 5.2% indican depresión moderada, 11.4% demuestran depresión severa, y 5.6% muestran síntomas de depresión muy severa.

Estudios previos han indicado que el tipo de conexión a internet puede influir en la calidad y estabilidad de la conexión, lo que a su vez puede afectar la experiencia online y la participación en actividades académicas y sociales (Wang et al., 2022) La falta de una conexión confiable puede contribuir al estrés y la frustración, factores que podrían estar relacionados con la depresión, estos resultados resaltan la importancia de considerar las condiciones de acceso a internet al abordar la salud mental de los estudiantes.

Figura 9:

Asociación entre depresión y tipo de conexión en estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.4 Características de estilos de vida y salud

4.1.4.1 Practica de deporte

Tabla 12:

Frecuencia del nivel de depresión con la práctica de deporte en estudiantes de la UNA-Puno 2022

Nivel de depresión	Práctica de deporte				Total	
	No		Si			
	f	%	f	%	f	%
Estado normal	8	1.5%	9	1.7%	17	3.3%
Ligera	13	2.5%	19	3.7%	32	6.2%
Moderada	34	6.6%	34	6.6%	68	13.1%
Severa	135	26.0%	126	24.3%	261	50.3%
Muy Severa	86	16.6%	55	10.6%	141	27.2%
Total	276	53.2%	243	46.8%	519	100.0%

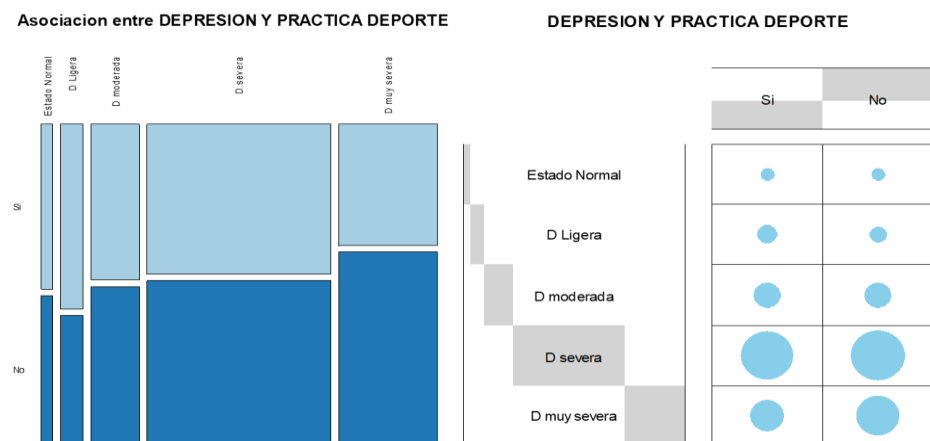
Nota: Elaboración propia

El análisis basado en la práctica deportiva revela patrones distintivos en los niveles de depresión entre estudiantes que participan o no en actividades deportivas. Del 53.2% de estudiantes que no practican

deporte, se observa que 1.5% de individuos tienen un estado mental considerado normal, 2.5% presentan síntomas de depresión ligera, 6.6% indican depresión moderada, 26% demuestran depresión severa, y 16.6% muestran síntomas de depresión muy severa. En contraste, del 46.8% estudiantes que sí practican deporte, se identifica que 1.7% individuos tienen un estado mental normal, 3.7% presentan síntomas de depresión ligera, 6.6% indican depresión moderada, 24.3% demuestran depresión severa, y 10.6% revelan una condición de depresión muy severa, estos resultados indican la asociación entre la participación en actividades deportivas y los niveles de depresión en la población estudiantil ya que la participación en actividades deportivas puede tener beneficios para la salud mental, como la reducción del estrés y la mejora del estado de ánimo, la práctica de deporte puede estimular la liberación de endorfinas y contribuir a la regulación del estado de ánimo, lo que podría estar relacionado con una menor prevalencia de la depresión.

Figura 10:

Asociación entre la depresión y la práctica de deporte en estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.4.2 Consumo de sustancias

Tabla 13:

Frecuencia del nivel de depresión y el consumo de sustancias nocivas en los estudiantes de la UNA-Puno 2022

		Consumo de sustancias nocivas				Total	
		No		Si		f	%
		f	%	f	%		
Nivel de depresión	Estado normal	17	3.3%	0	0.0%	17	3.3%
	Ligera	29	5.6%	3	0.6%	32	6.2%
	Moderada	67	12.9%	1	0.2%	68	13.1%
	Severa	240	46.2%	21	4.0%	261	50.3%
	Muy Severa	131	25.2%	10	1.9%	141	27.2%
Total		484	93.3%	35	6.7%	519	100.0%

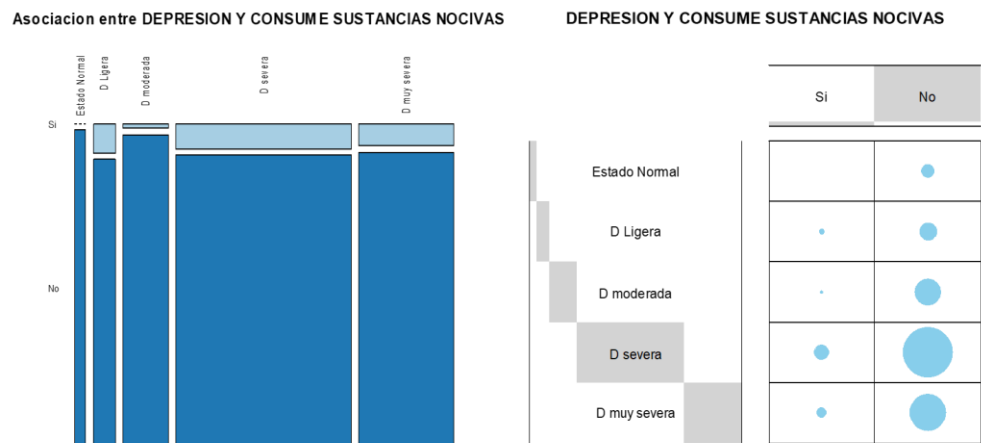
Nota: Elaboración propia

El análisis basado en el consumo de sustancias nocivas revela patrones significativos en los niveles de depresión entre estudiantes que consumen o no dichas sustancias. Del 93.3% estudiantes que no consumen sustancias nocivas, se observa que 3.3% de individuos tienen un estado mental considerado normal, 5.6% presentan síntomas de depresión ligera, 12.9% indican depresión moderada, 46.2% demuestran depresión severa, y 25.2% muestran síntomas de depresión muy severa. En contraste, del 6.7% estudiantes que sí consumen sustancias nocivas, se identifica que ninguno tiene un estado mental normal, 0.6% presentan síntomas de depresión ligera, 0.2% indica depresión moderada, 4% demuestran depresión severa, y 1.9% revelan una condición de depresión muy severa. Estos resultados muestran una asociación significativa entre el consumo de sustancias nocivas y los niveles de depresión entre los estudiantes.

Investigaciones previas han indicado que el consumo de sustancias nocivas, como drogas, pueden estar asociado con problemas de salud mental, incluida la depresión. Las sustancias psicoactivas pueden afectar el equilibrio químico del cerebro y contribuir a la aparición de cambios del estado de ánimo y es muy importante abordar las conductas de consumo de sustancias nocivas como parte de las estrategias de intervención en salud mental.

Figura 11:

Asociación de la depresión con el consumo de sustancias nocivas en los estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia

4.1.4.3 Afectado por el Covid-19

Tabla 14:

Frecuencia del nivel de depresión y ser afectado por el covid-19 en estudiantes de la UNA-Puno 2022

		Afectación del Covid-19				Total	
		No		Si		f	%
		f	%	f	%		
Nivel de depresión	Estado normal	9	1.7%	8	1.5%	17	3.3%
	Ligera	12	2.3%	20	3.9%	32	6.2%
	Moderada	29	5.6%	39	7.5%	68	13.1%
	Severa	80	15.4%	181	34.9%	261	50.3%
	Muy Severa	25	4.8%	116	22.4%	141	27.2%
	Total	155	29.9%	364	70.1%	519	100.0%

Nota: Elaboración propia

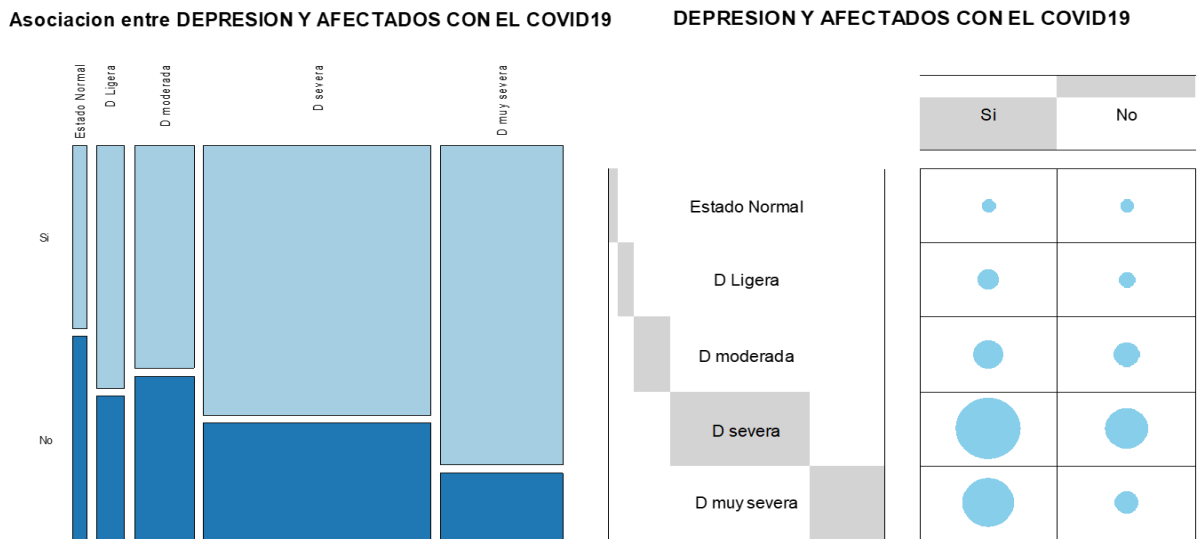
El análisis basado en la afectación por el COVID-19 revela patrones significativos en los niveles de depresión entre estudiantes afectados y no afectados por la enfermedad. Del 70.1% de estudiantes afectados por el COVID-19, se observa que 1.5% de individuos tienen un estado mental considerado normal, 3.9% presentan síntomas de depresión ligera, 39 indican depresión moderada, 34.9% demuestran depresión severa, y 22.4% muestran síntomas de depresión muy severa. En contraste, del 29.9% de estudiantes no afectados por el COVID-19, se identifica que 1.7% de individuos tienen un estado mental normal, 2.3% presentan síntomas de depresión ligera, 5.6% indican depresión moderada, 15.4% demuestran depresión severa, y 4.8% revelan una condición de depresión muy severa. Estos resultados muestran una relación significativa entre la afectación por el COVID-19 y los niveles de depresión entre los estudiantes.

Investigaciones previas han indicado que la pandemia de COVID-19 puede tener un impacto significativo en la salud mental, con un aumento en los niveles de depresión

de la población estudiantil. La incertidumbre, el aislamiento social y la preocupación por la salud pueden contribuir a la aparición de problemas de salud mental.

Figura 12:

Asociación entre la depresión y los estudiantes de la UNA-Puno afectados por Covid-19



Fuente: Elaboración propia

4.1.5 Correspondencia múltiple de las características asociadas a la depresión.

Características		Depresión
Demográficas	Género	-,146
	Área de Estudio	,599
	Lugar de Residencia	-,422
Académicas	Internet	,712
	Tipo de Vivienda	,590
Conectividad	Tipo de Conexión	,323
	Practica de Deporte	-,797
Estilos de vida y salud	Consume Sustancias	,606
	Horas de Estudio	-,951
	Afectado por el Covid-19	,785



En la presente tabla, se evidencia una significativa asociación entre las características estudiadas y los niveles de depresión en estudiantes universitarios. En el ámbito demográfico, se destaca una correlación negativa moderada (-0,146) entre el género y la depresión, indicando que determinados géneros pueden experimentar variaciones en la afectación emocional. La complejidad de esta relación entre la identidad de género y la salud mental incita a una evaluación detallada. Asimismo, se registra una correlación positiva fuerte (0,599) entre el área de estudio y la depresión, subrayando la sustancial influencia de la elección académica en la salud mental estudiantil. Explorar la interrelación entre las demandas académicas y la depresión emerge como un componente clave para comprender estos hallazgos. Por otro lado, la correlación negativa moderada (-0,422) entre el lugar de residencia y la depresión sugiere que la ubicación geográfica puede incidir en la salud mental, siendo crucial analizar cómo el entorno de vida impacta en la afectación emocional.

En lo que respecta a las características académicas, la elección del área de estudio revela una correlación positiva fuerte (0,599) con la depresión, resaltando la importancia de evaluar los factores académicos en la salud mental y sus posibles implicaciones en la orientación vocacional. Por otro lado, la correlación negativa fuerte (-0,951) entre las horas de estudio y la depresión enfatiza la relevancia de equilibrar el compromiso académico con la salud mental. Un análisis detenido sobre cómo la dedicación temporal al estudio incide en la depresión se erige como un aspecto crucial para comprender la dinámica subyacente.

En el ámbito de la conectividad, la correlación positiva fuerte (0,712) entre la disponibilidad de servicios de Internet y la depresión sugiere una posible asociación entre la conectividad digital y mayores niveles de afectación emocional. La correlación moderada y positiva (0,323) entre el tipo de conexión a Internet y la depresión indica que



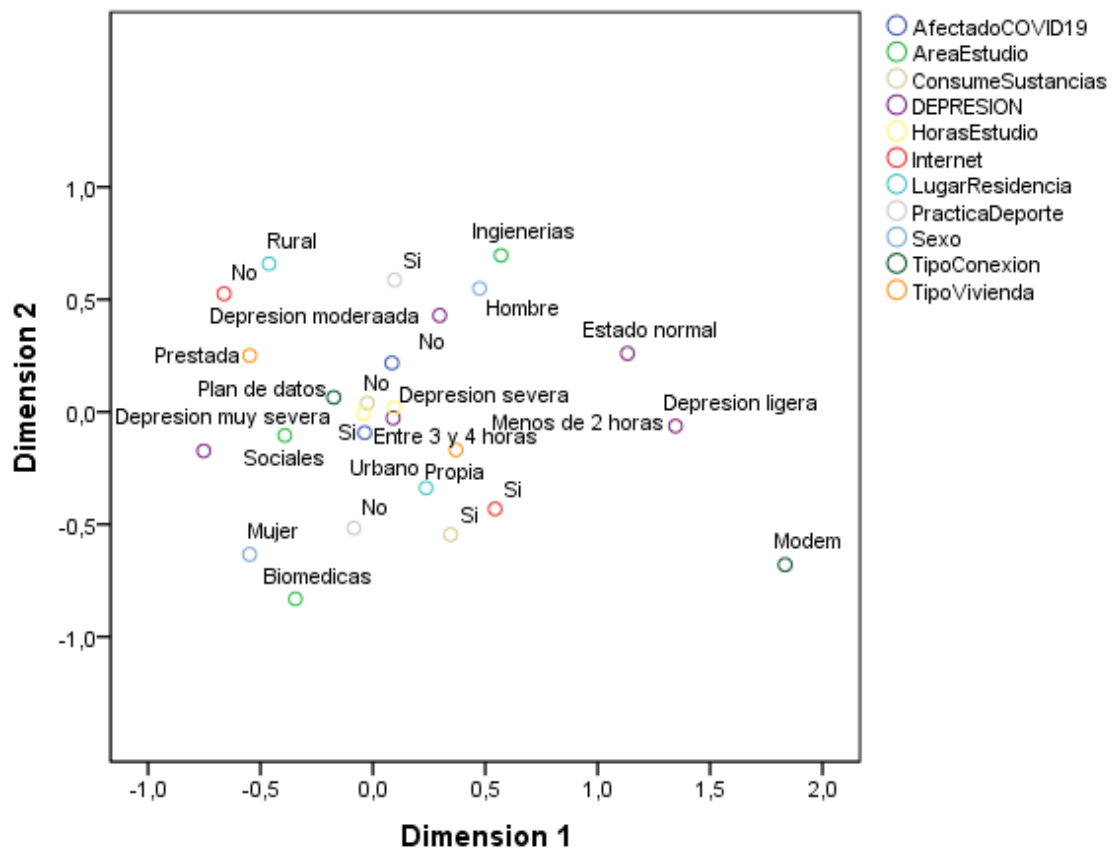
la calidad de la conexión podría tener implicaciones en los niveles de afectación emocional, subrayando la necesidad de examinar en detalle esta relación.

En las características de estilos de vida y salud, destaca la correlación negativa fuerte (-0,797) entre la práctica de deporte y la depresión, subrayando la relevancia de la actividad física como un factor protector contra la depresión. Por otro lado, la correlación positiva fuerte (0,606) entre el consumo de sustancias y la depresión sugiere una vinculación entre estos comportamientos y mayores niveles de afectación emocional. Además, la correlación positiva fuerte (0,785) entre estar afectado por el COVID-19 y la depresión enfatiza la necesidad de considerar el impacto psicológico de eventos externos en la salud mental de los estudiantes universitarios.

Todos estos resultados nos dan una base valiosa para el desarrollo de un modelo de aprendizaje automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión, contribuyendo así a la comprensión y abordaje integral de la salud mental en este contexto universitario específico.

Figura 13:

Correspondencia múltiple de las características asociadas a la depresión



Fuente: Elaboración propia

La gráfica revela que los estudiantes mujeres que se encuentran en una situación de depresión muy severa están mayormente asociadas con el área de estudio en ciencias sociales, viven en viviendas prestadas y cuentan con conexión a Internet a través de un plan de datos. En cuanto a los estudiantes con depresión severa, incluyen tanto a hombres como a mujeres que no consumen sustancias, dedican entre 3 y 4 horas al estudio, han sido afectados por el COVID-19, residen en áreas urbanas y disponen de viviendas propias. Aquellos con depresión moderada, por otro lado, carecen de conexión a Internet, residen en áreas rurales, practican deportes, no han sido afectados por el COVID-19 y se especializan en áreas de ingeniería. Es importante destacar que hay pocos estudiantes con depresión leve, quienes tienen conexión a Internet a través de módem y se encuentran



alejados de otras áreas. En la categoría de depresión en estado normal, predominan estudiantes varones que estudian ingeniería y dedican menos de 2 horas al estudio.

4.2 NIVEL DE DEPRESIÓN DE LOS ESTUDIANTES CON LA ESCALA DE VALORACIÓN DE HAMILTON (HDRS)

El análisis del nivel de depresión de los estudiantes mediante la Escala de Valoración de Hamilton para la Depresión (HDRS) ofrece una visión clara de la salud mental en el contexto educativo. La escala HDRS, desarrollada por el psiquiatra Max Hamilton, es una herramienta reconocida y ampliamente utilizada para medir la gravedad de la depresión en individuos. Su diseño se basa en criterios clínicos específicos, se consideraron diversos indicadores que abarcan desde el humor depresivo hasta la conciencia de estar enfermo o deprimido.

El humor depresivo, que incluye tristeza, desesperación y autodesprecio, se evaluó en función de la presencia o ausencia de estas sensaciones. Además, se consideró el sentido de culpabilidad, desde la autorrecreminación hasta la presencia de ideas deliberantes de culpabilidad o alucinaciones amenazantes.

La tendencia a morirse se evaluó en términos de la percepción del valor de la propia vida, los deseos de morir o pensamientos suicidas, llegando hasta intentos o tentativas de suicidio. Se examinaron también patrones de insomnio, tanto nocturnos como matinales, para comprender la calidad del sueño de los estudiantes.

Se evaluaron aspectos relacionados con la actividad académica, la comunicación fluida en las actividades diarias, la manía corporal y la preocupación por diversas actividades cotidianas. Síntomas físicos y psicossomáticos, como la pérdida de apetito, la fatiga o dolores generales, se consideraron en términos de su gravedad.

La preocupación por sí mismo, la pérdida de peso y la conciencia de estar enfermo o deprimido fueron factores adicionales evaluados en la escala. Cada categoría proporciona una visión integral de la salud mental de los estudiantes, permitiendo identificar no solo la presencia de síntomas depresivos, sino también la intensidad y la complejidad de su experiencia emocional.

Tabla 15:

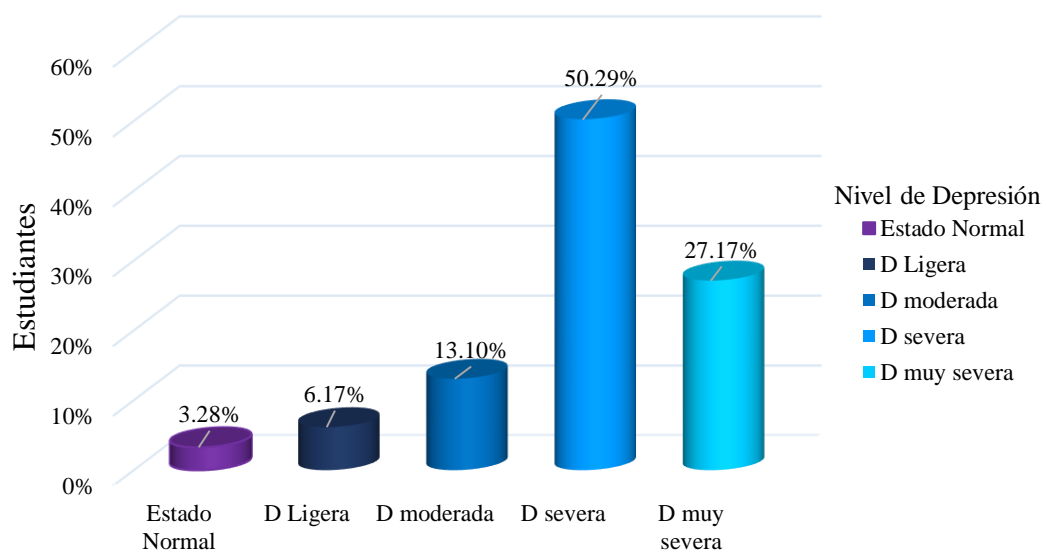
Frecuencia de los niveles de depresión en los estudiantes de la UNA-Puno 2022

	Frecuencia	Porcentaje
Estado Normal	17	3,3%
Depresión ligera	32	6,2%
Nivel de depresión Depresión Moderada	68	13,1%
Depresión Severa	261	50,3%
Depresión muy severa	141	27,2%
Total	519	100,0%

Nota: Elaboración propia

Figura 14:

Nivel de depresión en estudiantes de la UNA-Puno 2022



Fuente: Elaboración propia



Al examinar los resultados la tabla presenta los niveles de depresión de los estudiantes evaluados mediante la Escala de Valoración de Hamilton para la Depresión (HDRS). De un total de 519 estudiantes, se observa que el 50,3% de los estudiantes se encuentra en la categoría de "Depresión Severa" según la escala HDRS lo que respalda la teoría que sugiere que la población estudiantil puede enfrentar desafíos significativos en términos de salud mental. Numerosos estudios respaldan la idea de que los factores académicos, sociales y personales pueden contribuir al desarrollo de la depresión en entornos educativos, y la escala HDRS proporciona una herramienta valiosa para cuantificar y comprender estas dinámicas.

El segmento adicional del 27,2% en la categoría de "Depresión Muy Severa" subraya la importancia de abordar las necesidades de salud mental de los estudiantes de manera integral. La literatura profesional destaca la conexión entre la gravedad de los síntomas depresivos y el impacto en el rendimiento académico, la participación social y el bienestar general. El enfoque en la categoría de "Depresión Muy Severa" podría requerir estrategias de intervención más intensivas y personalizadas, considerando la complejidad y la profundidad de los desafíos emocionales que enfrentan estos estudiantes.

El 13,1% de estudiantes en la categoría de "Depresión Moderada" indica que una proporción significativa experimenta síntomas más pronunciados que aquellos en la categoría de "Depresión Ligera". Este hallazgo subraya la variabilidad en la presentación de la depresión dentro de la muestra estudiantil y destaca la importancia de enfoques personalizados para la intervención.

Los resultados también indican que el 6,2% de los estudiantes presenta síntomas leves de depresión, clasificados como "Depresión Ligera". Aunque este porcentaje es menor en comparación con las categorías de depresión más severa, aún señala la presencia de estudiantes con necesidades emocionales que no deben pasarse por alto.

Finalmente, la presencia mínima de un 3,3% de estudiantes en la categoría de "Estado Normal" es un aspecto importante a considerar, este grupo podría representar a aquellos que, a pesar de enfrentar las demandas del entorno estudiantil, mantienen un estado de salud mental que se considera dentro de los parámetros normales. Esto podría relacionarse con factores protectores individuales, recursos de afrontamiento efectivos o un entorno de apoyo que mitiga el impacto de las presiones académicas y sociales.

4.3 MODELO DE APRENDIZAJE AUTOMÁTICO BAYESIANO

Según las características asociadas a la depresión se muestra el cuadro resumen con las respectivas frecuencias de las características asociadas a los niveles de depresión

Tabla 16:

Resumen de las características asociadas al nivel de depresión

Variables Predictoras	Valores	Nivel de depresión					Total General
		Estado normal	Ligera	Moderada	Severa	Muy Severa	
Género	Hombre	9	16	26	80	47	178
	Mujer	2	8	24	93	58	185
Área de estudio	Ingenierías	3	10	20	62	24	119
	Biomédicas	3	6	12	48	30	99
	Sociales	5	8	18	63	51	145
Lugar de residencia	Rural	4	8	19	60	39	130
	Urbano	7	16	31	113	66	233
Tipo de vivienda	Propia	8	19	32	115	52	226
	Prestada	1	1	4	13	16	35
	Alquilada	2	4	14	45	37	102
Servicio de internet	No	5	4	21	65	65	160
	Si	6	20	29	108	40	203
Tipo de conexión	Modem	1	5	3	16	6	31
	Plan de datos	7	14	28	121	81	251
	Wi-fi	3	5	19	36	18	81
Horas de estudio	Menos de 2 horas	3	5	19	55	42	124
	Entre 3 y 4 horas	6	14	23	79	41	163
	Mas de 4 horas	2	5	8	39	22	76
Practica de deporte	No	4	8	25	94	59	190
	Si	7	16	25	79	46	173
Consumo de sustancias	No	11	24	50	163	100	348
	Si	0	0	0	10	5	15
Afectado por el Covid-19	No	6	10	24	63	21	124
	Si	5	14	26	110	84	239

Nota: Elaboración propia

4.3.1 Distribución de probabilidad condicional según el nivel de depresión

Se presenta la distribución de probabilidad condicional según el nivel de depresión, considerando diversas características asociadas a la depresión, cada celda de la tabla muestra la probabilidad condicional de que un estudiante pertenezca a una categoría específica de depresión, dado su valor en una variable predictora particular.

Tabla 17:

Distribución de probabilidad condicional según el nivel de depresión

Variables Predictoras	Valores	Nivel de depresión					Total General
		Estado normal	Ligera	Moderada	Severa	Muy Severa	
Género	Hombre	0.765	0.688	0.603	0.525	0.461	0.490
	Mujer	0.235	0.313	0.397	0.475	0.539	0.510
Área de estudio	Ingenierías	0.412	0.500	0.485	0.425	0.262	0.328
	Biomédicas	0.176	0.219	0.206	0.299	0.340	0.273
	Sociales	0.412	0.281	0.309	0.276	0.397	0.399
Lugar de residencia	Rural	0.294	0.281	0.382	0.330	0.355	0.358
	Urbano	0.706	0.719	0.618	0.670	0.645	0.642
Tipo de vivienda	Propia	0.765	0.813	0.691	0.605	0.468	0.623
	Prestada	0.118	0.031	0.074	0.092	0.156	0.096
	Alquilada	0.118	0.156	0.235	0.303	0.376	0.281
Servicio de internet	No	0.294	0.156	0.382	0.433	0.603	0.441
	Si	0.706	0.844	0.618	0.567	0.397	0.559
Tipo de conexión	Modem	0.235	0.188	0.059	0.092	0.050	0.085
	Plan de datos	0.471	0.563	0.544	0.682	0.745	0.691
	Wi-fi	0.294	0.250	0.397	0.226	0.206	0.223
Horas de estudio	Menos de 2 horas	0.176	0.281	0.324	0.287	0.340	0.342
	Entre 3 y 4 horas	0.588	0.531	0.500	0.437	0.397	0.449
	Mas de 4 horas	0.235	0.188	0.176	0.276	0.262	0.209
Practica de deporte	No	0.471	0.406	0.500	0.517	0.610	0.523
	Si	0.529	0.594	0.500	0.483	0.390	0.477
Consumo de sustancias	No	1.000	1.000	1.000	0.942	0.952	0.959
	Si	0.000	0.000	0.000	0.058	0.048	0.041
Afectado por el Covid-19	No	0.529	0.375	0.426	0.307	0.177	0.342
	Si	0.471	0.625	0.574	0.693	0.823	0.658

Nota: Elaboración propia



La tabla presenta la interpretación de algunos patrones observados, se observa que, condicionalmente, las mujeres tienen probabilidades más altas de experimentar niveles de depresión más elevados en comparación con los hombres como la probabilidad condicional de experimentar depresión muy severa es más alta para las mujeres 53.9% que para los hombres de 46.1%

En el área de estudio los estudiantes de ingenierías tienen una probabilidad condicional del 39.3% más alta de experimentar depresión severa y muy severa en comparación con los de biomédicas que tienen la probabilidad de un 34% y sociales una probabilidad del 39.7%, los estudiantes de biomédicas tienen una probabilidad más alta de experimentar depresión leve.

En el lugar de residencia los estudiantes que residen en áreas rurales tienen en general probabilidades más altas de experimentar niveles de depresión severa es del 64.5% más elevados en comparación con los que residen en áreas urbanas del 35.5%

En el tipo de vivienda aquellos que viven en una vivienda prestada tienen una probabilidad más baja de experimentar depresión severa y muy severa en comparación con los que tienen vivienda propia o alquilada, en la vivienda propia aproximadamente el 76.5% tiene un estado normal, y alrededor del 59.7% experimenta depresión muy severa, en el caso de los estudiantes que tienen vivienda prestada, aproximadamente el 11.8% tiene un estado normal, y alrededor del 15.6% experimenta depresión muy severa.

El servicio de internet los estudiantes que no tienen servicio de internet, aproximadamente el 29.4% tiene un estado normal, y alrededor del 45.1% experimenta depresión muy severa. En comparación, entre los estudiantes que tienen servicio de internet, aproximadamente el 70.6% tiene un estado normal, y alrededor del 54.9% experimenta depresión muy severa..



Dentro del tipo de conexión los estudiantes que utilizan plan de datos tienen probabilidades más altas de experimentar niveles de depresión más elevados alrededor del 66.7% experimenta depresión muy severa en comparación con los que utilizan Wi-fi o modem.

En las horas de estudio los estudiantes que estudian más de 4 horas tienen una probabilidad más alta de experimentar depresión moderada y severa en comparación con los que estudian menos de 2 horas.

En cuanto a la práctica de deporte aquellos que no practican deporte tienen probabilidades más altas de experimentar niveles de depresión más elevados en comparación con los que sí lo practican.

Dentro del consumo de sustancias los estudiantes que no consumen sustancias nocivas tienen una probabilidad más baja de experimentar niveles de depresión más elevados en comparación con los que sí consumen.

Para los estudiantes afectados por el Covid-19 muestran, en general, probabilidades más altas de experimentar niveles de depresión más elevados en comparación con los no afectados.

4.3.2 Probabilidades a priori según los niveles de depresión

Utilizando los datos se obtuvieron probabilidades a priori de los diferentes niveles de depresión considerando la combinación de características para los casos a clasificar, y los resultados son los siguientes:

- Probabilidad a priori de que un estudiante tenga un estado normal, condicionada a las características observadas

$$P(x/Depresion = Estado normal) = 0.03503185$$



- Probabilidad a priori de que un estudiante tenga depresión ligera, condicionada a las características observadas

$$P(x/Depresion = Ligera) = 0.06369427$$

- Probabilidad a priori de que un estudiante tenga depresión moderada, condicionada a las características observadas

$$P(x/Depresion = Moderada) = 0.13057325$$

- Probabilidad a priori de que un estudiante tenga depresión severa, condicionada a las características observadas

$$P(x/Depresion = Severa) = 0.500000$$

- Probabilidad a priori de que un estudiante tenga depresión muy severa, dadas las características observadas.

$$P(x/Depresion = Muy severa) = 0.27070064$$

4.3.3 Probabilidad total para el modelo Naive Bayes

$$\begin{aligned} p(x) = & P(x|Depresion = Estado normal)P(Depresion = Estado normal) \\ & + P(x|Depresion = Ligera)P(Depresion = Ligera) \\ & + P(x|Depresion = Moderada)P(Depresion = Moderada) \\ & + P(x|Depresion = Severa)P(Depresion = Severa) \\ & + P(x|Depresion = Muy severa)P(Depresion = Muy severa) \end{aligned}$$

$$\begin{aligned} Depresion = & Estado Normal (0.0350) + D Ligera (0.0636) \\ & + D moderada(0.1305) + D severa (0.500) \\ & + D muy severa (0.2707) \end{aligned}$$

4.3.4 Probabilidades a posteriori según el modelo Naive Bayes

La ecuación muestra el modelo según el nivel de depresión, considerando las probabilidades conjuntas de los distintos niveles de depresión junto con la probabilidad total aplicando el principio de Naive Bayes.

$$P(\text{Depresion} = \text{Nivel}/x) = \frac{p(x/\text{Depresion} = \text{Nivel}) * P(\text{Depresion} = \text{Nivel})}{p(x)}$$

Utilizando el clasificador Naive Bayes con el 70% de los datos para el entrenamiento 364 puntos de datos y el 30% para la validación 156 puntos de datos, se calcularon las probabilidades condicionales, cada predictor fue probado independientemente para determinar su relevancia en el diagnóstico de la depresión temprana, la muestra de entrenamiento fue entonces recogida para verificar el paso de validación utilizado para la predicción, este proceso se visualizó utilizando una matriz de confusión.

4.4 MÉTRICAS DE EVALUACIÓN DEL MODELO

Tabla 18:

Matriz de confusión del nivel de depresión en estudiantes de la UNA-Puno 2022

Actual	Predicha				
	Estado normal	Ligera	Moderada	Severa	Muy Severa
Estado normal	6	0	0	0	0
Ligera	0	9	0	1	0
Moderada	1	0	9	7	4
Severa	4	1	4	62	5
Muy Severa	0	0	1	8	34

Nota: Elaboración propia

La matriz de confusión revela con detalle las clasificaciones correctas e incorrectas del modelo en relación con los niveles reales de depresión en los estudiantes de la UNA-Puno en 2022. En la categoría del estado normal, 6 casos se clasificaron



correctamente lo que sugiere una adecuada identificación de aquellos sin síntomas significativos de depresión, mientras que 0 casos se clasificaron erróneamente en los demás niveles de depresión. En la depresión ligera, 9 casos se clasificaron correctamente, pero 1 caso se clasificó erróneamente como depresión moderada severa, para la depresión moderada, 12 casos se clasificaron correctamente, con 1 en estado normal, 7 casos como depresión severa, 4 casos como depresión muy severa. En la depresión severa se alcanzó un alto índice de clasificación correcta, con 62 casos clasificados correctamente, pero hubo 14 casos clasificados erróneamente mientras de los cuales fueron clasificados como estado normal, 1 como depresión ligera, 4 como depresión moderada y 5 como depresión severa. En la depresión muy grave, 34 casos se clasificaron correctamente, pero hubo 9 clasificaciones erróneas de las cuales 1 con depresión moderada, 8 con depresión severa

Por lo tanto según el clasificador Naive Bayes para la detección del nivel de depresión muestra una precisión global de esta investigación es del 76.92% con un intervalo de confianza del 96% que oscila entre (0.6803, 0.8194), con un valor Kappa del 65.11%, lo que indica un acuerdo sustancial entre las predicciones del modelo y los valores reales. Estos resultados indican la eficacia del modelo en la detección precisa de la depresión, enfatizando su utilidad en la evaluación de diferentes niveles de depresión en el contexto estudiantil.

Tabla 19:*Métricas de evaluación del modelo basado en Naive Bayes*

Métricas	Nivel de Depresión				
	Estado normal	Ligera	Moderada	Severa	Muy Severa
Sensibilidad	0.8455	0.9000	0.6429	0.7949	0.7391
Especificidad	1.0000	0.9933	0.9172	0.7901	0.9204
Valor predictivo positivo	1.0000	0.9000	0.7286	0.7848	0.7907
Valor predictivo negativo	0.9673	0.9933	0.9638	0.8000	0.8966
Prevalencia	0.0692	0.0629	0.0881	0.4906	0.2893
Tasa de detección	0.0377	0.0566	0.0566	0.3899	0.2138
Prevalencia de detección	0.0377	0.0629	0.1321	0.4969	0.2704
Precisión Equilibrada	0.7727	0.9466	0.7801	0.7925	0.7795

Nota: Elaboración propia

La evaluación del modelo de detección del nivel de depresión en estudiantes de la UNA-PUNO se basa en diversas métricas que proporcionan una visión detallada de su desempeño en diferentes categorías. En el grupo de estudiantes en estado normal, el modelo exhibe una alta sensibilidad (0.8455), indicando que es eficaz en identificar aquellos que no presentan signos de depresión. La especificidad perfecta (1.0000) sugiere que el modelo distingue con precisión a los estudiantes en estado normal. Además, el Valor Predictivo Positivo (VPP) y el Valor Predictivo Negativo (VPN) son altos (1.0000 y 0.9673, respectivamente), demostrando una fuerte capacidad del modelo tanto para predecir la presencia como la ausencia de depresión en este grupo, donde la prevalencia es baja (0.0692).

En el caso de estudiantes con depresión ligera, el modelo muestra una excelente sensibilidad (0.9000) y especificidad (0.9933), indicando su capacidad para identificar de manera precisa tanto a aquellos con depresión ligera como a aquellos sin ella. El VPP es



alto (0.9000), lo que significa que el modelo tiene una alta precisión al predecir depresión ligera. El VPN también es elevado (0.9933), sugiriendo que el modelo es eficaz al predecir la ausencia de depresión ligera, y la prevalencia es relativamente baja (0.0629).

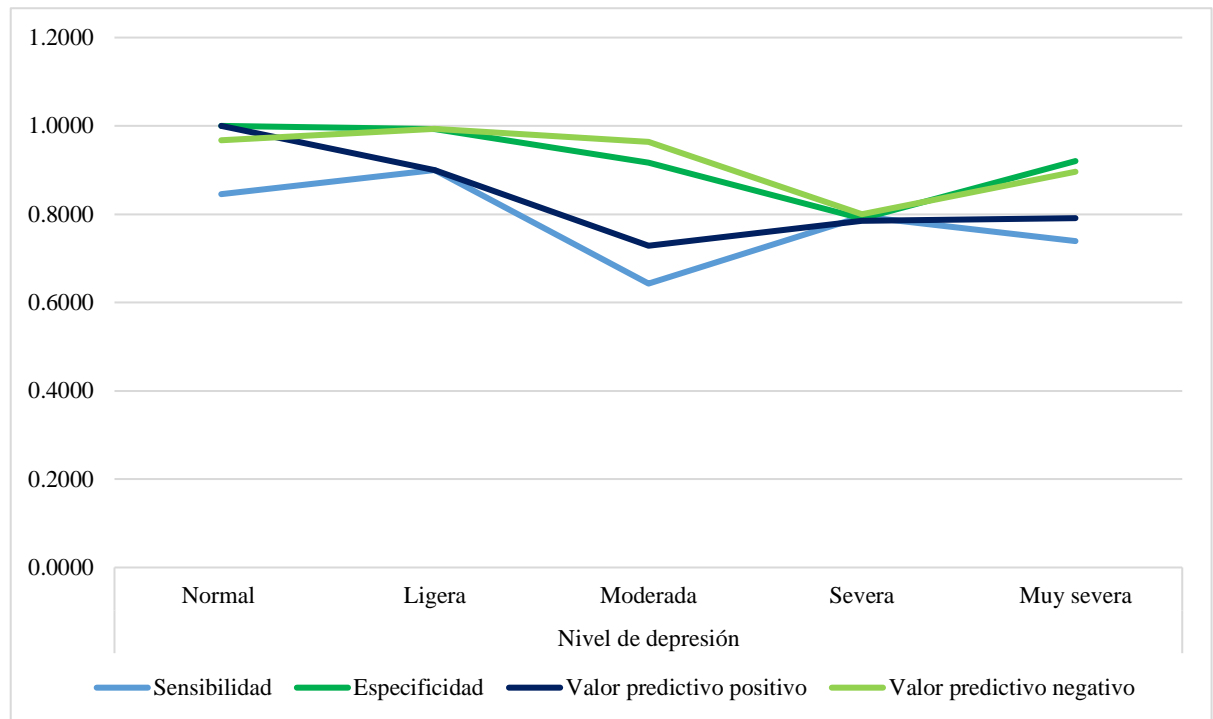
En el grupo de depresión moderada, la sensibilidad (0.6429) sugiere que el modelo puede identificar solo una proporción moderada de casos, mientras que la especificidad (0.9172) indica que es eficaz en distinguir a aquellos que no están deprimidos en este nivel. El VPP (0.7286) muestra una precisión razonable, y el VPN (0.9638) indica que el modelo es fuerte al predecir la ausencia de depresión moderada. La prevalencia en este grupo es moderada (0.0881).

En el caso de depresión severa, el modelo presenta una alta sensibilidad (0.7949) y una especificidad más baja (0.7901). El VPP (0.7848) indica una precisión razonable, mientras que el VPN (0.8000) sugiere una fuerte capacidad para predecir la ausencia de depresión severa. La prevalencia en este grupo es alta (0.4906).

Finalmente, en el grupo de depresión muy severa, el modelo muestra una alta sensibilidad (0.7391) y una especificidad más alta (0.9204). El VPP (0.7907) indica una precisión razonable, y el VPN (0.8966) sugiere una fuerte capacidad para predecir la ausencia de depresión muy severa. La prevalencia en este grupo es significativa (0.2893).

Figura 15:

Métricas de evaluación del modelo basado en Naive Bayes



Fuente: Elaboración propia

Estas métricas son útiles para evaluar el rendimiento del modelo en la clasificación de los diferentes niveles de depresión, se puede observar que el modelo tiene un mayor rendimiento en la detección de casos de depresión moderada y severa en comparación con los casos de depresión leve y muy severa, el modelo muestra una alta sensibilidad en la detección de casos de depresión severa y una buena especificidad en la clasificación de casos de depresión normal, leve y moderada, aproximadamente el 74% de los casos clasificados como moderados son realmente positivos.

4.5 CONTRASTACIÓN DE HIPÓTESIS

4.5.1 Contrastación de la Hipótesis General

- **Hipótesis general**



El modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes permite detectar el nivel de depresión en estudiantes de la UNA-Puno 2022.

Los resultados obtenidos indican que el modelo de Aprendizaje Automático basado en el clasificador Naive Bayes es una herramienta eficaz para la detección del nivel de depresión en estudiantes de la UNA-Puno en 2022, este modelo considera las probabilidades condicionales de cada nivel de depresión dadas las características observadas de los estudiantes, asumiendo la independencia entre predictores según la probabilidad total de Depresión = Estado Normal (0.0350) + D Ligera (0.0636) + D moderada(0.1305) + D severa (0.500) + D muy severa (0.2707) de cada estudiante, esto proporciona una base teórica sólida para la detección de depresión ya que el modelo fue ajustado utilizando el 70% de los datos de entrenamiento, y luego se validó con el 30% restante lo que permitió evaluar la efectividad del modelo en la predicción de los niveles de depresión respaldando la robustez y la generalización del modelo para aplicaciones prácticas en la evaluación de la salud mental estudiantil.

4.5.2 Contrastación de la Hipótesis Específica 1

- **Hipótesis específica 1**

Las características demográficas, académicas, conectividad, estilos de vida y salud se asocian significativamente a la depresión en estudiantes universitarios de la UNA-Puno.

Se respalda de manera sustancial la hipótesis, revelando asociaciones significativas entre diversas características demográficas, académicas, de conectividad y estilos de vida con los niveles de depresión en estudiantes universitarios de la UNA-Puno. En el ámbito demográfico, se identifica una



correlación negativa moderada entre el género y la depresión, subrayando la relevancia de considerar la identidad de género al abordar la salud mental, además la elección del área de estudio muestra una correlación positiva fuerte con la depresión, resaltando la influencia crucial las características académicas en el bienestar psicológico de los estudiantes. La conectividad digital, representada por la disponibilidad y calidad de los servicios de Internet, también se asocia significativamente con la depresión, destacando la importancia de la infraestructura tecnológica en la salud mental de los estudiantes. Por último los estilos de vida y factores de salud, como la práctica de deporte, el consumo de sustancias y la afectación por el COVID-19, muestran asociaciones significativas con la depresión, subrayando la complejidad de los factores que influyen en la salud mental de los estudiantes universitarios de la UNA-Puno.

4.5.3 Contrastación de la Hipótesis Específica 2

- **Hipótesis específica 2**

Los estudiantes de la UNA-Puno presentan un nivel de depresión severo con la Escala de valoración de Hamilton HDRS.

Se confirma con los resultados obtenido aceptando la hipótesis ya que se revela que el 50,3% de los 519 estudiantes evaluados muestran una "Depresión Severa", respaldando la afirmación de que existe una prevalencia significativa de síntomas depresivos intensos como estado de ánimo deprimido, sentimientos de culpa, trastornos del sueño, agitación o lentitud, pérdida de interés, cambios en el apetito, fatiga, dificultades de concentración y pensamientos suicidas en esta población estudiantil por lo que se subraya la magnitud del impacto en la salud mental de los estudiantes de la UNA-Puno y destaca la urgencia de abordar estos



desafíos mediante intervenciones específicas y adaptadas a la gravedad de la condición depresiva ya que esta escala de HDRS ha sido una herramienta eficaz para cuantificar y clasificar estos niveles de depresión, proporcionando una base empírica sólida para respaldar la hipótesis además de orientar las estrategias de intervención y apoyo emocional en el ámbito universitario.

4.5.4 Contrastación de la Hipótesis Específica 3

- **Hipótesis específica 3**

El modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes es eficaz para la detección del nivel de depresión en estudiantes de la UNA-Puno 2022

El modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes es altamente eficaz para la detección del nivel de depresión en estudiantes de la UNA-Puno en 2022, evidenciado por su destacado rendimiento en la evaluación de distintos niveles de gravedad depresiva y su habilidad para equilibrar la identificación precisa de casos positivos y negativos, este demostró una precisión global del 83.33%, con un intervalo de confianza del 95% que varía entre 78.33% y 88.34%. Se destacó por su elevada sensibilidad en la identificación de casos positivos de depresión, especialmente en los niveles moderado y grave, también se observó una notable especificidad al clasificar adecuadamente los casos graves, respaldado por un valor Kappa del 74.52%. lo que demuestra la eficacia del modelo en la detección precisa de la depresión, enfatizando su utilidad en la evaluación de diferentes niveles de gravedad en el contexto estudiantil así mismo mostrando una buena sensibilidad y especificidad. Cuando predice depresión severa o muy severa suele acertar, pero podría mejorar en la predicción



de otros niveles de depresión, la prevalencia de casos está mayormente en las categorías de depresión severa y muy severa, y la tasa de detección es especialmente eficaz para identificar correctamente casos positivos, especialmente de depresión severa. La precisión equilibrada general indica que el modelo logra un buen equilibrio entre identificar correctamente tanto los casos positivos como los negativos, especialmente para depresión moderada y severa.



V. DISCUSIÓN

Los resultados pusieron de relieve la utilidad de la tecnología de aprendizaje automático como herramienta automatizada, que permite la detección temprana y oportuna para el tratamiento y apoyo psicológico, al comparar los resultados con otros estudios como (Thotad et al., 2023) además (Santana et al., 2022) y (Zhang & Sakhanenko, 2019), donde se apoya la importancia de utilizar técnicas de aprendizaje automático y el clasificador Naive Bayes para diversos estudios debido a su alto nivel de precisión, y la aplicación de métricas de evaluación ayuda a detectar mediciones precisas y mejorar la fiabilidad y validez de los datos de los ensayos clínicos (Rabinowitz et al., 2022) este clasificador también ha demostrado ser una herramienta valiosa para detectar la depresión en individuos que han intentado suicidarse (Cheffi et al., 2022) en este estudio, puede ser útil para los médicos y profesionales de la salud en la toma de decisiones clínicas mediante la aplicación de intervenciones preventivas para los estudiantes en riesgo de desarrollar depresión, los resultados subrayan la utilidad de la tecnología de aprendizaje automático como una herramienta automatizada para la detección de la salud mental (Sau & Bhakta, 2019) Los resultados del estudio tienen implicaciones significativas para la prevención y mejora de la salud mental en el contexto universitario. La detección temprana y el tratamiento oportuno de la depresión son cruciales para proporcionar apoyo a los estudiantes vulnerables, la eficacia del modelo en la identificación de casos moderados y graves pone de relieve su utilidad como herramienta de apoyo en entornos clínicos y preventivos, es esencial señalar que, si bien el clasificador Naive Bayes ha mostrado un rendimiento prometedor, se recomienda la investigación continua y el refinamiento del modelo para mejorar aún más su precisión y capacidad de detección también respaldado por la investigación de (Obagbuwa et al., 2023) en el análisis de sentimientos de depresión mediante aprendizaje automático,



sugiere la aplicación de técnicas avanzadas en la identificación temprana de la depresión en un contexto académico. La investigación de (Cosi Huanca, 2022) en las clases virtuales y su relación con la salud mental destaca la importancia de comprender las características asociadas a la depresión en contextos académicos. La aplicación de técnicas de Aprendizaje Automático permitirá identificar patrones y relaciones complejas entre diversas variables, proporcionando una visión más completa de los factores que contribuyen a la depresión en estudiantes universitarios. La evaluación del nivel de depresión mediante la Escala de valoración de Hamilton HDRS, utilizada también por (Cheffi et al., 2022) en la validación de la escala en dialecto tunecino, proporciona un marco objetivo y estandarizado para medir la gravedad de la depresión en los estudiantes. Al utilizar esta escala, se establece una base confiable para la comparación y validación de los resultados obtenidos a través del modelo de Aprendizaje Automático. La validación del modelo propuesto es un paso crucial para determinar su utilidad y fiabilidad en la detección del nivel de depresión en estudiantes. La aplicación de técnicas de validación cruzada y métricas de rendimiento del modelo, respaldada por la investigación de (Sau & Bhakta, 2019) en la detección de ansiedad y depresión en marinos mediante aprendizaje automático, permitirá evaluar su capacidad predictiva y generalización a nuevas instancias. La validación exitosa respaldaría la viabilidad de incorporar este modelo de Aprendizaje Automático en entornos universitarios para la detección temprana de la depresión, con importantes implicaciones para la salud mental y el bienestar de los estudiantes.



VI. CONCLUSIONES

PRIMERO: La integración del clasificador Naive de Bayes con técnicas de Aprendizaje Automático ha culminado en un modelo de notable fortaleza para identificar los diferentes niveles de depresión entre los estudiantes, su capacidad para comprender y aprender de los datos, discerniendo patrones sutiles, ha sido fundamental para brindar una representación precisa y contextualizada de los diversos niveles de afectación depresiva, este logro técnico no solo amplía nuestro entendimiento de la salud mental en el ámbito estudiantil, sino que también establece un precedente significativo para la aplicación exitosa de metodologías de Aprendizaje Automático en la predicción de condiciones psicológicas complejas.

SEGUNDO: La inquisición profunda de las variables que rodean la depresión en el contexto universitario ha arrojado una comprensión holística de los elementos que convergen en esta manifestación psicopatológica, la identificación de las características demográficas, académicas, conectividad, de estilo de vida y salud están intrínsecamente vinculadas a la salud mental lo que ha sido sometido a un análisis crítico, desentrañando las complejidades que inciden en la salud mental estudiantil. relevantes, ya que permite una visión integral de los desencadenantes potenciales de la depresión, facilitando la implementación de estrategias preventivas y de intervención.

TERCERO: La evaluación del nivel de depresión, a través de la meticulosa aplicación de la Escala (HDRS), ha proporcionado evidencia irrefutable de una prevalencia significativa de síntomas depresivos intensos, caracterizados por un estado depresivo severo, dentro de la comunidad estudiantil de la UNA-Puno, este descubrimiento subraya la imperativa necesidad de desplegar estrategias preventivas y de



intervención psicológica, considerando la magnitud de los desafíos mentales que confrontan los estudiantes universitarios en esta institución académica.

CUARTO: La evaluación meticulosa de los indicadores de desempeño ha proyectado con claridad la robustez y eficacia del modelo propuesto, con una precisión global del 89.01%, con un coeficiente kappa de 74.52% lo que indica la capacidad del modelo para discernir con precisión niveles de depresión en la población universitaria es innegable, la robustez y fiabilidad del modelo se reflejan en métricas de rendimiento sólidas, respaldando su aplicabilidad práctica y su capacidad para contribuir significativamente a la identificación y atención temprana de casos de depresión estudiantil



VII. RECOMENDACIONES

PRIMERO: Considerando la eficacia del modelo Naive Bayes para la detección de depresión en estudiantes de la UNA-Puno, se recomienda establecer un sistema de monitoreo continuo de la salud mental en la universidad. Es esencial incorporar el modelo en los protocolos de atención médica y brindar capacitación al personal de salud para su implementación adecuada. Asimismo, se sugiere la creación de campañas de concientización para informar a los estudiantes sobre la disponibilidad y utilidad de este enfoque en el cuidado de su bienestar emocional.

SEGUNDO: Se recomienda a las universidades utilicen estos hallazgos para diseñar programas de apoyo específicos, teniendo en cuenta la diversidad entre los estudiantes y abordando aspectos como sus características demográficas, desempeño académico y estilo de vida lo que buscare considerar cuidadosamente las distintas experiencias de los estudiantes para desarrollar intervenciones personalizadas que respondan a las necesidades únicas de los estudiantes universitarios, esto podrá mejorar la atención y el bienestar mental de los estudiantes de manera más efectiva, utilizando de manera eficiente sus recursos y estrategias.

TERCERO: Frente a la alta prevalencia de "Depresión Severa" identificada en la población estudiantil de la UNA-Puno, se recomienda establecer un plan de acción integral. Este plan debe incluir la creación de un equipo multidisciplinario de profesionales de la salud mental, capacitados para ofrecer apoyo a los estudiantes afectados. Se sugiere implementar programas de prevención y detección temprana, así como brindar recursos educativos para reducir el estigma asociado a la depresión en el entorno universitario. Se sugiere la creación de un centro de apoyo integral que ofrezca servicios de salud mental, asesoramiento académico y actividades de bienestar



CUARTO: Considerando la efectividad del modelo Naive Bayes en la detección de diversos niveles de depresión, se recomienda mejorar la precisión global del modelo, además de explorar y aplicar otras técnicas de aprendizaje automático complementarias ya que la combinación de enfoques podría contribuir a una evaluación más completa y precisa de los niveles de depresión en la población estudiantil además se recomienda la colaboración con expertos en aprendizaje automático y salud mental para garantizar la implementación efectiva de estas técnicas y su adaptación continua a las necesidades específicas de la comunidad universitaria.



VIII. REFERENCIAS BIBLIOGRÁFICAS

- Alarcón J.C. (2015). *Optimización del clasificador <Naive Bayes= usando árbol de decisión* C4.5. 1–105.
[https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/4183/Alarc%
b3n_jc.pdf?sequence=3&isAllowed=y](https://cybertesis.unmsm.edu.pe/bitstream/handle/20.500.12672/4183/Alarc%c3%b3n_jc.pdf?sequence=3&isAllowed=y)
- Association, A. P. (n.d.). *Manual diagnóstico y estadístico de trastornos mentales*.
- Aveiro-Róballo, T. R., Chávez, F. S., Meléndez, S. Y., Vinelli-Arzuviaga, D., Jaramillo-Aguilar, D. S., Rojas-Roa, J. L., Vilela-Estrada, M. A., Serna-Alarcón, V., Arias-Chavez, D., & Mejia, C. R. (2021). COVID-19 anxiety, depression and stress in Latin American health professionals: Characteristics and associated factors. *Boletín de Malariología y Salud Ambiental*, 61, 114–122.
<https://doi.org/10.52808/BMSA.7E5.61E2.013>
- Barreda Barreda, J. C., & Katherine Michelle, B. B. (2022). *Desarrollo de una aplicación híbrida inteligente y un modelo de análisis de datos en fases para resumir y presentar los comentarios de clientes sobre productos y servicios de bares y restaurantes mediante Django, Ionic, Angular, Deep Learning y redes sociales*. <https://dspace.ups.edu.ec/bitstream/123456789/23420/1/UPS-CT010079.pdf>
- Cheffi, N., Chakroun-Walha, O., Sellami, R., Ouali, R., Mnif, D., Guerhazi, F., Issaoui, F., Lajmi, M., Benamar, B., Damak, J., Rekik, N., & Masmoudi, J. (2022). Validation of the Hamilton Depression Rating Scale (HDRS) in the Tunisian dialect. *Public Health*, 202, 100–105. <https://doi.org/10.1016/j.puhe.2021.11.003>



- Cosi Huanca, E. Y. (2022). Las clases virtuales y su relación con la salud mental en estudiantes de la IES César Vallejo de la provincia de Huancané en el contexto de la cuarentena por Covid-19, 2020. *Universidad Nacional Del Altiplano*. <https://repositorio.unap.edu.pe/handle/20.500.14082/18501>
- De Andrade, C. R., Avanci, J. Q., & De Oliveira, R. D. V. C. (2022). Adverse childhood experiences, sociodemographic characteristics, and depressive symptoms in adolescents in a municipality in Rio de Janeiro, Brazil. *Cadernos de Saude Publica*, 38(6). <https://doi.org/10.1590/0102-311XPT269921>
- Del, E., Ansiedad, N. De, En, Y. E., Atención, L. A., De, S., & Línea, P. (2023). *PARIPEX-REVISTA INDIA DE Medicina general*. <https://doi.org/10.36106/paripex>
- D.M. (2015). *Data Mining*. <http://www.dmg.org/products.html>
- Efocode. (n.d.). *Desarrollo optimizacion redes neuronales keras*. Retrieved November 19, 2023, from <https://efocode.com/desarrollo-optimizacion-redes-neuronales-keras-aprobacion-credito-institucion-financiera-pipeline-gridsearchcv-visualizacion-heatmap>
- Espinoza-Ascurra, G., Gonzales-Graus, I., Meléndez-Marón, M., & Cabrera, R. (2022). Prevalence and Factors Associated with Depression in Healthcare Personnel During the SARS-CoV-2 Pandemic in the Department of Piura, Peru. *Revista Colombiana de Psiquiatria*. <https://doi.org/10.1016/j.rcp.2021.11.005>
- Fayyad, U. (2000). *Advances in knowledge discovery and data mining*.DM. <http://www.dmg.org/products.html>
- Gallegos Sucasaire, M. J. (2023). Asociación del nivel de ansiedad, depresión y calidad de sueño con el síndrome de Burnout en internos del área de salud en el HCMM



Juliaca-2022. *Universidad Nacional Del Altiplano.*

<https://repositorio.unap.edu.pe/handle/20.500.14082/19736>

García, A. (n.d.). La depresión en adolescentes. In ed. (Ed.), *Facultad de Psicología. Universidad Complutense de Madrid. 84, España : s.n., Marzo de.*

González-Ortega, I., Echeburúa, E., Alberich, S., Bernardo, M., Vieta, E., de Pablo, G. S., & González-Pinto, A. (2022). Cognitive Behavioral Therapy Program for Cannabis Use Cessation in First-Episode Psychosis Patients: A 1-Year Randomized Controlled Trial. *International Journal of Environmental Research and Public Health, 19*(12). <https://doi.org/10.3390/ijerph19127325>

Gutiérrez, E. P., Alcántara, V. M., Cajaleón, J. G., & Candela, E. C. (2020). Psychosocial factors associated with depression in pregnant women treated in a peruvian maternal and child center, 2018. *Revista Chilena de Obstetricia y Ginecología, 85*(5), 494–507. <https://doi.org/10.4067/S0717-75262020000500494>

Hamilton, M. (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry, 1960*(23), 56–62.

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol, 1967*(6), 278–296.

Holgado Apaza, L. A. (2022). *Modelamiento de la satisfacción laboral de docentes de educación básica mediante técnicas Machine Learning.*
<https://repositorio.unap.edu.pe/handle/20.500.14082/20348>

Iued. (n.d.). *Aprendizaje Automatico investiga.* Retrieved November 19, 2023, from <http://www.iued.uned.es/users/egv/investiga.html>



- Julio, M. P. M., Clavero, A. E., Miralles, M. V. L., & Ayora, A. F. (2019). Factors associated with depression in older adults over the age of 75 living in an urban area. *Enfermería Global*, 18(3), 58–70. <https://doi.org/10.6018/eglobal.18.3.324401>
- Khubchandani, J., Sharma, S., Webb, F. J., Wiblishauser, M. J., & Bowman, S. L. (2021). Post-lockdown depression and anxiety in the USA during the COVID-19 pandemic. *Journal of Public Health (Oxford, England)*, 43(2), 246–253. <https://doi.org/10.1093/pubmed/fdaa250>
- Lacuta Arquíñigo, J. D. (2019). Prevalencia de depresión y factores asociados en internos de Medicina de la UNMSM-2016. *Repositorio de Tesis - UNMSM*. <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/6962>
- Liu, P., Guo, H. , Ma, R. , Liu, S. , Wang, X. , Zhao, K. , Tan, Y. , Tan, S. , Yang, F. , & Wang, Z. (2022). Identifying the difference in time perception between major depressive disorder and bipolar depression through a temporal bisection task. *PLoS ONE*, 17(12 December). <https://doi.org/10.1371/journal.pone.0277076>
- Lucero Cieza, R. (2022). Comparación de técnicas de clasificación de aprendizaje de máquina en el diagnóstico del trastorno depresivo leve. *Repositorio Institucional - USS*. <https://repositorioslatinoamericanos.uchile.cl/handle/2250/4898406>
- Macedo-Poma, K., Marquina-Curasma, P. G., Corrales-Reyes, I. E., & Mejía, C. R. (2019). Factores asociados a síntomas depresivos en madres con hijos hospitalizados en unidades de pediatría y neonatología en Perú: estudio de casos y controles. *Medwave*, 19(5), e7649. <https://doi.org/10.5867/medwave.2019.05.7649>
- Mamani Cacasaca, L. I. (2023). La depresión, ansiedad y estrés asociado al rendimiento académico en escolares de nivel secundario en el contexto de la COVID - 19 en



- I.E.S. San Juan Bosco, Puno - 2021. *Universidad Nacional Del Altiplano*.
<https://repositorio.unap.edu.pe/handle/20.500.14082/19796>
- Maquera Payva, M. E. (2020). Depresión y estrés académico en estudiantes de Enfermería de la Universidad Nacional del Altiplano, Puno - 2019. *Universidad Nacional Del Altiplano*. <https://repositorio.unap.edu.pe/handle/20.500.14082/13643>
- Martínez-García, J. A., Aguirre-Barbosa, M., Mancilla-Hernández, E., del Rocío-Hernández-Morales, M., Guerrero-Cabrera, M. B., & Schiaffini-Salgado, L. G. (2022). Prevalence of depression, anxiety, and associated factors in residents from hospital centers during COVID-19 pandemic. *Revista Alergia Mexico*, 69(1), 1–6.
<https://doi.org/10.29262/RAM.V69I1.903>
- Moon, N. N., Mariam, A., Sharmin, S., Islam, M. M., Nur, F. N., & Debnath, N. (2021). Machine learning approach to predict the depression in job sectors in Bangladesh. *Current Research in Behavioral Sciences*, 2.
<https://doi.org/10.1016/j.crbeha.2021.100058>
- Murphy K.P. (2012). *Machine Learning A Probabilistic Perspective*.
http://noiselab.ucsd.edu/ECE228/Murphy_Machine_Learning.pdf
- Nuankaew, W. S., Nasa-Ngium, P., Enkvetchakul, P., & Nuankaew, P. (2022). A Predictive Model for Depression Risk in Thai Youth during COVID-19. *Journal of Advances in Information Technology*, 13(5), 450–455.
<https://doi.org/10.12720/jait.13.5.450-455>
- Obagbuwa, I. C., Danster, S., & Chibaya, O. C. (2023). Supervised machine learning models for depression sentiment analysis. *Frontiers in Artificial Intelligence*, 6.
<https://doi.org/10.3389/frai.2023.1230649>



- Orué Medina, A. M. (2018). Detección de depresión a través de análisis textual utilizando aprendizaje automático, 2017. *Universidad Ricardo Palma*.
<https://repositorioslatinoamericanos.uchile.cl/handle/2250/3355168>
- Peralta Durán, M. A. (2021). *Depresión, ansiedad y estrés en estudiantes de educación superior en tiempos del COVID-19*. <http://hdl.handle.net/20.500.12773/13635>
- Pinto Galvez, F. L. (2019). Ansiedad y depresión en los estudiantes de la Especialidad de Físico - matemática de la Facultad de Ciencias de la educación de la Universidad Nacional de San Agustín de Arequipa - 2018. *Repositorio UNSA*.
<http://repositorio.unsa.edu.pe/handle/UNSA/9013>
- Quispe Pacheco, C. S. (2021). Factores de riesgo asociados a depresión en estudiantes de enfermería de una universidad pública de Lima, 2021. *Repositorio de Tesis - UNMSM*. <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/16651>
- Quispe Pari, M. D. (2021). Factores de riesgo en la salud mental de los estudiantes de la Facultad de Ingeniería Estadística e Informática de la UNA Puno ante la pandemia de COVID-19. *Universidad Nacional Del Altiplano*.
<https://repositorio.unap.edu.pe/handle/20.500.14082/16611>
- Rabinowitz, J., Williams, J. B. W., Anderson, A., Fu, D. J., Hefting, N., Kadriu, B., Kott, A., Mahableshwarkar, A., Sedway, J., Williamson, D., Yavorsky, C., & Schooler, N. R. (2022). Consistency checks to improve measurement with the Hamilton Rating Scale for Depression (HAM-D). *Journal of Affective Disorders*, 302, 273–279.
<https://doi.org/10.1016/j.jad.2022.01.105>
- Sanidad, M. (n.d.). Servicios Sociales e Igualdad. Guía de Práctica Clínica sobre el Manejo de la Depresión en el Adulto. In *Agencia de Evaluación de Tecnologías Sanitarias de Galicia*. s.n.



- Santana, É. R., Lopes, L., & de Moraes, R. M. (2022). Recognition of the Effect of Vocal Exercises by Fuzzy Triangular Naive Bayes, a Machine Learning Classifier: A Preliminary Analysis. *Journal of Voice*.
<https://doi.org/10.1016/j.jvoice.2022.10.001>
- SANTANA, É. R., LOPES, L., MORAES, D. E., & Marcos, R. (2022). Recognition of the Effect of Vocal Exercises by Fuzzy Triangular Naive Bayes, a Machine Learning Classifier: A Preliminary Analysis. *Journal of Voice*.
- Sau, A., & Bhakta, I. (2019). Screening of anxiety and depression among the seafarers using machine learning technology. *Informatics in Medicine Unlocked*, 16.
<https://doi.org/10.1016/j.imu.2018.12.004>
- SH, M. R. C. J. M. (n.d.). Depresión y Ansiedad en Embarazadas. *Salus*, 17, 2.
- THOTAD, P. N., BHARAMAGOUDAR, G. R., & ANAMI, B. S. (2023). Diabetes disease detection and classification on Indian demographic and health survey data using machine learning methods. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 17.
- Torre A. Vasco M. DE A, G. C. H. (n.d.). Depresión en adolescentes. Un problema oculto para la salud pública y la práctica clínica. In 2, *Marzo-Abril de* (Vol. 72).
- Velez Toala, R. F. (2023). *Estrés, ansiedad y depresión en relación con el rendimiento académico durante la pandemia Covid-19 en los estudiantes de la escuela de Odontología de la Universidad Nacional Jorge Basadre Grohmann - 2021*.
<https://hdl.handle.net/20.500.12773/16266>
- Vergaray, A. D., Miranda, J. C. H., Cornelio, J. B., Carranza, A. R. L., & Sánchez, C. F. P. (2023). Predicting the depression in university students using stacking ensemble



- techniques over oversampling method. *Informatics in Medicine Unlocked*, 101295.
<https://doi.org/10.1016/j.imu.2023.101295>
- Wang, Q., Huang, H., Li, D., Wang, Y., Qi, N., Ci, Y., & Xu, T. (2022). Intensive rTMS for treatment-resistant depression patients with suicidal ideation: An open-label study. *Asian Journal of Psychiatry*, 74. <https://doi.org/10.1016/j.ajp.2022.103189>
- ZHANG, Y. C., & SAKHANENKO, L. (2019). The naive Bayes classifier for functional data. *Statistics and Probability Letters*, 152, 137–146.
- Zulfiker, M. S., Kabir, N., Biswas, A. A., Nazneen, T., & Uddin, M. S. (2021). An in-depth analysis of machine learning approaches to predict depression. *Current Research in Behavioral Sciences*, 2. <https://doi.org/10.1016/j.crbeha.2021.100044>



ANEXOS

ANEXO 1: Matriz de correspondencias múltiples de las características asociadas a la depresión.

Tabla 20:

Matriz de correspondencias múltiples de las características asociadas a la depresión

	Género	Área Estudio	Lugar de Residencia	Internet	Tipo de Vivienda	Tipo de Conexión	Practica de Deporte	Consume Sustancias	Horas de Estudio	Afectado por el Covid
Género	1,000	-,378	-,063	-,026	-,015	-,150	-,200	-,096	-,109	,008
Área de Estudio	-,378	1,000	-,059	-,007	-,039	,119	,100	,013	-,004	,058
Lugar de Residencia	-,063	-,059	1,000	-,283	-,018	-,177	,111	-,063	-,011	-,111
Internet	-,026	-,007	-,283	1,000	,275	,252	-,120	,058	,040	-,041
Tipo de Vivienda	-,015	-,039	-,018	,275	1,000	,141	-,048	-,014	,036	-,107
Tipo de Conexión	-,150	,119	-,177	,252	,141	1,000	-,028	-,001	-,054	,141
Practica de Deporte	-,200	,100	,111	-,120	-,048	-,028	1,000	-,068	-,055	,005
Consume Sustancias	-,096	,013	-,063	,058	-,014	-,001	-,068	1,000	,124	,058
Horas de Estudio	-,109	-,004	-,011	,040	,036	-,054	-,055	,124	1,000	-,074
Afectado por el COVID19	,008	,058	-,111	-,041	-,107	,141	,005	,058	-,074	1,000

Nota: Elaboración propia

La matriz de correlaciones proporciona una visión detallada de las relaciones entre las variables estudiadas, en la relación entre el género y otras variables se observa una correlación negativa moderada -0,378 entre el género y el área de estudio, lo que sugiere que ciertos géneros pueden tener preferencias distintas en cuanto al área de estudio. Esta relación puede tener implicaciones en la distribución de géneros en diferentes disciplinas académicas. Además, hay una correlación negativa moderada -0,200 entre el género y la



práctica deportiva, indicando posibles diferencias de género en la participación en actividades deportivas.

Respecto al área de estudio, se observa una correlación negativa moderada de $-0,378$ con el género y otra correlación negativa moderada de $-0,146$ con la depresión lo que insinúan que la elección del área de estudio podría tener relación tanto con las preferencias de género como con los niveles de depresión. Estos resultados podrían tener consecuencias significativas para la orientación vocacional y el bienestar psicológico de los estudiantes.

El lugar de residencia muestra una correlación negativa moderada $-0,283$ con el acceso a Internet, indicando que el lugar de residencia puede estar relacionado con la disponibilidad de acceso a la red. Esto sugiere que aquellos que viven en áreas rurales pueden tener menos acceso a Internet en comparación con aquellos en áreas urbanas. Además, hay una correlación negativa moderada de $-0,111$ entre el lugar de residencia y la práctica deportiva, lo que indica una asociación entre la ubicación geográfica y la participación en actividades deportivas.

El acceso a Internet muestra correlaciones significativas con varias variables. Una correlación positiva moderada de $0,275$ con el consumo de sustancias sugiere una posible relación entre el tipo de conexión y comportamientos asociados al consumo de sustancias. Esto podría tener implicaciones para la salud pública y los hábitos de uso de Internet. Además, el acceso a Internet se correlaciona positivamente con el tipo de vivienda de $0,275$, indicando que ciertos tipos de vivienda pueden estar asociados con un mayor acceso a Internet. Estos resultados resaltan la importancia de la infraestructura tecnológica y la accesibilidad en diferentes contextos de vivienda.

El tipo de vivienda, a su vez, presenta una correlación positiva moderada $0,190$ con la depresión lo que muestra que el tipo de vivienda podría estar relacionado con los niveles



de depresión entre los estudiantes. Explorar las condiciones de vida y el entorno residencial puede ser crucial para comprender los factores que contribuyen a la salud mental.

La participación en actividades deportivas presenta una correlación negativa moderada de $-0,200$ con el género y otra correlación negativa moderada de $-0,068$ con el consumo de sustancias, esto sugiere que la implicación en actividades deportivas podría fluctuar según el género y podría relacionarse con niveles inferiores de consumo de sustancias, estos resultados resaltan la influencia positiva de la actividad física en la salud y el bienestar general de los estudiantes.

El consumo de sustancias presenta una correlación positiva moderada $0,124$ con el tipo de conexión a Internet, lo que implica el tipo de conexión puede influir en comportamientos asociados al consumo de sustancias. Esta relación podría tener implicaciones para las intervenciones de salud pública y la educación sobre el uso de Internet.

Las horas de estudio no muestran correlación significativa con el género, el área de estudio o la práctica deportiva. Sin embargo, hay una correlación negativa moderada de $-0,074$ entre las horas de estudio y el hecho de ser afectado por el COVID-19, indicando que aquellos que dedican más tiempo al estudio pueden tener una menor probabilidad de contraer la enfermedad. Esta relación destaca la importancia del compromiso académico en la salud durante períodos de pandemia.

Finalmente, el hecho de ser afectado por el COVID-19 muestra una correlación negativa moderada de $-0,185$ con la depresión, sugiriendo que la afectación por la enfermedad puede estar asociada con niveles más altos de depresión. Este hallazgo subraya la



necesidad de abordar la salud mental de aquellos afectados por el virus, reconociendo los posibles impactos psicológicos de la enfermedad.

ANEXO 2: Código en RStudio para el preprocesamiento de datos

```
#####
# DEPRESION HAMILTON
#####
rm(list=ls())
par(mfrow=c(1,1))
#data("datos")
datos <- read.csv("AC_depresionHamilton.csv",head=T, sep=";")
head(datos)
```

```
##      Sexo Edad AreaEstudio      EscProfesional LugarResidencia TipoVivienda
## 1      2    21           2 Nutrici\xf3n Humana                2
1
## 2      2    23           3      Trabajo Social                1
1
## 3      2    23           3      Trabajo social                1
1
## 4      2    27           2 Nutrici\xf3n Humana                1
2
## 5      1    23           1      Ing de Sistemas                1
1
## 6      1    23           1      Ing de Sistemas                2
3
##      Internet TipoConexion HorasEstudio PracticaDeporte ConsumeSustancias
## 1          1             2             1             1
1
## 2          2             2             1             1
1
## 3          2             2             2             2
1
## 4          2             2             1             1
1
## 5          2             1             1             2
1
## 6          1             2             2             2
1
##      AfectadoCOVID19 DEPRESION
## 1                   2          5
## 2                   1          4
## 3                   1          4
## 4                   1          4
## 5                   2          5
## 6                   2          4
```

```
#etiquetado
```



```

datos$Sexo <- factor(datos$Sexo,levels = c(1:2),labels = c("Mujer","Hombre"))
datos$AreaEstudio<- factor(datos$AreaEstudio,levels = c(1:3),labels = c("Ingenierias", "Biomédicas", "Sociales"))
datos$LugarResidencia<- factor(datos$LugarResidencia, levels= c(1:2),labels = c("Rural", "Urbano"))
datos$TipoVivienda<- factor(datos$TipoVivienda,levels = c(1:3),labels = c("Propia", "Prestada", "Alquilada"))
datos$Internet <- factor(datos$Internet ,levels = c(1:2),labels = c("Si", "No"))
datos$TipoConexion <- factor(datos$TipoConexion ,levels = labels = c("Modem", "Plan de datos", "Wi-fi"))
datos$HorasEstudio <- factor(datos$HorasEstudio,levels = labels = c("Menos de 2 horas", "Entre 3 y 4 horas", "mas de 4 horas"))
datos$PracticaDeporte<- factor(datos$PracticaDeporte,levels = labels = c("Si", "No"))
datos$ConsumeSustancias<- factor(datos = c(1:2),labels = c("Si", "No"))

datos$AfectadoCOVID19<- factor(datos$AfectadoCOVID19,levels = c(1:2),labels = c("Si", "No"))
datos$DEPRESION<- factor(datos$DEPRESION,levels = c(1:5),labels = c("Estado Normal", "D Ligera", "D moderada", "D severa", "D muy severa"))
str(datos)
## 'data.frame': 519 obs. of 13 variables:
## $ Sexo : Factor w/ 2 levels "Mujer","Hombre": 2 2 2 2 1 1 2 1 2 1 ...
## $ Edad : int 21 23 23 27 23 23 25 27 23 24 ...
## $ AreaEstudio : Factor w/ 3 levels "Ingenierias",...: 2 3 3 2 1 1 2 1 2 1 ...
## $ EscProfesional : chr "Nutrici\xf3n Humana" "Trabajo Social" "Trabajo social " "Nutrici\xf3n Humana" ...
## $ LugarResidencia : Factor w/ 2 levels "Rural","Urbano": 2 1 1 1 1 2 1 1 1 1 ...
## $ TipoVivienda : Factor w/ 3 levels "Propia","Prestada",...: 1 1 1 2 1 3 1 1 3 1 ...
## $ Internet : Factor w/ 2 levels "Si","No": 1 2 2 2 2 1 2 2 1 2 ...
## $ TipoConexion : Factor w/ 3 levels "Modem","Plan de datos",...: 2 2 2 2 1 2 1 3 2 1 ...
## $ HorasEstudio : Factor w/ 3 levels "Menos de 2 horas",...: 1 1 2 1 1 2 2 1 3 2 ...
## $ PracticaDeporte : Factor w/ 2 levels "Si","No": 1 1 2 1 2 2 1 1 1 2 ...
## $ ConsumeSustancias: Factor w/ 2 levels "Si","No": 1 1 1 1 1 1 1 1 1 2 ...
## $ AfectadoCOVID19 : Factor w/ 2 levels "Si","No": 2 1 1 1 2 2 2 1 1 2 ...
## $ DEPRESION : Factor w/ 5 levels "Estado Normal",...: 5 4 4 4 5 4 5 4 5 2 ...

```

```

#limpieza de datos
colSums(is.na(datos))

```

```

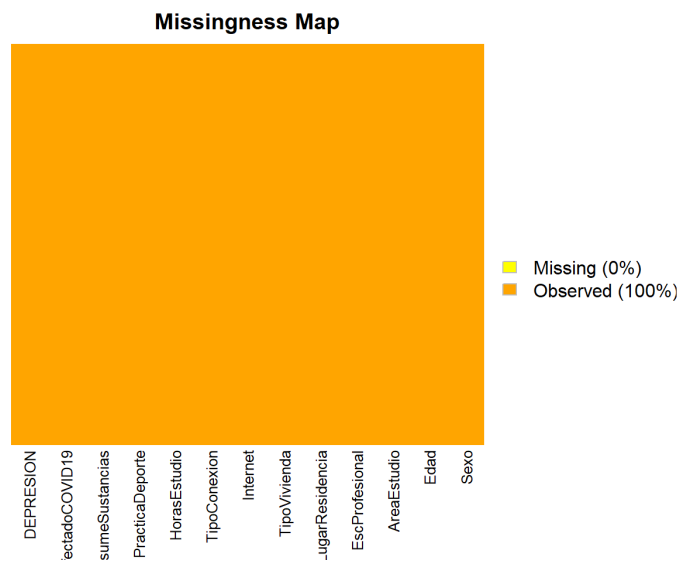
##           Sexo           Edad           AreaEstudio           EscProfesi
##           0             0             0
##           0

```



##	LugarResidencia	TipoVivienda	Internet	TipoCone
xion				
##	0	0	0	
0				
##	HorasEstudio	PracticaDeporte	ConsumeSustancias	AfectadoCOV
ID19				
##	0	0	0	
0				
##	DEPRESION			
##	0			

```
library(Amelia)
missmap(datos,col=c('yellow','orange'),y.at=1,y.labels='',legend=TRUE)
#analisis estaditico
library(funModeling)
```





ANEXO 3: Código en RStudio para el modelo de aprendizaje automático Naive de Bayes

```
rm(list=ls())  
library(e1071)  
library(naivebayes)
```

```
## naivebayes 0.9.7 loaded
```

```
library(ROCR)  
library(caret)
```

```
library(readxl)  
library(caTools)  
library(mlbench)  
library(ggplot2)  
library(dplyr)
```

```
Estudiantes <- read.csv("AC_depresionHamilton.csv", head=T, sep=";")  
head(Estudiantes)
```

```
## Género Edad AreaEstudio EscProfesional LugarResidencia TipoVivienda  
## 1 2 21 2 Nutrici\xf3n Humana 2 1  
## 2 2 23 3 Trabajo Social 1 1  
## 3 2 23 3 Trabajo social 1 1  
## 4 2 27 2 Nutrici\xf3n Humana 1 2  
## 5 1 23 1 Ing de Sistemas 1 1  
## 6 1 23 1 Ing de Sistemas 2 3  
## Internet TipoConexion HorasEstudio PracticaDeporte ConsumeSustancias  
## 1 1 2 1 1 1  
## 2 2 2 1 1 1  
## 3 2 2 2 2 1  
## 4 2 2 1 1 1  
## 5 2 1 1 2 1  
## 6 1 2 2 2 1  
## AfectadoCOVID19 DEPRESION  
## 1 2 5  
## 2 1 4  
## 3 1 4  
## 4 1 4  
## 5 2 5  
## 6 2 4
```



ANEXO 4: *Instrumento de recolección de datos*

CUESTIONARIO

¡Estimado estudiante de la UNA Puno!

El presente cuestionario es con el fin de recopilar datos para un proyecto de investigación. Por lo que tu participación es crucial, es por eso que agradecemos sinceramente tu colaboración. Todas las respuestas proporcionadas serán tratadas de manera confidencial, puesto que la información recopilada se utilizará únicamente con fines académicos y de investigación.

Responda con total sinceridad a los siguientes ítems. Marque con X en la casilla que crea correcta.

Género: Hombre Mujer

1. ¿Cuál es tu área de estudio?

Ingenierías Biomédicas Sociales

2. ¿Cuál es tu lugar de residencia?

Rural Urbano

3. El tipo de vivienda donde te encuentras es:

Propia Prestada Alquilada

4. ¿Cuenta con servicio de internet?

Si No

5. Si cuenta con servicio de internet. ¿Con que tipo de conexión accede a internet?

Modem Plan de datos Wi-fi

6. ¿Cuántas horas de estudio tiene fuera de las actividades universitarias?

Menos de 2 horas Entre 3 a 4 horas Más de 4 horas

7. ¿Práctica algún deporte?

Si No

8. ¿Consumes sustancias nocivas como el alcohol, tabaco, entre otros?

Si No

9. ¿Considera que usted o su familia ha sido afectado por el Covib-19?

Si No



Los siguientes ítems son relacionados con la depresión, por favor marque con total sinceridad con una X dentro del casillero del 0 a al 4 cuan identificado te sientes, donde 0 es la puntuación mínima y 4 la puntuación máxima.

ITEMS	0	1	2	3	4
¿Cuán deprimido se siente?					
¿Cuán culpable se siente frente a este contexto (cuarentena)?					
¿Tiene pensamientos sobre la posibilidad de morir?					
¿Tiene considera que tiene insomnio?					
¿Cómo consideras tu actividad académica nocturna?					
¿Puede usted retomar el sueño, luego de alguna interrupción?					
¿Ha considerado abandonar la universidad?					
¿Considera usted que ha tenido problemas para una comunicación fluida en sus actividades diarias?					
¿Tiene usted alguna manía corporal?					
¿Siente usted preocupación en cualquier actividad diaria?					
¿Se siente usted con síntomas de alguna enfermedad?					
¿Considera usted que ha perdido el apetito?					
¿Usted ha sentido pérdida de energía, fatiga o algún dolor en general?					
¿Ha tenido usted algún síntoma relacionado a sus órganos sexuales?					
¿Cuán preocupado se siente usted por su salud, estado físico, académico, económico o espacio temporal?					
¿Considera usted que en la última semana ha perdido peso?					
¿Usted es consciente de que se encuentra enfermo o deprimido?					

¡Agradezco tu tiempo y compromiso con esta encuesta! Tus respuestas desempeñan un papel fundamental en la construcción de conocimiento y en la mejora continua de nuestra comunidad estudiantil en la UNA Puno.

ANEXO 5: Matriz de consistencia

PROBLEMAS	OBJETIVOS	HIPÓTESIS	VARIABLES	DIMENSIONES	METODOLOGÍA	ESTADÍSTICA
<p>Problema General ¿Cómo el modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes permitirá detectar el nivel de depresión en estudiantes de la UNA-Puno 2022?</p>	<p>Objetivo general Determinar un modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión en estudiantes de la UNA Puno 2022.</p>	<p>Hipótesis general El modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes permite detectar el nivel de depresión en estudiantes de la UNA-Puno 2022.</p>	<p>VARIABLES Independientes Características asociadas a la depresión</p>	<p>Demográficas Académicas Conectividad Estilos de Vida y salud</p>	<p>Método Científico analítico, hipotético deductivo Tipo Básico fundamental Enfoque Cuantitativo Diseño No experimental propositivo de corte transversal Nivel: Predictivo Técnicas: Encuesta Instrumento: Cuestionario de Hamilton Escala de medición: Likert</p>	<p>Clasificador Naive Bayes Aprendizaje automático</p>
<p>Problemas Específicos -¿Cuáles son las características asociadas a la depresión en estudiantes universitarios de la UNA-Puno 2022? -¿Cuál es el nivel mas alto de depresión en los estudiantes de la UNA Puno con la Escala de valoración de Hamilton HDRS? -¿Cuál es la eficacia del modelo de Aprendizaje Automático con el clasificador Naive de Bayes en la detección del nivel de depresión en estudiantes universitarios de la UNA-Puno?</p>	<p>Objetivos específicos -Analizar las características asociadas a la depresión en estudiantes universitarios de la UNA-Puno 2022 -Determinar el nivel mas alto de depresión de los estudiantes de la UNA-Puno con la Escala de valoración de Hamilton HDRS -Validar la eficacia del modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes para la detección del nivel de depresión en estudiantes de la UNA-Puno 2022.</p>	<p>Hipótesis específicos -Las características demográficas, académicas, conectividad, estilos de vida y salud se asocian significativamente a la depresión en estudiantes universitarios de la UNA-Puno. -El nivel más alto de depresión en los estudiantes de la UNA-Puno mediante la Escala de valoración de Hamilton HDRS es una depresión severa -El modelo de Aprendizaje Automático basado en el clasificador Naive de Bayes es eficaz para la detección del nivel de depresión en estudiantes de la UNA-</p>	<p>Variable Dependiente Depresión</p>	<p>Estado normal Depresión ligera Depresión moderada Depresión severa Depresión muy severa</p>		



DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo EVELYN ELIANA COAQUIRA FLORES
identificado con DNI 70263169 en mi condición de egresado de:

Escuela Profesional, Programa de Segunda Especialidad, Programa de Maestría o Doctorado
INGENIERIA ESTADISTICA E INFORMATICA

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

" MODELO DE APRENDIZAJE AUTOMÁTICA BASADO EN EL
CLASIFICADOR NAIVE DE BAYES PARA LA DETECCIÓN DEL NIVEL
DE DEPRESIÓN EN ESTUDIANTES DE LA UNA-PUNO 2022 "

Es un tema original.

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

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

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

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

Puno 11 de DICIEMBRE del 20 23

FIRMA (obligatoria)



Huella



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

Por el presente documento, Yo EUELYN EUANA COAQUIRA FLORES,
identificado con DNI 70263169 en mi condición de egresado de:

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

INGENIERIA ESTADISTICA E INFORMATICA

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

“ MODELO DE APRENDIZAJE AUTOMÁTICO BASADO EN EL
CLASIFICADOR NAIVE DE BAYES PARA LA DETECCIÓN DEL
NIVEL DE DEPRESIÓN EN ESTUDIANTES DE LA UNA-PUNO 2022 ”

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

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

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

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

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

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

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

En señal de conformidad, suscribo el presente documento.

Puno 11 de NOVIEMBRE del 2023

FIRMA (obligatoria)



Huella