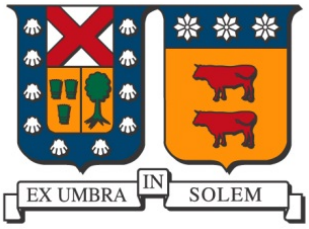
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INFORMÁTICA
SANTIAGO - CHILE



“MODELOS DE MACHINE LEARNING PARA PREDECIR IDEACIÓN SUICIDA: UN ANÁLISIS DE FACTORES PSICOLÓGICOS EN ESTUDIANTES UNIVERSITARIOS.”

FERNANDA RUBIO VALENZUELA

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL EN INFORMÁTICA

Profesor Guía: Mauricio Olivares
Profesor Correferente: José Luis Martí

Junio - 2024

DEDICATORIA

Para todo aquel que no encuentre salida, todo estará bien. Aunque demore y no lo parezca, así será.

AGRADECIMIENTOS

A mi familia, amigos y a mi Iggy, por su presencia incondicional y su ayuda. A la Fernanda del pasado por la valentía y a la vida por sus vueltas.

RESUMEN

Resumen— En este estudio se implementan y evalúan diversos modelos de machine learning para predecir la ideación suicida en estudiantes universitarios, utilizando datos psicológicos y demográficos en sus versiones numéricas y categóricas. Los modelos utilizados incluyen Regresión Logística, Random Forest, SVM, Redes Neuronales y XGBoost, aplicados con y sin técnicas de reducción de características (RFECV) y balanceo de clases. Los principales resultados indican que los modelos de Regresión Logística optimizada con datos numéricos y XGBoost con RFECV fueron las más efectivas en términos de equilibrar las métricas de recall y accuracy. El análisis de importancia de características destaca la relevancia de variables como la depresión, la desesperanza, la soledad y un diagnóstico de neurodivergencia en la predicción de riesgo suicida. Específicamente, se encontró que tener ”Depresión Normal” podría considerarse un factor protector en comparación con tener niveles más elevados de depresión, así como también la ausencia de soledad o desesperanza. Estos hallazgos son cruciales para mejorar la detección temprana de ideación suicida, proporcionando herramientas más efectivas para la intervención temprana y la prevención en el contexto universitario. Además, se sientan las bases para futuras investigaciones que podrían explorar técnicas de machine learning más avanzadas y personalizadas, optimizando aún más la capacidad predictiva y la efectividad de las intervenciones preventivas.

Palabras Clave— Machine learning; Ideación suicida; Análisis predictivo; Salud mental; Modelos de clasificación.

ABSTRACT

***Abstract*—** In this study, various machine learning models are implemented and evaluated to predict suicidal ideation in university students, using psychological and demographic data in their numerical and categorical forms. The models used include Logistic Regression, Random Forest, SVM, Neural Networks, and XGBoost, applied with and without feature reduction techniques (RFECV) and class balancing. The main results indicate that optimized Logistic Regression models with numerical data and XGBoost with RFECV were the most effective in terms of balancing recall and accuracy metrics. The feature importance analysis highlights the relevance of variables such as depression, hopelessness, loneliness, and a neurodivergence diagnosis in predicting suicide risk. Specifically, it was found that having ”Normal Depression” could be considered a protective factor compared to having higher

levels of depression, as well as the absence of loneliness or hopelessness. These findings are crucial for improving the early detection of suicidal ideation, providing more effective tools for early intervention and prevention in the university context. Additionally, this study lays the foundation for future research that could explore more advanced and personalized machine learning techniques, further optimizing predictive capacity and the effectiveness of preventive interventions.

Keywords— Machine learning; Suicidal ideation; Predictive analysis; Mental health; Classification models.

GLOSARIO

ML: Machine learning.
CSV: Del inglés Comma separated value, un tipo de archivo de texto.
RFECV: Recursive Feature Elimination with Cross-Validation. Técnica utilizada para seleccionar características importantes en modelos de aprendizaje automático.
SVM: Support Vector Machine. Algoritmo utilizado para clasificación y regresión.
XGBoost: Extreme Gradient Boosting. Algoritmo utilizado para clasificación y regresión.
VIF: Variance Inflation Factor. Índice que mide la colinealidad entre variables.
NHST: Null Hypothesis Significance Testing. Prueba de significancia de hipótesis nula, utilizada para evaluar hipótesis estadísticas.

ÍNDICE DE CONTENIDOS

RESUMEN iv

ABSTRACT iv

GLOSARIO vi

ÍNDICE DE FIGURAS ix

ÍNDICE DE TABLAS ix

INTRODUCCIÓN 1

CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA 2
1.1 Situación Actual y Motivación . . . 2
1.2 Limitaciones de los Modelos Explicativos en Psicología . . . 3
1.3 Objetivos . . . 5
1.3.1 Objetivo General . . . 5
1.3.2 Objetivos Específicos . . . 5

CAPÍTULO 2: MARCO CONCEPTUAL 6
2.1 Salud Mental en estudiantes universitarios . . . 6
2.2 Variables Psicológicas . . . 7
2.2.1 Estrés . . . 7
2.2.2 Ansiedad . . . 7
2.2.3 Depresión . . . 8
2.2.4 Soledad . . . 9
2.2.5 Desesperanza . . . 9
2.2.6 Neurodivergencia . . . 10
2.2.7 Ideación suicida y su relación con otros factores psicológicos . . . 10
2.3 Modelos de Machine Learning . . . 12
2.4 Machine Learning en estudios de riesgo suicida . . . 13
2.5 Desafíos y Limitaciones del Machine Learning . . . 14
2.5.1 Sesgos en ML . . . 14
2.5.2 Interpretabilidad en ML . . . 16
2.6 Modelos de Machine Learning seleccionados . . . 17
2.6.1 Regresión logística . . . 18
2.6.2 Random forest . . . 19
2.6.3 Máquinas de Soporte Vectorial (SVM) . . . 21
2.6.4 Redes neuronales . . . 22
2.6.5 Gradient Boosting . . . 26
2.7 Relevancia de atributos . . . 27
2.8 Métricas de evaluación de modelos de clasificación . . . 29
2.8.1 Accuracy . . . 29

2.8.2 Recall (Sensibilidad) . . . 30
2.8.3 Matriz de Confusión . . . 30
2.9 Python . . . 31
2.9.1 Librerías de ciencia de datos y machine learning . . . 31
2.10 Jupyter Notebook . . . 32
2.11 Psytoolkit . . . 32

CAPÍTULO 3: PROPUESTA DE SOLUCIÓN **33**
3.1 Metodología Propuesta . . . 33
3.2 Recopilación de datos . . . 34
3.2.1 Creación del Instrumento de Medición . . . 34
3.2.2 Difusión del Cuestionario . . . 36
3.3 Limpieza y transformación de los datos . . . 36
3.4 Análisis Estadístico . . . 39
3.4.1 Datos demográficos . . . 39
3.4.2 Variables psicológicas . . . 40
3.4.3 Estadística inferencial . . . 42
3.5 Diseño e implementación de los modelos de ML . . . 43

CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN **45**
4.1 Metodología de evaluación . . . 45
4.2 Resultados de la Validación . . . 45
4.2.1 Baseline . . . 45
4.2.2 Regresión Logística . . . 46
4.2.3 Random Forest . . . 48
4.2.4 SVM (Máquinas de Soporte Vectorial) . . . 50
4.2.5 Redes Neuronales . . . 51
4.2.6 XGBoost . . . 52
4.3 Comparación y Análisis de los Modelos . . . 54

CAPÍTULO 5: CONCLUSIONES **57**
5.1 Conclusiones generales . . . 57
5.2 Limitaciones del estudio . . . 60
5.3 Trabajos futuros . . . 61

REFERENCIAS BIBLIOGRÁFICAS **63**

ÍNDICE DE FIGURAS

1 Algoritmo Random Forest. 20
2 Ejemplo de una red neuronal de 1 capa oculta. 25
3 Ejemplo de funciones de activación. 25
4 Evolución de los algoritmos basados en árboles de decisión. 27
5 Diseño metodológico de la propuesta de solución. 34
6 Interfaz para la implementación del instrumento. 35
7 Página de bienvenida al cuestionario en PsyToolKit. 36
8 Página de bienvenida al cuestionario: Extracto del consentimiento informado. 37
9 Descarga de datos desde la plataforma. 37
10 Distribución de las edades de los participantes. 39
11 Distribución de depresión, ansiedad y estrés medidas por el test Dass-21. . . 41
12 Distribución de soledad, desesperanza y consumo de alcohol. 42

ÍNDICE DE TABLAS

1 Matriz de Confusión 30
2 Distribución por género 39
3 Recuento de Neurodivergencia 40
4 Recuento de Tipos de Trabajo 40
5 Recuento de Áreas de Estudio 40
6 Datos agrupados por presencia de ideación suicida 41
7 Valores de diversas métricas por género 41
8 Recuento de ideación suicida en estudiantes universitarios 41
9 Resultados del modelo base de Regresión Logística para las versiones numérica y categórica de los datos 46
10 Resultados de la regresión logística para la versión numérica de los tests . . . 47
11 Resultados de la regresión logística para la versión categórica de los tests . . 48
12 Resultados del Random Forest para la versión numérica de los tests 49
13 Resultados del Random Forest para la versión categórica de los tests 49
14 Resultados del Random Forest para las versiones numérica y categórica de los tests con RFECV 50
15 Resultados del SVM para las versiones numérica y categórica de los tests . . 51
16 Resultados de Redes Neuronales para las versiones numérica y categórica de los tests 52
17 Resultados de XGBoost para las versiones numérica y categórica de los tests . 53
18 Resultados de XGBoost para las versiones numérica y categórica de los tests con RFECV 54

INTRODUCCIÓN

La salud mental de los estudiantes universitarios es un área de creciente preocupación debido al aumento sostenido en la incidencia de trastornos mentales en este grupo. Estudios previos han revelado una alta prevalencia de condiciones como la depresión, la ansiedad y la ideación suicida entre los estudiantes universitarios, comparado con la población general y jóvenes que no están en la universidad. Este estudio se enmarca en la necesidad de mejorar las herramientas para la evaluación y prevención de riesgos de salud mental en este grupo demográfico, con un enfoque en la detección temprana y efectiva de la ideación suicida.

El problema central que se aborda es la identificación de la ideación suicida en estudiantes universitarios utilizando técnicas avanzadas de machine learning. Estas técnicas se seleccionaron debido a su capacidad para manejar y modelar relaciones complejas en los datos, lo que las hace adecuadas para capturar las interacciones no lineales entre las variables psicológicas y demográficas que contribuyen al riesgo suicida.

La propuesta de solución incluye la creación de un dataset comprensivo con información sobre los niveles de depresión, estrés, ansiedad y otros factores relevantes, obtenidos a través de tests y cuestionarios validados. Este dataset sirve como base para el entrenamiento y evaluación de diversos modelos de machine learning, incluidos la regresión logística, random forest, SVM, redes neuronales y XGBoost, aplicados tanto a datos numéricos como categóricos. Se emplearon técnicas de reducción de características (RFECV) y balanceo de clases para optimizar el rendimiento de los modelos.

La metodología empleada comprende la recopilación de datos, su limpieza y transformación, un análisis estadístico descriptivo e inferencial, y la implementación y evaluación de los modelos de machine learning. Se evaluaron diferentes métricas de desempeño, como el recall y el accuracy, para seleccionar los modelos más efectivos en la predicción de la ideación suicida.

En términos de estructura, este documento se organiza en varios capítulos. En el Capítulo 1, se define el problema y se establecen los objetivos de la investigación. El Capítulo 2 presenta el marco conceptual, incluyendo una revisión de la literatura sobre salud mental en estudiantes universitarios y las técnicas de machine learning utilizadas. El Capítulo 3 describe la propuesta de solución, detallando la metodología empleada. En el Capítulo 4, se presentan los resultados de la validación de los modelos, y finalmente, en el Capítulo 5, se discuten las conclusiones, limitaciones del estudio y posibles trabajos futuros.

Esta investigación pretende no solo mejorar la capacidad predictiva de los modelos para la detección de ideación suicida, sino también proporcionar herramientas más efectivas para la intervención temprana y la prevención en el contexto universitario, sentando las bases para futuras investigaciones en esta área crítica de la salud mental.

CAPÍTULO 1
DEFINICIÓN DEL PROBLEMA

1.1. Situación Actual y Motivación

Según el Ministerio de Educación, los estudiantes universitarios han reportado una alta presencia de problemáticas relacionadas con la salud mental en los últimos años. Asimismo, ha aumentado la demanda por su atención y abordaje por parte de sus instituciones, lo que ha desencadenado el reconocimiento de la salud mental como una arista fundamental en los procesos formativos y académicos [Ministerio de Educación de Chile, 2023].

Una revisión de salud mental en universitarios chilenos donde se analizaron 32 estudios publicados hasta 2019 reportó una prevalencia combinada de síntomas psicológicos, con rangos que variaban entre los 22,9 % a 40,7 % para malestar psicológico, 16,5 % a 38,8 % para síntomas depresivos, 16,5 % a 23,7 % para síntomas ansiosos, 19,7 % a 29,7 % para consumo de cannabis en los últimos 12 meses, y 84,0 % a 92,6 % para consumo de alcohol en los últimos 12 meses. Entre otras variables, el estrés académico y el consumo problemático de alcohol se asociaron directamente con mayor prevalencia de síntomas depresivos y la conducta suicida se relacionó directamente con una mayor percepción de estrés [Martínez *et al.*, 2021].

El estrés y la ansiedad son causas comunes de problemas de salud más severos como la depresión y la presencia de ideación suicida, lo cual puede llegar a consecuencias extremadamente peligrosas para los individuos afectados. La raíz de estos problemas, en el caso de los estudiantes universitarios puede estar en la gran sobrecarga académica a la cual se enfrentan, proyectos grupales, competitividad, carencia de recursos tecnológicos, mal manejo del tiempo para las actividades académicas y personales, falta de supervisión, sumado a la complejidad natural de estudiar una carrera universitaria [Pozos-Radillo *et al.*, 2014]. Se suma a esto el hecho de que muchos estudiantes deben trabajar o realizar otras labores fuera de la universidad y no cuentan con el espacio y el tiempo para relajarse y lidiar adecuadamente con el estrés.

Como consecuencia de lo anterior, muchos estudiantes deciden desertar sus carreras, cambiar de universidad o en casos más lamentables y severos, poner en riesgo sus vidas ante la presión percibida. Se ha estudiado además, que existe una relación inversamente proporcional entre los logros académicos y la presencia de depresión en estudiantes universitarios [Lama, 2011]. Por otra parte, se ha estudiado que un estado de ánimo positivo está vinculado a una mejor capacidad de desempeño, mejor rendimiento, cognición y memoria en personas tanto sanas como deprimidas, al enfrentarse a una tarea con carga emocional [Sano *et al.*, 2015].

Según la OMS (Organización Mundial de la Salud), entre el año 2000 y 2019 disminuyeron las tasas de suicidio a nivel mundial en un 36 %, con zonas como el Pacífico Occi-

dental que vieron descensos incluso del 49 %. Sin embargo, en la región de las Américas sucedió lo contrario, incrementando esta cifra en un 17 % durante el mismo lapso de tiempo[Organización Mundial de la Salud, 2021].

Si analizamos los casos de defunciones por suicidio en Chile entre los años 2000 a 2017, se documenta un total de 6,292 muertes por suicidio en adolescentes y jóvenes entre los 10 a 24 años de edad considerando todas las regiones del país [Araneda *et al.*, 2021]. Siendo Chile unos de los países con mayor índice de suicidio en América [Quinlan-Davidson *et al.*, 2014].

Teniendo en cuenta lo expuesto, surge la pregunta de si podemos adoptar medidas preventivas para mitigar estas cifras en lugar de simplemente registrarlas una vez que se han materializado, convirtiéndose en datos estadísticos que reflejan el pasado. ¿Cómo podemos mejorar nuestra comprensión del comportamiento de estas variables para anticiparnos a ellas? ¿De qué manera se aborda el estudio de la predicción de variables de esta magnitud desde el ámbito de la psicología o la psiquiatría?

1.2. Limitaciones de los Modelos Explicativos en Psicología

El campo de la psicología se ha dedicado históricamente a entender el comportamiento humano y su sentir. Para lograrlo se cree necesario poder explicar dicho comportamiento encontrando sus causas y además poder predecirlo antes de que se manifieste. Sin embargo, ambos enfoques se fusionan en esta disciplina, considerando que al encontrar un modelo que explique lo suficientemente bien un comportamiento y donde se encuentren las causas y los procesos mentales que lo producen, se está encontrando un modelo que además logre predecir dicho comportamiento en un futuro [Yarkoni y Westfall, 2017]. Asumir esto es peligroso, dado que poder explicar un comportamiento pasado no asegura necesariamente que se pueda extrapolar lo encontrado para modelar el mismo comportamiento a futuro. Solo nos habla de cómo se relacionaron las variables en un determinado conjunto de datos para orquestar la conducta de interés.

Para conseguir dicho poder explicativo los investigadores se han valido comúnmente de técnicas como regresión de mínimos cuadrados, donde se busca significancia estadística para poner a prueba las hipótesis de estudio y determinar el efecto que tienen las variables predictoras en el fenómeno observado [Rosenbusch *et al.*, 2021], siendo esta la esencia de la estadística inferencial en psicología.

La prueba de significancia de hipótesis nula (NHST por sus siglas en inglés) es la técnica favorita de los investigadores en esta disciplina para sacar conclusiones a partir de los datos de una muestra [Cumming, 2008]. La idea es generar una “hipótesis nula”, la cual se quiere refutar. Luego, asumiendo que esta hipótesis es verdadera, se calcula el valor p, el cual representa la probabilidad de obtener resultados por lo menos igual de extremos que los observados. Un valor p pequeño, se traduce a una mayor probabilidad de que la hipótesis nula sea falsa. Cuando el estadístico Ronald Fisher introdujo el concepto del valor p, lo con-

sideraba un medio informal para explorar la significacia de las hipótesis, sin embargo, en la actualidad y desde ese entonces (1920s) se ha usado como prueba definitiva para sustentar todo tipo de relaciones entre variables y respaldar la toma de decisiones importantes en la psicología y otras disciplinas [Nuzzo, 2014]. Esto ocurre aún cuando al intentar replicar un estudio de esta naturaleza, el valor p encontrado sea dramáticamente diferente al original, incluso en muestras de gran tamaño, dando cuenta de esta métrica como una medida muy poco confiable sobre la veracidad de los hallazgos [Cumming, 2008]. Y es que estas pruebas estadísticas evalúan más que las hipótesis de estudio. Sino que se sostienen en base a un sin fin de supuestos de investigación y un contexto experimental acotado que muchas veces no es reportado en las publicaciones, interfiriendo en la posibilidad de una futura replicación del estudio. Por esto, no se debería utilizar la estadística inferencial como un motor para crear generalizaciones sobre el comportamiento manifestado en los datos de muestra, sino que como una ayuda auxiliar para la interpretación local de los datos obtenidos [Amrhein *et al.*, 2019].

Esta falla al replicar con datos nuevos los resultados obtenidos con el modelo original, se puede atenuar utilizando las herramientas que ofrece el machine learning, cuya esencia radica en asegurar cierto poder predictivo al incluir en su metodología una fase de testeo del modelo con datos que no se usan en las fases previas de entrenamiento [Yarkoni y Westfall, 2017]. Sin embargo, a pesar de los beneficios que ofrece, el uso generalizado del machine learning aún no ha sido adoptado por la comunidad de investigadores en psicología u otras ciencias sociales, ni se ha integrado de manera sistemática en los programas de formación académica. Según lo consultado en la literatura, existen varios esfuerzos por comprender y estudiar la salud mental de los estudiantes universitarios a nivel global usando herramientas de inteligencia artificial, pero muchos de estos solo teorizan sobre su uso o reportan el uso de programas de computadora o aplicaciones de smartphones y no real uso de herramientas de inteligencia artificial. Tanto es así, que en un review donde se analizaron 296 artículos entre 2014 a 2020 donde se habla de “riesgo suicida” e “inteligencia articicial”, “deep learning” o algún concepto de esta naturaleza, solo 17 se consideraron relevantes [Lejeune *et al.*, 2022].

Por otra parte, muchos de estos estudios en América latina utilizan técnicas de machine learning para predecir los niveles de una variable clínica como la depresión [Gutiérrez *et al.*,] o el estrés [Vilca Masco, 2020] a través de cuestionarios diseñados para medir dichas variables. Estos cuestionarios se utilizan como datos de entrenamiento para el algoritmo de forma supervisada, donde las respuestas a las preguntas se usan como etiquetas para clasificar los niveles de estrés o depresión. Es evidente que al conocerse de antemano las etiquetas que representan el nivel de estrés o depresión del paciente la herramienta pierde su valor como ayuda auxiliar al diagnóstico.

1.3. Objetivos

1.3.1. Objetivo General

Diseñar y evaluar algoritmos de machine learning para predecir la presencia de ideación suicida en estudiantes universitarios y analizar la relación de esta variable con otros factores psicológicos como depresión, ansiedad, estrés y otros datos demográficos.

1.3.2. Objetivos Específicos

- Crear un dataset con información de los estudiantes universitarios sobre sus niveles de depresión, estrés, ansiedad y otros factores psicológicos y demográficos a través de tests y cuestionarios validados por psicólogos y expertos en el área.
- Analizar estadísticamente la relación entre las variables involucradas para conocer cuáles se asocian más fuertemente con la presencia de ideación suicida.
- Construir diversos modelos de machine learning usando distintas técnicas de clasificación para predecir la presencia de ideación suicida.
- Evaluar los modelos diseñados a través de diversas métricas (F1-Score, accuracy, recall) y seleccionar aquel que mejor se desempeñe en la tarea dada.

CAPÍTULO 2
MARCO CONCEPTUAL

2.1. Salud Mental en estudiantes universitarios

Estudios de prevalencia con estudiantes universitarios, tanto en Chile como a nivel internacional, muestran un aumento sostenido en la incidencia de trastornos mentales en este grupo. Además, se ha notado una alta prevalencia de comorbilidades y una mayor incidencia de trastornos como la depresión, ansiedad y el riesgo suicida en comparación con la población general y los jóvenes que no están en la universidad. Esta evidencia pone en duda la idea previamente aceptada de que los años de escolaridad actúen como un factor protector en la salud mental [Barrera-Herrera y San Martín, 2021]. La prevalencia de problemas psicológicos en estudiantes puede variar desde un 2 % hasta un 50 %. Es más, si en cualquier momento se estudia de manera aleatoria el bienestar de un estudiante, se descubre que 1 de cada 10 estudiantes experimentará problemas emocionales lo suficientemente graves como para requerir la asistencia de un profesional [Kumaraswamy, 2013]. Además, diversos estudios han demostrado que los estudiantes tienen una mayor incidencia de problemas psicológicos en comparación con individuos de la misma edad o con cualquier otra población en general [Saleh *et al.*, 2017].

Debido al incremento del acceso a la educación universitaria y la fuerza laboral, además del aplazamiento de hitos como el matrimonio y la paternidad en las últimas décadas, se ha acuñado el término de adultez emergente para referirse al periodo evolutivo entre los 18 y 29 años de edad, donde muchos jóvenes se sienten entre la adolescencia y la adultez y se caracteriza por ser un periodo de exploración y construcción de la identidad, separación progresiva del hogar y la mantención de un equilibrio entre centrarse en si mismo y nutrir sus relaciones interpersonales [Barrera-Herrera y Vinet, 2017]. Sumado a lo anterior, la mayoría de estos jóvenes están cursando una carrera técnica y/o universitaria durante este mismo período. Esto agrega a los desafíos inherentes a esta nueva etapa de la vida los retos de adaptarse a un sistema académico exigente, muy distinto al experimentado durante la educación secundaria [Barrera-Herrera y San Martín, 2021].

De acuerdo con la literatura revisada y las recomendaciones de expertos en el campo, se han identificado las siguientes variables para estudio: estrés, ansiedad, depresión, niveles de soledad, desesperanza y riesgo suicida, teniendo en cuenta la ideación e intentos suicidas previos. La relevancia de estas variables se abordará en el siguiente apartado.

2.2. Variables Psicológicas

2.2.1. Estrés

El estrés ha sido objeto de estudio en el campo de la psicología desde la década de 1950, lo que ha generado una amplia cantidad de antecedentes, análisis, experimentación y resultados sobre este fenómeno.[Putwain, 2007]. Actualmente, se podría definir el estrés como una crisis en el estilo de vida de una persona, independientemente de su edad o etapa de desarrollo [Reddy *et al.*, 2018] y su asociación con otras enfermedades o problemas de salud se considera ampliamente aceptada en la literatura científica [Putwain, 2007]. A pesar de ello, durante mucho tiempo se creyó que los estudiantes, de cualquier edad, eran los menos propensos a experimentar estrés, ya que la única actividad esperada de ellos era estudiar, lo cual, a su vez, no se consideraba estresante [Reddy *et al.*, 2018]. Sin embargo, la evidencia ha demostrado que el trabajo académico y las tareas asociadas representan una fuente significativa de estrés y preocupación para los estudiantes [Putwain, 2007].

Hoy en día, los estudiantes universitarios reportan problemas más complejos en comparación con décadas anteriores. Para muchos, la transición de la escuela a la universidad representa un cambio significativo que genera estrés. Además, para algunos el hecho de tener que mudarse de sus hogares para asistir a la universidad también puede ser una fuente importante de estrés. Asimismo, se ha observado dificultades para establecer relaciones íntimas y desarrollar un sentido de identidad en esta población, lo cual puede ser un estresor relevante [Kumaraswamy, 2013]. Aunque cierto nivel de estrés es esperable y puede ser beneficioso para la motivación y la acción, niveles elevados pueden tener efectos negativos en el desempeño y el bienestar. Esto es especialmente relevante para muchos estudiantes universitarios, que identifican las demandas académicas, la conciliación entre trabajo, estudios y vida personal, así como los problemas económicos, como sus principales estresores [Stallman y Hurst, 2016].

2.2.2. Ansiedad

Otra variable de interés en este estudio es la ansiedad, que se define como una reacción psicológica, fisiológica y conductual. Al igual que el estrés, la ansiedad es una respuesta esperable y saludable en niveles adecuados, pero niveles elevados pueden afectar negativamente la capacidad de pensar, percibir y aprender de una persona. La presencia de ansiedad también puede causar problemas de concentración y dificultades para recordar información. Sin embargo, a pesar de ser uno de los trastornos mentales diagnosticados más comunes, la ansiedad tiende a ser subestimada e invisibilizada [Sakin Ozen *et al.*, 2010]. Más específicamente, la ansiedad se define como una sensación desagradable de temor ante la anticipación del peligro o lo desconocido, que se caracteriza por una profunda incomodidad [Dias Lopes *et al.*, 2020].

La prevalencia de la ansiedad en la población general suele oscilar entre el 10 % y el 30 %, y su nivel entre estudiantes universitarios tiene consecuencias tanto a corto como a largo plazo, relacionándose con una fuerte preocupación sobre el futuro y las limitaciones educativas [Sakin Ozen *et al.*, 2010]. Según algunos autores [Bamber y Schneider, 2016], la incapacidad para manejar el estrés de manera adecuada puede ser una causa de ansiedad, la cual tiende a empeorar a medida que el individuo se expone más a la fuente de estrés.

Aquellos estudiantes que presentan simultáneamente altos niveles de ansiedad y deficientes aptitudes académicas están más propensos a tener habilidades de estudio deficientes y resultados académicos pobres, lo que los coloca en un alto riesgo de fracaso. Se ha mostrado además que la ansiedad no solo se desencadena por un bajo rendimiento en exámenes y tareas, sino que, de manera inversa, altos niveles de ansiedad pueden resultar en calificaciones más bajas y dificultades en la realización de tareas [Bamber y Schneider, 2016].

Cuando se descuidan los efectos de la ansiedad, es posible que se desarrollen trastornos de ansiedad, lo cual puede ser extremadamente peligroso tanto para la persona afectada como para su círculo cercano. Se ha observado que los trastornos de ansiedad son muy comunes entre la población universitaria y, al afectar el estilo de vida de la persona en general, pueden manifestarse de diversas formas. En particular, el alcoholismo y el abuso de sustancias son problemas frecuentes entre los estudiantes que no reciben atención adecuada [Teelhawod *et al.*, 2021].

2.2.3. Depresión

Otra variable de gran importancia para este estudio es la depresión, que constituye uno de los problemas de salud más frecuentes entre los estudiantes universitarios. Este trastorno afecta significativamente las habilidades interpersonales y sociales de los individuos. La depresión se manifiesta a través de una reducción en el afecto positivo, lo cual puede resultar en una serie de problemas como descuido personal, trastornos del sueño, falta de concentración, ansiedad y un desinterés general en actividades diarias [Ibrahim *et al.*, 2013].

De manera similar a las variables mencionadas anteriormente, diversos estudios han establecido una relación entre la depresión y varios resultados negativos, como el deterioro en los logros académicos, la inestabilidad en las relaciones personales, la presencia de pensamientos o intentos suicidas, y un rendimiento laboral deficiente [Ibrahim *et al.*, 2013]. Además, los estudios indican que es común que la ansiedad y la depresión ocurran simultáneamente entre los estudiantes universitarios, siendo estos el primer y tercer motivo más frecuente de consulta psicológica, respectivamente. Las inquietudes relacionadas con el ámbito académico y laboral se sitúan como la segunda causa más común. Estas dificultades emocionales y psicológicas pueden llevar a la deserción estudiantil, especialmente entre los estudiantes de primer año, quienes tienen el doble de probabilidad de abandonar sus estudios comparados con los alumnos de años superiores [Lama, 2011]. Dada la importancia de los estudiantes como futuro de nuestra comunidad, es crucial abordar sus problemas de salud mental, que

no solo afectan al individuo sino que reflejan el estado mental de la sociedad en general. Sería ventajoso para las instituciones educativas identificar los factores de la vida estudiantil que se asocian con niveles reducidos de depresión, ansiedad y estrés, para así fomentar activamente estos aspectos positivos entre sus estudiantes [Beiter *et al.*, 2015].

2.2.4. Soledad

La investigación en torno al suicidio y la depresión frecuentemente apunta a la calidad insuficiente de las relaciones sociales e interpersonales como un factor determinante. Se ha identificado que la soledad, entendida como la discrepancia entre las relaciones que los individuos anhelan y las que efectivamente poseen, juega un rol crucial en estos contextos [Helm *et al.*, 2020]. Además, estudios específicos han profundizado en cómo la soledad no solo afecta la salud mental de manera general, sino que también se asocia directamente con aspectos más graves como la ideación suicida. Se ha establecido un vínculo significativo entre la soledad y tanto la ideación como los intentos de suicidio, afectando tanto a la población en general como a los pacientes con trastornos psiquiátricos [Stravynski y Boyer, 2001]. Por otra parte, aunque la soledad se ha vinculado persistentemente con la depresión, aún no se ha determinado con certeza cuál de estas condiciones causa la otra. Algunos estudios proponen que la soledad puede inducir la depresión, mientras que otros argumentan que la depresión podría ser un factor predisponente que intensifica los niveles de soledad experimentados por los individuos [Diehl *et al.*, 2018].

En este contexto, es relevante considerar los efectos de la conectividad en la salud mental. Algunos autores han demostrado que la conectividad, definida como la conexión no solo con otras personas sino también con actividades significativas como el trabajo o proyectos, actúa como un factor protector que mitiga la ideación suicida. Esta capacidad de conectar de manera significativa podría, por lo tanto, contrarrestar los efectos adversos de la soledad [Klonsky y May, 2015]. Esta observación subraya la severidad del impacto negativo que la soledad puede tener sobre la salud mental, enfatizando la necesidad urgente de abordar este problema en nuestras comunidades y sistemas de apoyo.

2.2.5. Desesperanza

Otra variable que se ha relacionado con la ideación y riesgo suicida de manera persistente es la desesperanza, la cual se define como un estado transitorio y profundo caracterizado por emociones negativas, pensamientos y deseos que afectan la percepción del futuro [Chang, 2017]. En varios estudios, la desesperanza se ha vinculado con problemas de salud mental como la depresión y la soledad, así como con un incremento en el riesgo de suicidio, abuso de sustancias, ansiedad y bajo rendimiento académico [Keshoofy *et al.*, 2023]. Además, algunos autores sostienen que la desesperanza podría incluso tener una mayor relevancia que la depresión a la hora de explicar la suicidalidad, subrayando su papel crítico en

la manifestación de comportamientos asociados al suicidio [Williams *et al.*, 2008].

2.2.6. Neurodivergencia

La neurodivergencia se considera un término general que incluye diversos diagnósticos, tales como la dispraxia, dislexia, trastorno por déficit de atención e hiperactividad, síndrome de Tourette y el espectro autista. El término surge como respuesta a la perspectiva patologizante de estos diagnósticos y la aceptación de un tipo en especifico de funcionamiento cognitivo y mental como normal o sano [Flores-Fernández *et al.*, 2022].

La neurodiversidad es un término que incluye tanto a las personas neurotípicas, cuyo procesamiento mental se encuentra dentro de los parámetros considerados normales, como a las personas neurodivergentes, cuyo procesamiento mental difiere de estos parámetros [Otárola *et al.*, 2023].

En este estudio se evalúa el impacto de un diagnóstico de neurodivergencia en la medición de la variable dependiente y su asociación con otras características, tanto psicológicas como demográficas.

2.2.7. Ideación suicida y su relación con otros factores psicológicos

Como se ha destacado previamente, las repercusiones de enfrentar desafíos psicológicos y emocionales durante los años universitarios pueden ser diversas entre los estudiantes. Estos problemas pueden manifestarse en diversas formas, incluyendo trastornos físicos, alteraciones del sueño, consumo excesivo de alcohol o drogas, trastornos alimentarios y dificultades en las relaciones interpersonales. En algunos casos, estos problemas pueden llevar a la deserción escolar o evolucionar hacia la ideación, planificación e incluso intentos de suicidio[Reddy *et al.*, 2018]. Esta problemática constituye una de las principales preocupaciones en la salud pública global, lo que ha motivado la implementación de numerosas estrategias de prevención y detección temprana. La identificación de los factores de riesgo clave es fundamental para el desarrollo de estas estrategias. En este sentido, investigaciones recientes han identificado los trastornos del estado de ánimo y la ansiedad como factores de riesgo significativos para el comportamiento suicida. No obstante, la relación entre estos y el comportamiento suicida no proporciona una base suficientemente sólida para aplicaciones clínicas, dado que su capacidad predictiva apenas supera a la del azar [van Mens *et al.*, 2020]. Por otra parte, se observa que la depresión incrementa significativamente la probabilidad de comportamientos suicidas, aunque es importante señalar que no todas las personas que presentan ideación suicida sufren de depresión. La American Foundation for Suicide Prevention (AFSP) clasifica los factores asociados al suicidio en tres categorías principales: factores de salud, factores ambientales y factores históricos. Dentro de estas categorías, investigaciones han explorado aspectos más detallados, incluyendo problemas de salud mental, consumo de

sustancias, variaciones en la personalidad, experiencias negativas de vida, y factores sociales y cognitivos [Ji *et al.*, 2020]. En relación a esto, un estudio reveló que, aunque la ideación suicida tiene limitaciones estadísticas como herramienta de detección del suicidio, la manifestación de dicha ideación refleja claramente el malestar psicológico del individuo. Por tanto, una detección temprana y precisa de los signos de ideación suicida puede mejorar significativamente la identificación de personas con pensamientos suicidas y facilitar una intervención más rápida y efectiva por parte de los profesionales de la salud mental, activando las redes de apoyo necesarias para cuidar adecuadamente a estos individuos [Ji *et al.*, 2020].

Un estudio llevado a cabo por la Universidad Austral junto con la Universidad de la Frontera, que incluyó a 804 estudiantes, reveló que el 13.7 % de ellos cumplía con los criterios para un diagnóstico de depresión mayor, mientras que el 26.9 % mostró síntomas de algún tipo de depresión. La muestra tenía un promedio de edad de 21 años, con un 53.7 % de hombres y un 46.3 % de mujeres. Otras cifras importantes destacaron que un 10.4 % de los estudiantes fue diagnosticado con trastorno bipolar y el 5.3 % exhibió altos niveles de desesperanza, asociados con un riesgo moderado a severo de suicidio o autolesiones, medido mediante la Escala de desesperanza de Beck. Esta herramienta, validada en varios países latinoamericanos incluido Chile, evalúa las expectativas negativas hacia el futuro y sirve como un indicador confiable de riesgo suicida. Además, el 84 % de los estudiantes informó haber consumido alcohol en el último año y el 55.9 % en el último mes, de los cuales el 24.2 % indicó un consumo de alcohol riesgoso y probable dependencia. Este estudio subraya una alta prevalencia de psicopatologías entre los estudiantes universitarios, señalándolos como un grupo vulnerable a enfrentar problemas de salud mental más severos [Baader *et al.*, 2014].

Un estudio exploratorio llevado a cabo por la Pontificia Universidad Católica de Chile con una muestra de 460 estudiantes universitarios que acudían a un servicio de salud estudiantil encontró que el 92 % de los diagnósticos se centraron en trastornos adaptativos, trastornos del estado de ánimo y trastornos de ansiedad. Los trastornos adaptativos presentaron la incidencia más alta con un 48.3 %. Entre los trastornos del estado de ánimo, la depresión mayor de episodio único fue la más común, afectando al 7 % de los estudiantes. En cuanto a los trastornos de ansiedad, el 6.1 % correspondía a trastornos de ansiedad en estudio, el 3.3 % a trastorno de angustia sin agorafobia y el 3 % a fobia social. Se observó que un 12.6 % de los estudiantes reportó trastornos del sueño secundarios, principalmente asociados a diagnósticos principales de trastornos de ansiedad o del estado de ánimo. Respecto a la ideación suicida, con o sin plan, se encontró que el 6.1 % de los estudiantes reportó ideación en los últimos 6 meses y un 5.9 % en periodos anteriores. Además, el 6.3 % de los estudiantes había intentado suicidarse antes de los últimos 6 meses. Entre los estudiantes con ideación suicida y plan en los últimos 6 meses, el 75 % había intentado el suicidio antes de este periodo, mientras que el 16.7 % de aquellos sin plan previo también lo había intentado. No se reportaron diferencias significativas por género en la conducta suicida analizada [Micin y Bagladi, 2011]. Los resultados de este estudio subrayan la prevalencia significativa de trastornos psicológicos y conductas suicidas entre los estudiantes universitarios que acuden a servicios de salud. Estos hallazgos ponen de relieve la necesidad urgente de estrategias de intervención y prevención más efectivas en los campus universitarios.

2.3. Modelos de Machine Learning

El machine learning es un campo de la inteligencia artificial que está fuertemente relacionado con la estadística y las ciencias de la computación. Esta disciplina se centra en el desarrollo de algoritmos que aprenden a partir de la experiencia para optimizar su rendimiento en tareas específicas. Estos algoritmos adquieren conocimiento a partir de los datos, identificando patrones sin mucha intervención humana. Tradicionalmente, los programadores tenían que descubrir las relaciones entre variables y extraer las características importantes de los datos para que los algoritmos pudieran procesar esta información. Sin embargo, el enfoque moderno del machine learning permite que los algoritmos descubran por sí mismos patrones y relaciones complejas, muchas veces inaccesibles para el análisis humano. Una ventaja notable de estos algoritmos es su capacidad de mejorar continuamente su rendimiento con la incorporación de nuevos datos, ajustando sus resultados de manera progresiva [Linthicum *et al.*, 2019]. Dependiendo del tipo de problema y la naturaleza de los datos, se puede distinguir entre tres tipos de algoritmos de machine learning: aquellos de aprendizaje supervisado, aprendizaje no supervisado y aprendizaje reforzado [Janiesch *et al.*, 2021].

Los algoritmos de aprendizaje supervisado operan mediante un conjunto de datos de entrenamiento, los cuales incluyen ejemplos de entradas junto con sus correspondientes etiquetas o datos de salida, también llamadas targets. Estos ejemplos sirven para ajustar los parámetros del modelo. Tras completar el entrenamiento, la eficacia del modelo se evalúa utilizando nuevos conjuntos de datos. Este proceso permite verificar la precisión con la que el modelo puede predecir resultados basados en la información aprendida [Janiesch *et al.*, 2021]. El aprendizaje supervisado se utiliza para resolver dos tipos principales de problemas. Primero, los problemas de regresión, en los cuales el objetivo es predecir un valor numérico continuo, como por ejemplo, la cantidad estimada de personas que utilizarán un servicio o el valor de una deuda universitaria. Segundo, los problemas de clasificación, donde el objetivo es determinar valores discretos o categorías, como identificar si las características observadas en una imagen médica indican la presencia de cáncer [Linthicum *et al.*, 2019].

Los algoritmos de aprendizaje no supervisado son aplicados cuando el objetivo es que el modelo identifique patrones intrínsecos en los datos, sin que estos patrones hayan sido previamente definidos ni que se proporcionen etiquetas específicas. En este enfoque, solo se disponen de datos de entrada y no hay un target o salida predefinida. El propósito es descubrir estructuras subyacentes en el conjunto de datos, lo que puede resultar en la agrupación de elementos similares (clusterización) o en la simplificación de los datos a través de la reducción de su dimensionalidad [Janiesch *et al.*, 2021].

El aprendizaje reforzado se distingue por no requerir de datos de entrada y salida predefinidos. En su lugar, evalúa el estado actual de un sistema y, tras establecer un objetivo específico, proporciona un conjunto de acciones posibles y restricciones. El modelo de machine learning procede entonces a intentar alcanzar dicho objetivo mediante métodos de ensayo y error. Esta metodología permite al modelo aprender de las consecuencias de sus

acciones, ajustando su estrategia para optimizar los resultados hacia la meta establecida [Janiesch *et al.*, 2021]. El propósito de desarrollar sistemas de aprendizaje reforzado es capacitarlos para aprender y perfeccionar su rendimiento mediante el uso de la retroalimentación obtenida del entorno [Linthicum *et al.*, 2019].

El problema central de este estudio es la identificación de la presencia o ausencia de ideación suicida, lo que constituye un problema de clasificación binaria dentro del campo del aprendizaje supervisado. Este método se destaca por su utilidad en contextos donde las respuestas correctas están predefinidas, permitiendo al modelo aprender de ejemplos etiquetados para luego aplicar y generalizar este conocimiento a nuevos datos, no contemplados durante la fase de entrenamiento. La capacidad de los algoritmos de aprendizaje supervisado para generar predicciones precisas y fundamentadas en datos históricos validados y la eficacia de los métodos de clasificación en tareas donde las categorías de salida son explícitas los convierte en la opción más adecuada para abordar el problema identificado en esta investigación.

2.4. Machine Learning en estudios de riesgo suicida

Dada la complejidad de las variables involucradas y la influencia de múltiples factores externos en el análisis del comportamiento y emociones de los individuos, es lógico considerar que las interacciones entre estas variables no se ajustan a modelos lineales simples. Por esta razón, es crucial utilizar métodos de aprendizaje que puedan abarcar y modelar adecuadamente estas relaciones no lineales. Las técnicas de machine learning, en particular, ofrecen una capacidad superior para manejar y modelar estas complejidades, mostrando una flexibilidad significativa en la identificación de patrones y correlaciones no evidentes a través de métodos convencionales [van Mens *et al.*, 2020]. Aunque el uso de machine learning y la inteligencia artificial ofrece numerosas ventajas, también presenta desafíos éticos y legales que no pueden ignorarse. Es ampliamente reconocido que las técnicas de machine learning pueden estar sujetas a sesgos inherentes, lo que podría limitar la capacidad de generalizar los resultados obtenidos a poblaciones que difieren cultural, social o demográficamente de la muestra original utilizada para el entrenamiento de los datos. Adicionalmente, incluso con un modelo predictivo eficaz, siempre existirá la posibilidad de enfrentar casos de falsos negativos y falsos positivos, lo que plantea retos significativos debido a la delicadeza de la variable en estudio, como es la ideación suicida. Sin embargo, la aplicación de la inteligencia artificial para identificar patrones de ideación suicida podría representar un uso valioso de esta tecnología. Si se implementa correctamente, tiene el potencial de mejorar significativamente el bienestar individual y colectivo, ofreciendo herramientas proactivas para la intervención y prevención en salud mental [Ji *et al.*, 2020].

En el ámbito de la inteligencia artificial, los algoritmos de machine learning son fundamentales para desarrollar clasificadores avanzados que se fortalecen con el uso de grandes conjuntos de datos. La efectividad de estos algoritmos mejora con la cantidad de datos procesados, potenciando así su capacidad predictiva. Sin embargo, es crucial reconocer que los estudios sobre la prevención del suicidio intersectan múltiples disciplinas, incluyendo la medicina y la

informática, lo que incrementa la complejidad de su revisión y difusión [Bernert *et al.*, 2020].

Un meta-análisis de 365 estudios longitudinales llevados a cabo durante los últimos 50 años, centrados en la predicción de pensamientos y conductas suicidas, concluyó que no ha habido mejoras significativas en la capacidad de predecir el riesgo suicida a lo largo de este periodo. Además, se destacó que estos estudios raramente consideraban el efecto combinado de múltiples factores de riesgo. Esto subraya la importancia de transitar hacia un enfoque más innovador que incorpore modelos algorítmicos de análisis de riesgo basados en técnicas de machine learning para una evaluación más efectiva [Franklin *et al.*, 2017]. La aplicación de la inteligencia artificial ya ha mostrado resultados fructíferos en otras áreas áreas de la medicina (imágenes, patología, dermatología, etc), superando en velocidad a los expertos y alcanzando niveles de precisión comparables en el diagnóstico de ciertas enfermedades. Aunque la precisión diagnóstica de estas herramientas de inteligencia artificial no alcanza el 100 %, la combinación de su uso con las evaluaciones de profesionales experimentados podría ofrecer grandes beneficios, especialmente en la detección temprana de condiciones que requieren clasificación [Lejeune *et al.*, 2022].

En el ámbito de la investigación del suicidio, un estudio sistemático ha revelado que el machine learning sirve a tres propósitos esenciales. Primero, mejora la precisión en las predicciones de riesgo suicida, superando los métodos estadísticos tradicionales en términos de poder predictivo. Segundo, permite la identificación de nuevas variables relevantes y la exploración de interacciones entre predictores, enriqueciendo así la comprensión de los factores de riesgo. Tercero, facilita la modelación de subgrupos de pacientes, lo que puede conducir a intervenciones más personalizadas y efectivas. Estos avances subrayan la utilidad del machine learning no solo en la detección del riesgo suicida, sino también en la segmentación clínica, ofreciendo caminos prometedores para futuras investigaciones y aplicaciones prácticas en el campo de la salud mental [Burke *et al.*, 2019]. Los estudios recientes destacan que las herramientas de machine learning poseen un significativo potencial predictivo en la identificación del riesgo suicida, tanto a nivel individual como grupal. Individualmente, estos algoritmos son capaces de detectar a personas en crisis, facilitando intervenciones oportunas. A nivel grupal, permiten identificar colectivos en riesgo y determinar qué individuos dentro de estos grupos están más vulnerables. Esta dualidad de aplicaciones subraya la valiosa contribución del machine learning en la prevención del suicidio, ofreciendo nuevas vías para abordar esta problemática de salud pública [Lejeune *et al.*, 2022].

2.5. Desafíos y Limitaciones del Machine Learning

2.5.1. Sesgos en ML

Los algoritmos de machine learning dependen fundamentalmente de los datos con los que son entrenados. Si estos datos contienen sesgos, ya sean implícitos o explícitos, los modelos resultantes probablemente perpetuarán o incluso exacerbarán estos sesgos en sus predic-

ciones. Esto es particularmente preocupante en el contexto de la salud mental, donde decisiones incorrectas pueden tener consecuencias graves. En esta sección, exploraremos cómo los sesgos en los datos se manifiestan en modelos de machine learning, examinando tanto las causas subyacentes como las estrategias potenciales para mitigar estos efectos.

Suresh y Guttag [Suresh y Guttag, 2019], identificaron cinco sesgos que surgen con la utilización de ML que podrían causar consecuencias no deseadas o resultados poco favorables para algunos sectores de la sociedad:

- **Sesgo histórico:** Este constituye uno de los primeros y más fundamentales desafíos al trabajar con machine learning, ya que afecta tanto al proceso de generación de datos como a la definición de los objetivos del modelo. Aunque los datos se obtengan y seleccionen de manera óptima, el sesgo histórico puede seguir presente. Esto ocurre porque los patrones recurrentes en la sociedad, como estereotipos o prejuicios, se pueden reflejar en el modelo, perpetuando así discriminaciones que podrían perjudicar a grupos específicos.

- **Sesgo en la representación:** Este tipo de sesgo se presenta durante el proceso de selección de muestras de una población. Puede darse que ciertos grupos no estén tan representados como deberían, lo que afecta la efectividad del modelo. Este problema puede surgir si el método de muestreo no logra cubrir adecuadamente toda la población o si hay una discrepancia entre la población objetivo del estudio y la población utilizada durante la fase de entrenamiento del modelo. Por ejemplo, emplear datos que son demasiado antiguos o que provienen de diferentes regiones geográficas puede resultar en un modelo que no sea aplicable o efectivo para una población distinta.

- **Sesgo de medición:** Este tipo de sesgo se manifiesta cuando la calidad o detalle de los datos recogidos varía entre diferentes grupos demográficos. Por ejemplo, un fenómeno podría parecer más prevalente en un grupo específico simplemente porque ese grupo está sujeto a un monitoreo más intenso, y no necesariamente porque el fenómeno ocurra más frecuentemente en realidad. Además, en el contexto del aprendizaje supervisado, el sesgo de medición puede manifestarse cuando las definiciones utilizadas para la clasificación de los datos son excesivamente simplistas. Un ejemplo claro de esto es cuando se intenta clasificar el éxito académico utilizando como etiqueta target las calificaciones obtenidas. Esta aproximación falla en capturar la complejidad y la variedad de factores que constituyen el rendimiento estudiantil, resultando en un modelo que puede no reflejar con precisión las habilidades o conocimientos reales de los estudiantes.

- **Sesgo de Agregación:** Este tipo de sesgo surge cuando un modelo de machine learning se aplica de manera uniforme a todo un grupo, sin considerar las variaciones internas o la importancia diferencial de ciertas características en subgrupos específicos. Aunque estos subgrupos estén adecuadamente representados en los datos de muestra, el modelo puede fallar en predecir con precisión el comportamiento de todos los integrantes del grupo. Esto ocurre porque el modelo generaliza demasiado, ignorando

cómo ciertas características pueden influir de manera más significativa en algunos segmentos del grupo que en otros. Un ejemplo de esto se presenta en el contexto médico, donde las diferencias étnicas pueden llevar a que ciertas características o complicaciones de la misma enfermedad se manifiesten de manera diferente entre los grupos, lo cual será ignorado por el modelo.

- **Sesgo de evaluación:** Este sesgo se presenta cuando los datos utilizados como referencia o benchmark para evaluar la efectividad de un algoritmo no reflejan adecuadamente la población objetivo. La comparación de diferentes algoritmos requiere de un estándar común para medir objetivamente su rendimiento. No obstante, si los datos de referencia están sesgados o no representan fielmente la realidad, los modelos evaluados con estos benchmarks no detectarán ni serán penalizados por replicar estos mismos sesgos. Esto resulta en la perpetuación de prejuicios y la continuación del uso de conjuntos de datos sesgados como estándares de comparación.

Esta formulación subraya la importancia de la identificación precisa de sesgos como un paso esencial para mejorar la exactitud de los análisis y reducir los efectos negativos asociados con decisiones basadas en datos sesgados. Aunque técnicamente no siempre sea posible mitigar completamente sus efectos, es fundamental considerar y comunicar el impacto de estos sesgos. Además, es esencial informar las condiciones específicas bajo las cuales se obtuvieron las conclusiones del estudio para asegurar no solo la posibilidad de replicación a futuro, sino que también un entendimiento más acabado del contexto en el cual se obtienen dichas conclusiones.

Es esencial reconocer el impacto significativo de la inteligencia artificial y el machine learning en la vida contemporánea, desde algoritmos que determinan la elegibilidad para créditos hasta vehículos autónomos. Sin embargo, el entendimiento público sobre cómo se toman estas decisiones es limitado, lo que plantea desafíos éticos. Para asegurar la transparencia, es crucial involucrar a expertos de múltiples disciplinas en la toma de decisiones. La integración de perspectivas de ciencias sociales, filosofía, medicina y ciencia enriquece el análisis y la comprensión de las implicaciones éticas de la IA [Yapo y Weiss, 2018].

2.5.2. Interpretabilidad en ML

La transparencia en el uso de la inteligencia artificial es fundamental, pero para que esta sea efectiva, también es necesario abordar la interpretabilidad de los modelos de machine learning. Involucrar a expertos de diversas disciplinas ayuda a alinear las decisiones tecnológicas con valores éticos y sociales, pero entender cómo los algoritmos toman decisiones es igualmente importante. La interpretabilidad no solo refuerza la transparencia, sino que también facilita la validación y crítica constructiva de los modelos, permitiendo a usuarios y expertos comprender y cuestionar los procesos y resultados. Esto es esencial en sectores como la medicina, justicia y seguridad pública, donde explicar las decisiones de los modelos

mejora la confianza, fiabilidad y equidad, y permite identificar y corregir sesgos, alinearse con normativas éticas y legales, y facilitar la integración de la IA en la toma de decisiones diarias.

El término de interpretabilidad hace referencia a sistemas donde el usuario no sólo obtiene una salida del modelo, sino que también puede comprender y analizar cómo estas salidas se derivaron matemáticamente a partir de los datos de entrada [Doran *et al.*, 2017]. La dificultad para lograr esta interpretabilidad ha limitado el uso de estas herramientas en el campo de la salud, donde la capacidad predictiva de un modelo es tan crucial como la comprensión de sus predicciones [Stiglic *et al.*, 2020]. En este sentido, un bajo nivel de interpretabilidad podría ser aceptable cuando el algoritmo se utiliza, por ejemplo, en marketing o ventas. Sin embargo, si el algoritmo se emplea para realizar diagnósticos médicos o hacer predicciones donde un error puede ser perjudicial, debe tener una mayor interpretabilidad, ya que el potencial daño causado por una mala decisión es mucho mayor [Adadi y Berrada, 2018].

En el contexto de este estudio, la interpretabilidad del modelo no es tan crítica como lo sería en aplicaciones clínicas directas. El objetivo principal es identificar patrones que expliquen la ideación suicida, en lugar de utilizar el modelo como una herramienta para tomar decisiones clínicas inmediatas. Al enfocarse en la detección de tendencias y relaciones subyacentes en los datos, se puede ofrecer información valiosa para futuras investigaciones y potenciales intervenciones, sin que un bajo nivel de interpretabilidad presente un riesgo significativo para la salud de los individuos. Por lo general, los modelos más complejos que ofrecen mayor exactitud son menos explicables y más difíciles de entender, mientras que los modelos más sencillos estructuralmente son más interpretables [Adadi y Berrada, 2018]. Sin embargo, debido a la naturaleza de este estudio, es posible explorar ambos tipos de modelos. Se requiere una gran exactitud en la predicción para identificar con precisión los patrones de ideación suicida, pero también es necesaria una considerable explicabilidad para entender los patrones subyacentes a esta variable de interés. Esta dualidad permite aprovechar los beneficios de ambos enfoques, logrando un equilibrio entre precisión predictiva e interpretabilidad.

La importancia de manejar con cuidado y precisión los datos se vuelve crítica cuando estos son de naturaleza delicada, como es el caso de las variables psicológicas clínicas y el riesgo suicida abordados en el presente estudio. Por ello, es esencial comunicar claramente las condiciones bajo las cuales se realiza la investigación, los supuestos teóricos y metodológicos que la sustentan, y facilitar el acceso a los datos utilizados. Estas prácticas son fundamentales para apoyar los esfuerzos de mitigar las limitaciones inherentes al estudio y garantizar que los hallazgos puedan ser validados y replicados de manera independiente.

2.6. Modelos de Machine Learning seleccionados

Según una importante revisión de estudios sobre suicidio utilizando machine learning, los métodos empleados para predecir el riesgo suicida varían ampliamente e incluyen tanto aprendizaje supervisado como no supervisado. Los métodos supervisados, como random

forest, clasificadores naïve Bayes, árboles de decisión, regresión logística, regresión de mínimos cuadrados y máquinas de soporte vectorial (SVM), son los más comúnmente utilizados. Por otro lado, los estudios que emplean técnicas no supervisadas se centran en algoritmos de clusterización, redes neuronales, mapas auto-organizados, análisis de componentes principales (PCA) y árboles de decisión. Algunos estudios combinan ambos enfoques. Además, muchos de estos estudios informan el uso de validación cruzada y métodos de particionamiento de datos en conjuntos de entrenamiento y prueba para evaluar el desempeño de los modelos. Solo unos pocos estudios utilizaron conjuntos de datos distintos y separados temporalmente para el entrenamiento y la prueba de los modelos. Las muestras utilizadas variaban ampliamente, desde 55 hasta 975,057 participantes, siendo en su mayoría adultos. En general, hubo una sobre representación de estudios exploratorios, con muy pocos estudios que proponían nuevas configuraciones de modelos predictivos o utilizaban muestras de participantes diferentes [Bernert *et al.*, 2020].

Considerando lo encontrado en la literatura, se decide optar por los siguientes modelos:

- **Regresión logística**
- **Random forest**
- **Máquinas de Soporte Vectorial (SVM)**
- **Redes neuronales**
- **Gradient Boosting**

2.6.1. Regresión logística

La regresión logística es un método estadístico utilizado principalmente para tareas de clasificación binaria. Este método se basa en probabilidades para determinar si los datos de entrada pertenecen a una categoría específica [Padmasree y Muthyam,]. Además, es el modelo de clasificación “más popular para predecir la presencia o ausencia de un atributo usando covariantes” [Amini *et al.*, 2016] y se encuentra dentro de la categoría de los modelos altamente interpretables [Stiglic *et al.*, 2020].

En este caso, al haber más de un predictor, se utiliza la regresión logística múltiple, que tiene la siguiente expresión matemática (ecuación 1):

$$p(X) = \frac{e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X_1+\cdots+\beta_p X_p}} \tag{1}$$

Cuyos parámetros son:

- $\beta_0, \beta_1, \beta_2, \ldots, \beta_p$: Coeficientes de regresión de los predictores. Representan la influencia de cada predictor en la probabilidad de que la variable dependiente sea 1.
- $X_1, X_2, \ldots, X_p$: Variables independientes o predictores. Son las variables que se consideran pueden influir en la variable dependiente binaria.
- $p(X)$: Probabilidad de que la variable dependiente Y sea igual a 1 dado el conjunto de predictores X.

La salida de la función $p(X)$ representa la probabilidad de que la clase objetivo sea 1 (o ”sí”). Si la probabilidad es mayor que un umbral (típicamente 0.5), se predice que la observación pertenece a la clase 1; de lo contrario, se predice que pertenece a la clase 0. En este caso, la función entregará la probabilidad de la existencia de ideación suicida a partir del resto de variables independientes o predictores.

Dentro de las ventajas de este método se encuentran la facilidad de implementación, alta interpretabilidad y transparecia del modelo. Entre sus desventajas se encuentra su bajo desempeño con variables no lineales, ya que si las relaciones en los datos son complejas y no lineales, la regresión logística puede no capturar adecuadamente estas interacciones, lo que puede llevar a un rendimiento subóptimo del modelo.

2.6.2. Random forest

Este método se basa en la construcción de múltiples árboles de decisión durante el entrenamiento del modelo y la combinación de sus resultados (calculo de la moda para clasificación y la media para regresión) para mejorar la precisión y controlar el sobreajuste (overfitting) [Padmasree y Muthyam,]. Es un método robusto y efectivo que logra detectar valores atípicos (outliers) y minimizar falsos positivos [Abdullah y Bouke,]. Es capaz de manejar datos de alta dimensionalidad y puede capturar relaciones no lineales entre características. Este enfoque proporciona varias ventajas, incluyendo una mayor precisión en comparación con un solo árbol de decisión, robustez contra el overfitting y la capacidad de manejar grandes conjuntos de datos con muchas características. Además, Random Forest ofrece medidas de importancia de las características, lo que permite entender mejor la contribución de cada variable en la predicción. Sin embargo, aunque cada árbol individual es interpretable, el conjunto de muchos árboles es menos interpretable que en modelos más simples de clasificación.

Como se trata de un problema de clasificación, los árboles de decisión creados en la fase de entrenamiento emitirán cada uno un voto. La predicción final se hace determinando la clase más popular entre los árboles [Breiman, 2001]. Esto se puede ilustrar en la ecuación 2:

$$\hat{y}(\mathbf{x}) = \text{moda}\left\{\hat{y}_1(\mathbf{x}), \hat{y}_2(\mathbf{x}), \ldots, \hat{y}_B(\mathbf{x})\right\} \tag{2}$$

Donde:

- $\hat{y}(\mathbf{x})$ es la predicción final para la observación **x**.
- moda $\{\hat{y}_1(\mathbf{x}), \hat{y}_2(\mathbf{x}), \ldots, \hat{y}_B(\mathbf{x})\}$ es la moda de las predicciones de todos los árboles del bosque.
- $\hat{y}_i(\mathbf{x})$ es la predicción del i-ésimo árbol de decisión para la observación **x**.
- B es el número total de árboles en el Random Forest.
- **x** es un vector de características (o variables predictoras) correspondientes a un solo individuo o caso en el conjunto de datos. Es una observación específica para la cual queremos hacer una predicción

La figura 1 ilustra lo anterior:

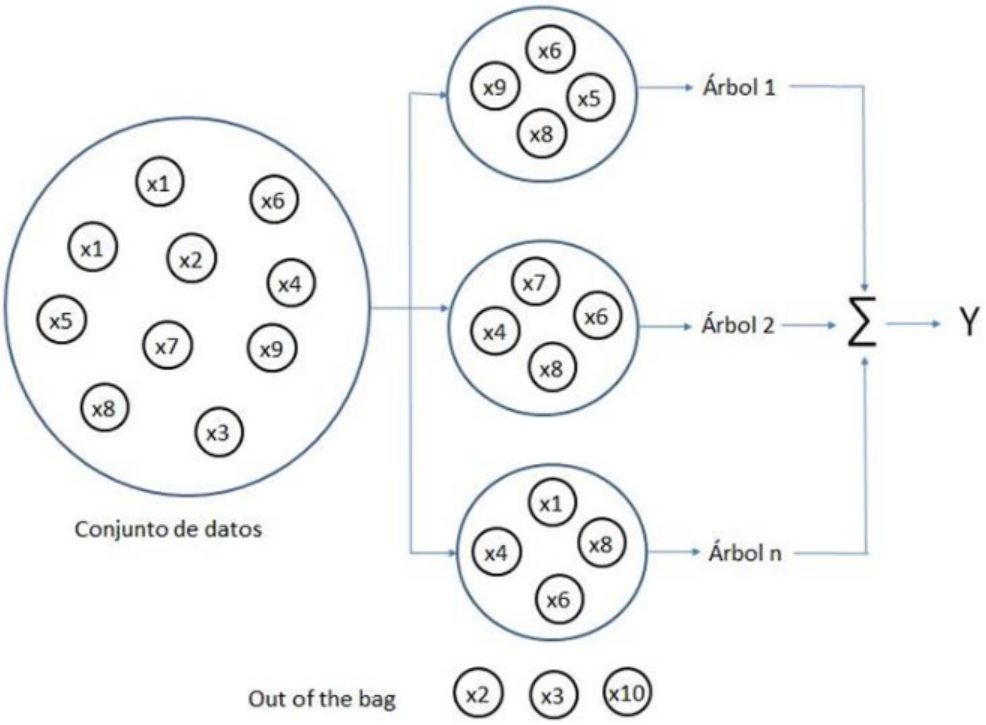


Figura 1: Algoritmo Random Forest.
Fuente:[Espinosa-Zúñiga, 2020].

Cada árbol de decisión creado en la etapa de entrenamiento utiliza un subconjunto de los datos de entrada seleccionados de manera aleatoria y emite un voto sobre la clasificación a la que corresponden. La clase final se determina mediante el cálculo de la moda de los votos emitidos por todos los árboles (en problemas de clasificación) o por el promedio de

las predicciones (en problemas de regresión). Adicionalmente, el modelo se valida utilizando los datos no utilizados en la fase de entrenamiento, conocidos como “out of the bag”, para evaluar su rendimiento de manera robusta [Espinosa-Zúñiga, 2020].

2.6.3. Máquinas de Soporte Vectorial (SVM)

El método de Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés) es un potente y popular algoritmo de aprendizaje supervisado utilizado tanto para clasificación como para regresión. El objetivo principal de SVM es encontrar el hiperplano que mejor separa las clases en un espacio de características N-dimensional [Paraıso *et al.*,]. Este modelo “se destaca especialmente en situaciones donde las instancias de datos caen en dos clases distintas y requieren un límite bien definido entre ellas” [Padmasree y Muthyam,].

En términos sencillos, la idea central de SVM es encontrar una línea (en 2D), un plano (en 3D) o un hiperplano (en dimensiones superiores) que separe las diferentes clases de datos con el mayor margen posible. Un hiperplano es una superficie que divide el espacio en dos partes. En el caso de un problema de clasificación binaria, el hiperplano separa las dos clases. Las etiquetas de estas clases suelen ser +1 o -1 para clasificación binaria.

La formulación matemática de este método, según lo planteado por [Dubaish y Jaber, 2024] se explicará a continuación.

En primer lugar, la ecuación de un hiperplano es:

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \tag{3}$$

Donde:

- **w** es el vector de pesos.
- **x** es el vector de características.
- b es el sesgo.

SVM resuelve un problema de optimización para encontrar los valores de w y b que maximizan el margen. El margen es la distancia entre el hiperplano y los puntos más cercanos de cada clase, llamados vectores de soporte.

La función de decisión para una observación **x** clasificada en +1 o -1 es:

$$f(x) = \text{sign}(w \cdot x + b)$$

Donde sign() corresponde al signo de la clase (+1 o -1).

No siempre es posible encontrar un hiperplano lineal que separe las clases en el espacio original de características. Para hacer que estos puntos sean separables, los transformamos a un espacio de mayor dimensión, para lo cual se puede usar una función de Kernel que calcula el producto interno en este nuevo espacio sin necesidad de realizar la transformación explícitamente.

Al realizar esto, la función de decisión se convierte en:

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right)$$

Donde:

- α_i son multiplicadores de Lagrange.
- y_i Etiqueta de clase de la i-ésima observación (+1 o-1).
- $K(x_i, x)$ es la función de kernel.

Sujeto a:

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

SVM es efectivo en espacios de alta dimensión y sigue siendo eficiente incluso cuando el número de dimensiones es mayor que el número de muestras. Efectivas en problemas de alta dimensionalidad y pueden utilizar diferentes kernels para capturar relaciones no lineales. Presenta robustez contra el overfitting. Sin embargo, pueden ser lentas de entrenar en grandes conjuntos de datos y difíciles de interpretar. Además, la selección del kernel y sus parámetros puede ser compleja y requiere experimentación.

2.6.4. Redes neuronales

Las redes neuronales son un modelo de aprendizaje automático inspirado en el funcionamiento del cerebro humano. Están compuestas por capas de nodos, o neuronas, que transforman los datos de entrada para producir una salida. Los nodos, Los nodos, a excepción de los de la capa de entrada, aplican una función de activación previamente determinada que introduce no linealidad al modelo. Además, los nodos están interconectados a través de pesos que representan la fuerza de las señales que se transmiten entre ellos. Cada conexión

tiene un peso asociado que se ajusta durante el proceso de entrenamiento para mejorar la precisión del modelo [Wu y Feng, 2018].

Existen distintos tipos de redes neuronales, que pueden variar en la cantidad de capas ocultas y en su arquitectura, dependiendo de la complejidad del problema, las características y naturaleza de los datos y los recursos computacionales disponibles. Algunas de las más utilizadas se mencionan a continuación [James *et al.*, 2023]:

- **Redes Neuronales Artificiales (ANN):** son las más básicas y generalmente consisten en una o dos capas ocultas. Son adecuadas para problemas sencillos de clasificación y regresión.
- **Redes Neuronales Profundas (DNN):** estas redes se utilizan en el campo del Deep Learning (aprendizaje profundo), donde se emplean estructuras que contienen múltiples capas ocultas. Esta arquitectura permite a las redes neuronales profundas (DNN) aprender representaciones más complejas de los datos y capturar relaciones no lineales que pueden ser demasiado sofisticadas para una red simple.
- **Redes Neuronales Convolucionales (CNN):** utilizadas principalmente para el procesamiento y clasificación de imágenes, las CNN tienen capas convolucionales que detectan características espaciales y patrones en los datos de entrada. La cantidad de capas ocultas y convolucionales puede variar dependiendo de la complejidad de la tarea de reconocimiento de imágenes.
- **Redes Neuronales Recurrentes (RNN):** estas redes están diseñadas para procesar secuencias de datos, como series temporales o texto. Las RNN tienen una estructura que permite la retroalimentación de la salida de una capa a la entrada de la misma capa en el siguiente paso temporal. La cantidad de capas ocultas depende de la longitud y complejidad de las secuencias.

Para la presente investigación, una red neuronal con una o dos capas ocultas podría ser suficiente dado que se cuenta con datos tabulares estructurados y de escasa complejidad. Las redes neuronales profundas pueden ser excesivas y no proporcionar un beneficio significativo adicional [Kadra *et al.*, 2021].

La formulación matemática de una red neuronal simple es:

Capa de Entrada, que recibe los datos de entrada o inputs:

$$\mathbf{x} = [x_1, x_2, \ldots, x_n]$$

Capa Oculta, donde cada neurona calcula una combinación lineal de las entradas seguida de una función de activación. Para una neurona j en la capa oculta h:

$$z_j^{(h)} = \sum_{i=1}^{n} w_{ij}^{(h)} x_i + b_j^{(h)}$$

Donde:

- $w_{ij}^{(h)}$ son los pesos de las conexiones entre las neuronas de la capa anterior y la capa oculta h.
- $b_j^{(h)}$ es el sesgo de la neurona j en la capa oculta h.

Aplicando la función de activación ϕ:

$$a_j^{(h)} = \phi(z_j^{(h)})$$

Capa de Salida:

Para una neurona k en la capa de salida o:

$$z_k^{(o)} = \sum_{j=1}^{m} w_{jk}^{(o)} a_j^{(h)} + b_k^{(o)}$$

Aplicando la función de activación (puede ser una función diferente en la capa de salida, como softmax para clasificación):

$$y_k = \phi(z_k^{(o)})$$

La figura 2 ilustra la estructura de una red neuronal de 1 capa oculta de 5 nodos y 4 características de entrada o predictores.

Las funciones de activación introducen no linealidad en el modelo, permitiendo a las redes neuronales aprender patrones complejos. Algunas de las funciones de activación más comunes son:

- **ReLU (Rectified Linear Unit):** una de las más utilizadas en redes neuronales profundas debido a su simplicidad y efectividad. Se define como $f(x) = \text{máx}(0, x)$. Por lo que para cualquier valor de entrada x, la salida será x si x es mayor que 0; de lo contrario, la salida será 0.

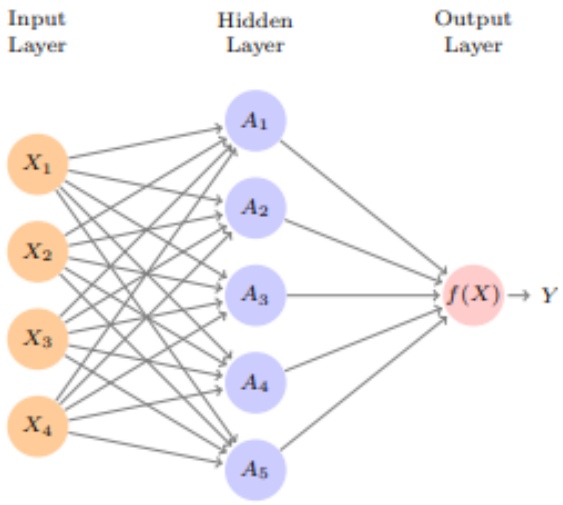


Figura 2: Ejemplo de una red neuronal de 1 capa oculta.
Fuente:[James *et al.*, 2023].

- **Función de Activación Sigmoid:** otra función comúnmente utilizada, especialmente en la capa de salida de modelos de clasificación binaria. Se define como $f(x) = \frac{1}{1+e^{-x}}$. Esta función transforma cualquier valor de entrada en un rango entre 0 y 1, lo que facilita la interpretación como una probabilidad.

ReLU (Rectified Linear Unit) y Sigmoid se ilustran en la figura 3:

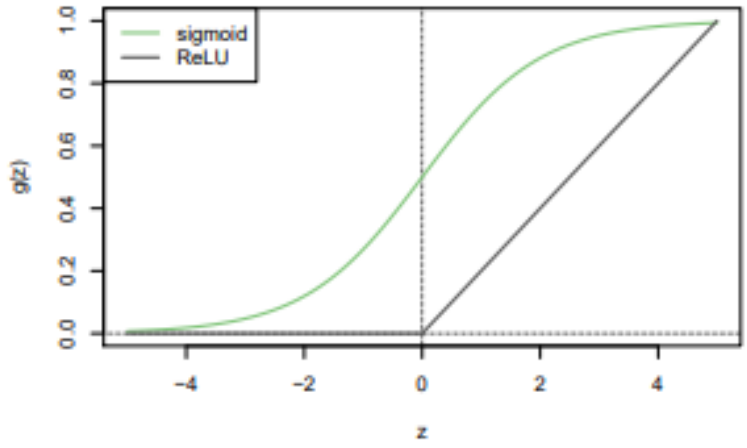


Figura 3: Ejemplo de funciones de activación.
Fuente:[James *et al.*, 2023].

2.6.5. Gradient Boosting

Los gradient boosting machines (GBMs) corresponden a un conjunto de técnicas de ML que funcionan entrenando modelos secuencialmente, de manera que cada nuevo modelo intenta corregir los errores cometidos por los modelos anteriores. Al minimizar la función de pérdida de forma iterativa, Gradient Boosting asegura que cada modelo subsecuente complemente y mejore la capacidad predictiva del conjunto, logrando una gran precisión y robustez en la predicción [Goyal *et al.*,]. Una de las implementaciones más populares y eficientes de Gradient Boosting es XGBoost (Extreme Gradient Boosting).

La formulación matemática de la función de actualización de cada modelo corresponde a:

$$F_m(x) = F_{m-1}(x) + \nu h_m(x)$$

Donde:

- $F_m(x)$ es la predicción del m-ésimo modelo del conjunto.
- $F_{m-1}(x)$ es la predicción del modelo en la iteración $m-1$, obtenida de la agregación de los m-1 modelos anteriores.
- ν La tasa de aprendizaje, también conocida como shrinkage, que controla cuánto influye el nuevo modelo $h_m(x)$ en la actualización. Es un valor escalar que típicamente está entre 0 y 1.
- $h_m(x)$ es el modelo base (o débil) en la iteración m, que está ajustado a los residuos de la iteración anterior. En el caso de Gradient Boosting, este suele ser un árbol de decisión.

En este estudio, se utilizará el modelo XGBoost, del inglés Extreme Gradient Boosting que es una implementación avanzada y optimizada de la técnica de Gradient Boosting. Desarrollado para mejorar tanto la velocidad como el rendimiento de los modelos de Gradient Boosting, XGBoost ha ganado popularidad rápidamente en la comunidad de ciencia de datos y ha sido ampliamente utilizado en competiciones de aprendizaje automático y minería de datos[Nalluri *et al.*, 2020]. El principal impulsor del éxito de XGBoost es su escalabilidad en una gran variedad de escenarios. Esta característica, junto con su flexibilidad y adaptabilidad, explica su rendimiento robusto y eficiente.[Goyal *et al.*,]. XGBoost incorpora varias mejoras, como la paralelización del entrenamiento, la regularización para prevenir el sobreajuste, una gestión eficiente de datos faltantes y un método de poda basado en la reducción de pérdidas. Entre sus desventajas está el alto consumo de recursos computacionales con grandes bases de datos. Además, el ajuste de parámetros puede no ser trivial y requiere

un conocimiento detallado del modelo. También, XGBoost solo trabaja con datos numéricos, por lo que es necesario realizar una transformación de los datos no numéricos antes de utilizarlos [Espinosa-Zúñiga, 2020].

En la figura 4 se puede ver la progresión de los algoritmos basados en árboles de decisión, de los cuales se utilizarán 2 en este estudio, Random Forest y XGBoost:

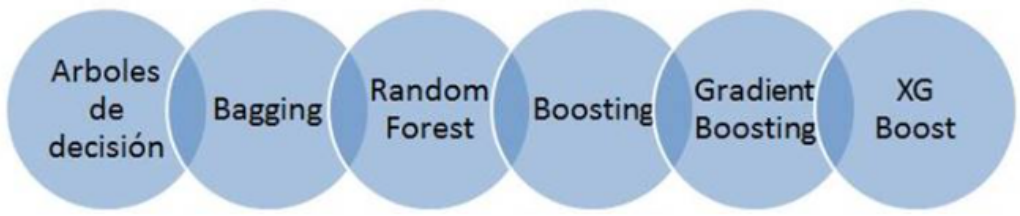


Figura 4: Evolución de los algoritmos basados en árboles de decisión.
Fuente:[Espinosa-Zúñiga, 2020].

2.7. Relevancia de atributos

En el ámbito del machine learning, los atributos o características de los datos juegan un papel crucial en la construcción de modelos predictivos efectivos. La importancia de los atributos se refiere a la influencia que cada uno de ellos tiene sobre la variable dependiente que se intenta predecir. Identificar y entender cuáles atributos impactan más en las relaciones con la variable dependiente es fundamental para mejorar la precisión, interpretabilidad y eficiencia de los modelos.

La importancia de atributos en modelos lineales como la regresión lineal o logística se mide a través de los coeficientes asociados a cada uno. Un coeficiente alto sugiere que el atributo tiene un impacto significativo en la variable dependiente. Esto permite una interpretación directa de la importancia de las características de los datos de entrada con respecto a la variable de salida [Saarela y Jauhiainen, 2021].

En el caso de los modelos basados en árboles de decisión, como Random Forest o XGBoost utilizan técnicas como el índice de Gini para evaluar la importancia de los atributos. Estas métricas cuantifican cuánto reduce un atributo la impureza en una partición de los datos. Durante la construcción de un árbol de decisión, el algoritmo evalúa posibles divisiones de los datos en cada nodo. Para cada posible división, el índice de Gini se calcula para los nodos resultantes. La división que resulta en la menor suma ponderada del índice de Gini en los nodos hijos se selecciona como la mejor división. Mientras menor sea la impureza, más importante esa característica en el modelo. En el caso de la clasificación binaria, este índice tiene la siguiente formulación matemática [Saarela y Jauhiainen, 2021]:

$$\text{Gini} = p_1(1 - p_1) + p_2(1 - p_2)$$

Donde p_1 y p_2 representan las proporciones de las dos clases.

El índice de Gini toma valores entre 0 y 1, donde un índice de 0 indica pureza total, lo que significa que todas las muestras en el nodo pertenecen a una sola clase y un índice cercano a 1 indica una distribución uniforme entre las clases, lo que significa alta impureza.

Por otra parte, para modelos complejos y no lineales como las redes neuronales, existen técnicas avanzadas que proporcionan interpretaciones detalladas de la importancia de los atributos. Una de ellas es LIME (Local Interpretable Model-agnostic Explanations), un algoritmo que produce interpretaciones locales e independientes del modelo. El término ”model-agnostic”significa que una técnica o método es independiente del tipo de modelo de machine learning utilizado, es decir, lo considera una caja negra de la cual no se pueden hacer suposiciones [Mishra *et al.*, 2017].

La formulación matemática de este algoritmo, basada en [Mishra *et al.*, 2017] y [Saarela y Jauhiainen, 2021], se explica a continuación:

Se tiene un modelo complejo f y una instancia x cuya predicción se desea interpretar. LIME genera un modelo interpretable g (por ejemplo, una regresión lineal) que aproxima f en el vecindario de x.

Para ello se crea un conjunto de datos sintéticos Z, donde cada punto $z_i \in Z$ es una variación de x a la cual se le asigna un peso dependiendo de la distancia entre ambos.

En seguida, se entrena el modelo g usando los datos sintéticos Z y los pesos $\pi_x(z)$. El objetivo es minimizar la siguiente función de pérdida:

$$\xi(g, f, \pi_x) = \sum_{z_i \in Z} \pi_x(z_i) \cdot (f(z_i) - g(z_i))^2 + \Omega(g)$$

Donde:

- $\pi_x(z_i)$ es el peso asignado a z_i.
- $f(z_i)$ es la predicción del modelo complejo para z_i.
- $g(z_i)$ es la predicción del modelo interpretable para z_i.
- $\Omega(g)$ es un término de regularización que favorece la simplicidad del modelo g.

La interpretación se realiza utilizando los coeficientes del modelo interpretable g para explicar la predicción de f en el vecindario de x. Los coeficientes indican la importancia de cada característica en la predicción del modelo original.

En el caso particular de esta investigación, resulta fundamental comprender qué variables o combinación de variables explican mejor el riesgo suicida. La identificación de estos atributos clave no solo mejorará la precisión de los modelos predictivos, sino que también proporcionará valiosos insights que pueden ser utilizados para la prevención y la intervención temprana. Utilizando técnicas de evaluación de la importancia de los atributos, como las mencionadas anteriormente, se pueden determinar las variables que tienen el mayor impacto en la predicción del riesgo suicida, lo que es esencial para desarrollar estrategias efectivas y basadas en datos.

2.8. Métricas de evaluación de modelos de clasificación

Para determinar qué tan buen desempeño ha tenido un modelo en comparación con otro, es fundamental utilizar medidas cuantitativas que evalúen su rendimiento. Estas métricas permiten no solo cuantificar la precisión y efectividad de un modelo, sino también comparar diferentes modelos bajo los mismos criterios de evaluación. En esta sección, se discutirán las métricas de evaluación más comúnmente utilizadas en modelos de clasificación [Kumar, 2024].

2.8.1. Accuracy

Accuracy (exactitud) es una de las métricas de evaluación más comúnmente utilizadas para modelos de clasificación. Se define como la proporción de predicciones correctas realizadas por el modelo sobre el total de predicciones. Es decir, mide qué tan a menudo el clasificador es correcto [Kumar, 2024]. Es un valor que va de 0 a 1, siendo 1 el mejor caso posible. Lo anterior se calcula como la suma de las predicciones acertadas dividido por el total de predicciones, como se ilustra en la siguiente fórmula [Vujovic, 2021]:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde:

- TP (True Positives) son los casos verdaderamente positivos que el modelo ha predicho correctamente como tal.
- TN (True Negatives) son los casos verdaderamente negativos que el modelo ha predicho correctamente como tal.
- FP (False Positives) son los casos negativos que el modelo ha predicho incorrectamente como positivos. Esto se conoce como error tipo I.

- FN (False Negatives) son los casos positivos que el modelo ha predicho incorrectamente como negativos. Esto se conoce como error tipo II.

2.8.2. Recall (Sensibilidad)

El recall, también conocido como sensibilidad o tasa de verdaderos positivos, es una métrica de evaluación utilizada en problemas de clasificación que mide la capacidad del modelo para identificar correctamente todas las instancias positivas. Es útil cuando la existencia de falsos negativos, o error tipo II, es más preocupante que la cantidad de falsos positivos. En este estudio en particular, se considera más peligroso no detectar un caso de ideación suicida cuando existe, que levantar una falsa alarma sobre ideación cuando en realidad no la había [Kumar, 2024].

El recall se calcula utilizando la siguiente fórmula:

$$\text{Recall} = \frac{TP}{TP + FN}$$

Donde:

- TP (True Positives) son los casos verdaderamente positivos que el modelo ha predicho correctamente como tal.
- FN (False Negatives) son los casos positivos que el modelo ha predicho incorrectamente como negativos. Esto se conoce como error tipo II.

2.8.3. Matriz de Confusión

La matriz de confusión es una herramienta esencial para evaluar el rendimiento de un modelo de clasificación donde las salidas son dos o más clases. Proporciona una visión detallada de las predicciones del modelo, mostrando cómo se distribuyen las predicciones correctas e incorrectas en cada clase [Kumar, 2024]. Lo anterior se puede visualizar en la tabla 1.

	Predicción Positiva	**Predicción Negativa**
Real Positiva	TP	FN
Real Negativa	FP	TN

Tabla 1: Matriz de Confusión

Donde:

- TP (True Positives) son los casos verdaderamente positivos que el modelo ha predicho correctamente como tal.
- TN (True Negatives) son los casos verdaderamente negativos que el modelo ha predicho correctamente como tal.
- FP (False Positives) son los casos negativos que el modelo ha predicho incorrectamente como positivos. Esto se conoce como error tipo I.
- FN (False Negatives) son los casos positivos que el modelo ha predicho incorrectamente como negativos. Esto se conoce como error tipo II.

2.9. Python

Python es un lenguaje de programación de alto nivel, interpretado y de uso general, conocido por su simplicidad y legibilidad. Ha ganado una enorme popularidad en la comunidad de ciencia de datos y aprendizaje automático gracias a su variada colección de bibliotecas y herramientas especializadas para todo tipo de tareas. Además de estas bibliotecas que proporcionan poderosas funcionalidades para tareas complejas, Python tiene incorporada una librería estándar implementada en lenguaje C, por lo que incluso sus funciones primitivas son eficientes [Challenger-Pérez *et al.*, 2014]. Considerando que el presente estudio requiere de estas capacidades, se ha optado por Python como el lenguaje a utilizar.

2.9.1. Librerías de ciencia de datos y machine learning

En esta sección, exploraremos las principales características y ventajas de las librerías de Python que más popularmente se utilizan para tareas de ciencia de datos y machine learning.

- **Pandas:** Es la librería más utilizada para la etapa de preparación de los datos. Cuenta con una gran variedad de opciones para la entrada, salida y manipulación de datos, lo cual la convierte en la mejor opción para esta tarea [Stančin y Jović, 2019].
- **Scikit-Learn:** Es una librería de machine learning, muy popular para la implementación de algoritmos clásicos de ML. Se integra fácilmente con Pandas y puede utilizarse para análisis de datos [Saabith *et al.*, 2020].
- **Keras:** Es una librería de machine learning especializada en la implementación de redes neuronales. Funciona sobre TensorFlow (una librería para deep learning) y permite construir cada capa de la red, las funciones de activación y otros parámetros de forma personalizada [Gevorkyan *et al.*, 2019].
- **Matplotlib:** Es una poderosa librería para visualización de datos [Saabith *et al.*, 2020].

- **Rapidfuzz:** Es una librería especializada en la búsqueda de coincidencias aproximadas en cadenas de texto (fuzzy matching)[github,].

2.10. Jupyter Notebook

Otra gran herramienta que se utilizará en este estudio es Jupyter Notebook, una plataforma que funciona como laboratorio virtual para trabajar con datos, código (en lenguajes como Python o R), visualizaciones y flujos de trabajo que facilitan la creación y el uso compartido de documentos. Estos documentos se guardan en repositorios online y pueden interactuar con otras herramientas disponibles [Randles *et al.*, 2017]. Además, Jupyter Notebook ofrece un entorno interactivo donde se puede ejecutar y modificar el código en tiempo real, lo que facilita la experimentación y el análisis. Otra ventaja es la facilidad de realizar debug del código dado que se puede trabajar en distintas celdas, lo que permite identificar errores más fácilmente.

2.11. Psytoolkit

Esta plataforma gratuita ofrece una variedad de herramientas para la creación, administración y análisis de experimentos psicológicos. Los participantes pueden acceder al experimento a través del enlace proporcionado y comple- tarlo en línea. PsyToolkit registra automáticamente las respuestas de los participantes y las almacena en una base de datos segura.

La elección de PsyToolkit como plataforma para alojar los instrumentos de medición se basa en diversas consideraciones. En primer lugar, su interfaz simple y fácil de usar ofrece una experiencia más intuitiva tanto para los participantes como para el investigador. A diferencia de otras herramientas como Google Docs, PsyToolkit permite una mayor personalización de la apariencia de los cuestionarios, lo que resulta fundamental para mantener la coherencia visual con el diseño general del estudio. Además, PsyToolkit ofrece la posibilidad de obtener estadísticas básicas, lo que facilita el análisis preliminar de los datos recopilados. El uso de PsyToolkit implica un pequeño grado de programación; no obstante, la plataforma ofrece una amplia documentación que guía a los usuarios en el proceso. Esta documentación detallada proporciona instrucciones claras sobre cómo trabajar con la plataforma, lo que facilita el aprendizaje y la implementación de las funciones necesarias para personalizar los cuestionarios y realizar análisis de datos [PsyToolkit,].

CAPÍTULO 3
PROPUESTA DE SOLUCIÓN

3.1. Metodología Propuesta

La solución propuesta para abordar el problema se centra en explorar la posible relación predictiva existente entre las variables estudiadas y la ideación suicida, así como cualquier otra asociación relevante entre ellas, utilizando algoritmos de machine learning. Dichos modelos son comparados para determinar cuál se ajusta mejor al objetivo de estudio y al tipo de datos utilizados. Todo este proceso se lleva a cabo en Jupyter Notebook utilizando el lenguaje Python y sus librerías especializadas en tareas de machine learning.

A continuación, se especifican las etapas que componen la metodología de trabajo y el diseño metodológico, junto con un diagrama explicativo ilustrado en la figura 5:

1. **Recopilación de datos:** En esta etapa, se crea un instrumento que comprende varios tests psicológicos para medir las variables de interés y preguntas para recopilar datos demográficos de los participantes. Una vez completada la creación del instrumento, se solicita la participación de los estudiantes para que respondan el cuestionario. Posteriormente, se recopilan las respuestas para obtener la base de datos con la cual se lleva a cabo el estudio.

2. **Limpieza y transformación de los datos:** En esta etapa, se realiza una selección de los cuestionarios que no contienen datos faltantes o nulos. Posteriormente, se aplican las transformaciones necesarias a los datos para prepararlos adecuadamente para el análisis posterior.

3. **Análisis estadístico:** Se lleva a cabo un análisis utilizando estadística descriptiva para comprender el comportamiento de los datos. Se exploran las relaciones entre variables y su prevalencia en la muestra estudiada, además de otros aspectos relevantes.

4. **Diseño e implementación de los modelos de ML:** Se desarrollan y entrenan algoritmos de ML para predecir la ideación suicida y encontrar asociaciones relevantes. Esto implica seleccionar los algoritmos adecuados, dividir los datos en conjuntos de entrenamiento y prueba, optimizar los parámetros de los modelos y evaluar su rendimiento.

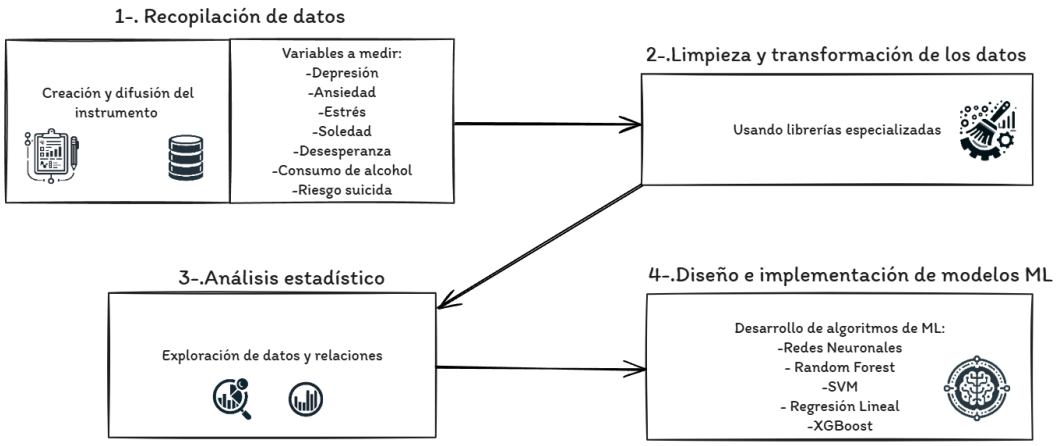


Figura 5: Diseño metodológico de la propuesta de solución.
Fuente: Elaboración propia.

3.2. Recopilación de datos

3.2.1. Creación del Instrumento de Medición

El primer paso crucial en la implementación de la solución es la selección de las variables que se medirán y los instrumentos para hacerlo. Esta decisión se fundamenta en una revisión exhaustiva de la literatura científica, donde se identifican las variables más relevantes para abordar el problema de investigación y los instrumentos más utilizados en la practica considerando la validez de su utilización en población chilena y la extensión medida en cantidad de preguntas. Además, se solicita la retroalimentación de académicos y psicólogos expertos en el área para validar la selección de variables e instrumentos y asegurar su idoneidad para el estudio en cuestión. Finalmente, las variables a considerar y sus respectivos instrumentos son:

- Depresión, estrés y ansiedad: Escala Dass-21
- Sensación de Soledad: Escala de soledad de UCLA versión corta
- Desesperanza: Escala de desesperanza de Beck versión corta
- Nivel de riesgo del consumo de alcohol: Test de AUDIT
- Riesgo suicida (ideación e intento suicida): Escala de Suicidalidad de Okasha

Estos instrumentos son posteriormente alojados en la plataforma en línea especializada en investigación en psicología, Psytoolkit. Para realizar esto, se deben ingresar manualmente las preguntas de los distintos cuestionarios utilizando un lenguaje de scripting propio de la

plataforma. Este lenguaje es de fácil uso y cuenta con una gran cantidad de documentación y ejemplos para guiar al investigador. Además, Psytoolkit ofrece un repositorio con una amplia variedad de cuestionarios previamente implementados que se pueden copiar y adaptar para su uso. En este trabajo en particular, las preguntas se ingresan manualmente, y se ajustan parámetros como la cantidad de preguntas o cuestionarios por página, los colores y destacados de algunas palabras, la inclusión de imágenes, entre otros (figura 6).

También se puede personalizar la vigencia del cuestionario, ya sea poniendo un límite de participantes o de tiempo. En este caso, se deja libre ya que se espera conseguir la mayor cantidad posible de participantes durante un tiempo arbitrario. Además, se debe configurar el idioma español y el esquema de colores a utilizar, junto con otros detalles como la presencia de un contador de páginas o de una cuenta regresiva, en caso de ser necesario que el cuestionario sea respondido en un tiempo determinado por cada participante.

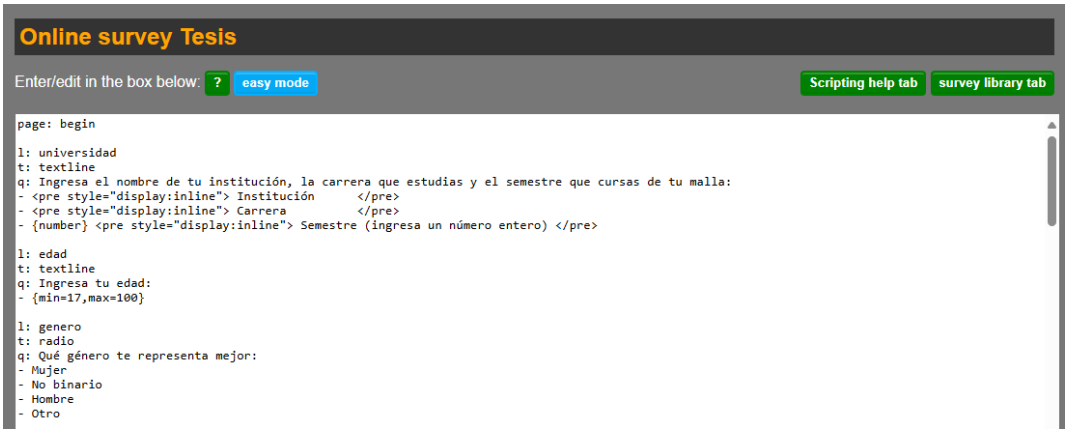


Figura 6: Interfaz para la implementación del instrumento.
Fuente: Elaboración propia.

La página de bienvenida (figura 7) incluye información básica sobre el estudio, los datos de la estudiante a cargo y el consentimiento informado que se debe aceptar antes de comenzar el cuestionario.

El consentimiento informado (figura 8) es un aspecto fundamental en cualquier estudio científico que involucre la participación de personas. Garantiza que los participantes comprendan plenamente los propósitos, procedimientos, riesgos y beneficios de su participación, y les otorga la libertad de tomar una decisión informada sobre si desean participar o no. Además de ser un requisito ético y legal en la investigación, el consentimiento informado protege los derechos y la privacidad de los participantes, promoviendo la transparencia, la confianza y la integridad en el proceso de investigación.



Figura 7: Página de bienvenida al cuestionario en PsyToolKit.
Fuente: Elaboración propia.

3.2.2. Difusión del Cuestionario

En esta fase del estudio, se lleva a cabo la solicitud de colaboración a diversas instituciones educativas mediante el envío de correos electrónicos a directivos. El objetivo de esta iniciativa es solicitar la colaboración de dichas instituciones para difundir el cuestionario entre sus estudiantes universitarios de pregrado. La participación de estos estudiantes resulta fundamental para obtener una muestra representativa y diversa.

En los correos electrónicos enviados a los directivos, se proporciona una breve descripción del estudio, informando acerca de la confidencialidad de los datos y destacando su relevancia en el ámbito de la salud mental por su potencial impacto en la mejora de herramientas de detección temprana. Además, se explica el procedimiento para acceder al cuestionario y se solicita ayuda con su difusión entre los alumnos a través de los medios de comunicación internos de la institución, como correos electrónicos, plataformas educativas o redes sociales.

3.3. Limpieza y transformación de los datos

Una vez transcurrida la etapa de recopilación de datos, se puede descargar el archivo CSV desde la página de Psytoolkit. Para hacerlo, primero debemos decidir si incluiremos aquellas encuestas que tengan valores faltantes o estén incompletas. En este caso, solo se trabaja con las encuestas totalmente respondidas y sin datos faltantes, no sólo para facilitar el proceso de limpieza posterios, sino que además para respetar la decisión de los participantes de dejar de

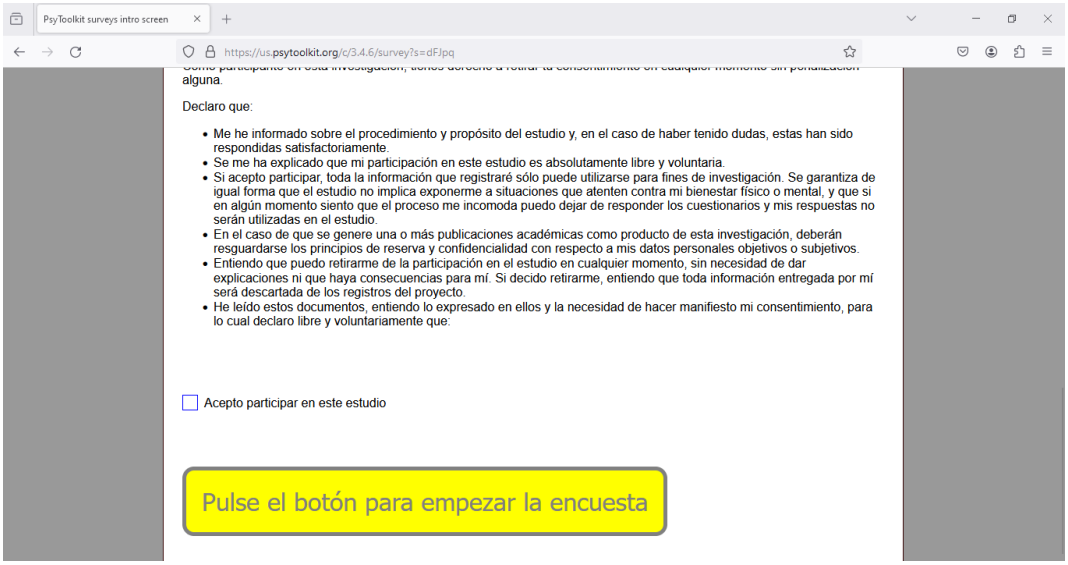


Figura 8: Página de bienvenida al cuestionario: Extracto del consentimiento informado.
Fuente: Elaboración propia.

responder, lo cual fue descrito en el consentimiento informado. Luego, se presiona el botón de preparación de datos para que la plataforma los prepare para la descarga, lo cual toma unos segundos. Finalmente, se descarga el archivo comprimido en formato ZIP 9.

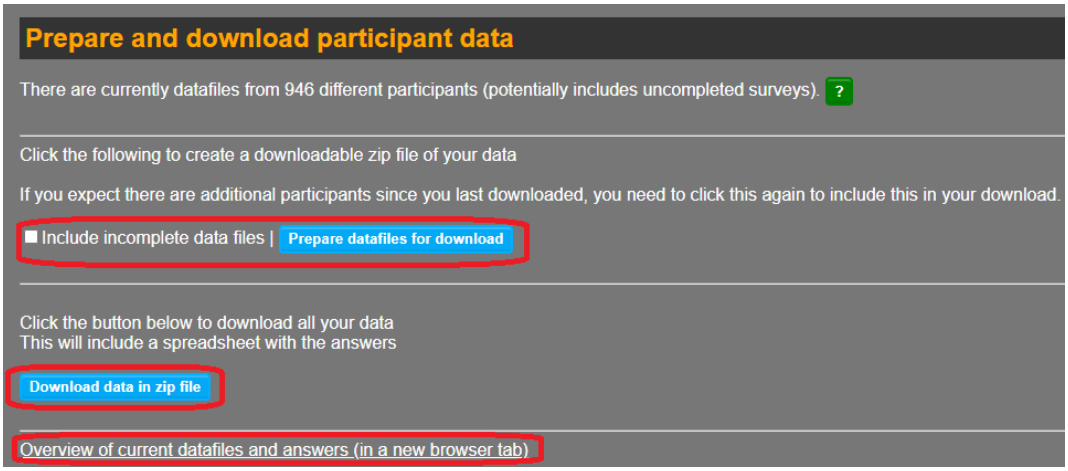


Figura 9: Descarga de datos desde la plataforma.
Fuente: Elaboración propia.

Una vez obtenido el archivo CSV, se cargan los datos en un documento de Jupyter Notebook utilizando la librería pandas de Python. Pandas permite manejar los datos mediante estructuras denominadas DataFrames (DF), lo que facilita su manipulación y análisis.

La primera tarea es estandarizar los datos entregados manualmente por los participantes. Estos datos incluyen la universidad a la que pertenecen, la carrera que estudian y el semestre

de la malla que están cursando. La estandarización es crucial para garantizar la consistencia y precisión del análisis posterior. Este proceso implica convertir todos los valores a un formato uniforme, corregir errores tipográficos y unificar distintas variantes de los mismos nombres. Además, es importante manejar cualquier dato faltante o inconsistente para asegurar la integridad del conjunto de datos. Para facilitar esta tarea, primero se convierten todos los nombres a minúsculas y se eliminan espacios en blanco.

Para la estandarización de los nombres de universidades y carreras, se utiliza la librería rapidfuzz, la cual permite realizar coincidencias aproximadas (fuzzy matching) de manera eficiente. Este método se prefiere sobre las técnicas tradicionales de pandas debido a su capacidad para manejar variaciones en la entrada de datos, tales como errores tipográficos y abreviaciones inconsistentes. La función extractOne de rapidfuzz se emplea para encontrar la mejor coincidencia entre los nombres ingresados por los participantes y una lista predefinida de nombres estandarizados. Por ejemplo, para la Universidad Técnica Federico Santa María se prepara el siguiente par clave-valor:

```
'universidad tecnica federico santa maria': ['universidad
  tecnica federico santa maria', 'usm', 'utfsm']
```

Donde la clave es el nombre estandarizado y los valores son posibles variantes ingresadas por los participantes. Lo mismo se realiza para todas las universidades y carreras participantes.

Se establece un umbral de similitud del 80 %, lo que permite asegurar que solo se acepten coincidencias con un alto grado de similitud. Este enfoque no solo reduce la necesidad de enumerar exhaustivamente todas las variantes posibles, como requería el uso de pandas, sino que también mejora la precisión y robustez del proceso de estandarización al capturar y corregir variaciones imprevistas en los datos. Además, se crea una lista donde el algoritmo almacena los nombres para los que no se encuentra una coincidencia satisfactoria, permitiendo que estos sean revisados posteriormente de forma manual.

La revisión manual se lleva a cabo en Microsoft Excel, utilizando los filtros de las columnas de interés (universidad y carrera) para identificar y eliminar duplicados con alguna variante. Además, se eliminan las filas que contienen datos nulos, datos de prueba y datos de personas que no cumplen con el requisito de ser estudiantes de pregrado, como los estudiantes de magíster o doctorado, para asegurar que la base de datos solo contenga información pertinente al la población de interés. Una vez finalizada esta tarea, se obtiene un archivo con los datos de 591 estudiantes de pregrado de 28 universidades del país, pertenecientes a 71 carreras diferentes.

3.4. Análisis Estadístico

Con el objetivo de conocer la distribución de los datos y obtener una mejor comprensión del perfil de los participantes y la diversidad de la muestra, se lleva a cabo un análisis estadístico descriptivo. Este análisis permite identificar características clave de los datos, así como patrones y tendencias dentro de la muestra. A continuación, se presentan los resultados de este análisis.

3.4.1. Datos demográficos

La edad de los participantes varía entre 17 y 60 años, con una mediana de 21 años. La distribución de edades muestra un sesgo positivo, concentrando la mayoría de los datos en el rango inferior de edades, con algunos valores extremos hacia la derecha. La distribución detallada de las edades se puede observar en la figura 10.

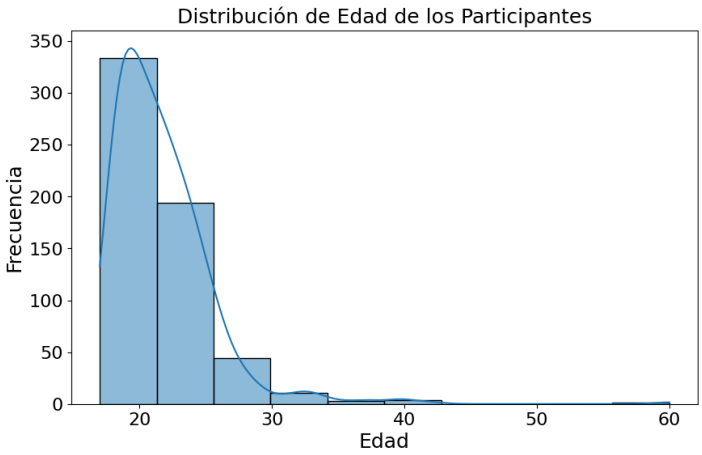


Figura 10: Distribución de las edades de los participantes.
Fuente: Elaboración propia.

Respecto al género, la presencia de diagnóstico de neurodivergencia o trabajo, las distribuciones se pueden observar en las tablas 2, 3 y 4.

Género	Cantidad
Mujer	323
No binario	21
Hombre	238
Otro	9

Tabla 2: Distribución por género

Las carreras se dividen por área o campo de estudio, para lo cual se crea una nueva columna en el archivo que indica este dato y se visualiza en la tabla 5.

Neurodivergencia	Cantidad
Sí	124
No	467

Tabla 3: Recuento de Neurodivergencia

Trabajo	Cantidad
No	398
Part time	91
Independiente	70
Full time	32

Tabla 4: Recuento de Tipos de Trabajo

Área	Cantidad
Ingeniería y Tecnología	218
Ciencias de la Salud	104
Ciencias Sociales	78
Otros	69
Ciencias Exactas y Naturales	46
Humanidades y Artes	30
Educación	26
Desconocida	20

Tabla 5: Recuento de Áreas de Estudio

Finalmente, se determina que la mediana de los estudiantes se encuentra cursando el quinto semestre de su malla.

3.4.2. Variables psicológicas

Se lleva a cabo una primera observación agrupando los datos según la variable objetivo, la ideación suicida, con el fin de identificar las principales diferencias a priori. En la tabla 8, se puede observar que los estudiantes con ideación suicida presentan puntajes promedio más altos en las variables que miden sintomatología psicopatológica, así como también en la medición de consumo problemático de alcohol. Por otro lado, la edad y el semestre en curso muestran valores bastante similares entre los grupos con y sin ideación suicida. Esta comparación preliminar ayuda a identificar las áreas donde existen diferencias notables, proporcionando una base para un análisis más detallado.

Como se muestra en la tabla 7, los valores de diversas métricas también varían según el género.

Las variables psicólogicas medidas por el instrumento Dass-21 (estrés, depresión y ansiedad)

Ideación	Depresión	Ansiedad	Estrés	Soledad	Desesperanza	Consumo	Edad	Semestre
No	5.5	5.4	8.4	8.1	0.4	2.8	22.0	5.5
Sí	11.1	9.4	11.5	10.1	1.3	3.9	21.3	4.7

Tabla 6: Datos agrupados por presencia de ideación suicida

Género	Depresión	Ansiedad	Estrés	Soledad	Desesperanza	Consumo	Edad	Semestre
Hombre	6.8	5.3	8.0	8.6	0.6	3.5	22.0	5.6
Mujer	8.1	8.1	10.6	9.0	0.7	3.1	21.6	4.9
No binario	10.8	10.1	12.7	10.5	1.8	2.9	21.4	3.8
Otro	13.0	9.0	12.2	11.3	1.3	4.8	21.4	4.8

Tabla 7: Valores de diversas métricas por género

muestran las siguientes prevalencias (figura 11).

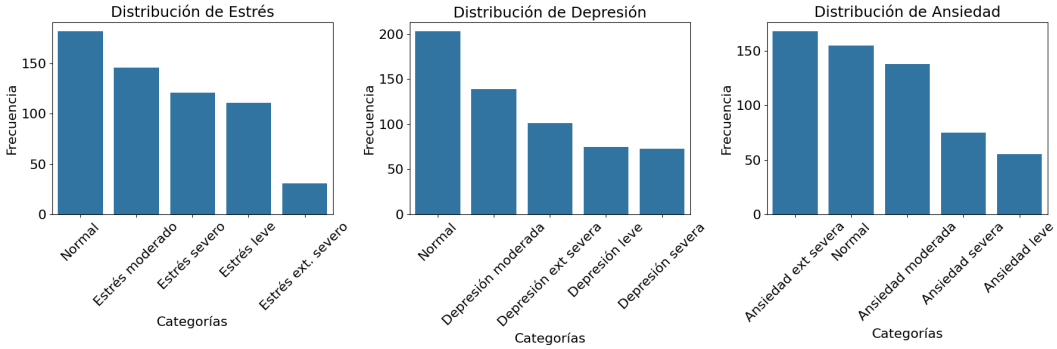


Figura 11: Distribución de depresión, ansiedad y estrés medidas por el test Dass-21. Fuente: Elaboración propia.

El resto de las variables medidas a través de instrumentos se visualizan en la siguiente figura 12:

Como se muestra en la tabla 8, las clases que representan la presencia o ausencia de ideación suicida se encuentran bastante balanceadas, lo cual se tendrá en cuenta para la implementación de los modelos de aprendizaje automático.

Ideación	Cantidad
No	349
Sí	242

Tabla 8: Recuento de ideación suicida en estudiantes universitarios

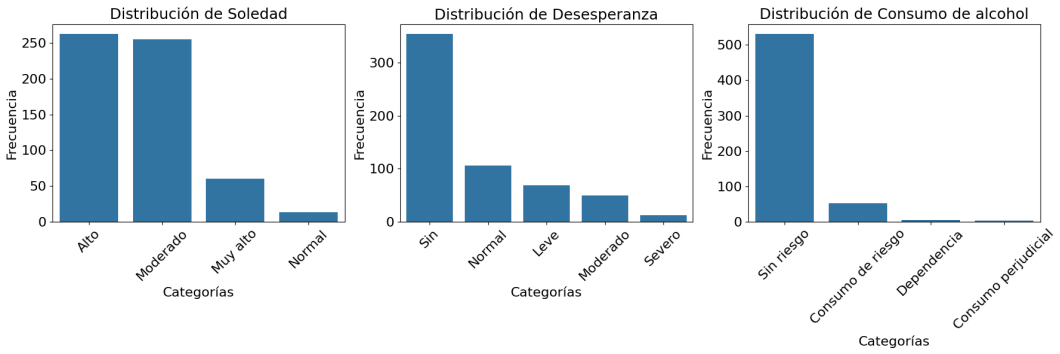


Figura 12: Distribución de soledad, desesperanza y consumo de alcohol.
Fuente: Elaboración propia.

3.4.3. Estadística inferencial

Para entender el comportamiento de las variables en el estudio, es fundamental explorar posibles correlaciones o colinealidades entre ellas. Estas relaciones pueden interferir significativamente en el desempeño de los modelos de aprendizaje automático, como la regresión logística. Identificar y manejar la colinealidad es crucial para asegurar que los modelos sean precisos y generalizables.

La primera etapa de este análisis consiste en determinar si la distribución de los datos es normal. La normalidad de los datos puede influir en la elección de los métodos estadísticos adecuados para evaluar las correlaciones. Para esto, se emplean pruebas estadísticas como Shapiro-Wilk y Kolmogorov-Smirnov, consideradas unas de las más poderosas para detectar desviaciones de la normalidad.

Las pruebas mencionadas rechazaron la hipótesis de distribución normal para todas las variables numéricas: depresión, ansiedad, estrés, soledad, desesperanza, consumo de alcohol, edad y semestre en curso. Con esto en mente, se procede a elegir los estadísticos de correlación adecuados.

Para medir la correlación entre datos númericos se utiliza el coeficiente de correlación de Spearman. Para las variables categóricas se utiliza Chi-cuadrado.

Spearman arrojó correlaciones altas y significativas entre las variables ansiedad y estrés, y entre las variables edad y semestre, naturalmente. También se reportaron correlaciones moderadas y significativas entre las variables depresión y estrés, así como entre depresión y ansiedad.

Repitiendo el mismo test de Spearman con las variables numéricas y el riesgo suicida, se encontró que el nivel de riesgo suicida tiene una correlación alta y significativa con la sintomatología depresiva. Además, tiene una correlación moderada y significativa con la ansiedad, el estrés, la soledad y la desesperanza.

Para detectar asociaciones entre variables categóricas, primero se debe abordar el desbalance en las categorías de las variables. Por ejemplo, la variable Región está altamente desbalanceada, ya que la mayoría de los participantes se concentran en la Región Metropolitana, mientras que las demás regiones tienen casos atípicos. En este caso, conviene eliminar la variable por completo para evitar que su dominancia en la muestra se confunda con una dominancia en la población real.

Por otra parte, existe un desbalance en la variable género, donde las categorías dominantes son Hombre y Mujer, con pocas ocurrencias en No binario y Otros. En este caso, se decide dejar solo las categorías Hombre y Mujer durante la prueba de Chi cuadrado, pero se vuelven a considerar para la implementación de los modelos de aprendizaje automático.

Con esto en cuenta, la prueba de Chi cuadrado muestra una asociación muy significativa entre el área de estudio y el género. También existe una asociación significativa entre el área de estudio y la presencia de neurodiversidad, así como entre el área de estudio y el trabajo. Finalmente, se detecta una asociación significativa entre el género y la presencia de un diagnóstico de neurodivergencia.

Repitiendo el análisis con las variables categóricas y la ideación suicida. Se encuentra una asociación altamente significativa con el género y la presencia de un diagnóstico de neurodivergencia, además de una asociación significativa con el área de estudio.

Luego del análisis de correlaciones y asociaciones, que nos ayuda a conocer cómo interactúan los datos, se debe realizar un análisis de colinealidad para explorar la posibilidad de que algunas variables se puedan explicar como la combinación lineal de otras variables, lo cual entorpecería la regresión logística. Para ello, se utiliza el Variance Inflation Factor (VIF), que nos proporciona información sobre la ocurrencia de multicolinealidad entre las variables. A través de esta medida, se descubre una alta colinealidad en la variable edad, la cual se decide eliminar. Una vez eliminada esta variable, se procede a realizar el análisis nuevamente, lo que revela una alta colinealidad en la variable estrés, que también se elimina.

3.5. Diseño e implementación de los modelos de ML

Como se mencionó anteriormente, los modelos a utilizar corresponden a regresión logística, random forest, SVM, redes neuronales y XGBoost. La programación de todos estos modelos se realiza utilizando librerías de Python especializadas para esta tarea. Respecto a las variables independientes, se decide no incorporar la región de procedencia debido al gran desbalance encontrado en los datos. Por otra parte, se decide entrenar los modelos con dos tipos de datos diferentes: los puntajes en bruto obtenidos en los tests por los participantes y las categorías en las que dichos puntajes se encuentran (“Normal”, “Leve”, “Moderada”, entre otras). De esta manera, se espera evaluar cuál de ellos proporciona un mejor rendimiento y qué tipo de representación de los datos es más efectiva para predecir la ideación suicida.

Dado que la regresión logística no se desempeña de la mejor manera ante la presencia de colinealidad en sus variables independientes, se decide eliminar del conjunto de predictores las variables “Edad” y “Estrés”, que muestran cifras altas de colinealidad. Aunque el resto de los modelos no se ven afectados por la colinealidad de la misma manera que la regresión logística, se decide eliminar estas variables para todos los modelos, con el fin de que la comparación de sus desempeños se realice en igualdad de condiciones.

Respecto al pipeline de trabajo, primero se divide el dataset dejando un 80 % para la fase de entrenamiento y un 20 % de los datos para la fase de prueba. Luego, se realiza un procedimiento para escalar los datos a valores estandarizados para todas las variables. Esto se hace con el objetivo de evitar sesgos en los modelos debido a la diferencia en la representación de algunos datos numéricos.

Después de la estandarización, se procede al entrenamiento de los modelos. En el caso de las redes neuronales, se utiliza RandomizedSearchCV para la búsqueda de hiperparámetros, permitiendo la inclusión de parámetros como el número de neuronas por capa, el número de capas, el optimizador y la tasa de aprendizaje. Además, se incluye la opción de dropout, regularizadores y batch normalization para mejorar el rendimiento del modelo. Para el resto de los modelos se utiliza GridSearchCV para encontrar los mejores hiperparámetros, también con Cross-Validation de 5 folds. En cada iteración se obtienen las métricas de evaluación de desempeño y se calcula la importancia de las características. Luego, se realiza una selección de características utilizando Recursive Feature Elimination con Validación Cruzada (RFECV) para elegir las más influyentes y eliminar aquellas que producen ruido en el modelo. Esta selección de características se realiza siempre para la regresión logística, y solo en algunas ocasiones para los modelos de random forest y XGBoost, con el objetivo de explorar la forma en que cambia el desempeño de estos modelos al realizar este proceso de priorización de características.

Finalmente, se evalúan todos los modelos en el conjunto de prueba, utilizando las métricas de accuracy, recall (sensibilidad), precisión, matriz de confusión y el reporte de clasificación para comparar su desempeño y determinar cuál modelo es el más adecuado para predecir la ideación suicida en la población estudiada. Se termina el proceso obteniendo la importancia de las características en el modelo final.

CAPÍTULO 4
VALIDACIÓN DE LA SOLUCIÓN

Para validar la solución propuesta, se implementan y evalúan los modelos de machine learning desarrollados en este estudio utilizando el conjunto de datos obtenido. Esta validación se realiza tanto en la versión numérica de los tests como en su representación categórica. El objetivo de esta validación es demostrar que los modelos pueden predecir de manera efectiva la ideación suicida en estudiantes universitarios, empleando las variables psicológicas y demográficas seleccionadas.

Cada modelo se implementa con diversas combinaciones de parámetros y técnicas de ingeniería de características, además de considerar el balance de las clases en algunos casos. Esto permite contrastar las propiedades de los modelos y encontrar el mejor equilibrio entre la minimización de falsos negativos y la precisión general del modelo. A continuación, se presentan los resultados más destacados de la fase de experimentación.

4.1. Metodología de evaluación

La evaluación se realiza mediante la comparación de los resultados obtenidos por cada modelo en el conjunto de prueba, que representa el 20 % de los datos totales. Se utilizan las siguientes métricas para evaluar el desempeño de los modelos:

- **Recall (Sensibilidad):** Proporción de verdaderos positivos sobre el total de positivos reales.
- **Accuracy:** Proporción de predicciones correctas sobre el total de predicciones.
- **F1-Score:** Media armónica entre la precisión y el recall.
- **Tiempo de ejecución:** Se comparará el tiempo de ejecución de cada modelo.

4.2. Resultados de la Validación

4.2.1. Baseline

Se elige la regresión logística sin reducción de características ni optimización de parámetros como modelo de base para realizar la validación de la propuesta de investigación.

Para el análisis de la versión numérica de los datos, ambas clases tienen un recall de 0.82, lo que indica que el modelo es igualmente bueno para identificar correctamente las instancias negativas (clase 0) y las positivas (clase 1). El F1-score es de 0.84 para la clase 0 y de 0.80 para la clase 1, reflejando un buen equilibrio entre precisión y recall. El accuracy es de 0.82, indicando que el 82 % de las predicciones fueron correctas. Por otra parte, al poner a prueba el modelo con la versión categórica de los datos, vemos que el recall es de 0.74 para la clase 0 y de 0.78 para la clase 1, mostrando una buena capacidad para identificar correctamente las instancias positivas, aunque ligeramente inferior para las negativas en comparación con la versión numérica. El F1-score es de 0.78 para la clase 0 y de 0.73 para la clase 1, lo que indica un buen equilibrio entre precisión y recall, aunque algo menor en comparación con la versión numérica. El accuracy es de 0.76, indicando que el 76 % de las predicciones fueron correctas (9).

Métrica	**Numérica**	**Categórica**
Recall Clase 0	0.82	0.74
Recall Clase 1	0.82	0.78
F1-score Clase 0	0.84	0.78
F1-score Clase 1	0.80	0.73
Accuracy	0.82	0.76
Tiempo de Ejecución	0.033 s	0.019 s

Tabla 9: Resultados del modelo base de Regresión Logística para las versiones numérica y categórica de los datos

Los resultados del modelo base de Regresión Logística muestran un buen rendimiento inicial tanto en la versión numérica como en la versión categórica de los datos. La versión numérica presenta un mayor accuracy y recall en comparación con la versión categórica. Estos resultados servirán como referencia para comparar y evaluar las mejoras obtenidas mediante técnicas más avanzadas y optimizadas. En particular, se observa que la versión numérica es más precisa y tiene un mejor equilibrio en la identificación de las clases, mientras que la versión categórica, aunque ligeramente menos precisa, sigue siendo eficaz y rápida.

4.2.2. Regresión Logística

La versión optimizada de la regresión logística muestra resultados más exitosos al evaluar el modelo con reducción de características y optimización de parámetros, tanto con como sin balanceo de clases, para las versiones numéricas y categóricas de los tests. Por lo tanto, se presentan ambos enfoques en este apartado.

Los resultados obtenidos en el conjunto de prueba utilizando el modelo de regresión logística con los datos numéricos de los instrumentos son los siguientes:

Para el modelo sin considerar los pesos de las clases, se obtiene un recall para la clase 1 de 0.84 y una accuracy de 0.85. Esto indica que el modelo tiene una alta capacidad para

identificar correctamente las instancias positivas (clase 1), con un buen equilibrio general en la clasificación correcta de todas las instancias.

Por otro lado, el modelo que balancea las clases mediante la asignación de pesos logra un recall para la clase 1 de 0.88 y una accuracy de 0.83. Este ajuste en los pesos mejora la identificación de la clase minoritaria (clase 1), aumentando la sensibilidad del modelo a estas instancias a costa de una ligera disminución en el accuracy del modelo.

En términos de f1-score, el modelo sin pesos de clase logra 0.87 para la clase 0 y 0.83 para la clase 1, reflejando un equilibrio entre precisión y recall. Mientras tanto, el modelo con pesos de clase obtiene un f1-score de 0.84 para la clase 0 y 0.82 para la clase 1, lo que sigue siendo alto, pero indica una pequeña compensación en la precisión general.(10).

Métrica	**Con pesos de clase**	**Sin pesos de clase**
Recall Clase 0	0.79	0.85
Recall Clase 1	0.88	0.84
F1-score Clase 0	0.84	0.87
F1-score Clase 1	0.82	0.83
Accuracy	0.83	0.85
Tiempo de Ejecución (s)	0.54	0.37

Tabla 10: Resultados de la regresión logística para la versión numérica de los tests

Los resultados obtenidos en el conjunto de prueba utilizando el modelo de regresión logística con los datos categóricos de los instrumentos son los siguientes:

Para el modelo sin considerar los pesos de las clases, se obtiene un recall para la clase 1 de 0.92 y una accuracy de 0.65. Esto indica una alta capacidad para identificar correctamente las instancias positivas (clase 1), aunque la precisión general es moderada.

Por otro lado, el modelo que balancea las clases mediante la asignación de pesos logra también un recall para la clase 1 de 0.92 y una accuracy de 0.65. Este ajuste en los pesos no mejora la identificación de la clase minoritaria (clase 1) en términos de recall y mantiene la misma precisión general del modelo.

En términos de f1-score, ambos modelos lograron un f1-score de 0.59 para la clase 0 y 0.69 para la clase 1. Esto refleja un equilibrio entre precisión y recall en ambos enfoques (11).

La importancia de las características se determina utilizando los coeficientes del modelo entrenado. Se observa que las variables “Depresión”, tener un diagnóstico de “Neurodivergencia” y “Soledad” muestran los coeficientes más altos en la versión numérica de los tests, tanto en el modelo balanceado como en el tradicional.

Por otra parte, al considerar la versión categórica de los tests, se encuentra que el modelo no balanceado utiliza solo dos variables para realizar las predicciones. En primer lugar, la categoría “Otro” de la variable género actúa como un predictor directo, aunque moderado, de

Métrica	Con pesos de clase	Sin pesos de clase
Recall Clase 0	0.44	0.44
Recall Clase 1	0.92	0.92
F1-score Clase 0	0.59	0.59
F1-score Clase 1	0.69	0.69
Accuracy	0.65	0.65
Tiempo de Ejecución (s)	0.45	0.34

Tabla 11: Resultados de la regresión logística para la versión categórica de los tests

la variable objetivo (su presencia se asocia con una mayor incidencia de ideación suicida). La categoría "Normal" de la variable Depresión se manifiesta de manera negativa y significativa, es decir, su presencia reduce la probabilidad del evento objetivo en comparación con su ausencia.

Finalmente, se observa que el modelo balanceado destaca la importancia de tener un diagnóstico de "Neurodivergencia", y presenta una relación negativa con la categoría "Normal" de la variable Depresión y la ausencia de "Desesperanza". En otras palabras, pertenecer al rango normal de depresión y no mostrar signos de elevada soledad reduce la probabilidad de presentar ideación suicida en comparación con otras combinaciones de características.

4.2.3. Random Forest

El modelo Random Forest mostró sus mejores métricas al balancear las clases de la variable objetivo para todos los casos.

Los resultados obtenidos en el conjunto de prueba con los datos numéricos de los instrumentos muestran un recall de 0.68 para la clase 0 y de 0.84 para la clase 1. Esto indica que el modelo tiene una alta capacidad para identificar correctamente las instancias positivas (clase 1), aunque la precisión para la clase 0 es moderada.

En términos de f1-score, el modelo logró 0.75 para la clase 0 y 0.74 para la clase 1, lo que refleja un equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.75, lo que indica que tres cuartas partes de las predicciones fueron correctas (12).

La importancia de las características se determina utilizando el índice Gini, destacando las variables "Depresión", "Soledad" y "Desesperanza" como las tres más influyentes. Sus índices indican que las decisiones basadas en estas características ayudan a reducir la impureza en los árboles del bosque aleatorio, contribuyendo a la precisión del modelo.

Los resultados en el conjunto de prueba utilizando los datos categóricos de los instrumentos alcanzan un recall de 0.56 para la clase 0 y de 0.90 para la clase 1. Esto indica que el mo-

Métrica	
Recall Clase 0	0.68
Recall Clase 1	0.84
F1-score Clase 0	0.75
F1-score Clase 1	0.74
Accuracy	0.75
Tiempo de Ejecución	7 min 58 s

Tabla 12: Resultados del Random Forest para la versión numérica de los tests

delo tiene una capacidad moderada para identificar correctamente las instancias negativas (clase 0), mientras que tiene una alta capacidad para identificar correctamente las instancias positivas (clase 1).

En términos de f1-score, el modelo logró 0.68 para la clase 0 y 0.72 para la clase 1, lo que refleja un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.71, lo que indica que el 71 % de las predicciones fueron correctas (13).

Métrica	**Valor**
Recall Clase 0	0.56
Recall Clase 1	0.90
F1-score Clase 0	0.68
F1-score Clase 1	0.72
Accuracy	0.71
Tiempo de Ejecución	7 min 58 s

Tabla 13: Resultados del Random Forest para la versión categórica de los tests

Nuevamente, la importancia de las características se determina utilizando el índice Gini, obteniendo esta vez que la categoría “Normal” de la variable Depresión fue la más influyente, seguida por la “Ausencia” de Desesperanza y la categoría “Moderada” de la variable Soledad.

Se quiso explorar el comportamiento del modelo Random Forest al hacer una reducción de características utilizando la técnica de selección RFECV. A continuación, se presentan los resultados obtenidos en las versiones numérica y categórica de los tests.

Para la versión numérica del modelo, se obtuvo un recall de 0.60 para la clase 0 y de 0.80 para la clase 1. Esto indica que el modelo tiene una capacidad moderada para identificar correctamente las instancias negativas (clase 0) y una alta capacidad para identificar correctamente las instancias positivas (clase 1).

En términos de f1-score, el modelo logró 0.69 tanto para la clase 0 como para la clase 1, reflejando un equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.69, lo que indica que el 69 % de las predicciones fueron correctas.

Para la versión categórica del modelo, se obtuvo un recall de 0.44 para la clase 0 y de 0.92 para la clase 1. Esto indica una capacidad moderada para identificar correctamente las instancias negativas y una alta capacidad para identificar correctamente las instancias positivas.

En términos de f1-score, el modelo logró 0.59 para la clase 0 y 0.69 para la clase 1, lo que también refleja un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.65, lo que indica que el 65 % de las predicciones fueron correctas (14).

Métrica	**Numérica**	**Categórica**
Recall Clase 0	0.60	0.44
Recall Clase 1	0.80	0.92
F1-score Clase 0	0.69	0.59
F1-score Clase 1	0.69	0.69
Accuracy	0.69	0.65
Tiempo de Ejecución	8 min 12 s	8 min 3 s

Tabla 14: Resultados del Random Forest para las versiones numérica y categórica de los tests con RFECV

Estos resultados muestran cómo la reducción de características mediante RFECV afecta el rendimiento del modelo Random Forest en diferentes versiones de los datos. En la versión numérica, el modelo muestra un buen equilibrio entre precisión y recall, con una precisión general del 69 %. En la versión categórica, el modelo tiene una alta capacidad para identificar instancias positivas, con una precisión general del 65 %. En ambos modelos la reducción de características resultó en la selección de una sola variable predictora. En el caso del modelo entrenado con las variables numéricas, se subraya la relevancia de la variable “Depresión”, mientras que el modelo entrenado con las variables categóricas nos revela la importancia de la categoría “Normal” de la variable “Depresión” en específico como la más influyente para predecir la presencia de ideación suicida. Cabe destacar, que este aumento en la capacidad de detectar casos positivos de ideación suicida ocurre a costa de un sobre diagnóstico de este fenómeno y un elevado numero de falsos positivos. Aún así, la técnica de selección de características RFECV permite mejorar la eficiencia del modelo, manteniendo tiempos de ejecución razonables.

4.2.4. SVM (Máquinas de Soporte Vectorial)

Para la versión numérica del modelo SVC, se obtuvo un recall de 0.81 para la clase 0 y de 0.78 para la clase 1. Esto indica que el modelo tiene una alta capacidad para identificar correctamente tanto las instancias negativas (clase 0) como las positivas (clase 1).

En términos de f1-score, el modelo logró 0.82 para la clase 0 y 0.77 para la clase 1, reflejando un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo,

medida por el accuracy, fue de 0.80, lo que indica que el 80 % de las predicciones fueron correctas.

Para la versión categórica del modelo SVC, se obtuvo un recall de 0.71 para la clase 0 y de 0.84 para la clase 1. Esto indica una capacidad moderada para identificar correctamente las instancias negativas y una alta capacidad para identificar correctamente las instancias positivas.

En términos de f1-score, el modelo logró 0.77 para la clase 0 y 0.75 para la clase 1, lo que refleja un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.76, lo que indica que el 76 % de las predicciones fueron correctas (15).

Métrica	**Numérica**	**Categórica**
Recall Clase 0	0.81	0.71
Recall Clase 1	0.78	0.84
F1-score Clase 0	0.82	0.77
F1-score Clase 1	0.77	0.75
Accuracy	0.80	0.76
Tiempo de Ejecución	0 min 57 s	2 min 57 s

Tabla 15: Resultados del SVM para las versiones numérica y categórica de los tests

Respecto a la importancia de las características, se muestra que para el modelo SVM en su versión numérica, las tres variables más influyentes son “Depresión”, “Soledad” y poseer un diagnóstico de “Neurodivergencia”. Por otra parte, el modelo en su versión categórica encuentra que la variable que más contribuye a la precisión del modelo es la presencia de un diagnóstico de “Neurodivergencia”, seguida de la ausencia de “Desesperanza” y tener niveles de depresión en su categoría “Normal”.

4.2.5. Redes Neuronales

Para la versión numérica del modelo de Redes Neuronales, se obtuvo un recall de 0.74 para la clase 0 y de 0.86 para la clase 1. Esto indica que el modelo tiene una alta capacidad para identificar correctamente tanto las instancias negativas (clase 0) como las positivas (clase 1).

En términos de f1-score, el modelo logró 0.80 para la clase 0 y 0.78 para la clase 1, reflejando un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.79, lo que indica que el 79 % de las predicciones fueron correctas.

Para la versión categórica del modelo de Redes Neuronales, se obtuvo un recall de 0.49 para la clase 0 y de 0.96 para la clase 1. Esto indica una capacidad moderada para identificar

correctamente las instancias negativas y una alta capacidad para identificar correctamente las instancias positivas.

En términos de f1-score, el modelo logró 0.64 para la clase 0 y 0.73 para la clase 1, lo que refleja un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.69, lo que indica que el 69 % de las predicciones fueron correctas (16).

Métrica	**Numérica**	**Categórica**
Recall Clase 0	0.74	0.49
Recall Clase 1	0.86	0.96
F1-score Clase 0	0.80	0.64
F1-score Clase 1	0.78	0.73
Accuracy	0.79	0.69
Tiempo de Ejecución	5 min 34 s	5 min 35 s

Tabla 16: Resultados de Redes Neuronales para las versiones numérica y categórica de los tests

Respecto a la importancia de las características, se obtiene que la más influyente en el modelo de redes neuronales en su versión numérica es la “Depresión”, seguida de la variable “Área ciencias de la salud” y “Área ingeniería y tecnología”. Mientra que el modelo en su versión categórica de los datos, muestra la influencia específica de la categoría “Normal” de la variable depresión, seguida de “Área ingeniería y tecnología”’ y “Área ciencias de la salud”.

4.2.6. XGBoost

Para la versión numérica del modelo XGBoost, se obtuvo un recall de 0.62 para la clase 0 y de 0.90 para la clase 1. Esto indica que el modelo tiene una capacidad moderada para identificar correctamente las instancias negativas (clase 0) y una alta capacidad para identificar correctamente las instancias positivas (clase 1).

En términos de f1-score, el modelo logró 0.73 para la clase 0 y 0.75 para la clase 1, reflejando un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.74, lo que indica que el 74 % de las predicciones fueron correctas.

Para la versión categórica del modelo XGBoost, se obtuvo un recall de 0.47 para la clase 0 y de 0.90 para la clase 1. Esto indica una capacidad moderada para identificar correctamente las instancias negativas y una alta capacidad para identificar correctamente las instancias positivas.

En términos de f1-score, el modelo logró 0.61 para la clase 0 y 0.69 para la clase 1, lo que refleja un buen equilibrio entre precisión y recall para ambas clases. La precisión general del

modelo, medida por el accuracy, fue de 0.66, lo que indica que el 66 % de las predicciones fueron correctas (17).

Métrica	Numérica	Categórica
Recall Clase 0	0.62	0.47
Recall Clase 1	0.90	0.90
F1-score Clase 0	0.73	0.61
F1-score Clase 1	0.75	0.69
Accuracy	0.74	0.66
Tiempo de Ejecución	13 min 10 s	11 min 58 s

Tabla 17: Resultados de XGBoost para las versiones numérica y categórica de los tests

Respecto a la importancia de las características, se observa que “Depresión” es la más influyente en el modelo XGBoost en su versión numérica. Le siguen “Desesperanza” y tener un diagnóstico de “Neurodivergencia”. Además, se muestra que en la versión categórica de los datos, la categoría “Normal” de la variable Depresión es la más importante, seguida por la ausencia de “Desesperanza”.

Se realizó también un procedimiento de reducción de características para este modelo, revelando que para la versión numérica solo se requieren tres variables para predecir ideación suicida. Estas son “Depresión”, “Desesperanza” y poseer un diagnostico de “Neurodivergencia”. Por otra parte, la versión categórica con reducción de características muestra que la categoría “Normal” de la variable depresión por sí sola puede predecir ideación suicida. Para la versión numérica del modelo XGBoost utilizando RFECV, se obtuvo un recall de 0.71 para la clase 0 y de 0.86 para la clase 1. Esto indica que el modelo tiene una capacidad moderada para identificar correctamente las instancias negativas (clase 0) y una alta capacidad para identificar correctamente las instancias positivas (clase 1).

En términos de f1-score, el modelo logró 0.78 para la clase 0 y 0.77 para la clase 1, reflejando un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.77, lo que indica que el 77 % de las predicciones fueron correctas. El tiempo de ejecución del modelo fue de aproximadamente 10 minutos y 47 segundos.

Para la versión categórica del modelo XGBoost utilizando RFECV, se obtuvo un recall de 0.44 para la clase 0 y de 0.92 para la clase 1. Esto indica una capacidad moderada para identificar correctamente las instancias negativas y una alta capacidad para identificar correctamente las instancias positivas.

En términos de f1-score, el modelo logró 0.59 para la clase 0 y 0.69 para la clase 1, lo que refleja un buen equilibrio entre precisión y recall para ambas clases. La precisión general del modelo, medida por el accuracy, fue de 0.65, lo que indica que el 65 % de las predicciones fueron correctas. El tiempo de ejecución del modelo fue de aproximadamente 11 minutos y 38 segundos (18).

Métrica	Numérica (RFECV)	Categórica (RFECV)
Recall Clase 0	0.71	0.44
Recall Clase 1	0.86	0.92
F1-score Clase 0	0.78	0.59
F1-score Clase 1	0.77	0.69
Accuracy	0.77	0.65
Tiempo de Ejecución	10 min 47 s	11 min 38 s

Tabla 18: Resultados de XGBoost para las versiones numérica y categórica de los tests con RFECV

Estos resultados muestran el rendimiento del modelo XGBoost en diferentes versiones de los datos, aplicando la técnica de RFECV. En la versión numérica, el modelo muestra un buen equilibrio entre precisión y recall, con una precisión general del 77 %. En la versión categórica, el modelo tiene una alta capacidad para identificar instancias positivas, con una precisión general del 65 %. El tiempo de ejecución es considerablemente alto en ambos casos, siendo más largo para los datos categóricos. El tiempo de ejecución es considerablemente alto en ambos casos, siendo más largo para los datos categóricos.

4.3. Comparación y Análisis de los Modelos

En esta sección se comparan los diferentes modelos de machine learning implementados para predecir la ideación suicida en estudiantes universitarios, considerando sus desempeños en términos de recall, accuracy, F1-score y tiempos de ejecución, tanto en las versiones numéricas como categóricas de los datos, con y sin la técnica de reducción de características RFECV.

Para la Regresión Logística, se observan algunas diferencias entre los modelos con y sin balanceo de clases. En la versión numérica, los resultados indican que el balanceo de clases mejora la identificación de la clase minoritaria (clase 1), aunque con una ligera disminución en la precisión general del modelo. En la versión categórica, el impacto del balanceo de clases es mínimo, mostrando métricas casi idénticas entre ambos enfoques. Estos resultados demuestran que, mientras el balanceo de clases es crucial para la versión numérica, su efecto es menos pronunciado en la versión categórica, lo que sugiere que otros factores podrían estar influyendo más en el rendimiento del modelo categórico.

El modelo Random Forest muestra un mejor rendimiento en la identificación de instancias positivas en la versión categórica, con un recall más alto, aunque el accuracy general es menor comparado con la versión numérica. La aplicación de RFECV en Random Forest no mejora la identificación de la clase 1 en la versión numérica, y resulta en una disminución del balance entre precisión y recall. Sin embargo, RFECV mejora el rendimiento en la versión categórica, con un ligero aumento en el recall de la clase 1, aunque a costa de una menor precisión general del modelo.

El SVM muestra un buen equilibrio entre precisión y recall en ambas versiones (numérica y categórica), con un rendimiento ligeramente mejor en la versión categórica en términos de recall para la clase 1.

Las Redes Neuronales muestran un alto recall para la clase 1 en ambas versiones, especialmente en la versión categórica, aunque con una precisión general más baja.

XGBoost muestra un alto rendimiento en la identificación de instancias positivas en ambas versiones (numérica y categórica), con un mejor accuracy en la versión numérica. La aplicación de RFECV en XGBoost mejora ligeramente el recall en la versión categórica, aunque nuevamente a costa de una menor precisión general del modelo. En la versión numérica, RFECV no proporciona mejoras significativas.

1. **Mejor Recall:** La versión categórica de Redes Neuronales y XGBoost con RFECV, uestran un alto recall para la clase 1.
2. **Mejor Accuracy:** La Regresión Logística optimizada sin considerar los pesos de las clases, muestra un alto accuracy, especialmente en la versión numérica.
3. **Mejor Equilibrio:** Para un buen equilibrio entre recall y precisión, la versión numérica optimizada de la Regresión Logística y XGBoost con RFECV fueron las más efectivas.
4. **Mejor Tiempo de Ejecución:** La Regresión Logística base y optimizada tuvieron los mejores tiempos de ejecución en comparación con otros modelos.

Los modelos más complejos como XGBoost y Redes Neuronales demuestran ser muy efectivos para la identificación de la clase 1, mientras que la Regresión Logística, especialmente en su versión numérica optimizada, ofrece un buen equilibrio entre precisión y recall con tiempos de ejecución más eficientes. La aplicación de RFECV muestra mejoras selectivas, siendo más beneficiosa en ciertos contextos de datos categóricos, mientras que en datos numéricos su impacto es más limitado y costoso en términos de tiempo de procesamiento.

Comparando con el modelo baseline de Regresión Logística, los modelos optimizados, especialmente XGBoost y Redes Neuronales, muestran una mejor capacidad para identificar la clase 1 (recall alto). Sin embargo, en términos de precisión general (accuracy), la Regresión Logística base y optimizada se desempeñan bien, especialmente en la versión numérica.

En el análisis de importancia de características a lo largo de los diferentes modelos evaluados (Regresión Logística, Random Forest, SVM, Redes Neuronales y XGBoost), algunas variables se destacan consistentemente, siendo estas las que tienen un mayor impacto en la predicción de la ideación suicida en estudiantes universitarios.

En la **versión numérica** de los tests, donde se utilizan los puntajes de estos como datos de entrenamiento, se identifica la **Depresión** como la variable más influyente en la mayoría de

los modelos. Otras variables importantes para la predicción de ideación suicida son la **Desesperanza**, la presencia de un diagnóstico de **Neurodivergencia** y los niveles de **Soledad**.

Por otra parte, al analizar los datos en su **versión categórica**, se define la categoría **Normal** de la depresión como la más influyente a la hora de predecir ideación suicida en la mayoría de los modelos. A esta le siguen la **ausencia de Desesperanza** y la presencia de un diagnóstico de **Neurodivergencia**.

Los resultados de esta validación sugieren que la selección cuidadosa de técnicas y modelos, junto con la consideración de las características más influyentes, puede mejorar significativamente la capacidad predictiva y la precisión de los modelos de machine learning en contextos críticos como la predicción de la ideación suicida en estudiantes universitarios. Además, los resultados indican que siempre existe una compensación (trade-off) entre la mejora de una métrica en desmedro de otra. La elección final del modelo depende de la tarea y de las consecuencias de priorizar la detección de una clase por sobre otra en problemas de clasificación.

CAPÍTULO 5
CONCLUSIONES

5.1. Conclusiones generales

El presente estudio se centra en abordar el delicado tema de la salud mental en estudiantes universitarios y sus principales problemáticas. Surge como respuesta a la necesidad de entender las posibles causas de un problema cada vez más habitual en la comunidad universitaria: la ideación suicida. Para ello, se busca analizar de manera objetiva las aflicciones más comúnmente asociadas a este fenómeno, así como explorar otras posibilidades menos comunes y hallar el conjunto de características que mejor expliquen su elevada presencia entre los estudiantes. Además, se aborda la necesidad de mejorar las herramientas tradicionales de investigación en campos como la psicología, donde resulta crucial elevar el poder predictivo de los modelos y adoptar buenas prácticas tanto en las fases de preprocesamiento de los datos como en el análisis matemático de las variables de estudio, actualizando los actuales modelos explicativos para diseñar e implementar modelos de carácter predictivo.

La psicología y las ciencias cognitivas en general han sido fundamentales para el desarrollo de los primeros modelos de aprendizaje automático, donde se intentaba emular el comportamiento del cerebro y de las neuronas para obtener formas de procesamiento similares a las de un humano en distintas tareas que demandaban un robusto análisis y capacidad analítica. Estos aportes lograron mejorar algunos modelos y despertar una revolución de la computación que no se ha detenido hasta el momento. Este estudio puede considerarse un intento por retribuir a la psicología con las herramientas que ayudó a crear y devolver la mano a esta disciplina en su búsqueda de entender de mejor manera los fenómenos de la mente y el comportamiento.

Con respecto a los objetivos específicos de este trabajo, se propuso en primer lugar la creación de un dataset actualizado con información tanto psicológica como demográfica de estudiantes universitarios de pregrado. Esto con la finalidad de realizar el análisis principal con datos actuales y específicos de la población de interés. Se considera que este punto se cumplió en gran medida, dado que se logró obtener una muestra de 591 estudiantes de pregrado pertenecientes a 28 distintas universidades a lo largo del país y de 71 carreras diferentes. Esto revela una gran diversidad y amplia representación de las distintas realidades de la comunidad estudiantil.

El dataset recopilado es robusto e incluye información detallada sobre los niveles de depresión, estrés, ansiedad y otros factores psicológicos y demográficos de los estudiantes universitarios, utilizando tests y cuestionarios validados. Este dataset no solo fue la base del análisis estadístico y predictivo realizado, sino que además sirvió para ilustrar la prevalencia de las variables medidas y otros datos relevantes que facilitaron la tarea.

Por otra parte, se cumplió con el objetivo de analizar estadísticamente las variables y sus relaciones, tanto entre ellas como con la variable dependiente, para conocer cómo se comportan en conjunto y distinguir las asociaciones más comunes. Al respecto, se pudo encontrar varias relaciones significativas que aportan a la comprensión del fenómeno de la ideación suicida en estudiantes universitarios.

En dicho análisis se reveló que la ansiedad y el estrés están ampliamente relacionados. En particular, se encontró una alta colinealidad en la variable estrés, por lo que fue eliminada de las características que aportaban información al modelo. Asimismo, se encontró una asociación importante entre el área de estudio y otras variables como el género, el diagnóstico de neurodivergencia y la modalidad de trabajo. Respecto a la ideación suicida, se encontró una asociación significativa con variables como la depresión, el género y la presencia de diagnóstico de neurodivergencia. Finalmente, tras un análisis de colinealidad, se decidió eliminar del estudio también la variable de edad.

En conjunto, estos hallazgos proporcionan una visión comprensiva de cómo interactúan estas variables entre sí y con la ideación suicida, permitiendo identificar los factores de riesgo más relevantes. Esto no solo valida la importancia de incluir estas variables en los modelos predictivos, sino que también resalta áreas clave donde se pueden dirigir los esfuerzos de intervención y prevención.

Otros objetivos centrales del presente estudio eran diseñar e implementar diversos algoritmos de machine learning para la tarea de clasificación de ideación suicida en estudiantes universitarios y evaluar su desempeño. Este objetivo se considera cumplido a cabalidad, ya que se aplicaron cinco algoritmos diferentes para este propósito: regresión logística, random forest, SVM, redes neuronales y XGBoost.

Cada uno de estos algoritmos ofreció una perspectiva única y mostró diferentes grados de precisión y eficacia:

1. **Regresión Logística:** Este modelo tradicional de clasificación se utilizó tanto en su versión base como optimizada. Se destacó por su capacidad para mantener un buen equilibrio entre accuracy y recall, especialmente en la versión optimizada y numérica de los datos. Por otra parte, la versión categórica de los datos mejoró la detección de casos positivos al aumentar el recall, pero esto resultó en una disminución del accuracy del modelo, dejando muchos casos de falsos positivos.

2. **Random Forest:** Este algoritmo de ensamble mostró robustez en la identificación de instancias positivas en la versión categórica de los datos, aunque el uso de técnicas de reducción de características (RFECV) no siempre mejoró su desempeño. De hecho, el uso de RFECV redujo todas las métricas del modelo en su versión numérica. Para la versión categórica, se observó un leve aumento en el recall, alcanzando un 92 % en la identificación de casos positivos, pero disminuyendo el accuracy del modelo, lo que sugiere un overfitting que limita su eficacia como herramienta predictiva.

3. **SVM (Support Vector Machine):** El modelo SVM demostró ser eficaz en la clasificación con una alta accuracy y recall en ambas versiones de los datos, aunque su tiempo de ejecución fue considerablemente mayor en la versión categórica. Logra un buen equilibrio en la detección de casos positivos y negativos dependiendo del formato de los datos. En su versión categórica, destaca una alta capacidad de predicción de casos positivos sin dejar de tener un buen accuracy.

4. **Redes Neuronales:** En la versión categórica, las redes neuronales lograron un 96 % de recall para la detección de casos positivos, pero a costa de un 49 % de recall para detectar casos negativos, por lo cual se considera demasiado especializado para la detección de casos positivos. Esto resulta en un gran sobrediagnóstico de la variable dependiente. Por otro lado, la versión numérica presentó métricas más equilibradas, logrando un 86 % de recall y un buen accuracy. La mejor performance con datos numéricos podría deberse a la mayor información contenida en los valores continuos, lo que permite al modelo capturar mejor las sutilezas y variaciones en los datos, mientras que las variables categóricas pueden perder algunas de estas características informativas .

5. **XGBoost:** Este algoritmo de boosting mostró un rendimiento sobresaliente en la identificación de instancias positivas en ambas versiones de los datos. Destacando especialmente en la versión numérica, donde se logró un 90 % de recall para casos positivos sin comprometer excesivamente la detección de casos negativos. Por otra parte, la versión categórica mostró una capacidad similar para identificar casos positivos, pero con un desempeño deficiente en la detección de casos negativos. La aplicación de RFECV en la versión numérica resultó en un modelo mejor equilibrado en la detección de casos positivos y negativos, mientras que en la versión categórica aumentó la capacidad de detección de casos positivos a costa de una considerable disminución en la detección de casos negativos. A pesar de su notable capacidad para predecir la ideación suicida, el modelo en su versión categórica no se considera un buen modelo debido a que esta habilidad se logra a costa de sobrediagnosticar la variable dependiente.

Los resultados muestran que la elección del modelo depende de las prioridades del análisis y otros factores relevantes. Si el objetivo es maximizar el recall para la clase 1, XGBoost y Regresión Logística optimizada con datos categóricos son buenas opciones, aunque, como se mencionó anteriormente, su desempeño es muy pobre en la detección de casos negativos. Para un mejor balance entre accuracy y recall, la Regresión Logística optimizada con datos numéricos, redes neuronales con datos numéricos o SVM, también con datos numéricos, ofrecen un rendimiento robusto. En términos de accuracy general, la Regresión Logística sigue siendo competitiva, demostrando que los modelos más simples pueden ser muy efectivos con una adecuada optimización. Además, la Regresión Logística y el SVM presentan los mejores tiempos de ejecución, siendo adecuados cuando se busca una implementación rápida. Si se prioriza la interpretabilidad, la Regresión Logística es la opción más clara, ya que permite entender fácilmente el impacto de cada variable en la predicción, lo cual es especialmente importante en campos como la salud y las ciencias sociales, donde es crucial entender y comunicar cómo las variables independientes influyen en el resultado.. Por úl-

timo, en términos de recursos computacionales, la Regresión Logística y el SVM son menos demandantes comparados con Redes Neuronales y XGBoost, lo que los hace ideales para entornos con limitaciones de recursos.

Los resultados del análisis de importancia de características sugieren que la presencia de "Depresión", "Desesperanza", "Neurodivergencia" y "Soledad" serían las variables más influyentes en la mayoría de los modelos. Al analizar la versión categórica de los datos, se sabe que en especifico la "Depresión Normal" se asocia con una disminución en la probabilidad del resultado objetivo del modelo. En términos prácticos, esto implica que tener "Depresión Normal" podría considerarse un factor protector en comparación con tener niveles más elevados de depresión. Específicamente, la importancia negativa significativa de esta variable indica que los individuos con "Depresión Normal" están menos propensos a experimentar el resultado adverso predicho por el modelo, en contraste con aquellos que presentan niveles más severos de depresión. También se encuentra que la ausencia de "Desesperanza" y "Soledad" funcionaron como categorías importantes para descartar la presencia de ideación suicida, es decir, al igual que ocurre con la categoría de depresión normal, la ausencia de estas variables disminuyen la probabilidad de padecer ideación suicida.

Este análisis proporciona información fundamental para la detección correcta de ideación suicida. Aunque la literatura ya reconoce la importancia de la depresión y la desesperanza como indicadores de riesgo suicida, no se había analizado en detalle qué niveles específicos de estas variables actúan como facilitadores del fenómeno o como factores protectores. Además, este estudio destaca la capacidad de la soledad para predecir ideación suicida, una variable que no ha sido tan estudiada como la depresión o la desesperanza. Se descubrió que es la ausencia de soledad, más que su presencia, la que ayuda a descartar casos de riesgo suicida. Finalmente, se muestra consistentemente la relevancia de un diagnóstico de neurodivergencia, una cualidad que no se ha estudiado suficientemente en relación con la ideación suicida y que podría ser interesante de explorar más a fondo.

5.2. Limitaciones del estudio

A pesar de la significativa participación de estudiantes universitarios, se considera que la muestra no es lo suficientemente grande para utilizar esta información como base para la creación de herramientas que puedan realizar la detección en tiempo real o realizar un diagnóstico confiable de casos nuevos. En particular, esta conclusión se fundamenta en que más del 50 % de los participantes provenían de la región Metropolitana. Además, es posible que la procedencia de los estudiantes haya sido entendida como el lugar de residencia actual, lo cual también aumenta la probabilidad de errores al medir este dato.

Adicionalmente, la muestra está desequilibrada en términos de representación de las distintas carreras universitarias. Algunas áreas de estudio están significativamente más representadas que otras, lo que podría sesgar los resultados y limitar la generalización de los hallazgos a la población estudiantil en su totalidad. Este sesgo hacia ciertas carreras implica que

las conclusiones extraídas podrían no reflejar con precisión las experiencias y riesgos de los estudiantes de disciplinas subrepresentadas, afectando la validez externa de los resultados. Por lo tanto, para futuros estudios, sería crucial asegurar una mayor diversidad y equilibrio en la muestra, tanto en términos de procedencia geográfica como de representación de las distintas carreras universitarias, para obtener conclusiones más robustas y generalizables.

5.3. Trabajos futuros

Este estudio abre varias líneas interesantes para futuras investigaciones. En primer lugar, se podría ampliar la muestra para incluir una mayor diversidad geográfica y disciplinar, lo cual permitiría validar los hallazgos de este estudio en un contexto más amplio y diverso. Esta ampliación ayudaría a mejorar la representatividad y la generalización de los resultados.

Además, sería valioso investigar la implementación de técnicas de machine learning más avanzadas y específicas, como modelos de deep learning o el uso de técnicas de ensamblado que combinan múltiples modelos para mejorar la precisión y el recall. Estas técnicas podrían aprovecharse para crear herramientas predictivas más robustas que puedan ser utilizadas en tiempo real. En particular, se recomienda explorar la implementación de técnicas de ensamblado como el stacking o el blending, que permiten combinar diferentes modelos para determinar si esta arquitectura proporciona un rendimiento superior en comparación con cada modelo individual. Esta combinación de modelos puede aprovechar las fortalezas de cada uno y mitigar sus debilidades, proporcionando una solución más equilibrada y eficaz para la predicción de ideación suicida.

Otra área de investigación futura podría centrarse en la personalización de las intervenciones preventivas. Utilizando los hallazgos de este estudio, se podrían desarrollar programas de intervención específicos para estudiantes con diferentes perfiles de riesgo, optimizando los recursos disponibles y mejorando la efectividad de las intervenciones. Aunque esta herramienta no puede usarse como método diagnóstico, sí puede emplearse la información sobre las variables más importantes en la detección del riesgo suicida para priorizar aquellos casos que manifiesten estas características y dar mayor visibilidad a casos que podrían pasar desapercibidos. Esto permitiría una intervención más oportuna y focalizada, incrementando la capacidad de respuesta ante situaciones de riesgo y contribuyendo a la reducción de casos de ideación suicida en el ámbito universitario.

También se recomienda realizar estudios longitudinales que sigan a los estudiantes a lo largo del tiempo para observar cómo evolucionan sus niveles de estrés, ansiedad, depresión y otros factores psicológicos. Esto permitiría identificar patrones y precursores de ideación suicida con mayor precisión, proporcionando datos valiosos para intervenciones tempranas y personalizadas.

Por último, sería interesante explorar la integración de datos adicionales, como el uso de redes sociales, actividad física y hábitos de sueño, para construir modelos predictivos más

completos y precisos. Estos datos complementarios podrían ofrecer una visión más holística del bienestar de los estudiantes y mejorar la detección temprana de riesgos de salud mental.

En resumen, la continuación de esta línea de investigación no solo contribuirá a una mejor comprensión de la ideación suicida en estudiantes universitarios, sino que también ayudará a desarrollar herramientas más efectivas y personalizadas para la prevención e intervención, mejorando significativamente la salud mental y el bienestar de la comunidad estudiantil.

REFERENCIAS BIBLIOGRÁFICAS

[Abdullah y Bouke,] Abdullah, A. y Bouke, M. A. Towards image-based network traffic pattern detection for ddos attacks in cloud computing environments: A comparative study.

[Adadi y Berrada, 2018] Adadi, A. y Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access*, 6:52138–52160.

[Amini *et al.*, 2016] Amini, P., Ahmadinia, H., Poorolajal, J., y Amiri, M. M. (2016). Evaluating the high risk groups for suicide: a comparison of logistic regression, support vector machine, decision tree and artificial neural network. *Iranian journal of public health*, 45(9):1179.

[Amrhein *et al.*, 2019] Amrhein, V., Trafimow, D., y Greenland, S. (2019). Inferential statistics as descriptive statistics: There is no replication crisis if we don't expect replication. *The American Statistician*, 73(sup1):262–270.

[Araneda *et al.*, 2021] Araneda, N., Sanhueza, P., Pacheco, G., y Sanhueza, A. (2021). Suicidio en adolescentes y jóvenes en chile: riesgos relativos, tendencias y desigualdades. *Revista Panamericana de Salud Pública*, 45.

[Baader *et al.*, 2014] Baader, T., Rojas, C., Molina, J. L., Gotelli, M., Alamo, C., Fierro, C., Venezian, S., y Dittus, P. (2014). Diagnóstico de la prevalencia de trastornos de la salud mental en estudiantes universitarios y los factores de riesgo emocionales asociados. *Revista chilena de neuro-psiquiatría*, 52(3):167–176.

[Bamber y Schneider, 2016] Bamber, M. D. y Schneider, J. K. (2016). Mindfulness-based meditation to decrease stress and anxiety in college students: A narrative synthesis of the research. *Educational Research Review*, 18:1–32.

[Barrera-Herrera y San Martín, 2021] Barrera-Herrera, A. y San Martín, Y. (2021). Prevalencia de sintomatología de salud mental y hábitos de salud en una muestra de universitarios chilenos. *Psykhe (Santiago)*, 30(1):1–16.

[Barrera-Herrera y Vinet, 2017] Barrera-Herrera, A. y Vinet, E. V. (2017). Adultez emergente y características culturales de la etapa en universitarios chilenos. *Terapia psicológica*, 35(1):47–56.

[Beiter *et al.*, 2015] Beiter, R., Nash, R., McCrady, M., Rhoades, D., Linscomb, M., Clarahan, M., y Sammut, S. (2015). The prevalence and correlates of depression, anxiety, and stress in a sample of college students. *Journal of affective disorders*, 173:90–96.

[Bernert *et al.*, 2020] Bernert, R. A., Hilberg, A. M., Melia, R., Kim, J. P., Shah, N. H., y Abnousi, F. (2020). Artificial intelligence and suicide prevention: a systematic review of machine learning investigations. *International journal of environmental research and public health*, 17(16):5929.

[Breiman, 2001] Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

[Burke *et al.*, 2019] Burke, T. A., Ammerman, B. A., y Jacobucci, R. (2019). The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. *Journal of affective disorders*, 245:869–884.

[Challenger-Pérez *et al.*, 2014] Challenger-Pérez, I., Díaz-Ricardo, Y., y Becerra-García, R. A. (2014). El lenguaje de programación python. *Ciencias Holguín*, 20(2):1–13.

[Chang, 2017] Chang, E. C. (2017). Hope and hopelessness as predictors of suicide ideation in hungarian college students. *Death studies*, 41(7):455–460.

[Cumming, 2008] Cumming, G. (2008). Replication and p intervals: p values predict the future only vaguely, but confidence intervals do much better. *Perspectives on psychological science*, 3(4):286–300.

[Dias Lopes *et al.*, 2020] Dias Lopes, Luis Felipe and Chaves, Bianca Michels and Fabrício, Adriane and Porto, Adriana and Machado de Almeida, Damiana and Obregon, Sandra Leonara and Pimentel Lima, Mauren and Vieira da Silva, Wesley and Camargo, Maria Emilia and da Veiga, Claudimar Pereira and others (2020). Analysis of well-being and anxiety among university students. *International journal of environmental research and public health*, 17(11):3874.

[Diehl *et al.*, 2018] Diehl, K., Jansen, C., Ishchanova, K., y Hilger-Kolb, J. (2018). Loneliness at universities: determinants of emotional and social loneliness among students. *International journal of environmental research and public health*, 15(9):1865.

[Doran *et al.*, 2017] Doran, D., Schulz, S., y Besold, T. R. (2017). What does explainable ai really mean? a new conceptualization of perspectives. *arXiv preprint arXiv:1710.00794*.

[Dubaish y Jaber, 2024] Dubaish, A. y Jaber, A. (2024). Comparative analysis of svm and ann for machine condition monitoring and fault diagnosis in gearboxes. *Mathematical Modelling of Engineering Problems*.

[Espinosa-Zúñiga, 2020] Espinosa-Zúñiga, J. J. (2020). Aplicación de algoritmos random forest y xgboost en una base de solicitudes de tarjetas de crédito. *Ingeniería, investigación y tecnología*, 21(3).

[Flores-Fernández *et al.*, 2022] Flores-Fernández, C., OLIVARES, C., PINTO-TRONCOSO, G., y REYES-NARVÁEZ, C. (2022). Neurodivergencia en la biblioteca académica: adaptaciones en colecciones y servicios. *Ibersid*, 16(2).

[Franklin *et al.*, 2017] Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., y Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological bulletin*, 143(2):187.

[Gevorkyan *et al.*, 2019] Gevorkyan, M. N., Demidova, A. V., Demidova, T. S., y Sobolev, A. A. (2019). Review and comparative analysis of machine learning libraries for machine learning. *Discrete and Continuous Models and Applied Computational Science*, 27(4):305–315.

[github,] github. Rapidfuzz 3.9.2 documentation. Accedido: Junio, 2024.

[Goyal *et al.*,] Goyal, D., Sheth, F., Mathur, P., y Gupta, A. K. Discrete mathematical models for enhancing cybersecurity: A mathematical and statistical analysis of machine learning approaches in phishing attack detection.

[Gutiérrez *et al.*,] Gutiérrez, S. S. M., Reséndiz, J. L. A., y López, M. G. Machine learning y redes neuronales artificiales para predicción y prevención de trastorno depresivo. *Transformación Digital de las Instituciones Educativas*, p. 122.

[Helm *et al.*, 2020] Helm, P. J., Medrano, M. R., Allen, J. J., y Greenberg, J. (2020). Existential isolation, loneliness, depression, and suicide ideation in young adults. *Journal of Social and Clinical Psychology*, 39(8):641–674.

[Ibrahim *et al.*, 2013] Ibrahim, A. K., Kelly, S. J., Adams, C. E., y Glazebrook, C. (2013). A systematic review of studies of depression prevalence in university students. *Journal of psychiatric research*, 47(3):391–400.

[James *et al.*, 2023] James, G., Witten, D., Hastie, T., Tibshirani, R., y Taylor, J. (2023). *An introduction to statistical learning: With applications in python*. Springer Nature.

[Janiesch *et al.*, 2021] Janiesch, C., Zschech, P., y Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 31(3):685–695.

[Ji *et al.*, 2020] Ji, S., Pan, S., Li, X., Cambria, E., Long, G., y Huang, Z. (2020). Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Transactions on Computational Social Systems*, 8(1):214–226.

[Kadra *et al.*, 2021] Kadra, A., Lindauer, M., Hutter, F., y Grabocka, J. (2021). Regularization is all you need: Simple neural nets can excel on tabular data. *arXiv preprint arXiv:2106.11189*, 536.

[Keshoofy *et al.*, 2023] Keshoofy, A., Pearl, D. L., Lisnyj, K., Thaivalappil, A., y Papadopoulos, A. (2023). Risk and protective factors associated with hopelessness among canadian postsecondary students. *International Journal of Mental Health and Addiction*, pp. 1–16.

[Klonsky y May, 2015] Klonsky, E. D. y May, A. M. (2015). The three-step theory (3st): A new theory of suicide rooted in the "ideation-to-action" framework. *International Journal of Cognitive Therapy*, 8(2):114–129.

[Kumar, 2024] Kumar, S. (2024). Metrics to evaluate your classification model to take the right decisions. `https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/`. Accedido: Junio, 2024.

[Kumaraswamy, 2013] Kumaraswamy, N. (2013). Academic stress, anxiety and depression among college students: A brief review. *International review of social sciences and humanities*, 5(1):135–143.

[Lama, 2011] Lama, M. A.-Q. (2011). The relation of depression and anxiety in academic achievement among group of university students. *International journal of psychology and counselling*, 3(5):96–100.

[Lejeune *et al.*, 2022] Lejeune, A., Le Glaz, A., Perron, P.-A., Sebti, J., Baca-Garcia, E., Walter, M., Lemey, C., y Berrouiguet, S. (2022). Artificial intelligence and suicide prevention: a systematic review. *European psychiatry*, pp. 1–22.

[Linthicum *et al.*, 2019] Linthicum, K. P., Schafer, K. M., y Ribeiro, J. D. (2019). Machine learning in suicide science: Applications and ethics. *Behavioral sciences & the law*, 37(3):214–222.

[Martínez *et al.*, 2021] Martínez, P., Jiménez-Molina, Á., Mac-Ginty, S., Martínez, V., y Rojas, G. (2021). Salud mental en estudiantes de educación superior en chile: una revisión de alcance con meta-análisis. *Terapia psicológica*, 39(3):405–426.

[Micin y Bagladi, 2011] Micin, S. y Bagladi, V. (2011). Salud mental en estudiantes universitarios: Incidencia de psicopatología y antecedentes de conducta suicida en población que acude a un servicio de salud estudiantil. *Terapia psicológica*, 29(1):53–64.

[Ministerio de Educación de Chile, 2023] Ministerio de Educación de Chile (2023). Recomendaciones y orientaciones en salud mental para la educación superior. Accedido: Junio, 2024.

[Mishra *et al.*, 2017] Mishra, S., Sturm, B. L., y Dixon, S. (2017). Local interpretable model-agnostic explanations for music content analysis. En *ISMIR*, volumen 53, pp. 537–543.

[Nalluri *et al.*, 2020] Nalluri, M., Pentela, M., y Eluri, N. R. (2020). A scalable tree boosting system: Xg boost. *Int. J. Res. Stud. Sci. Eng. Technol*, 7(12):36–51.

[Nuzzo, 2014] Nuzzo, R. (2014). Statistical errors. *Nature*, 506(7487):150.

[Organización Mundial de la Salud, 2021] Organización Mundial de la Salud (2021). One in 100 deaths is by suicide. `https://www.who.int/es/news/item/17-06-2021-one-in-100-deaths-is-by-suicide`. Consultado el 14 de marzo de 2023.

[Otárola *et al.*, 2023] Otárola, L. G., Aguirre, E., y Ganga-Contreras, F. (2023). Apreciaciones sobre la neurodivergencia de docentes y estudiantes de una entidad educativa pública chilena. *Journal of the Academy*, (8):5–26.

[Padmasree y Muthyam,] Padmasree, R. y Muthyam, K. Enhancing iot network security through prompt intrusion detection using machine learning.

[Paraıso *et al.*,] Paraıso, E. H., da Silva, A. C., y Nobre, C. N. Underdiagnosed depression in older adults: Analysis of the national health survey and other aggregate factors.

[Pozos-Radillo *et al.*, 2014] Pozos-Radillo, B. E., de Lourdes Preciado-Serrano, M., Acosta-Fernández, M., de los Ángeles Aguilera-Velasco, M., y Delgado-García, D. D. (2014). Academic stress as a predictor of chronic stress in university students. *Psicología educativa*, 20(1):47–52.

[PsyToolkit,] PsyToolkit. The world-leading free platform for online psychological research and information. Accedido: Junio, 2024.

[Putwain, 2007] Putwain, D. (2007). Researching academic stress and anxiety in students: some methodological considerations. *British Educational Research Journal*, 33(2):207–219.

[Quinlan-Davidson *et al.*, 2014] Quinlan-Davidson, M., Sanhueza, A., Espinosa, I., Escamilla-Cejudo, J. A., y Maddaleno, M. (2014). Suicide among young people in the americas. *Journal of Adolescent Health*, 54(3):262–268.

[Randles *et al.*, 2017] Randles, B. M., Pasquetto, I. V., Golshan, M. S., y Borgman, C. L. (2017). Using the jupyter notebook as a tool for open science: An empirical study. En *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pp. 1–2. IEEE.

[Reddy *et al.*, 2018] Reddy, K. J., Menon, K. R., y Thattil, A. (2018). Academic stress and its sources among university students. *Biomedical and pharmacology journal*, 11(1):531–537.

[Rosenbusch *et al.*, 2021] Rosenbusch, H., Soldner, F., Evans, A. M., y Zeelenberg, M. (2021). Supervised machine learning methods in psychology: A practical introduction with annotated r code. *Social and Personality Psychology Compass*, 15(2):e12579.

[Saabith *et al.*, 2020] Saabith, A. S., Vinothraj, T., y Fareez, M. (2020). Popular python libraries and their application domains. *International Journal of Advance Engineering and Research Development*, 7(11).

[Saarela y Jauhiainen, 2021] Saarela, M. y Jauhiainen, S. (2021). Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences*, 3(2):272.

[Sakin Ozen *et al.*, 2010] Sakin Ozen, N., Ercan, I., Irgil, E., y Sigirli, D. (2010). Anxiety prevalence and affecting factors among university students. *Asia Pacific Journal of Public Health*, 22(1):127–133.

[Saleh *et al.*, 2017] Saleh, D., Camart, N., y Romo, L. (2017). Predictors of stress in college students. *Frontiers in psychology*, 8:19.

[Sano *et al.*, 2015] Sano, A., Amy, Z. Y., McHill, A. W., Phillips, A. J., Taylor, S., Jaques, N., Klerman, E. B., y Picard, R. W. (2015). Prediction of happy-sad mood from daily behaviors and previous sleep history. En *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 6796–6799. IEEE.

[Stallman y Hurst, 2016] Stallman, H. M. y Hurst, C. P. (2016). The university stress scale: measuring domains and extent of stress in university students. *Australian Psychologist*, 51(2):128–134.

[Stančin y Jović, 2019] Stančin, I. y Jović, A. (2019). An overview and comparison of free python libraries for data mining and big data analysis. En *2019 42nd International convention on information and communication technology, electronics and microelectronics (MIPRO)*, pp. 977–982. IEEE.

[Stiglic *et al.*, 2020] Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., y Cilar, L. (2020). Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5):e1379.

[Stravynski y Boyer, 2001] Stravynski, A. y Boyer, R. (2001). Loneliness in relation to suicide ideation and parasuicide: A population-wide study. *Suicide and life-threatening behavior*, 31(1):32–40.

[Suresh y Guttag, 2019] Suresh, H. y Guttag, J. V. (2019). A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002*, 2(8):73.

[Teelhawod *et al.*, 2021] Teelhawod, B. N., Akhtar, F., Heyat, M. B. B., Tripathi, P., Mehrotra, R., Asfaw, A. B., Al Shorman, O., y Masadeh, M. (2021). Machine learning in e-health: a comprehensive survey of anxiety. En *2021 International Conference on Data Analytics for Business and Industry (ICDABI)*, pp. 167–172. IEEE.

[van Mens *et al.*, 2020] van Mens, Kasper and de Schepper, CWM and Wijnen, Ben and Koldijk, Saskia J and Schnack, Hugo and de Looff, Peter and Lokkerbol, Joran and Wetherall, Karen and Cleare, Seonaid and O'Connor, Rory C and others (2020). Predicting future suicidal behaviour in young adults, with different machine learning techniques: A population-based longitudinal study. *Journal of affective disorders*, 271:169–177.

[Vilca Masco, 2020] Vilca Masco, H. (2020). Predicción del nivel de estrés en estudiantes universitarios utilizando técnicas de machine learning.

[Vujovic, 2021] Vujovic, Z. (2021). Classification model evaluation metrics. *International Journal of Advanced Computer Science and Applications*, 12(6):599–606.

[Williams *et al.*, 2008] Williams, C. B., Galanter, M., Dermatis, H., y Schwartz, V. (2008). The importance of hopelessness among university students seeking psychiatric counseling. *Psychiatric Quarterly*, 79:311–319.

[Wu y Feng, 2018] Wu, Y.-c. y Feng, J.-w. (2018). Development and application of artificial neural network. *Wireless Personal Communications*, 102:1645–1656.

[Yapo y Weiss, 2018] Yapo, A. y Weiss, J. (2018). Ethical implications of bias in machine learning.

[Yarkoni y Westfall, 2017] Yarkoni, T. y Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12(6):1100–1122.