

**MODELO MACHINE LEARNING/APRENDIZAJE AUTOMÁTICO PARA DETECCIÓN
DE SEÑALES DE ALERTA DE PROBLEMAS ASOCIADOS A SALUD MENTAL EN
POBLACIÓN INFANTIL ESCOLAR DE LA CIUDAD DE MANIZALES**

Integrantes

Gloria Esperanza Giraldo Vélez
Angélica María González Martínez
Andrea Muñoz Calderón

Ejecutores

Natalia Betancur Herrera
Frank Yesid Zapata Castaño
Margarita María Orozco García

**Universidad de Antioquia, Universidad de Caldas
Talento TECH
BOOTCAMP Inteligencia Artificial
Marzo 2025**

Tabla de contenido

Capítulo 1 Introducción	2
Capítulo 2 Planteamiento del problema	4
Capítulo 3 Objetivos	6
Objetivo general	6
Objetivos específicos	6
Capítulo 4 Justificación	7
Capítulo 5 Alcance	9
Capítulo 6 Antecedentes	10
Aplicación de Machine Learning en la detección de trastornos mentales ...	10
Capítulo 7 Marco Teórico	15
Influencia del Estadio de Desarrollo en la Salud Mental Infantil	16
Capítulo 8 Metodología	19
Recopilación de Datos	19
Descripción de las variables	19
Preprocesamiento de Datos	20
Desarrollo del Modelo de Machine Learning	20
Transformación de variables	21
Entrenamiento del modelo	21
Evaluación de desempeño	21
Balance de clases	22
Capítulo 9 Presupuesto	23
Capítulo 10 Resultados	24
Capítulo 11 Conclusiones.....	25
Capítulo 12 Recomendaciones.....	26
Capítulo 13 Líneas futuras.....	28
Capítulo 14 Bibliografía.....	29

Introducción

La salud mental infantil es un aspecto fundamental del bienestar general de los niños, niñas y adolescentes en su desarrollo; a pesar de su importancia, sigue siendo un tema frecuentemente subestimado en muchas sociedades, incluidas las de América Latina.

En particular, la población infantil escolar de Manizales, al igual que en muchas otras ciudades de Colombia, enfrenta desafíos relacionados con trastornos de salud mental que pueden impactar negativamente su desarrollo académico, social y emocional.

Factores como el entorno familiar, las condiciones socioeconómicas, las experiencias de abuso o violencia, y el estrés derivado de la vida escolar son solo algunos de los elementos que pueden contribuir a la aparición de problemas mentales en esta población.

La detección temprana de señales de alerta es crucial para implementar intervenciones adecuadas que promuevan un desarrollo saludable. Sin embargo, el diagnóstico temprano de problemas de salud mental en niños suele ser complejo debido a la falta de herramientas accesibles y a la variabilidad de los síntomas. Es aquí donde las tecnologías avanzadas, como el aprendizaje automático (Machine Learning), pueden jugar un papel esencial, teniendo en cuenta, que estos modelos informáticos tienen la capacidad de analizar grandes volúmenes de datos, identificar patrones y predecir posibles riesgos de manera más eficiente que los métodos tradicionales.

Este proyecto tiene como objetivo la creación de un modelo de Machine Learning para la detección temprana de señales de alerta relacionadas con problemas de salud mental en la población infantil escolar de Manizales. A través de la recopilación y análisis de datos demográficos y emocionales obtenidos del SIMAT (Sistema Integrado de Matrícula) en las instituciones educativas públicas de la ciudad de Manizales, el

modelo busca identificar patrones que permitan reconocer factores de riesgo asociados al desarrollo de trastornos mentales durante la etapa escolar. La implementación de este sistema podría transformar la forma en que los profesionales de la salud y la educación abordan la salud mental infantil, permitiendo intervenciones más oportunas y específicas, y contribuyendo a una sociedad más consciente y proactiva en la protección del bienestar emocional de los niños.

Planteamiento del Problema

La salud mental infantil es una preocupación creciente a nivel global, especialmente en poblaciones escolares, donde los niños y adolescentes se ven afectados por una serie de factores estresantes que pueden afectar su bienestar emocional y psicológico. En Manizales, al igual que en otras ciudades de Colombia, la población infantil escolar enfrenta una serie de desafíos relacionados con trastornos de salud mental, como la ansiedad, la depresión, el trastorno por déficit de atención e hiperactividad (TDAH), y los trastornos de conducta. Estos problemas, si no se detectan y tratan de manera temprana, pueden tener consecuencias graves para el desarrollo académico, social y emocional de los niños, afectando su rendimiento escolar, relaciones interpersonales y, en muchos casos, su futuro bienestar integral.

Uno de los principales desafíos en la atención de la salud mental infantil radica en la detección temprana de señales de alerta, ya que los síntomas de los trastornos mentales en niños no siempre son fácilmente identificables y a menudo se confunden con comportamientos típicos de la infancia. Además, muchos de estos trastornos se presentan en un contexto familiar y social específico, donde factores como el estrato socioeconómico, las condiciones familiares, el entorno escolar y los antecedentes demográficos pueden desempeñar un papel importante.

A pesar de la creciente conciencia sobre la importancia de la salud mental infantil, la detección temprana sigue siendo un desafío debido a la falta de herramientas accesibles y eficaces para los profesionales de la salud y la educación. En este contexto, los modelos de Machine Learning (aprendizaje automático) ofrecen una solución innovadora y prometedora. A través del análisis de grandes volúmenes de

datos, incluidos datos demográficos y emocionales, estos modelos pueden identificar patrones ocultos y predecir riesgos de salud mental en la población escolar.

Sin embargo, en la ciudad de Manizales, aún no existe una implementación de modelos de machine learning aplicados específicamente para la detección de problemas de salud mental infantil, lo que limita las intervenciones tempranas y la personalización del apoyo a los estudiantes en riesgo. Por lo tanto, surge la necesidad de desarrollar un modelo capaz de analizar datos relevantes, como el rendimiento académico, los factores demográficos (edad, género, estrato socioeconómico), antecedentes familiares y otros indicadores, con el objetivo de detectar señales de alerta que permitan identificar a los niños en riesgo de desarrollar trastornos mentales. La implementación de este modelo no solo contribuiría a una mejor comprensión de los factores de riesgo, sino que también mejoraría la capacidad de los profesionales en salud y educación para intervenir de manera más temprana y efectiva.

Este proyecto busca abordar esta necesidad y proporcionar una herramienta eficiente para la prevención de problemas de salud mental infantil en la población escolar de Manizales, con la meta de mejorar el bienestar emocional y académico de los niños y reducir el impacto a largo plazo de los trastornos mentales no tratados.

Objetivos

Objetivo General

Desarrollar un modelo de aprendizaje automático (machine learning) que permita identificar y predecir señales de alerta de problemas asociados a la salud mental en niños de la población escolar de Manizales, para promover intervenciones tempranas y adecuadas.

Objetivos Específicos:

Recopilar datos relevantes sobre la población de instituciones educativas públicas de Manizales, que incluyan factores socioeconómicos y emocionales de los estudiantes, con el fin de estructurar un conjunto de datos adecuado para el entrenamiento del modelo de machine learning.

Realizar un análisis estadístico y exploratorio de los datos recolectados para identificar patrones y correlaciones significativas que puedan indicar señales de alerta de posibles problemas de salud mental en los niños.

Implementar diferentes algoritmos de machine learning, para crear un modelo de árboles de decisión que sea capaz de predecir la probabilidad que un estudiante muestre señales de alerta de problemas de salud mental.

Validar la precisión y efectividad del modelo desarrollado utilizando técnicas de validación y métricas de desempeño.

Justificación

Durante el desarrollo infantil, la salud mental es un componente fundamental del bienestar general de los niños, con un impacto directo en su desarrollo emocional, social y académico. En este contexto, los problemas de salud mental pueden manifestarse de diversas formas, como trastornos de ansiedad, depresión, problemas de conducta o dificultades en la relación con los compañeros. Estos trastornos no solo afectan la vida cotidiana de los niños, sino también pueden tener efectos a largo plazo en su rendimiento académico y en su integración social. Sin embargo, en muchos contextos, incluido el de la ciudad de Manizales, los problemas de salud mental infantil pasan desapercibidos por falta de herramientas eficientes y sistemas de monitoreo adecuados.

Por consiguiente, este proyecto surge del creciente interés por abordar los problemas de salud mental en la población infantil, un tema que ha ganado relevancia en la última década a nivel global. En Colombia, y particularmente en Manizales, el enfoque en la salud mental en el ámbito escolar es aún incipiente, lo que subraya la importancia de desarrollar herramientas innovadoras que puedan facilitar la identificación temprana de trastornos emocionales. La creación de un modelo de *machine learning* para la detección de señales de alerta en niños no solo responde a la necesidad de mejorar el bienestar de los estudiantes, sino también a un interés más amplio por optimizar el uso de tecnologías en salud pública y educación.

Se tiene en cuenta, que el uso de *machine learning* en la salud mental infantil es una propuesta innovadora que, aunque está comenzando a implementarse en algunos países, aún es desconocido en muchas regiones de Colombia. La novedad de este

proyecto radica en aplicar técnicas avanzadas de aprendizaje automático para analizar datos complejos provenientes de múltiples fuentes (comportamiento escolar, información demográfica y socioemocional etc.), lo que permite detectar patrones y señales de alerta que a menudo no son evidentes para los observadores humanos. Esta metodología no solo mejora la precisión en la detección, sino que introduce un enfoque tecnológico que, además, puede ser escalable y replicable en otros contextos educativos y de salud pública a nivel local, nacional e internacional.

La utilidad de este modelo es múltiple y de gran impacto en diversos niveles. Para las instituciones educativas, este sistema representaría una herramienta proactiva que permitirá identificar a los estudiantes que evidencien señales de alerta en problemas de salud mental, facilitando intervenciones tempranas por parte de profesionales capacitados. Además, los resultados de este modelo ayudarán a los padres y educadores a comprender mejor las necesidades emocionales de los niños, lo que puede mejorar el ambiente escolar y la dinámica familiar.

Desde el punto de vista de la salud pública, el uso de *machine learning* permitiría optimizar los recursos disponibles para la atención de salud mental, mejorando la eficiencia en la distribución de los servicios y asegurando que los niños que más lo necesitan reciban la atención adecuada. Este enfoque puede contribuir a la reducción de la carga social y económica asociada a los trastornos mentales no tratados, promoviendo una comunidad escolar más saludable y equitativa; además, reducir el estigma asociado a los trastornos mentales, creando un ambiente más comprensivo y saludable.

Alcance

El alcance de este proyecto se centra en el desarrollo e implementación de un modelo de machine learning para la detección temprana de señales de alerta relacionadas con problemas de salud mental en la población infantil escolar de la ciudad de Manizales. Este modelo utilizará los datos de los niños matriculados en las instituciones educativas oficiales de la ciudad con el fin de identificar patrones y factores de riesgo asociados a posibles trastornos mentales. Se espera que el modelo permita detectar posibles signos tempranos de problemas de salud mental; usar datos demográficos que incorporen variables como el género, edad, estrato socioeconómico, zona de residencia, para mejorar la precisión de la predicción de riesgos asociados a problemas de salud mental; crear un sistema de clasificación que permita categorizar a los niños según el nivel de riesgo, permitiendo la intervención temprana de profesionales en salud y educación de las instituciones educativas públicas de la ciudad de Manizales, con el objetivo de ofrecer soluciones locales a una problemática global.

Antecedentes

En los últimos años, se ha incrementado el interés en la utilización de tecnologías avanzadas, como machine learning, en el ámbito de la salud mental, particularmente en la detección temprana de trastornos psicológicos. A continuación, se destacan algunos antecedentes relevantes en esta área:

Aplicación de Machine Learning en la detección de trastornos mentales:

En diversas investigaciones, el uso de algoritmos de aprendizaje automático ha demostrado ser efectivo para identificar patrones que podrían pasar desapercibidos por observadores humanos. Sánchez Torres, A. (2023). de la Universidad Politécnica de Valencia. Desarrollo un modelo de aprendizaje automático para la mejora de la detección temprana de casos de TDAH en niños y adolescentes, donde se utilizaron varios modelos predictivos, como los árboles de decisión o redes neuronales, generando una plataforma que permite un seguimiento más personalizado y efectivo del TDAH, mejorando así la calidad de vida de los pacientes y proporcionando información valiosa para la investigación y el tratamiento de la enfermedad.

Así mismo, Castillo, G. (2023). En su investigación, Contribución de las Tecnologías de la Información y las Comunicaciones para ayudar a la prevención del Comportamiento Suicida en Castilla y León: Un enfoque de Machine Learning y Salud Digital. Utiliza estos enfoques desde las TIC para ayudar a la prevención del Comportamiento Suicida y contribuir a nivel técnico, tecnológico y social en la toma de decisiones sobre las estrategias de manejo de estos pacientes.

Una investigación de actualidad siguiendo la línea de nuevos términos y diagnósticos de salud mental se identifica un estudio de gran relevancia. Publicado por Rodríguez, et

al. (2025). Titulado, Nomofobia y rendimiento académico: prevalencia y correlación en los jóvenes. De la Universidad de Granada en España. Menciona que el uso continuo e inadecuado del smartphone puede provocar problemas de adicción y dependencia que pueden desencadenar otras dificultades como el miedo irracional a no tener el teléfono disponible, así como las posibilidades que nos ofrece en cualquier momento. Este es el caso de la nomofobia (no-mobile-phobia). El objetivo de esta investigación fue analizar la prevalencia de la nomofobia en jóvenes de entre 12 y 20 años, así como determinar la relación entre todas las variables que identifican los niveles de nomofobia, y corroborar la incidencia del rendimiento académico en la mayor o menor prevalencia de este problema. Para ello se llevó a cabo una investigación cuantitativa en la que se utilizó el cuestionario estandarizado NMP-Q. La muestra está compuesta por un total de 1.630 estudiantes de distintas etapas educativas. Los resultados determinan que la imposibilidad de contactar y ser contactado son las variables en las que se ha observado una mayor prevalencia de esta fobia entre la población estudiada. Además, los estudiantes que afirman que su rendimiento académico es menor debido al uso excesivo de su teléfono inteligente muestran mayor ansiedad, nerviosismo, malestar, miedo o molestias al no poder hacer uso inmediato de su teléfono y de sus diferentes posibilidades. Siendo esta investigación de gran aporte a nivel de salud mental, uso de la tecnología y rendimiento académico en niños y jóvenes.

En Colombia, algunas iniciativas han comenzado a integrar tecnologías para abordar la salud mental en población escolar y universitaria, aunque aún son limitadas.

Un proyecto reciente en Bogotá liderado por Montes, et al. (2024), ha logrado la integración de IA y salud mental, por medio de la implementación de un sistema de

asistencia virtual para la detección y atención de problemas psicológicos. Este estudio se centra en el diseño de un asistente virtual basado en inteligencia artificial, con el propósito de identificar rasgos de depresión y ansiedad en los estudiantes de primera matrícula de la Universidad Nacional Abierta y a Distancia (UNAD). Buscó integrar herramientas digitales de seguimiento psicológico en escuelas para mejorar la detección de trastornos en la población estudiantil. La implementación de esta tecnología tiene el potencial de mejorar el acceso a la salud mental, reducir el estigma asociado a la búsqueda de apoyo psicológico y ofrecer una herramienta complementaria a la atención profesional.

Lozano, et al. (2021). Diseñó una solución tecnológica para suministrar apoyo psicológico a la sociedad colombiana. Se identificaron diferentes medios tecnológicos que permiten la inclusión de la tecnología en temas de salud, por lo cual se tuvieron en cuenta muchas aplicaciones ya existentes como base para la selección de la mejor alternativa y así dar solución a la problemática evidenciada en este proyecto.

Confirmando que existen medios que permiten integrar a los futuros profesionales, estudiantes de psicología para que tengan un acercamiento a la realidad laboral mediante el uso de aplicaciones emergentes que permiten el desarrollo de sus prácticas profesionales.

Saavedra, et al (2024). En su investigación. Análisis de depresión en adolescentes de básica secundaria y media usando algoritmos de machine learning. Un estudio reciente que analizó factores determinantes relacionados con la ansiedad, el estrés y la depresión en adolescentes de sexto a undécimo grado en una institución educativa del Tolima, Colombia, donde se aplicaron algoritmos de clasificación para identificar

variables predictoras de depresión, destacando factores como fatiga crónica, dificultad para tomar decisiones y conflictos interpersonales. Los resultados resaltan la importancia de la detección temprana y el uso de herramientas de machine learning para enfoques preventivos y personalizados.

Brand, et al. (2024). En Bucaramanga, Colombia, se llevó a cabo un estudio que utilizó algoritmos de Machine Learning para identificar patrones en casos de intentos suicidas reportados en el Sistema Nacional de Vigilancia en Salud Pública. El análisis de datos entre 2016 y 2023 permitió descubrir patrones previamente desconocidos, subrayando la utilidad de estas herramientas en la salud pública.

Noriega, et al. (2023). En su publicación. Resiliencia académica de adolescentes, factores asociados y metodología empleada: una revisión sistemática. Concluye que de acuerdo a estudios sobre la prevalencia de problemas de salud mental en niños en Colombia: La Encuesta Nacional de Salud Mental 2015 reveló que uno de cada cinco niños en Colombia presenta algún trastorno mental, siendo los más comunes la ansiedad y la depresión. Estos trastornos, si no se abordan de manera temprana, pueden afectar gravemente su rendimiento académico y su desarrollo emocional. Este dato resalta la urgencia de contar con sistemas de detección temprana en el ámbito escolar, como el que propone este proyecto.

A nivel municipal, un ejemplo relevante en el contexto de la salud mental infantil es el trabajo de Agudelo, et al. (2023). En su publicación. Salud mental pediátrica, experiencia de crianza y desempleo en un contexto latinoamericano. Donde desarrolla un estudio descriptivo de corte transversal, correlacional, con población de estrato uno y dos de Manizales. Se encontró una media de edad 9,28 años, una asociación

estadísticamente significativa entre depresión en cuidadores, dificultades en la experiencia de crianza y desempleo, con los síntomas psiquiátricos en la niñez y la juventud, además de una cantidad importante de cuidadores, con un desempleo del 55%.

Aunque no se identificaron estudios específicos sobre la aplicación de modelos de Machine Learning para la detección de problemas de salud mental en la población infantil escolar de Manizales, un estudio realizado en la ciudad por Mejía, et al. (2023). Analizó la correlación entre el perfil neuropsicológico y la creatividad en niños con riesgo de problemas de salud mental. Este estudio identificó correlaciones significativas que pueden informar futuras intervenciones.

Estos antecedentes destacan el potencial de los modelos de aprendizaje automático en la detección y prevención de problemas de salud mental en poblaciones infantiles y adolescentes, ofreciendo herramientas valiosas para profesionales y autoridades de salud y educación en la identificación temprana y atención oportuna.

Marco Teórico

Este proyecto se basa en varios conceptos clave que combinan los enfoques de salud mental infantil con las técnicas de Machine Learning aplicadas a la predicción y detección temprana. A continuación, se exponen los principales enfoques y teorías que sustentan este proyecto:

La salud mental infantil es un aspecto esencial del desarrollo integral de los niños, que influye en su bienestar emocional, social y académico. Según la Organización Mundial de la Salud (OMS), los trastornos mentales en la infancia pueden afectar el rendimiento escolar y las relaciones interpersonales, además de predisponer a problemas de salud mental en la adultez. Los trastornos más comunes incluyen la depresión, la ansiedad, el trastorno de déficit de atención e hiperactividad (TDAH) y los trastornos de conducta. La detección temprana y el tratamiento oportuno son fundamentales para prevenir efectos negativos a largo plazo.

El uso de modelos predictivos en salud mental busca identificar patrones ocultos en datos que puedan sugerir la presencia de un trastorno o problema. Estos modelos utilizan técnicas estadísticas y de Machine Learning para analizar la información, tales como características demográficas, comportamentales y psicológicas.

El uso de Machine Learning en entornos escolares está en auge. Según De la Hoz, et al. (2019), los sistemas de aprendizaje automático permiten realizar un análisis exhaustivo de los comportamientos de los estudiantes y sus interacciones, lo que puede ser útil para identificar factores de riesgo en salud mental, como el aislamiento social, la agresividad, o las fluctuaciones en el rendimiento académico. La información extraída de las interacciones diarias, las encuestas sobre bienestar emocional, y los

resultados académicos pueden ser procesada para generar alertas tempranas de problemas emocionales.

La detección temprana de problemas de salud mental es fundamental para intervenir antes de que los trastornos se agudizan. Según el modelo de Loeber et al. (2001), citado por Córdova, (2008). cuanto más temprano se detecta un trastorno, mayores son las probabilidades de intervención exitosa. Este concepto es especialmente relevante en el contexto escolar, donde los problemas emocionales a menudo afectan el rendimiento académico y las relaciones interpersonales de los niños. Los sistemas automatizados basados en machine learning pueden hacer que la detección temprana sea más precisa y eficiente, facilitando la intervención temprana por parte de psicólogos y educadores.

Influencia del Estadio de Desarrollo en la Salud Mental Infantil:

Teniendo en cuenta algunos de los autores más destacados en las teorías del desarrollo humano, como Jean Piaget, Erik Erikson y Lev Vygotsky entre otros, se puede deducir que, el estadio de desarrollo de los niños es otro factor clave que influye en cómo experimentan y manifiestan los problemas de salud mental. En la infancia temprana, las emociones, las habilidades sociales y la regulación emocional están en pleno desarrollo, y cualquier alteración en este proceso puede tener efectos duraderos.

a. Etapa de la primera infancia (0-6 años):

Durante los primeros años de vida, el cerebro infantil es altamente moldeable. Los niños en esta etapa están aprendiendo a gestionar sus emociones, a desarrollar vínculos afectivos con los cuidadores y a regular sus respuestas emocionales.

Los problemas en el desarrollo emocional y social en esta etapa pueden tener un efecto a largo plazo. La falta de apego seguro o la exposición temprana a situaciones estresantes (como negligencia o abuso) pueden generar trastornos de apego, ansiedad y problemas de autorregulación emocional. Además, en esta etapa los niños pueden tener dificultades para comunicar sus emociones y necesidades, lo que puede dificultar la identificación temprana de problemas de salud mental.

b. Etapa preescolar (3-6 años):

En esta etapa, los niños desarrollan una mayor comprensión de sí mismos y de los demás. Aprenden a interactuar con sus compañeros, y sus habilidades para manejar las emociones y establecer relaciones sociales comienzan a consolidarse. Los niños en esta etapa también son más susceptibles a las influencias externas, como el ambiente familiar, escolar y social.

La inseguridad emocional, la agresividad, los problemas de socialización y la baja autoestima son más evidentes en los niños de esta edad si experimentan dificultades en el hogar o en la escuela. El abuso de sustancias o la violencia doméstica también pueden tener un impacto negativo en el bienestar emocional, llevando a comportamientos disruptivos, ansiedad o dificultades para concentrarse.

c. Etapa escolar (6-12 años):

A medida que los niños crecen y se desarrollan, comienzan a enfrentar desafíos más complejos relacionados con la competencia académica, las relaciones con sus compañeros y las expectativas sociales. A esta edad, los niños tienen una mayor conciencia de su entorno y pueden experimentar más presiones externas, como el rendimiento académico, las relaciones con amigos y las expectativas familiares.

Los niños que no tienen las habilidades emocionales necesarias para manejar el estrés, la frustración o los conflictos sociales pueden desarrollar trastornos de ansiedad, depresión o problemas de conducta. La presión para cumplir con las expectativas académicas o sociales puede resultar en un aumento de la ansiedad o incluso en trastornos como el trastorno obsesivo-compulsivo (TOC).

El desarrollo emocional y cognitivo de los niños no es lineal, y las experiencias de vida durante estas etapas influyen profundamente en la forma en que procesan y responden a las emociones y problemas. Las capacidades para la autorregulación emocional y la toma de decisiones comienzan a formarse en la primera infancia, y las experiencias de vida (familiares, escolares, sociales) durante este período moldean estos procesos. (Condemarín, et al 2016).

Las habilidades de autorregulación emocional, como la capacidad para calmarse después de una frustración, jugar con otros niños de manera cooperativa o hacer frente a los desafíos académicos, son fundamentales para el bienestar mental. La ausencia de apoyo emocional adecuado durante esta etapa puede predisponer a los niños a desarrollar trastornos emocionales. (Ayala, 2024).

Metodología

Para llevar a cabo el desarrollo de este modelo de machine learning, se utilizará una metodología que abarca las siguientes fases:

Recopilación de Datos:

La base de datos fue obtenida del Sistema Integrado de Matrícula del Ministerio de Educación Nacional solicitada a la Secretaría de Educación de Manizales teniendo en cuenta la matrícula del año 2025 en las instituciones educativas oficiales de Manizales, tanto urbanas como rurales y que abarcan desde el grado Prejardín hasta grado 11°, incluyendo las jornadas mañana, tarde, única, nocturna, fin de semana y programas complementarios.

Descripción de las variables:

VARIABLES	DESCRIPCIÓN	CLASIFICACIÓN	VALORES ÚNICOS
INSTITUCIÓN	Corresponde a la columna número 2, la cual corresponde a la institución educativa.	Nominal	50
SECTOR	Corresponde a la columna número 4, que corresponde al sector educativo en este caso el sector oficial.	Nominal	0
SEDE	Corresponde a la columna número 5 que se refiere a la sede de la institución educativa.	Nominal	106
ZONA_SEDE	Corresponde a la columna número 6 donde se ubican las instituciones como rurales o urbanas.	Nominal	2
JORNADA	Corresponde a la columna número 7, donde mencionan las jornadas escolares mañana, tarde, nocturna, jornada única y fin de semana.	Nominal	5
GRADO_COD	Corresponde a la columna número 8 que corresponde al grado escolar en el que están matriculados los estudiantes.	Nominal	26
ESTRATO	Corresponde a la columna número 9, en el cual se encuentra el estrato socioeconómico de los estudiantes	Nominal	8
SISBEN	Corresponde a la columna número 10, donde encontramos la clasificación del sisben para los estudiantes.	Nominal	48
GÉNERO	Corresponde a la columna número 11 donde describen el género femenino o masculino de los estudiantes.	Nominal	2
FECHA DE NACIMIENTO	Corresponde a la columna número 12, de las fechas de nacimiento de los estudiantes matriculados.	Nominal	6029
BARRIO	Corresponde a la columna número 13 describe los barrios de procedencia de los estudiantes.	Nominal	2117
DISCAPACIDAD	Corresponde a la columna número 14 donde se indica si el estudiante presenta alguna discapacidad.	Nominal	11
PAÍS -ORIGEN	Corresponde a la columna número 15, describe el país de origen del estudiante.	Nominal	24
EDAD	Corresponde a la columna 16 donde describe la edad de cada estudiante	Ordinal	16

Preprocesamiento de Datos:

Limpieza de datos:

Se realizó la limpieza de datos de manera manual en el documento en **Excel** ya que no tenía mayores dificultades, los datos faltantes se registró **SIN INFORMACIÓN** Se eliminaron algunas columnas que no presentaban relevancia en el diseño del modelo, como la columna del **AÑO**, ya que era en la vigencia de 2025; la columna de **ETC** (Entidad Territorial Certificada) ya que corresponde a Manizales y **SECTOR** que corresponde al oficial. Se hizo corrección de inconsistencias y se transformación de variables categóricas a variables numéricas,

ANO	ETC	INSTITUCION	SECTOR	SEDE	ZONA_SEDE	JORNADA	GRADO_COD	ESTRATO	SISBEN IV	GENERO	FECHA_NACIMIENTO	BARRIO	DISCAPACIDAD	PAIS_ORIGEN	
0	2025	MANIZALES	INSTITUCIÓN EDUCATIVA SIETE DE AGOSTO	OFICIAL	INSTITUCIÓN EDUCATIVA SIETE DE AGOSTO	URBANA	MAÑANA	8	ESTRATO 1	NO APLICA	MASCULINO	2011-02-01	CARMEN	NO APLICA	COLOMBIA
1	2025	MANIZALES	INSTITUCIÓN EDUCATIVA SIETE DE AGOSTO	OFICIAL	INSTITUCIÓN EDUCATIVA SIETE DE AGOSTO	URBANA	MAÑANA	4	ESTRATO 1	A1	MASCULINO	2014-05-05	EL CARMEN	NO APLICA	COLOMBIA
2	2025	MANIZALES	INSTITUCIÓN EDUCATIVA COLEGIO SAN JORGE	OFICIAL	INSTITUCIÓN EDUCATIVA COLEGIO SAN JORGE	URBANA	TARDE	2	ESTRATO 3	A3	MASCULINO	2017-11-10	SAN JORGE	NO APLICA	COLOMBIA

Tipos de datos por columna:

INSTITUCION

SEDE

ZONA_SEDE

JORNADA

GRADO_COD

ESTRATO

SISBEN IV

GENERO

FECHA_NACIMIENTO

BARRIO

DISCAPACIDAD

PAIS_ORIGEN

object

object

object

object

int64

object

object

object

datetime64[ns]

object

object

object

object

Tipos de datos por columna:

```

INSTITUCION      object
SEDE              object
ZONA_SEDE        object
JORNADA           object
GRADO_COD        int64
ESTRATO           object
SISBEN_IV         object
GENERO            object
FECHA_NACIMIENTO datetime64[ns]
BARRIO            object
DISCAPACIDAD     object
PAIS_ORIGEN      object
dtype: object

```

Desarrollo del Modelo de Machine Learning:

Para la construcción del modelo de **Machine Learning** se utilizaron datos demográficos y registros de discapacidad (trastornos socioemocionales y físicos) de los estudiantes para predecir señales de alerta asociadas a posibles problemas de salud mental, como ansiedad, depresión, estrés y TDAH.

El modelo debe ser capaz de identificar patrones y correlaciones en los datos que no son fácilmente detectables por un observador humano. El que mejor se adapta a nuestro proyecto es el modelo de clasificación **Árboles de Decisión, Random Forest**.

Transformación de Variables: Se seleccionaron las variables más relevantes para predecir problemas de salud mental y factores socioeconómicos.

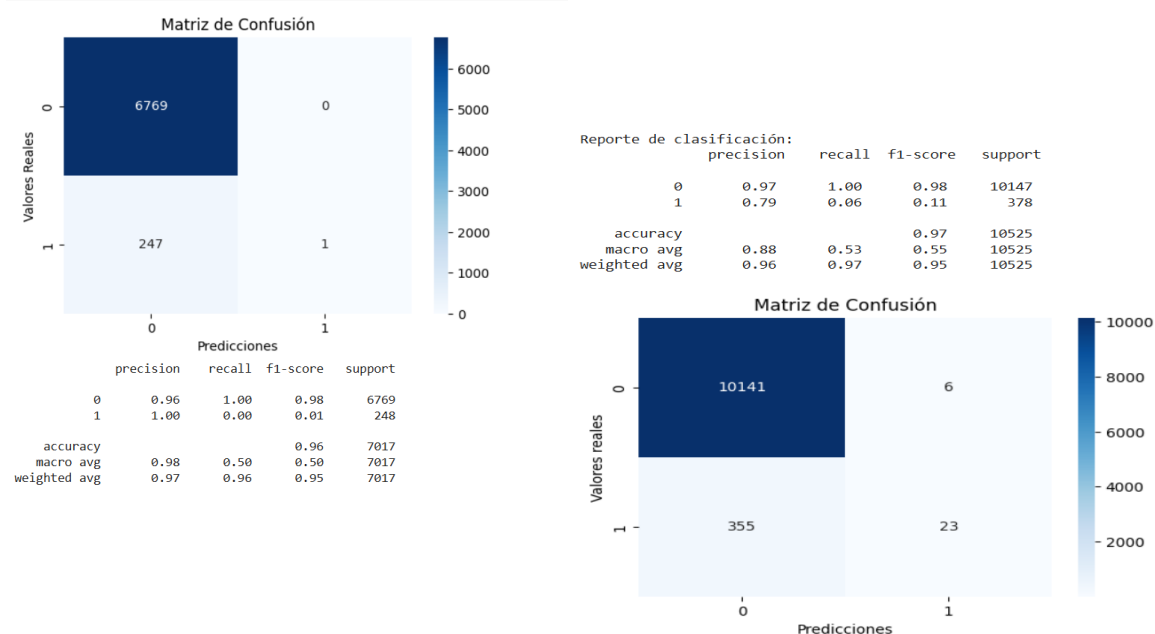
Además, se convirtieron las variables categóricas en variables numéricas mediante HyperTransformer

Entrenamiento del modelo: El modelo fue entrenado utilizando los datos disponibles., se realizaron diversas pruebas con distintos algoritmos y se seleccionó el que ofrecía el mejor desempeño en términos de Precisión, Recall, Accuracy.

Evaluación de desempeño: Se tuvieron en cuenta métricas como la precisión, recall, y F1-score para evaluar la efectividad del modelo en la detección de señales de alerta.

```
37 # 10 Separar en entrenamiento y prueba
38 X_train, X_test, y_train, y_test = train_test_split(X_transformed, y,
39 test_size=0.3, random_state=42)
```

Una vez que el modelo fue entrenado y se procedió a validarlo, para ello se hizo una división de los datos, 70% para entrenarlo y 30% restantes corresponden a la validación



Balance de clases:

En ambas matrices se evidencia un claro desbalance de clases, donde la clase "0" (negativa) tiene una cantidad significativamente mayor de muestras que la clase "1" (positiva). En la primera matriz, la clase "1" tiene un **recall de 0.00**, lo que indica que prácticamente no se detectan positivos. En la segunda matriz, la clase "1" tiene un **recall de 0.06**, lo que sigue siendo muy bajo, pero mejora respecto a la primera matriz. En la primera matriz, **solo 1 caso** de la clase "1" se predice correctamente, mientras que **247 casos positivos** son clasificados erróneamente como "0". En la segunda matriz, **23 casos positivos** fueron correctamente identificados, aunque **355 aún están mal clasificados**. La primera matriz tiene un **accuracy de 96%**, pero esto es engañoso porque casi todos los positivos se clasifican mal. La segunda matriz tiene un **accuracy de 97%**, con una leve mejora en la detección de la clase "1".

Ambos modelos predicen muy bien la clase 0, pero fallaron en la clase 1.

El desbalance de clases, donde la clase 1 es muy minoritaria, el modelo no aprende bien sus patrones.

Posibles soluciones: balanceo de datos, ajuste de umbral de decisión, uso de métricas alternativas como AUC-ROC.

Aunque la solución más segura sería robustecer la información de los datos.

Presupuesto

PRESUPUESTO					
ACTIVIDADES		CANTIDAD	COSTO MENSUAL	DURACIÓN EN MESES	COSTO TOTAL
Investigación y Planificación	Revisión de literatura y análisis de requisitos	1	8.000.000	1	8.000.000
	Definición de objetivos y métricas del modelo	1	5.000.000	2	10.000.000
	Consultoría con expertos en salud mental	3	3.000.000	2	6.000.000
Recolección y Preparación de Datos	Adquisición de datos	1	10.000.000	1	10.000.000
	Limpieza y preprocesamiento de datos	4	5.000.000	2	10.000.000
Desarrollo del Modelo	Selección de algoritmos y diseño del modelo	3	4.000.000	2	8.000.000
	Entrenamiento del modelo	5	5.000.000	3	15.000.000
	Optimización de hiperparámetros	4	2.000.000	3	6.000.000
Validación y Evaluación	Validación cruzada y pruebas del modelo	2	2.500.000	3	5.000.000
	Análisis de métricas (precisión, recall, F1-score, etc.)	2	2.000.000	2	4.000.000
	Ajustes finales basados en resultados	2	2.000.000	2	10.000.000
Implementación	Pruebas en entornos reales	3	5.000.000	3	15.000.000
Mantenimiento y Actualización	Monitoreo del modelo en producción	8	1.000.000	8	8.000.000
	Actualización de datos y reentrenamiento	8	1.000.000	8	8.000.000
	Soporte técnico	1	500.000	8	4.000.000
Otros Costos	Licencias de software y herramientas	1	4.000.000	1	4.000.000
	Infraestructura en la nube (almacenamiento y computación)	1	4.000.000	6	24.000.000
	Capacitación del equipo	3	8.000.000	2	16.000.000
TOTAL					171.000.000

Resultados

El modelo predictivo presentó dificultades para identificar a las personas con discapacidad, debido a una marcada disparidad en la proporción entre aquellas que presentan esta condición y aquellas para las que no aplica. Este resultado evidencia la necesidad de recopilar un volumen de datos mayor y mejor estructurado, con información más detallada y específica, que permita abordar estas limitaciones de manera efectiva.

Conclusiones

Podemos concluir que el modelo predictivo actual enfrenta serias limitaciones para identificar correctamente a las personas con discapacidad, debido al notable desbalance en los datos disponibles, donde la mayoría de las observaciones pertenecen a la categoría "No aplica". Esta situación subraya la necesidad de:

Mejorar la calidad de los datos, recopilando información más específica y representativa de las personas con discapacidad.

Aumentar la cantidad de datos disponibles para equilibrar las proporciones entre las distintas categorías.

Estas acciones no solo mejorarían el rendimiento del modelo, sino que también garantizarían que sus resultados sean más precisos e inclusivos. Esto es crucial en contextos donde la identificación precisa tiene un impacto significativo en la toma de decisiones.

Recomendaciones

Se recomienda fortalecer el monitoreo y la detección en las instituciones educativas e integrar herramientas tecnológicas dentro de las rutinas escolares para observar patrones de comportamiento y emociones que pueden permitir la identificación temprana de posibles señales de alerta, mejorando las intervenciones futuras.

Involucrar a los padres en el proceso de detección y prevención, desarrollando programas que involucren a los padres y cuidadores en la detección temprana de problemas de salud mental en la primera infancia. Mediante aplicaciones o herramientas basadas en Machine Learning, los padres pueden recibir recomendaciones personalizadas sobre cómo apoyar a sus hijos y ser capacitados en la detección de cambios significativos en el comportamiento, que pueden ser indicativos de trastornos emocionales.

Promover la formación y sensibilización de profesionales de la salud y la educación, donde reciban formación continua en el uso de herramientas tecnológicas para la detección y manejo de problemas de salud mental infantil. Esto no solo facilita la detección de trastornos en etapas tempranas, sino que también garantiza que las intervenciones sean apropiadas y efectivas.

El desarrollo del modelo de aprendizaje automático para la detección de señales de alerta en problemas asociados a la salud mental en la población infantil escolar de Manizales puede tener un impacto significativo en cuanto a la **sensibilización y políticas públicas**, los datos generados pueden servir como evidencia para sensibilizar a la comunidad educativa y las autoridades. Puede inspirar políticas públicas que prioricen la salud mental en programas escolares; **personalización del apoyo**, con un enfoque basado en datos, las intervenciones pueden adaptarse a las necesidades específicas de cada niño, mejorando así su efectividad; **mejorar la calidad educativa**, al abordar aspectos de salud mental, se favorece un entorno de aprendizaje más saludable y productivo, incrementando la participación y el desempeño académico; **optimización de recursos**, facilita a las instituciones educativas y de salud priorizar recursos y esfuerzos hacia los estudiantes con mayor riesgo. Reduce los costos asociados a diagnósticos tardíos o tratamientos más complejos; **intervención temprana**, el modelo permitirá identificar a los estudiantes que podrían estar en riesgo de desarrollar problemas de salud mental, lo que posibilita implementar intervenciones oportunas y preventivas. Esto reduciría el progreso de estos problemas, promoviendo el bienestar integral de la población infantil.

Líneas futuras

Este modelo no solo tiene el potencial de transformar la manera en que se aborda la salud mental infantil en las escuelas de Manizales, sino que también puede ser pionero en crear un estándar replicable en otras ciudades o países. Las líneas futuras de investigación y desarrollo lo posicionan como una herramienta esencial para construir comunidades escolares más inclusivas, sanas y resilientes. como:

Ampliación de la Base de Datos, integrar más datos demográficos, conductuales y académicos para aumentar la precisión y la aplicabilidad del modelo en otros contextos.

Adaptación a Contextos Sociales Diversos, desde diferentes regiones y niveles socioeconómicos, analizando cómo las características sociales influyen en los patrones detectados. **Enfoque Ético y de Privacidad**, profundizar en protocolos que aseguren un uso ético de los datos y garanticen la privacidad de los estudiantes.

Aplicación multimodal, integrar datos de voz, texto e imágenes para detectar señales emocionales y conductuales complejas. **Desarrollo de Herramientas Digitales**, crear aplicaciones o plataformas digitales que permitan a los docentes y psicólogos escolares

Estudio a Largo Plazo, evaluar el impacto del modelo en la salud mental de los estudiantes y el entorno escolar con un enfoque longitudinal. **Integración con**

Tecnologías Emergentes, explorar la combinación del modelo con tecnologías como Internet de las Cosas (IoT) para monitorear comportamientos en tiempo real, o incluso con realidad aumentada en entornos educativos. **Desarrollo de Herramientas**

Digitales, crear aplicaciones o plataformas digitales que permitan a los docentes y psicólogos escolares interactuar con el modelo y recibir recomendaciones prácticas.

Bibliografía

- Agudelo Hernández, A. F., Guapacha Montoya, M., & Giraldo Álvarez, A. B. (2023). Salud mental pediátrica, experiencia de crianza y desempleo en un contexto latinoamericano: Salud mental pediátrica y desempleo. *Tempus Psicológico*, 6(2). <https://doi.org/10.30554/tempuspsi.6.2.4677.2023>
- Ayala Figueroa, V. H. (2024). Estrategia pedagógica para mejorar las habilidades socioemocionales en los niños de nivel inicial de una institución educativa de Huacho.
- Brand, V. G., & García, L. E. G. (2024). Identificación de patrones a través de algoritmos de Machine Learning en los casos registrados de intentos suicidas en una ciudad de Colombia. *Psicoespacios: Revista virtual de la Institución Universitaria de Envigado*, 18(32), 3.
- Castillo Sánchez, G. A. (2023). Contribución de las Tecnologías de la Información y las Comunicaciones para ayudar a la prevención del Comportamiento Suicida en Castilla y León: Un enfoque de Machine Learning y Salud Digital.
- Condemarín, M., Gorostegui, M. E., Chadwick, M., & Milicic, N. (2016). *Madurez escolar*. Ediciones UC.
- Córdova, H. M. (2008). Factores asociados y trayectorias del desarrollo del comportamiento antisocial durante la adolescencia: implicancias para la prevención de la violencia juvenil en América Latina. *Revista Interamericana de Psicología/Interamerican Journal of Psychology*, 42(1), 129-142.

- De-La-Hoz, E. J., De-La-Hoz, E. J., & Fontalvo, T. J. (2019). Metodología de aprendizaje automático para la clasificación y predicción de usuarios en ambientes virtuales de educación. *Información tecnológica*, 30(1), 247-254.
- Lozano Cely, L. A., & Martin Fúquene, J. E. (2021). Diseño de una solución tecnológica para suministrar apoyo psicológico a la sociedad colombiana.
- Mejía, M. C., Reyes, A. C. D., Aldana, L. A., & Hernández, F. A. (2023). Correlación entre el perfil neuropsicológico y creatividad en una muestra de niños con riesgo de salud mental de la ciudad de Manizales. *Revista Iberoamericana de Psicología*, 16(1), 26-27.
- Mental, O. S. (2022). Fortalecer nuestra respuesta. *Datos y Cifras*.
- Montes, A. J. & Pacheco, P. T. (2024). Integración de IA y salud mental: Implementación de un sistema de asistencia virtual para la detección y atención de problemas psicológicos. [Proyecto aplicado]. Repositorio Institucional UNAD. <https://repository.unad.edu.co/handle/10596/65481>
- Noriega, J. A. V., & Cervantes, J. C. G. (2023). Resiliencia académica de adolescentes, factores asociados y metodología empleada: una revisión sistemática. *Rastros Rostros*, 25(1), 1.
- Palacio A., C. A. (2018). Situación de salud mental en Colombia. *International Journal of Psychological Research*, 11(2), 6–7.
<https://doiorg.bdbiblioteca.universidadean.edu.co/10.21500/20112084.3712>
- Rodríguez-García, A.-M., García-Carmona, M., Marín-Marín, J.-A., & Moreno-Guerrero, A.-J. (2025). Nomofobia y rendimiento académico: prevalencia y correlación en

los jóvenes. Revista Fuentes, 27(1), 107–120.

<https://doi.org/10.12795/revistafuentes.2025.23121>

Saavedra, Y. M. G., Flores, J. W. S., & SAAVEDRA, M. D. P. G. (2024). Análisis de depresión en adolescentes de básica secundaria y media usando algoritmos de Machine Learning: Analysis of Depression in High School and Middle Basic Adolescents Using Machine Learning Algorithms. Psicogente, 27(52).

Sánchez Torres, A. (2023). Desarrollo de un modelo de aprendizaje automático para la mejora de la detección temprana de casos de TDAH en niños y adolescentes.

Universitat Politècnica de València. <http://hdl.handle.net/10251/194403>

Serrano Santoyo Arturo, Martínez Evelio. (2003). La Brecha Digital, Mitos y Realidades.

http://www.labrechadigital.org/labrecha/LaBrechaDigital_MitosyRealidades.pdf