

UNIVERSIDAD NACIONAL AMAZÓNICA DE MADRE DE DIOS

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS E INFORMÁTICA



TESIS

**“Detección del trastorno por déficit de atención e hiperactividad
mediante técnicas de *Machine Learning* en estudiantes de la Institución
Maria Reina, Puerto Maldonado 2025”**

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS
E INFORMÁTICA:

AUTORES:

Bachiller: YABAR YLLA, Fabiola

Bachiller: VARGAS FIGUEROA, Alvaro Jesus

ASESOR:

Dr. HOLGADO APAZA, Luis Alberto

CO-ASESORA:

Dra. ULLOA GALLARDO, Nelly Jacqueline

Puerto Maldonado, julio del 2026

UNIVERSIDAD NACIONAL AMAZÓNICA DE MADRE DE DIOS

FACULTAD DE INGENIERÍA

ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS E INFORMÁTICA



TESIS

**“Detección del trastorno por déficit de atención e hiperactividad
mediante técnicas de *Machine Learning* en estudiantes de la Institución
Maria Reina, Puerto Maldonado 2025”**

PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO DE SISTEMAS
E INFORMÁTICA:

AUTORES:

Bachiller: YABAR YLLA, Fabiola

Bachiller: VARGAS FIGUEROA, Alvaro Jesus

ASESOR:

Dr. HOLGADO APAZA, Luis Alberto

CO-ASESORA:

Dra. ULLOA GALLARDO, Nelly Jacqueline

RST-RI_Detección del trastorno por déficit de atención e hiperactividad mediante técnicas de Machine Learning en estudiantes de la Institución Maria Reina, Puerto Maldonado 2025

INFORME DE ORIGINALIDAD

5%	4%	4%	3%
INDICE DE SIMILITUD	FUENTES DE INTERNET	PUBLICACIONES	TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	repositorio.upch.edu.pe	1%
	Fuente de Internet	
2	repositorio.esan.edu.pe	1%
	Fuente de Internet	
3	rodin.uca.es	1%
	Fuente de Internet	
4	Submitted to Universidad Nacional Abierta y a Distancia, UNAD,UNAD	1%
	Trabajo del estudiante	
5	www.logopedasinrecursos.com	1%
	Fuente de Internet	
6	Holgado Apaza, Luis Alberto. "Modelamiento de la satisfacción laboral de docentes de educación básica mediante técnicas Machine Learning", Universidad Nacional del Altiplano de Puno (Peru)	1%
	Publicación	
7	Palma Ttito, Luis Beltran. "Predicción de cáncer en expresiones genéticas de microarrays mediante un ensamble de modelos heterogéneos de machine learning", Universidad Nacional del Altiplano de Puno (Peru)	1%
	Publicación	

Excluir citas	Activo	Excluir coincidencias	< 1%
Excluir bibliografía	Activo		

DEDICATORIA

Con todo mi corazón, dedico este trabajo, en primer lugar, a mi hija Antonella Morgan Vargas Alagon.

Gracias, mi niña, porque tú fuiste la inspiración para realizarlo. Todo esto lo hice por y para ti. Quiero que guardes estas palabras para siempre: si en algún momento de tu vida sientes que el camino es difícil o llegas a pensar que las cosas van mal, recuerda que tu papá hizo este trabajo pensando en ti, solo para demostrarte lo importante que eres en mi vida y lo mucho que te amo. Eres mi mayor fuerza y la que es, y será siempre, mi más grande creación eres tú.

A mis padres y familia, por su apoyo incondicional durante este camino.

Alvaro Jesus Vargas Figueroa

Dedico este trabajo, en primer lugar, a mi hijo Gael Alessandro, mi mayor fuerza, inspiración y motivo para alcanzar cada una de mis metas. Cada esfuerzo y logro alcanzado llevan el inmenso amor que siento por él.

A mis padres, por ser mis pilares fundamentales, por su esfuerzo, apoyo incondicional y confianza constante a lo largo de mi formación personal y profesional. Gracias por enseñarme el valor de la perseverancia y acompañarme en cada etapa de este camino.

Y a mi familia, por brindarme fortaleza y palabras de aliento durante cada etapa de mi vida.

Yabar Ylla, Fabiola

AGRADECIMIENTO

Agradecemos, en primer lugar, a Dios por brindarnos fortaleza, sabiduría y perseverancia para culminar esta importante etapa de nuestra formación profesional.

A nuestros padres y familias, por su amor, apoyo incondicional y confianza permanente, quienes fueron el principal impulso para seguir adelante y no rendirnos ante las dificultades presentadas durante este proceso académico.

A nuestros hijos, quienes representan nuestra mayor motivación e inspiración para superarnos cada día y alcanzar nuevas metas personales y profesionales.

Expresamos también nuestro profundo agradecimiento a nuestro asesor y co-asesora de tesis, por su orientación, dedicación y valiosos conocimientos brindados durante el desarrollo de esta investigación.

Asimismo, agradecemos a la Institución Educativa Privada María Reina de Puerto Maldonado y a todas las personas que contribuyeron de manera directa e indirecta en la realización de este trabajo de investigación.

Finalmente, agradecemos a todas aquellas personas que, con sus palabras de aliento, apoyo y comprensión, formaron parte de este importante logro académico.

Yabar Ylla, Fabiola
Vargas Figueroa, Alvaro Jesus

PRESENTACIÓN

El presente trabajo de investigación, titulado “Detección del trastorno por déficit de atención e hiperactividad mediante machine learning en estudiantes de la institución Maria Reina, Puerto Maldonado 2025”, se centró en el uso innovador de la inteligencia artificial para abordar una de las preocupaciones más significativas en el ámbito educativo y de la salud mental infantil: la detección temprana del Trastorno por Déficit de Atención e Hiperactividad (TDAH).

Este estudio tiene como objetivo principal evaluar cómo la implementación de técnicas avanzadas de Machine Learning mejoró la precisión y eficiencia en la identificación de estudiantes que presentan características compatibles con el TDAH, comparativamente a los métodos tradicionales. La relevancia de este trabajo radicó en su potencial para transformar las estrategias de diagnóstico actuales, ofreciendo una herramienta valiosa para profesionales de la educación y de la salud, y así, contribuir significativamente al bienestar y al desarrollo óptimo de los estudiantes afectados.

La metodología de la investigación fue de tipo aplicado con un enfoque cuantitativo, caracterizándose por un diseño no experimental y transversal. Se analizó la relación entre los datos obtenidos a través de observaciones conductuales detalladas y los resultados generados por el modelo de Machine Learning, con el fin de determinar la viabilidad y eficacia de esta tecnología en el contexto educativo específico de la Institución Educativa Privada Maria Reina en Puerto Maldonado.

Para llevar a cabo este estudio, se aplicaron cuestionarios y se recogieron datos conductuales de una muestra representativa de estudiantes, los cuales fueron posteriormente analizados mediante algoritmos de aprendizaje automático. Este enfoque permitió no solo la identificación de patrones asociados al TDAH sino también la personalización de intervenciones educativas y terapéuticas.

RESUMEN

La presente investigación tuvo como objetivo analizar la aplicación de técnicas de Machine Learning para la detección del Trastorno por Déficit de Atención e Hiperactividad (TDAH) en estudiantes de la Institución Educativa Privada María Reina de Puerto Maldonado durante el año 2025. El estudio se desarrolló ante la necesidad de fortalecer los procesos de identificación temprana del TDAH mediante herramientas tecnológicas que contribuyan al análisis eficiente de patrones conductuales en el contexto educativo.

La investigación fue de tipo aplicada, con enfoque cuantitativo y diseño no experimental de corte transversal. Para la recolección de datos se utilizaron cuestionarios y registros de observación dirigidos a estudiantes y docentes, permitiendo obtener información relacionada con conductas asociadas al TDAH. Posteriormente, los datos recopilados fueron procesados mediante técnicas de aprendizaje automático con la finalidad de identificar patrones conductuales y evaluar la capacidad de clasificación del modelo implementado.

Los resultados evidenciaron que la aplicación de técnicas de Machine Learning permitió optimizar el análisis de comportamientos asociados al TDAH, facilitando la identificación de estudiantes con posibles indicadores del trastorno. Asimismo, la integración de información proporcionada por docentes fortaleció la calidad y consistencia de los datos utilizados durante el proceso de análisis, mejorando la interpretación de los patrones conductuales observados.

Se concluyó que el uso de herramientas basadas en inteligencia artificial constituye una alternativa de apoyo para la detección temprana del TDAH en contextos educativos, contribuyendo al fortalecimiento de los procesos de evaluación y favoreciendo intervenciones oportunas orientadas al bienestar y desarrollo integral de los estudiantes.

Palabras clave: Machine Learning, TDAH, inteligencia artificial, detección temprana, aprendizaje automático, educación.

ABSTRACT

This research aimed to analyze the application of Machine Learning techniques for the detection of Attention Deficit Hyperactivity Disorder (ADHD) in students of the María Reina Private Educational Institution in Puerto Maldonado during 2025. The study was developed in response to the need to strengthen early ADHD identification processes through technological tools that contribute to the efficient analysis of behavioral patterns in the educational context.

The research was applied, with a quantitative approach and a non-experimental, cross-sectional design. Data collection involved questionnaires and observation records administered to students and teachers, allowing for the gathering of information related to behaviors associated with ADHD. Subsequently, the collected data were processed using machine learning techniques to identify behavioral patterns and evaluate the classification capacity of the implemented model.

The results showed that the application of Machine Learning techniques optimized the analysis of behaviors associated with ADHD, facilitating the identification of students with potential indicators of the disorder. Furthermore, the integration of information provided by teachers strengthened the quality and consistency of the data used during the analysis process, improving the interpretation of the observed behavioral patterns.

It was concluded that the use of artificial intelligence-based tools constitutes a support alternative for the early detection of ADHD in educational contexts, contributing to the strengthening of assessment processes and facilitating timely interventions aimed at the well-being and holistic development of students.

Keywords: Machine Learning, ADHD, artificial intelligence, early detection, machine learning, education.

INTRODUCCIÓN

El Trastorno por Déficit de Atención e Hiperactividad (TDAH) se configura como una condición crónica manifestada a través de la inatención, la impulsividad y la hiperactividad en la población infanto-juvenil. Rusca y Cortez (2020) advierten que, contrario a la percepción tradicional de que la sintomatología remite con la edad, el trastorno suele persistir durante la adolescencia. Esta continuidad genera impactos negativos en las dimensiones académica, social y emocional, lo cual exige estrategias de evaluación y tratamiento adaptadas a las necesidades específicas de esta etapa de desarrollo.

Frente a este panorama, la investigación se centró en la identificación temprana del TDAH en escolares de primero a sexto grado de primaria, un periodo crítico para la intervención. Diversas investigaciones previas han demostrado que las técnicas de Machine Learning constituyen una alternativa eficaz para este propósito, Sánchez (2023) reportó una exactitud del 90% empleando Random Forest, y Chen et al. (2021) lograron un 85% con algoritmos de clasificación supervisada en contextos clínicos. En este marco, el estudio propuso el uso de Aprendizaje Automático (Machine Learning), empleando algoritmos de alta precisión como Random Forest, Máquinas de Vector de Soporte (SVM) y K-Nearest Neighbors (KNN), dichos modelos se entrenaron con datos obtenidos mediante la Escala de Conners (Conners' Rating Scale), una herramienta de validación clínica reconocida. La finalidad fue integrar este soporte tecnológico al diagnóstico tradicional, permitiendo una atención temprana y personalizada que elevara la calidad de vida del alumnado.

El documento se organiza en cuatro capítulos; el Capítulo I aborda el planteamiento del problema, definiendo sus objetivos, variables y la justificación correspondiente; el Capítulo II establece las bases teóricas y los antecedentes investigativos que fundamentan la propuesta; el Capítulo III detalla la estrategia metodológica, especificando el diseño de investigación, la población de estudio y los procedimientos de recolección y análisis de datos; el Capítulo IV expone el desarrollo del modelo predictivo y los resultados obtenidos; finalmente, se presentan las conclusiones y recomendaciones derivadas del estudio.

ÍNDICE

DEDICATORIA	
AGRADECIMIENTO.....	
PRESENTACIÓN	
RESUMEN.....	
ABSTRACT	
INTRODUCCIÓN	
ÍNDICE DE TABLAS	1
ÍNDICE DE FIGURAS.....	1
CAPÍTULO I: PROBLEMA DE INVESTIGACIÓN	1
1.1. Descripción del problema.....	1
1.2. Formulación del problema.....	3
1.3. Objetivos de la investigación.....	4
1.4. Variables	4
1.5. Operacionalización de variables	5
1.6. Hipótesis	6
1.7. Justificación	6
1.8. Consideraciones Éticas.....	8
CAPÍTULO II: MARCO TEÓRICO.....	10
2.1. Antecedentes de estudio.....	10
2.2. Marco teórico	12
2.3. Definición de términos	16
CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN	19
3.1. Tipo de estudio.....	19
3.2. Diseño de estudio	20
3.3. Población y muestra	20
3.4. Métodos y técnicas	22
3.5. Tratamiento de los datos.....	23
CAPÍTULO IV: RESULTADOS DEL TRABAJO DE INVESTIGACION.....	25
4.1. Entendimiento del negocio	25
4.2. Comprensión de los datos.....	26
4.3. Preparación de los datos.....	29
4.4. Modelamiento.....	32

4.5. Despliegue (Análisis, diseño, desarrollo y pruebas).....	36
4.6. Análisis inferencial	45
CONCLUSIONES	54
SUGERENCIAS	55
REFERENCIAS BIBLIOGRÁFICAS	56
Anexo 1. Matriz de operacionalización de variables	1
Anexo 2. Matriz de consistencia.....	1
Anexo 3. Instrumento de recolección de datos.....	1
Anexo 4. Solicitud de autorización para realización de estudio.....	1
Anexo 5. Solicitud de validación de instrumento.....	1
Anexo 6. Ficha de validación.....	1
Anexo 7. Consentimiento informado.....	1
Anexo 8. Reporte de recolección de datos en base a cuestionario del Test de Conners.....	1
Anexo 9. Código fuente del sistema de detección de TDAH.....	1
Anexo 10. Panel fotográfico.....	1

ÍNDICE DE TABLAS

Tabla 1. Operacionalización de variables.	5
Tabla 2. Descripción de datos del conjunto de variables.	27
Tabla 3. Selección de datos según variables.	30
Tabla 4. Ajuste de hiperparámetros mediante Grid Search para los algoritmos de clasificación.	33
Tabla 5. Requisitos funcionales.	37
Tabla 6. Requisitos No Funcionales.	38
Tabla 7. Diccionario de datos de la entidad ESTUDIANTE.	40
Tabla 8. Diccionario de datos de la entidad EVALUACION.	40
Tabla 9. Diccionario de datos de la entidad RESULTADO_IA.	42
Tabla 10. Casos de prueba del sistema de detección de TDAH.	45
Tabla 11. Prueba de Kolmogorov-Smirnov y Shapiro-Wilk para la variable Accuracy.	47
Tabla 12. Resultados de la prueba t de Student para una muestra aplicada a la exactitud del modelo K-Nearest Neighbors.	48
Tabla 13. Resultados de la prueba de rangos con signo de Wilcoxon para una muestra aplicada a la exactitud del modelo Random Forest.	50
Tabla 14. Resultados de la prueba de rangos con signo de Wilcoxon para una muestra aplicada a la exactitud del modelo Support Vector Machine (SVM).	52

ÍNDICE DE FIGURAS

Figura 1. Registro y tabulación de datos del cuestionario de la Escala de Conners en formato Excel.	26
Figura 2. Proyección PCA del espacio de decisión KNN.	28
Figura 3. Muestra de conjunto de datos faltantes.	29
Figura 4. Datos sanitizados luego del proceso de limpieza.	31
Figura 5. Impacto del balanceo de datos.	32
Figura 6. Comparación de desempeño de los modelos según métricas.....	34
Figura 7. Validación cruzada de modelo KNN.....	36
Figura 8. Diagrama de casos de uso del sistema.	37
Figura 9. Diagrama de clases del sistema de detección de TDAH.	38
Figura 10. Diagrama entidad relación del sistema de detección de TDAH.	39
Figura 11. Interfaz principal del sistema experto para la evaluación de síntomas de TDAH.....	44
Figura 12. Visualización del resultado del análisis predictivo generado por el sistema experto para la detección de TDAH.....	44

CAPÍTULO I: PROBLEMA DE INVESTIGACIÓN

1.1. Descripción del problema

Desde una perspectiva neurobiológica, el Trastorno por Déficit de Atención e Hiperactividad (TDAH) se define como una condición heterogénea y de origen multifactorial (Weiss et al., 2023). Su etiología combina una significativa predisposición genética, con estimaciones de heredabilidad que oscilan entre 50 % y un 78 %, junto con factores de riesgo gestacionales y perinatales.

Entre estos últimos, Han et al. (2015) señalan el bajo peso al nacer, la prematuridad y la exposición fetal a sustancias tóxicas como tabaco o alcohol, así como a determinados fármacos incluidos benzodiacepinas y anticonvulsivantes. Como consecuencia de esta interacción de factores biológicos y ambientales, el trastorno suele manifestarse en la infancia temprana y puede prolongarse hasta la adultez, expresándose mediante patrones de inatención e hiperactividad e impulsividad que afectan el desempeño social, académico y familiar del individuo (Parke et al., 2021).

Según reportes de la Organización Mundial de la Salud (2023), el TDAH se posiciona como uno de los trastornos del neurodesarrollo más prevalentes a nivel global, afectando entre el 5% y el 8% de la población infantil. Este cuadro clínico compromete el desempeño académico y la estabilidad socioemocional del menor, con tendencia a mantener sintomatología persistente en la adolescencia e incluso hasta la adultez si no se implementan intervenciones oportunas.

En el contexto latinoamericano, particularmente en México, la insuficiente capacitación docente respecto al TDAH constituye una barrera significativa para su detección y manejo oportuno (Ramos et al., 2022). Esta carencia formativa propicia la estigmatización del estudiante, quien frecuentemente es etiquetado de manera peyorativa como “problemático”, dificultando su

integración social y educativa (Luna et al., 2023). Mesino (2023) advierte que dicha segregación, sumada a la ausencia de estrategias pedagógicas inclusivas, genera bajos niveles de rendimiento y eleva el riesgo de deserción escolar. En conjunto, el desconocimiento técnico del profesorado limita la implementación de intervenciones efectivas en el aula (Ramos et al., 2022).

En el contexto peruano, el Ministerio de Salud (Ministerio de Salud, 2020) estima que entre un 5 % y un 10 % de la población infantil presenta Trastorno por Déficit de Atención e Hiperactividad. Esta prevalencia se traduce en un impacto significativo en la funcionalidad y el bienestar de los menores, aspecto respaldado por Ricketts et al. (2020), quienes señalan que esta condición afecta la regulación emocional y las competencias sociales.

A este cuadro clínico se suman barreras estructurales. Retamozo (2023) advierte que las limitaciones socioeconómicas del país evidenciadas en la escasez de profesionales capacitados y materiales pedagógicos idóneos, sobre todo en zonas rurales, obstaculizan el diagnóstico y la intervención oportuna, situando a los escolares en condición de vulnerabilidad persistente.

El contexto espacial de la presente investigación se sitúa en la Institución Educativa Privada María Reina, ubicada en Puerto Maldonado. En dicha institución se identifican limitaciones en los procesos de detección y manejo del TDAH, asociadas a la insuficiencia de recursos especializados, la escasa capacitación docente específica y la limitada disponibilidad de profesionales de la salud mental en la región.

Estas condiciones podrían incrementar el riesgo de subdetección del trastorno en el entorno escolar, reduciendo la probabilidad de acceso oportuno a evaluación clínica y acompañamiento psicopedagógico. Como consecuencia, los estudiantes con sintomatología compatible pueden experimentar dificultades persistentes en su desempeño académico y en su adaptación socioemocional.

Frente a este escenario, diversas investigaciones han demostrado que las técnicas de Machine Learning constituyen una alternativa eficaz para la detección temprana del TDAH. Sánchez (2023) reportó que un modelo basado

en Random Forest alcanzó una exactitud del 90% en la identificación de casos en población escolar, mientras que Chen et al. (2021) lograron un 85% de exactitud empleando algoritmos de clasificación supervisada en contextos clínicos. Estos resultados evidencian el potencial de estas herramientas para complementar los procesos diagnósticos tradicionales, especialmente en entornos con recursos especializados limitados.

En respuesta a esta problemática, el estudio propuso desarrollar un modelo predictivo para la detección temprana de posibles casos de TDAH, basado en técnicas de Machine Learning. El objetivo fue mejorar la precisión y eficiencia en la identificación preliminar mediante el análisis de patrones característicos asociados al trastorno, utilizando datos obtenidos a partir de evaluaciones psicopedagógicas y conductuales.

La implementación de este sistema buscó constituirse como una herramienta de apoyo al proceso diagnóstico tradicional, facilitando la derivación oportuna a evaluación especializada. Asimismo, se proyectó que esta propuesta tecnológica pudiera adaptarse a otros contextos educativos tanto a nivel local como nacional.

1.2. Formulación del problema

1.2.1. Problema General

¿Cuál es el desempeño predictivo de los modelos de Machine Learning en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada Maria Reina, Puerto Maldonado 2025?

1.2.2. Problemas Específicos

- ¿Cuál es el desempeño predictivo del modelo K-Nearest Neighbors (KNN) en la detección de casos del trastorno por déficit de atención e hiperactividad en estudiantes de la Institución Educativa Privada Maria Reina, Puerto Maldonado 2025?
- ¿Cuál es el desempeño predictivo del modelo Random Forest en la detección de casos del trastorno por déficit de atención e hiperactividad en

estudiantes de la Institución Educativa Privada Maria Reina, Puerto Maldonado 2025?

- ¿Cuál es el desempeño predictivo del modelo Support Vector Machine (SVM) en la detección de casos del trastorno por déficit de atención e hiperactividad en estudiantes de la Institución Educativa Privada Maria Reina, Puerto Maldonado 2025?

1.3. Objetivos de la investigación

1.3.1. Objetivo General

Evaluar el desempeño predictivo de técnicas de Machine Learning en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado 2025.

1.3.2. Objetivos Específicos

- Evaluar el desempeño del modelo K-Nearest Neighbors (KNN) en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada Maria Reina, Puerto Maldonado 2025.
- Evaluar el desempeño del modelo Random Forest en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada Maria Reina, Puerto Maldonado 2025.
- Evaluar el desempeño del modelo Support Vector Machine (SVM) en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada Maria Reina, Puerto Maldonado 2025.

1.4. Variables

1.4.1. Variable Independiente

Técnicas de Machine Learning.

1.4.2. Variable Dependiente

Detección del trastorno por déficit de atención e hiperactividad

1.5. Operacionalización de variables

Tabla 1. Operacionalización de variables.

Variables	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores
Variable Independiente: Técnicas de Machine Learning	Las técnicas de Machine Learning son métodos algorítmicos que permiten a un sistema aprender patrones a partir de datos, mejorando su desempeño en una tarea específica sin necesidad de programación explícita (Mitchell, 1997).	Esta variable se evaluó mediante la implementación de tres algoritmos de aprendizaje automático: K-Nearest Neighbors (KNN), Random Forest y Support Vector Machine (SVM). Cada modelo fue entrenado utilizando los datos obtenidos del cuestionario aplicado a los estudiantes y posteriormente se comparó su desempeño en la clasificación de posibles casos de TDAH. El rendimiento de los modelos fue medido mediante métricas de evaluación utilizadas en problemas de clasificación.	K-Nearest Neighbors (KNN)	Accuracy, Precisión, Recall, F1-Score, ROC-AUC
			Random Forest	Accuracy, Precisión, Recall, F1-Score, ROC-AUC
			Support Vector Machine (SVM)	Accuracy, Precisión, Recall, F1-Score, ROC-AUC
Variable Dependiente: Detección del Trastorno Por Déficit de Atención e Hiperactividad	El TDAH, Trastorno por Déficit de Atención/Hiperactividad, se define en el DSM-5, como un trastorno del neurodesarrollo caracterizado por un patrón persistente de déficit de atención y/o hiperactividad que es desadaptativo para el paciente, está presente antes de los 12 años e interfiere en sus actividades diarias. En el DSM-5 se informa de una prevalencia del 5% en niños/as en la mayoría de las culturas y de un 2.5% en adultos (Aguirre et al., 2022)	Esta variable fue medida mediante la aplicación de la Escala de Conners dirigida a padres o tutores, instrumento utilizado para identificar comportamientos asociados al TDAH en niños. Las respuestas obtenidas permitieron determinar la presencia de indicadores conductuales relacionados con el trastorno y fueron empleadas como datos de entrada para el modelo de clasificación.	Patrones conductuales asociados al TDAH	Nivel de atención, Hiperactividad, Impulsividad.

Fuente: Elaboración propia.

1.6. Hipótesis

1.6.1. Hipótesis General

Los modelos de Machine Learning presentan un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado, 2025.

1.6.2. Hipótesis Específicas

- El modelo K-Nearest Neighbors (KNN) presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado, 2025.
- El modelo Random Forest presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado, 2025.
- El modelo Support Vector Machine (SVM) presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado, 2025.

El umbral del 75% de exactitud fue establecido como criterio mínimo de desempeño aceptable en concordancia con investigaciones previas en el campo. Chen et al. (2021) y Sánchez (2023) reportaron niveles de exactitud superiores al 75% en modelos de clasificación aplicados a la detección del TDAH, validando dicho valor como referencia para evaluar la viabilidad clínica de un modelo predictivo en contextos educativos

1.7. Justificación

La identificación del TDAH en la población estudiantil de la I.E.P. María Reina, ubicada en Puerto Maldonado, se ve seriamente restringida por la brecha de capacitación docente y la carencia de logística especializada; como resultado, el alto índice de subdiagnóstico repercute de manera perjudicial en el éxito académico y la estabilidad emocional del alumnado.

Para revertir esta situación, se postula el desarrollo de un sistema predictivo fundamentado en Machine Learning. La integración de esta tecnología pretende reducir la subjetividad inherente a los métodos tradicionales, favoreciendo procesos de detección más precisos que faciliten la derivación oportuna a evaluación especializada.

Adicionalmente, la arquitectura replicable de este modelo no solo promueve la innovación pedagógica, sino que también aporta una base metodológica para el desarrollo de herramientas tecnológicas orientadas a la detección temprana de posibles alteraciones del neurodesarrollo en contextos educativos.

1.7.1. Justificación Teórica

El marco conceptual de esta investigación articula los postulados de las ciencias cognitivas con los recientes avances computacionales; es innegable que la neuropsicología ha detallado las severas implicancias socioeducativas del TDAH; no obstante, los abordajes clínicos habituales continúan condicionados por la limitación de recursos y los sesgos interpretativos.

Para superar estas deficiencias procedimentales, los algoritmos de aprendizaje automático se posicionan como una herramienta de alta fiabilidad, diseñados para rastrear y abstraer patrones conductuales intrincados a partir de grandes volúmenes de información; de este modo, se promueve un trabajo colaborativo entre la ingeniería de datos y la psicología educativa, buscando instaurar paradigmas de detección temprana regidos por la objetividad algorítmica.

1.7.2. Justificación Práctica

Desde una perspectiva práctica, la investigación planteó la implementación operativa de un modelo de Machine Learning para la detección del TDAH en la I.E.P. Maria Reina de Puerto Maldonado; la capacidad de identificar patrones asociados al trastorno con precocidad y exactitud viabiliza el diseño de planes de intervención a medida, dotando a la plana docente y a los especialistas de una comprensión más profunda de los requerimientos estudiantiles.

Asimismo, al procesar información psicopedagógica y conductual, el sistema se constituyó como un instrumento de soporte funcional que agilizó los procesos de detección preliminar y contribuye a mejorar la toma de decisiones pedagógicas y de derivación a evaluación especializada.

1.8. Consideraciones Éticas

El desarrollo del estudio se ajustó a los estándares éticos universales para la investigación científica con personas, en concordancia con los protocolos y regulaciones aplicables. Las medidas adoptadas para garantizar la integridad del proceso incluyeron:

- **Consentimiento informado:**

Como requisito indispensable para la obtención de datos, se gestionó el consentimiento informado de los padres o responsables legales de los estudiantes participantes. Este proceso aseguró que los padres conozcan los fines del estudio, la aplicación de algoritmos de aprendizaje automático y los protocolos de privacidad, validando su derecho inalienable a retirar su autorización sin penalización alguna.

- **Confidencialidad y privacidad:**

Para salvaguardar la privacidad, se implementaron protocolos de seguridad que contemplaron la codificación de los datos y su almacenamiento restringido. La información recolectada fue tratada con fines exclusivamente académicos y de investigación, asegurando que en ninguna circunstancia se comprometieran detalles que revelen la identidad de los participantes.

- **Minimización del riesgo:**

La recolección de la información se realizó mediante la aplicación del Conners' Rating Scale a padres, tutores y docentes, evitando la participación directa de los menores. El cuestionario tiene una duración aproximada de 20 minutos y fue aplicado en condiciones adecuadas que favorezcan la comprensión y el bienestar de los participantes. Asimismo, se informó claramente que los resultados obtenidos no constituyen un diagnóstico clínico, sino una herramienta de apoyo para la detección preliminar de posibles indicadores asociados al TDAH. En caso de que

algún padre o tutor manifieste inquietudes respecto a los resultados, se recomendará la consulta con un profesional especializado.

- **Autenticidad y citación:**

Todas las fuentes utilizadas en el desarrollo de la investigación fueron citadas adecuadamente siguiendo la norma ISO 690, garantizando la transparencia académica y la originalidad del trabajo. Además, el estudio fue supervisado por el asesor principal y especialistas en la materia para validar los procesos metodológicos y los análisis.

CAPÍTULO II: MARCO TEÓRICO

2.1. Antecedentes de estudio

2.1.1. Antecedentes Internacionales

Chen et al. (2021) en su estudio, se propusieron implementar un modelo predictivo que evalúe las derivaciones según el riesgo del paciente de padecer TDAH y que pudiera implementarse en un entorno clínico. Los algoritmos empleados incluyeron máquinas de vector de soporte, regresión logística, árbol de decisiones, K-vecinos más cercanos, bosque aleatorio y Naive Bayes. El conjunto de datos utilizado fue obtenido de la Fundación del Servicio Nacional de Salud de la Asociación del Suroeste de Yorkshire – SWYPFT, que contenía los datos anonimizados de las evaluaciones realizadas a pacientes con TDAH entre 2014 y 2017, con un total de 69 pacientes. Los resultados demostraron que el algoritmo de árbol de decisiones obtuvo un 85% de exactitud. Los autores concluyen que es posible diagnosticar correctamente a pacientes con TDAH con una exactitud estadística prometedora.

Maniruzzaman et al. (2023), en su investigación buscaron desarrollar un sistema preciso para identificar TDAH mediante señales de EEG. Las técnicas utilizadas abarcaron máquinas de vector de soporte, Random Forest, redes neuronales, K-vecinos más cercanos y Naive Bayes. Los datos analizados provinieron del Centro de Investigación Neurológica de Dhaka, con registros anonimizados de señales EEG de pacientes examinados entre 2020 y 2022, abarcando 72 casos. Los hallazgos revelaron que Random Forest alcanzó un AUROC de 85.6%. Los investigadores concluyen que el análisis de EEG con Machine Learning ofrece un avance notable para diagnosticar TDAH con precisión.

Santarrosa-López et al. (2023), en su trabajo pretendieron determinar las técnicas más eficaces para clasificar el TDAH en contextos clínicos. Los métodos aplicados comprendieron máquinas de vector de soporte, regresión lineal, K-vecinos más cercanos, árbol de decisiones, CatBoost y redes neuronales convolucionales. La información empleada fue recolectada del Centro de Investigación Psiquiátrica de Bogotá, incluyendo datos anonimizados de evaluaciones clínicas de pacientes con TDAH entre 2021 y 2023, con 90 casos. El análisis mostró que CatBoost logró una exactitud del 92%. Los autores concluyen que los modelos de Machine Learning presentaron un desempeño superior en la clasificación del TDAH dentro del conjunto de datos analizados.

Sánchez (2023), en su estudio intentaron diseñar un modelo para agilizar la identificación temprana de TDAH en escolares. Los modelos implementados englobaron Random Forest, máquinas de vector de soporte, XGBoost, K-vecinos más cercanos y regresión logística. Los datos recopilados procedieron de la Asociación Nacional de TDAH de España, con cuestionarios anonimizados sobre síntomas de TDAH completados por niños y adolescentes entre 2020 y 2022, totalizando 80 casos. La evaluación indicó que Random Forest logró una exactitud del 90%. Los autores sostienen que el Machine Learning facilita una detección temprana eficaz del TDAH, mejorando el apoyo educativo.

Das y Khanna (2021), ambicionaron crear un método novedoso para detectar TDAH utilizando biomarcadores pupilométricos. Los sistemas adoptados integraron máquinas de vector de soporte, Random Forest, K-vecinos más cercanos, Naive Bayes y redes neuronales. Los datos extraídos procedieron del Instituto de Neurociencias de Bangalore, con registros anonimizados de mediciones pupilométricas de pacientes evaluados entre 2018 y 2020, comprendiendo 50 casos. El estudio evidenció que las máquinas de vector de soporte alcanzaron un AUROC de 85.6%. Los autores destacan que el Machine Learning con biomarcadores pupilométricos promete mejorar la precisión diagnóstica del TDAH.

2.1.2. Antecedentes Nacionales

Casas y Celis (2023), en su estudio desarrollaron una aplicación móvil que utilizó la gamificación y el método TCT para reforzar el aprendizaje de niños con TDAH. El estudio concluyó que la aplicación móvil mejoró significativamente la memoria de trabajo de los niños con TDAH, lo que sugiere que las herramientas tecnológicas pueden complementar los enfoques tradicionales en el manejo de este trastorno.

Dueñas y Mendoza (2023), en su investigación exploraron el uso de juegos con realidad aumentada (RA) como herramienta terapéutica en niños con TDAH. Los resultados mostraron que el uso de RA y juegos serios ayudó en la mejora de las capacidades cognitivas y el enfoque, lo que demuestra el potencial de las tecnologías emergentes para el tratamiento del TDAH en contextos educativos y clínicos.

2.2. Marco teórico

2.2.1. Técnicas de Machine Learning

El Aprendizaje Automático o Machine Learning se define como una rama de la inteligencia artificial que faculta a los sistemas para identificar patrones subyacentes en los datos, prescindiendo de una programación explícita para cada tarea (Russell et al., 2021); bajo este paradigma, la presente investigación adoptó algoritmos de clasificación supervisada, tales como Random Forest, Máquinas de Vector de Soporte (SVM) y K-Nearest Neighbors (KNN) con el propósito de procesar indicadores conductuales y facilitar la detección del TDAH en la población estudiantil.

Para la implementación de los algoritmos de clasificación supervisada, el conjunto de datos de entrenamiento se representó mediante una matriz $X \in \mathbb{R}^{n \times p}$, donde n corresponde al número de observaciones (estudiantes evaluados) y p al número de características o variables conductuales. Cada fila $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ representa el vector de atributos de un estudiante, y el vector objetivo $y \in \{0, 1\}^n$ indica la etiqueta de clase correspondiente (0 = sin indicadores de TDAH, 1 = posible TDAH). El objetivo del modelo de clasificación supervisada consiste en aprender una función $f: \mathbb{R}^p \rightarrow \{0, 1\}$ que

minimice el error de clasificación sobre datos no vistos, estimada mediante validación cruzada estratificada (Hastie, Tibshirani y Friedman, 2009).

El proceso de aprendizaje supervisado aplicado en la presente investigación comprende tres etapas diferenciadas. En la etapa de entrenamiento, el modelo recibe la submatriz $X_{\text{train}} \in \mathbb{R}^{n \times p}$ junto con el vector y_{train} y ajusta sus parámetros internos minimizando el error empírico sobre los datos disponibles. En la etapa de validación, se evaluó la capacidad de generalización mediante validación cruzada estratificada (*Stratified K-Fold*, $k = 5$), que divide el conjunto balanceado en cinco particiones manteniendo la proporción de clases en cada subconjunto, reduciendo así el sesgo en la estimación del desempeño. En la etapa de evaluación, el modelo entrenado generó predicciones $\hat{y} = f(X_{\text{test}})$ sobre datos no vistos, comparándolas con las etiquetas reales y_{test} mediante las métricas de clasificación: *accuracy*, precisión, *recall*, *F1-score* y ROC-AUC (Hastie, Tibshirani y Friedman, 2009).

El error de generalización se estimó mediante la siguiente expresión:

$$E_{\text{gen}} = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} 1[\hat{y}_i \neq y_i]$$

donde $1[\cdot]$ es la función indicadora que toma el valor 1 cuando la predicción difiere de la etiqueta real y 0 en caso contrario. La optimización algebraica específica de cada algoritmo se detalla en las subsecciones siguientes (Russell et al., 2021).

- K-Nearest Neighbors:

El algoritmo K-Nearest Neighbors (KNN) es un método de clasificación supervisada basado en la similitud entre datos. Su principio consiste en asignar una categoría a un nuevo caso según la clase predominante entre sus vecinos más cercanos dentro del conjunto de entrenamiento (Cover y Hart, 1967). La similitud entre instancias se calcula mediante la distancia euclidiana, definida como:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Donde x e y son dos instancias del conjunto de datos y p es el número de variables o características. La clase asignada al nuevo caso corresponde a la clase mayoritaria entre los k vecinos más cercanos, determinada por votación simple. Este algoritmo es ampliamente utilizado en problemas de clasificación debido a su simplicidad y eficacia en el reconocimiento de patrones.

- Random Forest:

El algoritmo Random Forest se define como un método de ensamblaje que integra múltiples árboles de decisión con el fin de optimizar la precisión predictiva y mitigar el sobreajuste (Breiman et al., 2017). Cada árbol se construye a partir de una muestra aleatoria del conjunto de entrenamiento mediante la técnica de Bagging, y la partición de cada nodo se determina por el criterio de impureza de Gini, definido como:

$$Gini(t) = 1 - \sum_{i=1}^c p_i^2$$

Donde c es el número de clases y p_i es la proporción de instancias de la clase i en el nodo t . La predicción final del modelo se obtiene por votación mayoritaria entre todos los árboles del ensamblaje. Dentro de la clasificación supervisada, destaca por su robustez ante datos ruidosos. Maniruzzaman et al. (2023) validan su idoneidad para el análisis de conjuntos de datos con variables heterogéneas.

- Support Vector Machine:

Support Vector Machine (SVM) es un algoritmo de aprendizaje supervisado utilizado para tareas de clasificación y regresión. El funcionamiento del algoritmo consiste en encontrar un hiperplano óptimo que separe las clases en un espacio multidimensional, maximizando el margen entre los puntos más cercanos de cada clase, denominados vectores de soporte (Cortes y Vapnik, 1995). La función de decisión se define como:

$$f(x) = w^T x + b$$

Donde w es el vector de pesos, x es el vector de características y b es el sesgo. La optimización busca maximizar el margen $\frac{2}{||w||}$, sujeto a la condición $y_i(w^T x_i + b) \geq 1$ para cada instancia de entrenamiento. Para datos no linealmente separables se emplea el kernel RBF, definido como:

$$K(x_i, x_j) = e^{-\gamma ||x_i - x_j||^2}$$

Donde γ es un parámetro que controla el radio de influencia de cada instancia. El algoritmo destaca por su alta capacidad de generalización y su eficiencia en conjuntos de datos de alta dimensionalidad.

2.2.2. Detección del Trastorno por Déficit de Atención e Hiperactividad

El TDAH es un trastorno del neurodesarrollo caracterizado por dificultades en la atención, impulsividad e hiperactividad inapropiada para la edad (American Psychiatric Association, 2022). La detección se realiza mediante instrumentos validados, como el Conners' Rating Scale, que evalúa los patrones conductuales asociados al TDAH.

- Estado de Atención

Capacidad de los estudiantes para mantener la atención sostenida y controlar comportamientos hiperactivos o impulsivos, según los criterios diagnósticos del TDAH (Barkley, 2014).

- Nivel de Atención:

Frecuencia de distracciones o dificultad para seguir instrucciones, medido por ítems del Conners' Rating Scale (Conners, 2008).

- Hiperactividad:

Movimientos físicos excesivos o inquietud, observados en contextos escolares (Barkley, 2014).

- Impulsividad:

Acciones sin premeditación, como interrupciones frecuentes, evaluadas por docentes y padres (Barkley, 2014).

- Indicadores:

La Escala de Conners, en sus versiones para padres y maestros, es un instrumento ampliamente usado para detectar síntomas de TDAH en niños. Evalúa inatención, hiperactividad e impulsividad, proporcionando puntuaciones que guían el diagnóstico clínico (Grao-Cruces, 2017).

2.2.3. Integración de Machine Learning en la Detección del Trastorno por Déficit de Atención e Hiperactividad

La incorporación de técnicas de Machine Learning en la detección del TDAH constituyó una estrategia eficaz para incrementar la precisión y objetividad diagnóstica, aspecto de vital importancia en entornos con recursos limitados como Puerto Maldonado (Retamozo, 2023). Al procesar información proveniente de la Escala de Conners y registros conductuales, estos algoritmos identificaron patrones predictivos complejos, logrando superar el sesgo inherente a los métodos de evaluación subjetiva (Santarrosa-López et al., 2024). En última instancia, esta integración tecnológica optimizó los procesos de intervención temprana en el ámbito educativo.

2.3. Definición de términos

2.3.1. Machine Learning

Según Russell y Norvig (2021), el Aprendizaje Automático (Machine Learning) constituye una disciplina derivada de la inteligencia artificial que faculta a los sistemas computacionales para aprender a partir de la experiencia de los datos. Su característica distintiva radica en la capacidad de ejecutar predicciones mediante la identificación de patrones complejos, prescindiendo de instrucciones de programación explícitas para cada tarea.

2.3.2. Algoritmo

Según Cormen et al. (2022), un algoritmo se define como una secuencia rigurosa de instrucciones computacionales estructuradas para resolver problemas específicos o ejecutar tareas determinadas. En el ámbito de la

inteligencia artificial, estos procedimientos lógicos constituyen el cimiento de los modelos de aprendizaje automático, permitiendo el procesamiento sistemático de datos para la generación de soluciones.

2.3.3. Clasificación Supervisada

La clasificación supervisada es una técnica de aprendizaje automático que entrena modelos a partir de datos etiquetados, es decir, datos cuya categoría de salida es conocida, con el objetivo de predecir la clase de nuevas instancias no vistas previamente (Russell et al., 2021). En la presente investigación, esta técnica fue aplicada para clasificar a los estudiantes en dos categorías: presencia o ausencia de indicadores asociados al TDAH.

2.3.4. Accuracy

La exactitud (Accuracy) se define como la proporción de aciertos que un modelo de Machine Learning logra sobre el universo total de predicciones realizadas. Según Russell y Norvig (2021), este indicador se determina matemáticamente dividiendo la cantidad de predicciones correctas entre el número total de casos evaluados.

2.3.5. Precisión

La precisión se define como la métrica que cuantifica la proporción de instancias correctamente identificadas como positivas respecto al total de casos clasificados como tales. De acuerdo con Russell y Norvig (2021), este indicador es fundamental para determinar la confiabilidad y calidad de las predicciones positivas generadas por el modelo.

2.3.6. Recall

La sensibilidad, técnicamente denominada (Recall), evalúa la capacidad del modelo para identificar la totalidad de los casos positivos existentes en la muestra. Para Russell y Norvig (2021), esta métrica es determinante en escenarios clínicos como la detección del TDAH, donde la prioridad es minimizar los falsos negativos para asegurar que ningún estudiante con el trastorno quede sin diagnosticar.

2.3.7. F1-Score

Para ponderar simultáneamente la precisión y la sensibilidad, se emplea el F1-score. Esta métrica, calculada como una media armónica, es esencial para evaluar el rendimiento de modelos en contextos de desbalance de clases, una característica habitual en los estudios de salud y diagnóstico clínico (Russell y Norvig, 2021).

2.3.8. ROC-AUC

El Área Bajo la Curva ROC (AUC-ROC) se constituye como una métrica fundamental para evaluar la competencia del modelo al discriminar entre distintas clases. De acuerdo con Russell y Norvig (2021), un coeficiente próximo a la unidad (1) denota una elevada capacidad de diferenciación por parte del algoritmo.

2.3.9. Árbol de Decisión

Un árbol de decisión es un modelo de aprendizaje supervisado que representa decisiones mediante una estructura jerárquica de nodos y ramas, donde cada nodo interno corresponde a una condición sobre una variable, cada rama representa el resultado de dicha condición y cada nodo hoja indica la clase predicha (BREIMAN et al., 2017). Este modelo constituye la unidad base del algoritmo Random Forest empleado en la presente investigación.

2.3.10. Hiperplano

Un hiperplano es una superficie matemática de dimensión $n - 1$ que separa clases en un espacio multidimensional de dimensión n . En el contexto del algoritmo Support Vector Machine, el hiperplano óptimo maximiza el margen entre las clases, siendo determinado por los vectores de soporte más cercanos a la frontera de decisión (Rusca-Jordán y Cortez-Vergara, 2020).

CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN

3.1. Tipo de estudio

La presente investigación se clasificó como aplicada, dado que aplicó conocimientos científicos existentes derivados del campo del aprendizaje automático para abordar un problema específico del ámbito educativo. Según Castro et al. (2018), la investigación aplicada se orienta al uso del conocimiento científico con el propósito de resolver problemas concretos de la realidad.

Asimismo, el estudio adoptó un enfoque cuantitativo, distinguiéndose por la primacía de la medición objetiva y el tratamiento numérico de la información. De acuerdo con Vizcaíno Zúñiga et al. (2023), esta metodología implica la recolección sistemática de datos cuantificables para su posterior procesamiento mediante análisis estadísticos rigurosos.

En este contexto, los datos obtenidos fueron analizados mediante modelos de aprendizaje automático, específicamente Random Forest, Support Vector Machine (SVM) y K-Nearest Neighbors (KNN), con el propósito de evaluar su desempeño predictivo en la detección de patrones asociados al TDAH en la población estudiada.

Adicionalmente, el estudio presentó un alcance predictivo, dado que mediante los modelos de aprendizaje automático implementados se buscó predecir la presencia de indicadores asociados al TDAH en la población estudiantil analizada (Hernández-Sampieri et al., 2014).

3.2. Diseño de estudio

El presente estudio adoptó un diseño no experimental, ya que las variables de investigación no fueron manipuladas deliberadamente por el investigador, sino que fueron observadas y analizadas en su contexto natural. Según Hernández-Sampieri et al. (2014), los estudios no experimentales se caracterizan por analizar los fenómenos tal como ocurren en la realidad, sin intervenir directamente en su comportamiento. Este tipo de diseño resultó pertinente dado que el objetivo fue evaluar el desempeño de modelos predictivos a partir de datos conductuales ya existentes, sin introducir ninguna intervención sobre los estudiantes ni sobre su entorno educativo.

Asimismo, la adopción de este diseño se fundamentó en la necesidad de evaluar el desempeño predictivo de modelos de Machine Learning utilizando datos observacionales obtenidos a partir de la Conners' Rating Scale y registros conductuales, sin alterar el entorno educativo de los estudiantes. Este enfoque permitió identificar patrones asociados al Trastorno por Déficit de Atención e Hiperactividad (TDAH) en condiciones reales del contexto escolar.

El estudio presentó además un diseño de corte transversal, ya que la recolección de los datos se realizó en un único momento temporal durante el periodo correspondiente al año 2025. Según Hernández-Sampieri et al. (2014), los estudios transversales permiten analizar variables y fenómenos en un momento determinado, facilitando la descripción y análisis de las características presentes en una población específica.

3.3. Población y muestra

3.3.1. Población

La población está compuesta por 300 estudiantes del nivel primario (6 a 12 años) matriculados en la Institución Educativa Privada Maria Reina, Puerto Maldonado durante el año 2025.

Los estudiantes constituyeron la unidad poblacional del estudio, mientras que la información necesaria para la investigación fue recopilada mediante la

aplicación del instrumento Conners Rating Scale, el cual fue completado por padres o tutores y docentes, quienes actuaron como informantes del comportamiento y estado de atención de los estudiantes. Asimismo, se consideraron registros de observación conductual en el aula como información complementaria para el análisis.

3.3.2. Muestra

Se utilizó un muestreo probabilístico estratégico con asignación proporcional por grado, con el propósito de garantizar la representatividad de todos los grados y edades del nivel primario de la Institución Educativa Privada María Reina de Puerto Maldonado. Este procedimiento permitió realizar una selección objetiva de los estudiantes, asegurando la representatividad de la población y fortaleciendo el rigor estadístico del estudio.

Para determinar el tamaño de la muestra se empleó la fórmula para poblaciones finitas:

$$= \frac{NZ^2p \cdot q}{(N - 1)E^2 + Z^2pq}$$

Donde:

n= tamaño muestra

N=tamaño de la población

Z= nivel de confianza

p= proporción de población deseada

q= proporción complementaria

E= margen de error permitido

Para el presente estudio se consideró:

Z = 95% (Z=1.96)

E = 5% (E=0.05)

p = 0.5

q = 0.5

N = 300

Sustituyendo los valores en la fórmula:

$$n = \frac{(300)(1.96)^2(0.50)(0.50)}{(300 - 1)(0.05)^2 + (1.96)^2(0.50)(0.50)}$$

$$n = 169$$

En consecuencia, el tamaño de la muestra es de 169 estudiantes, lo que permitió realizar un análisis representativo de los casos asociados al Trastorno por Déficit de Atención e Hiperactividad (TDAH) en la población estudiada.

3.4. Métodos y técnicas

- Cuestionarios Estandarizados:

Se empleó el instrumento Conners' Rating Scale (Conners, 2008), Una herramienta psicométrica validada internacionalmente para evaluación de síntomas asociados al TDAH en la población estudiantil. Este cuestionario fue completado por los padres o tutores, quienes proporcionaron información sobre el comportamiento de los estudiantes en el entorno familiar y en otros contextos fuera del aula.

La validez de contenido del instrumento fue evaluada mediante juicio de expertos, mientras que la confiabilidad fue determinada a través del coeficiente alfa de Cronbach, obteniendo un valor superior a 0.80, lo cual indica una consistencia interna adecuada para su aplicación en el presente estudio (Hernández-Nieto, 2002).

Asimismo, se recopiló información complementaria proporcionada por docentes, quienes evaluaron el comportamiento de los estudiantes dentro del contexto escolar. La integración de ambas fuentes de información permitió obtener una visión más íntegra de los patrones conductuales relacionados con el TDAH, fortaleciendo la calidad de los datos recopilados para el análisis posterior.

Es importante precisar que el instrumento no fue utilizado con fines de diagnóstico clínico, sino como una herramienta de recolección de datos para

la identificación de indicadores conductuales que posteriormente fueron analizados mediante los modelos de aprendizaje automático planteados.

- **Observaciones conductuales:**

Se realizaron observaciones conductuales estructuradas en el aula (Hernández-Sampieri et al., 2014), con el propósito de registrar la frecuencia de comportamientos asociados a inatención, hiperactividad e impulsividad en los estudiantes. Estas observaciones siguieron un protocolo previamente definido, que contempló sesiones de observación de 45 minutos durante las actividades académicas regulares.

Los observadores fueron previamente capacitados, y su intervención se realizó de manera discreta, evitando cualquier tipo de interferencia en el desarrollo normal de las actividades escolares o posibles situaciones de estigmatización hacia los estudiantes.

3.5. Tratamiento de los datos

Los datos se recolectaron mediante la Escala de Conners, aplicada a padres, tutores y docentes, además de registros de observación conductual en el aula. Posteriormente, la información fue sometida a un proceso de preparación y tratamiento previo antes de su análisis.

Las respuestas del cuestionario fueron transformadas a valores numéricos según la escala del instrumento (0 = Nunca, 1 = A veces, 2 = Frecuentemente, 3 = Muy frecuentemente). Luego se realizó la depuración de los datos, identificando valores faltantes y tratándolos mediante imputación con la mediana para reducir posibles sesgos en el análisis.

Asimismo, las variables fueron sometidas a un proceso de normalización dentro del rango $[0,1]$, con la finalidad de favorecer un mejor desempeño de los algoritmos de aprendizaje automático.

A continuación, los datos fueron analizados mediante algoritmos de clasificación supervisada como Random Forest, Support Vector Machine y K-Nearest Neighbors (KNN), implementados en Python utilizando la biblioteca scikit-learn. Adicionalmente, se realizó la optimización de hiperparámetros

mediante el método de búsqueda en cuadrícula (Grid Search) con el fin de mejorar el desempeño de los modelos.

El desempeño de los modelos se evaluó mediante métricas de clasificación como accuracy, precisión, recall, F1-score y AUC-ROC (Hastie, Tibshirani y Friedman, 2009), empleando un esquema de validación cruzada k-fold ($k=5$) para estimar su capacidad de generalización.

Finalmente, el procesamiento de los datos se llevó a cabo bajo la supervisión de un profesional especializado en psicología, quien aseguró que la interpretación de los resultados se realizó de manera ética y responsable, evitando cualquier tipo de estigmatización. Del mismo modo, toda la información recopilada fue resguardada de manera confidencial en una base de datos segura, respetando los principios de privacidad y protección de datos de los participantes.

CAPÍTULO IV: RESULTADOS DEL TRABAJO DE INVESTIGACION

4.1 Entendimiento del negocio

4.1.1 Determinación de objetivos del negocio

El objetivo principal de este estudio fue desarrollar e implementar una herramienta de software predictivo basada en algoritmos de aprendizaje automático (Machine Learning) que sirva como sistema de triaje para la detección temprana del Trastorno por Déficit de Atención e Hiperactividad (TDAH) en estudiantes de la Institución Educativa Privada María Reina de Puerto Maldonado. El sistema buscó apoyar al departamento de psicología escolar procesando los resultados del test de Conners para emitir alertas tempranas, reduciendo la subjetividad y priorizando la minimización de falsos negativos.

4.1.2 Evaluación de la situación

Actualmente, la detección del TDAH en la institución se realiza de forma manual y observacional, lo que conlleva una alta carga administrativa, susceptibilidad a sesgos humanos y riesgo de subdiagnóstico debido a la falta de herramientas tecnológicas automatizadas. El conjunto de datos utilizado (ingresado en hojas de cálculo) presentaba inconsistencias estructurales, como fórmulas rotas y datos incompletos. Desde la perspectiva clínica, el problema presentaba un desbalance entre la clase positiva y negativa, una inmensa mayoría de estudiantes neurotípicos frente a un porcentaje reducido de casos confirmados o sospechosos de TDAH.

4.2 Comprensión de los datos

4.2.1 Adquisición de los datos

La información fue recolectada a través de la digitalización de cuestionarios físicos basados en la Escala de Conners (adaptada), instrumento ampliamente utilizado para la evaluación de comportamientos asociados al Trastorno por Déficit de Atención e Hiperactividad en población infantil (Conners, 2008), respondidos por padres y docentes. Los datos fueron tabulados en formato Excel/CSV conteniendo el ID del participante, características demográficas y las puntuaciones ordinales (0 al 3) de 16 variables conductuales correspondientes a las subescalas de Inatención, Hiperactividad y Control Cognitivo. La figura 1 muestra el proceso de recopilación y organización de los datos.

Figura 1. Registro y tabulación de datos del cuestionario de la Escala de Conners en formato Excel.

The image shows a screenshot of an Excel spreadsheet. The spreadsheet has a wide range of columns, with the first few columns containing text labels and the remaining columns containing numerical data. The data is organized in a grid-like structure, with rows representing individual participants and columns representing different variables or items from the Conners scale. The spreadsheet is displayed in a standard Excel interface with a ribbon at the top and a status bar at the bottom.

Fuente: Elaboración propia.

4.2.2 Descripción de los datos

El conjunto de datos original cargado en el entorno de Python constaba inicialmente de variables sociodemográficas y reactivos psicométricos. La tabla 2 presenta la descripción de las variables del conjunto de datos.

Tabla 2. Descripción de datos del conjunto de variables.

Id	Variable	Descripción	Tipo de dato	Categoría
01	ID_Participante	Identificador único asignado a cada estudiante	Object	Identificación
02	Edad	Edad del estudiante evaluado	Int64	Demográfica
03	Sexo_M	Indicador binario de sexo masculino	Int64	Demográfica
04	Sexo_F	Indicador binario de sexo femenino	Int64	Demográfica
05	Informante_Padre	Indica si el cuestionario fue respondido por el padre	Int64	Informante
06	Informante_Madre	Indica si el cuestionario fue respondido por la madre	Int64	Informante
07	Informante_Docente	Indica si el cuestionario fue respondido por el docente	Int64	Informante
08	Informante_Autorreporte	Indica si el cuestionario fue respondido por el propio estudiante	Int64	Informante
09	Contexto_Casa	Contexto de evaluación en el hogar	Int64	Contexto
10	Contexto_Escuela	Contexto de evaluación en la escuela	Int64	Contexto
11	I1_Atencion_tareas	Dificultad para mantener la atención en tareas	Int64	Inatención
12	I2_Terminar_tareas	Dificultad para finalizar tareas asignadas	Int64	Inatención
13	I3_Evita_esfuerzo_mental	Evita actividades que requieren esfuerzo mental sostenido	Int64	Inatención
14	I4_Distraccion_externa	Se distrae fácilmente con estímulos externos	Int64	Inatención
15	I5_Pierde_objetos	Pierde objetos necesarios para sus actividades	Int64	Inatención
16	I6_No_escucha	Parece no escuchar cuando se le habla directamente	Int64	Inatención
17	H1_Habla_exceso	Habla en exceso en situaciones inapropiadas	Int64	Hiperactividad
18	H2_Actua_sin_pensar	Actúa de forma impulsiva sin reflexionar	Int64	Hiperactividad
19	H3_Esperar_turno	Dificultad para esperar su turno	Int64	Hiperactividad
20	H4_Levanta_asiento	Se levanta del asiento cuando debería permanecer sentado	Int64	Hiperactividad
21	Total_Mean	Promedio general de los puntajes conductuales	Float64	Variable derivada
22	ADHD_Diagnosis	Variable objetivo que indica presencia o ausencia de TDAH	Object	Variable objetivo

Fuente: Elaboración propia.

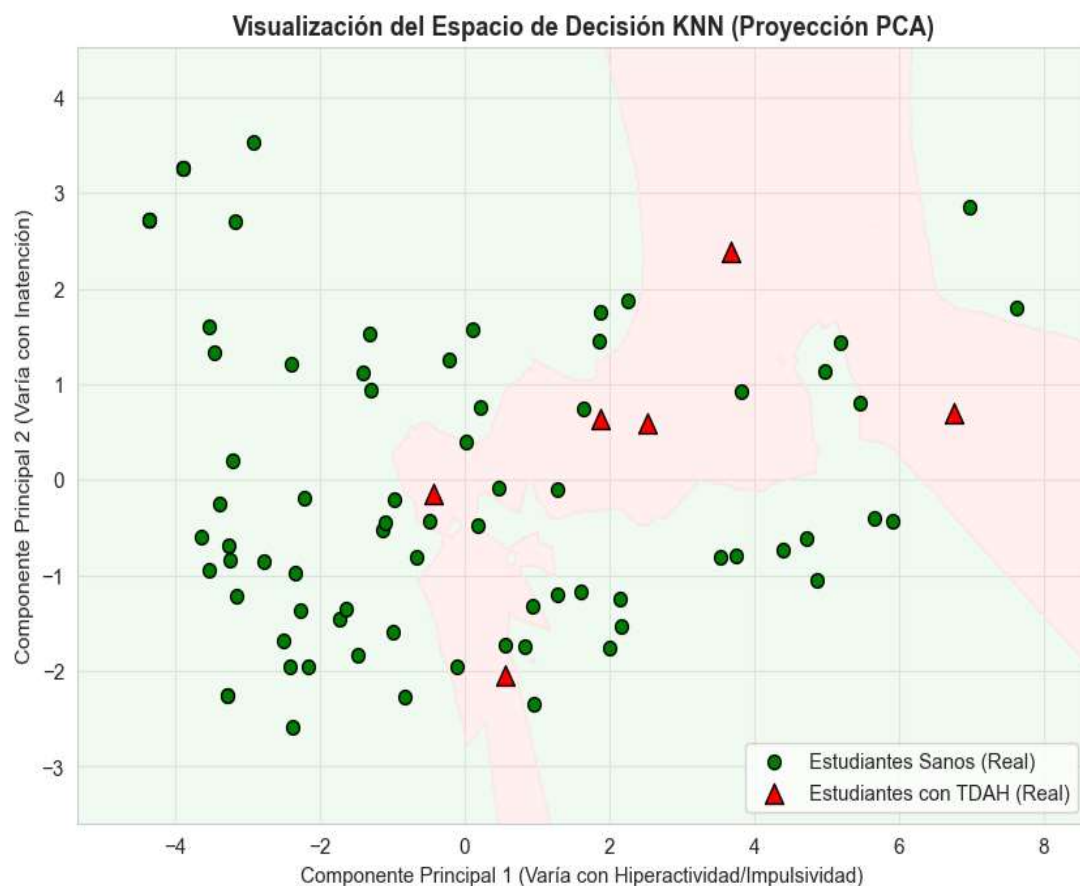
Las variables conductuales fueron registradas utilizando una escala ordinal de 0 a 3, donde 0 indica ausencia del comportamiento evaluado y 3 representa una frecuencia elevada del mismo. Estas variables corresponden a indicadores conductuales asociados a las dimensiones de inatención e hiperactividad evaluadas mediante la Escala de Conners.

4.2.3 Análisis exploratorio de los datos

Durante la exploración computacional usando Pandas y Seaborn, se evidenció la naturaleza altamente desbalanceada del problema clínico.

Se visualizó la proyección PCA (Análisis de Componentes Principales) en el notebook, la cual demostró que los casos de TDAH no se separan de los casos sanos mediante una línea recta (linealidad), sino que se agrupan en pequeñas "islas" o clústeres dispersos en el espacio multidimensional según la gravedad de sus síntomas combinados de inatención e hiperactividad. La figura 2 muestra la gráfica de análisis de componentes principales.

Figura 2. Proyección PCA del espacio de decisión KNN.

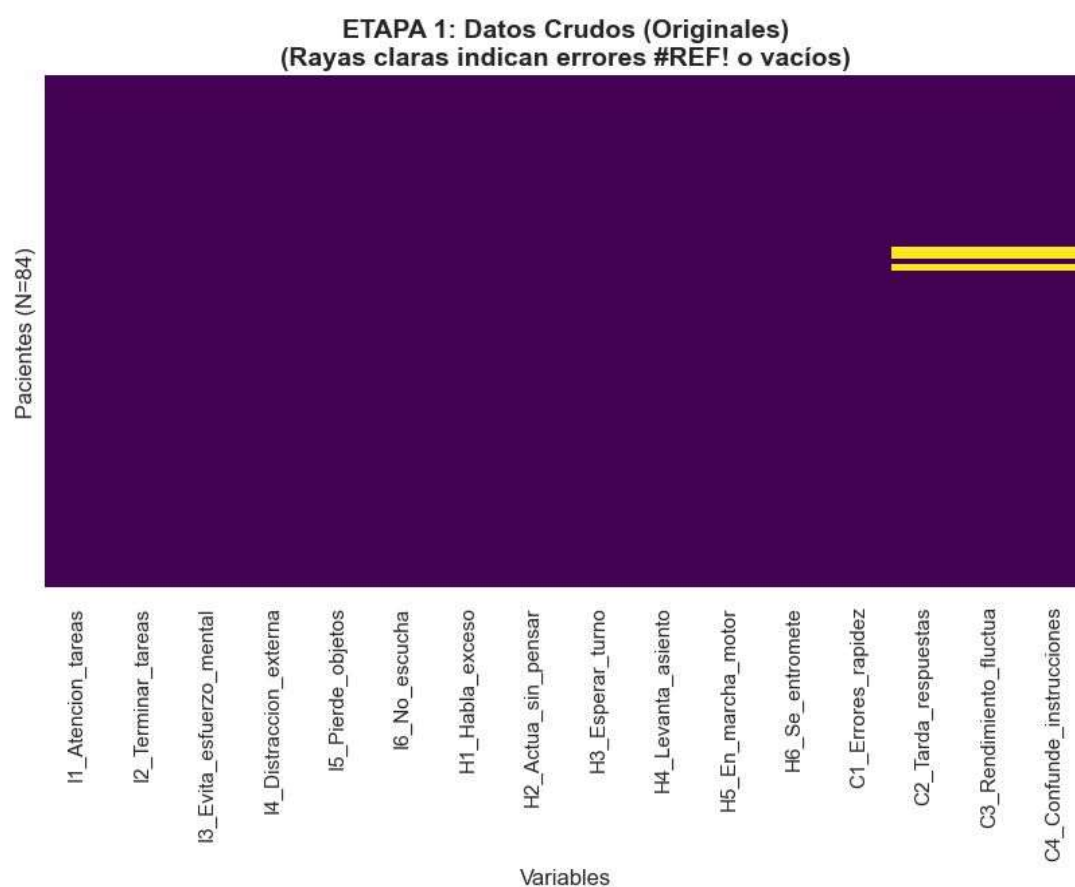


Fuente: Elaboración propia.

4.2.4 Verificación de la calidad de los datos

En esta tarea se verificó la presencia de valores faltantes NaN, celdas corruptas con error de referencia #REF!, estos problemas impedían el proceso de entrenamiento directo por parte de los algoritmos de Scikit-Learn, obligando a diseñar un pipeline de limpieza. La figura 3 muestra los datos faltantes del conjunto de datos.

Figura 3. Muestra de conjunto de datos faltantes.



Fuente: Elaboración propia.

4.3 Preparación de los datos

4.3.1 Selección de datos

Se eligieron las variables más relevantes del conjunto de datos, incluyendo edad, sexo y variables conductuales relacionadas con inatención e hiperactividad, con el fin de utilizarlas en el entrenamiento del modelo predictivo. La tabla 3 presenta la selección de las variables del conjunto de datos.

Tabla 3. Selección de datos según variables.

Id	Variable	Tipo de dato	Descripción
1	Edad	Int64	Edad del estudiante evaluado
2	Sexo_M	Int64	Indicador de sexo masculino
3	Sexo_F	Int64	Indicador de sexo femenino
4	I1_Atencion_tareas	Int64	Dificultad para mantener la atención en tareas
5	I2_Terminar_tareas	Int64	Dificultad para finalizar tareas
6	I3_Evita_esfuerzo_mental	Int64	Evita tareas que requieren esfuerzo mental sostenido
7	I4_Distraccion_externa	Int64	Se distrae fácilmente con estímulos externos
8	I5_Pierde_objetos	Int64	Pierde objetos necesarios para actividades
9	I6_No_escucha	Int64	Parece no escuchar cuando se le habla directamente
10	H1_Habla_exceso	Int64	Habla excesivamente en diferentes contextos
11	H2_Actua_sin_pensar	Int64	Actúa de manera impulsiva sin pensar
12	H3_Esperar_turno	Int64	Dificultad para esperar su turno
13	H4_Levanta_asiento	Int64	Se levanta del asiento en situaciones inapropiadas
14	H5_Inquietud_motora	Int64	Presenta movimientos constantes o inquietud
15	H6_Interrumpe	Int64	Interrumpe o se entromete en actividades de otros
16	C1_Control_conductual	Int64	Capacidad de regular el comportamiento
17	C2_Regulacion_emocional	Int64	Capacidad para controlar reacciones emocionales
18	C3_Persistencia	Int64	Capacidad de mantener esfuerzo en tareas
19	C4_Variabilidad_conductual	Int64	Variabilidad en el comportamiento observado

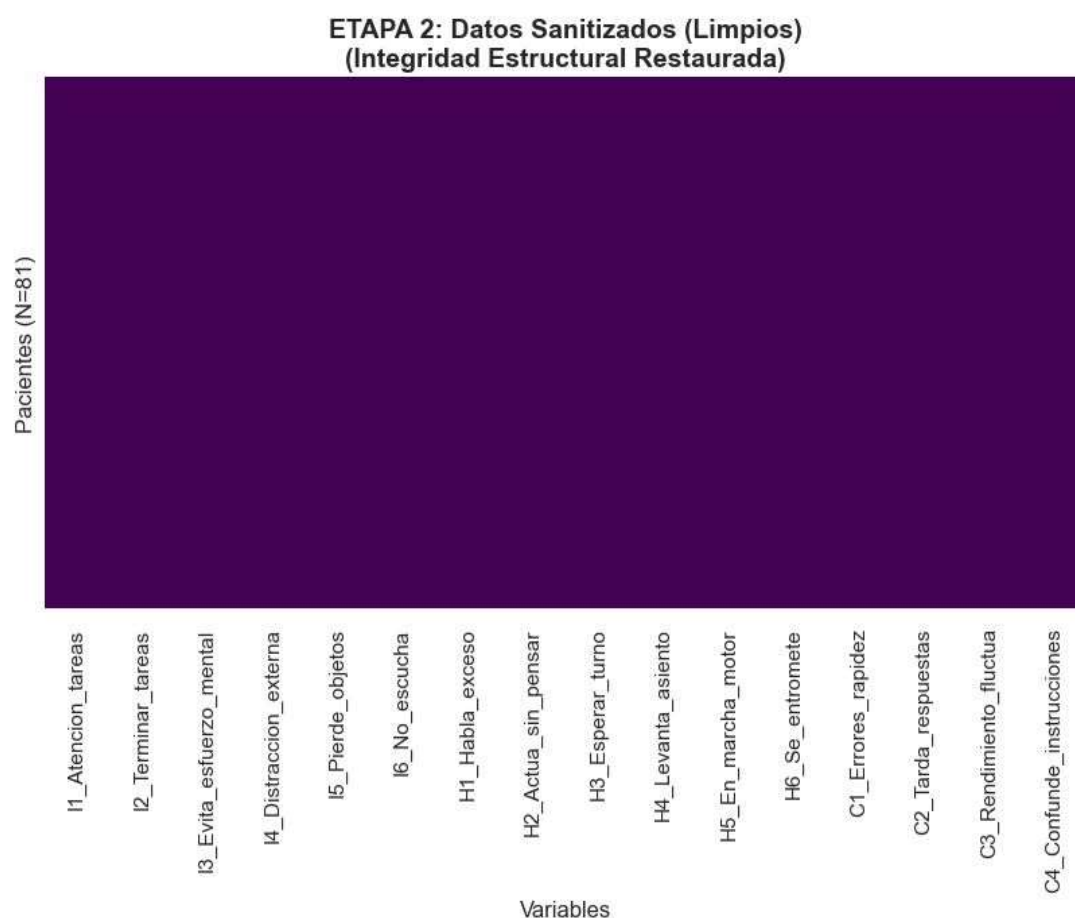
Fuente: Elaboración propia.

Las variables presentadas corresponden a los atributos seleccionados del conjunto de datos utilizados en el proceso de análisis y entrenamiento del modelo. Estas incluyen información demográfica y variables conductuales derivadas de los ítems evaluados en el cuestionario aplicado.

4.3.2 Limpieza y preprocesamiento de los datos

Se estudio implementó un script para automatizar la sanitización de los datos. Este proceso incluyó la coerción numérica, que forzó la conversión de cadenas de texto erróneas (por ejemplo, #REF!) a valores nulos (NaN). En relación con la variable objetivo, se realizó su codificación en formato binario (1 = Positivo, 0 = Negativo). Finalmente, se aplicó StandardScaler para estandarizar las magnitudes de las variables, el cual fue posteriormente exportado como escalador_tdah.pkl. La figura 4 muestra el conjunto de datos luego del proceso de limpieza.

Figura 4. Datos sanitizados luego del proceso de limpieza.

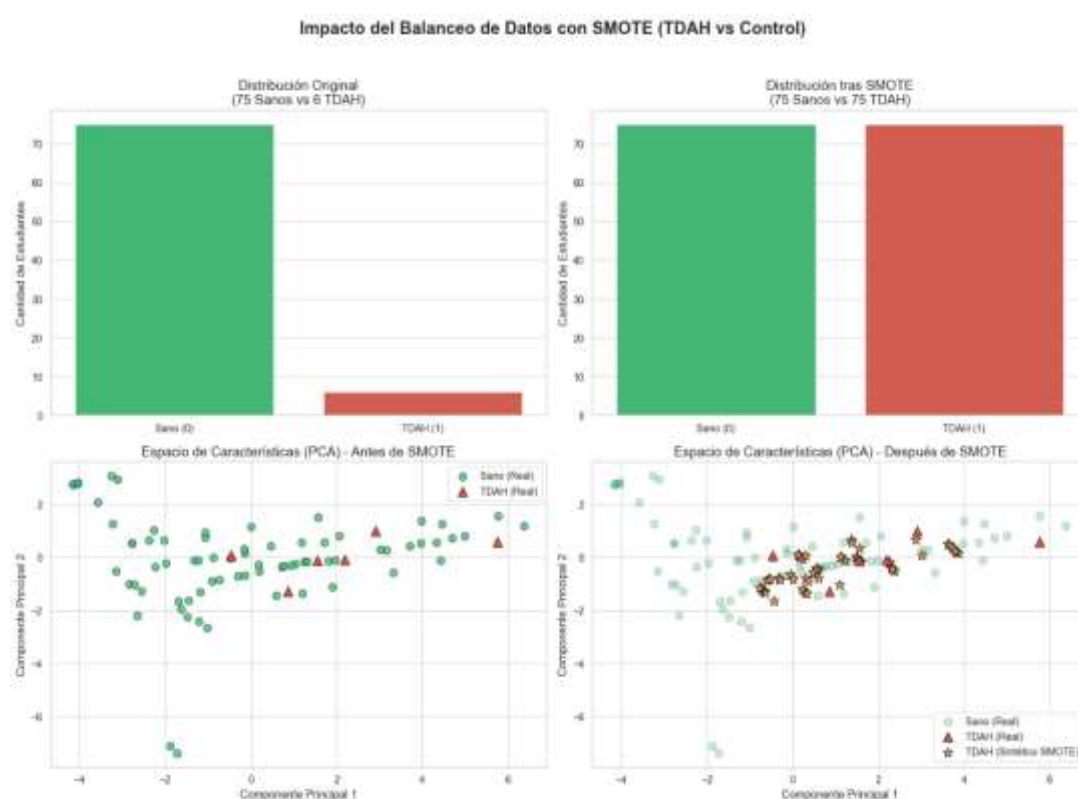


Fuente: Elaboración propia.

4.3.3 Aumento de datos (Balanceo de datos)

Debido al marcado desbalance entre clases, un modelo tradicional simplemente predeciría "sin indicadores de TDAH" siempre. Para corregir este sesgo, se aplicó la técnica SMOTE (Synthetic Minority Over-sampling Technique) (HASTIE, TIBSHIRANI y FRIEDMAN, 2009). El algoritmo generó casos sintéticos matemáticamente viables interpolando los vectores de los 6 pacientes reales con TDAH.

Figura 5. Impacto del balanceo de datos.



Fuente: Elaboración propia.

El resultado post-SMOTE mostró que el conjunto de datos de entrenamiento pasó a tener 150 registros perfectamente equilibrados (75 casos TDAH / 75 casos Control), permitiendo a los modelos aprender patrones sin sesgos de clase mayoritaria. Tal como muestra la figura 5 distribución de las clases en la variable objetivo, antes y después del proceso de balanceo.

4.4 Modelamiento

En esta fase se implementaron los algoritmos Random Forest, Máquinas de Vectores de Soporte (Support Vector Machine, SVM) y K-Nearest Neighbors

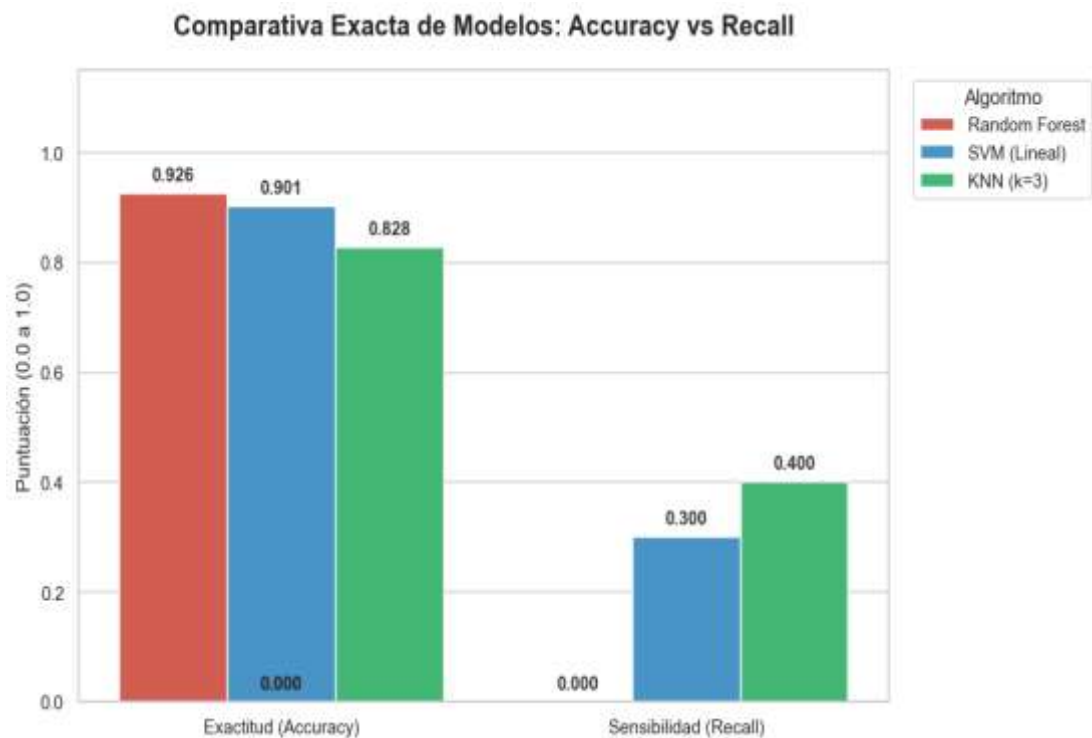
(KNN), debido a que estos métodos han demostrado un desempeño eficiente en tareas de clasificación dentro del ámbito del aprendizaje automático, siendo ampliamente utilizados en problemas de análisis predictivo (Hastie, Tibshirani y Friedman, 2009). Para el proceso de entrenamiento se empleó un esquema de validación cruzada estratificada (Stratified K-Fold, $k = 5$) con el propósito de garantizar una evaluación más robusta del modelo. Asimismo, se llevó a cabo un proceso de optimización de hiperparámetros mediante el método de búsqueda en cuadrícula (Grid Search), con el fin de identificar la configuración más adecuada para cada algoritmo. La tabla 4 presenta el ajuste de hiperparámetros aplicado a los modelos evaluados, mientras que la figura 6 muestra el desempeño comparativo de los algoritmos mediante gráficos de barras basados en las métricas de clasificación obtenidas.

Tabla 4. Ajuste de hiperparámetros mediante Grid Search para los algoritmos de clasificación.

Modelo	Hiperparámetro	Rango evaluado	Valor óptimo
Random Forest	n_estimators	[100, 200, 300]	100
	max_depth	[None, 10, 20]	None
	min_samples_split	[2, 5, 10]	2
Support Vector Machine (SVM)	C	[0.1, 1, 10]	1
	kernel	[linear, rbf]	rbf
	gamma	[scale, auto]	scale
K-Nearest Neighbors (KNN)	n_neighbors	[3, 5, 7]	3
	metric	[euclidean, manhattan]	euclidean
	weights	[uniform, distance]	uniform

Fuente: Elaboración propia.

Figura 6. Comparación de desempeño de los modelos según métricas.



Fuente: Elaboración propia.

4.4.1 Random Forest

Entrenamiento y ajuste de hiperparámetros.

Se entrenó un ensamblaje de múltiples árboles de decisión buscando aprovechar su capacidad teórica para manejar múltiples características de manera simultánea.

Métricas de evaluación.

- Exactitud General (Accuracy): 92.6% (0.926471)
- Sensibilidad (Recall para TDAH): 0.0

El modelo presentó limitaciones para generalizar adecuadamente los patrones clínicos asociados al TDAH. La elevada accuracy obtenida no reflejó una adecuada capacidad de detección de casos positivos, catalogó a casi todos los estudiantes como sanos (sobreajuste), lo que generó un 100% de falsos negativos en la población de riesgo (Hastie, Tibshirani y Friedman, 2009). Por esta razón fue descartado de inmediato.

4.4.2 Máquinas de vector de soporte (SVM)

Se evaluó el algoritmo con un kernel lineal, intentando trazar un hiperplano recto que separara las clases.

- Exactitud General (Accuracy): 90.1% (0.901471)
- Sensibilidad (Recall para TDAH): 0.3

El modelo SVM mostró un ligero avance respecto a Random Forest, logrando detectar algunos casos; sin embargo, la rigidez lineal geométrica del modelo le impidió adaptarse a las agrupaciones complejas y asimétricas ("islas") de síntomas que conforman el TDAH (Hastie, Tibshirani y Friedman, 2009).

4.4.3 K-Nearest Neighbors (KNN)

Se iteró el hiperparámetro k (número de vecinos) desde 1 hasta 15. La curva de estabilidad del modelo demostró que el rendimiento óptimo se lograba en k=3.

- Exactitud General (Accuracy): 82.7% (0.827941)
- Sensibilidad (Recall para TDAH): 0.4

Fue el modelo ganador seleccionado. Aunque sacrificó algunos puntos de exactitud global (aumentando levemente las alertas preventivas o falsos positivos), logró la mejor sensibilidad (0.4) para identificar casos de TDAH reales. Su lógica basada en distancia espacial fue ideal para este contexto clínico (Hastie, Tibshirani y Friedman, 2009). El modelo final fue serializado usando Joblib (modelo_tdah_knn.pkl). La figura 7 muestra la validación cruzada del conjunto de datos del modelo final.

Figura 7. Validación cruzada de modelo KNN.

```
# Validación cruzada estratificada (mantiene la proporción de casos en cada prueba)
cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)

modelos = {
    "Random Forest": RandomForestClassifier(n_estimators=100, random_state=42, class_weight='balanced'),
    "SVM (Kernel Lineal)": SVC(kernel='linear', probability=True, class_weight='balanced'),
    "KNN (k=3)": KNeighborsClassifier(n_neighbors=3)
}

resultados = []
print("Iniciando entrenamiento y validación cruzada...")

for nombre, modelo in modelos.items():
    try:
        # Pipeline: Escalar, Generar sintético (SMOTE), Entrenar Modelo
        pipeline = ImbPipeline([
            ('scaler', StandardScaler()),
            ('smote', SMOTE(random_state=42, k_neighbors=2)),
            ('model', modelo)
        ])

        # Métricas clave
        recall = cross_val_score(pipeline, X, y, cv=cv, scoring='recall').mean()
        accuracy = cross_val_score(pipeline, X, y, cv=cv, scoring='accuracy').mean()

        resultados.append([
            "Algoritmo": nombre,
            "Sensibilidad (Detectar TDAH)": recall,
            "Precisión General": accuracy
        ])
    except:
```

Fuente: Elaboración propia.

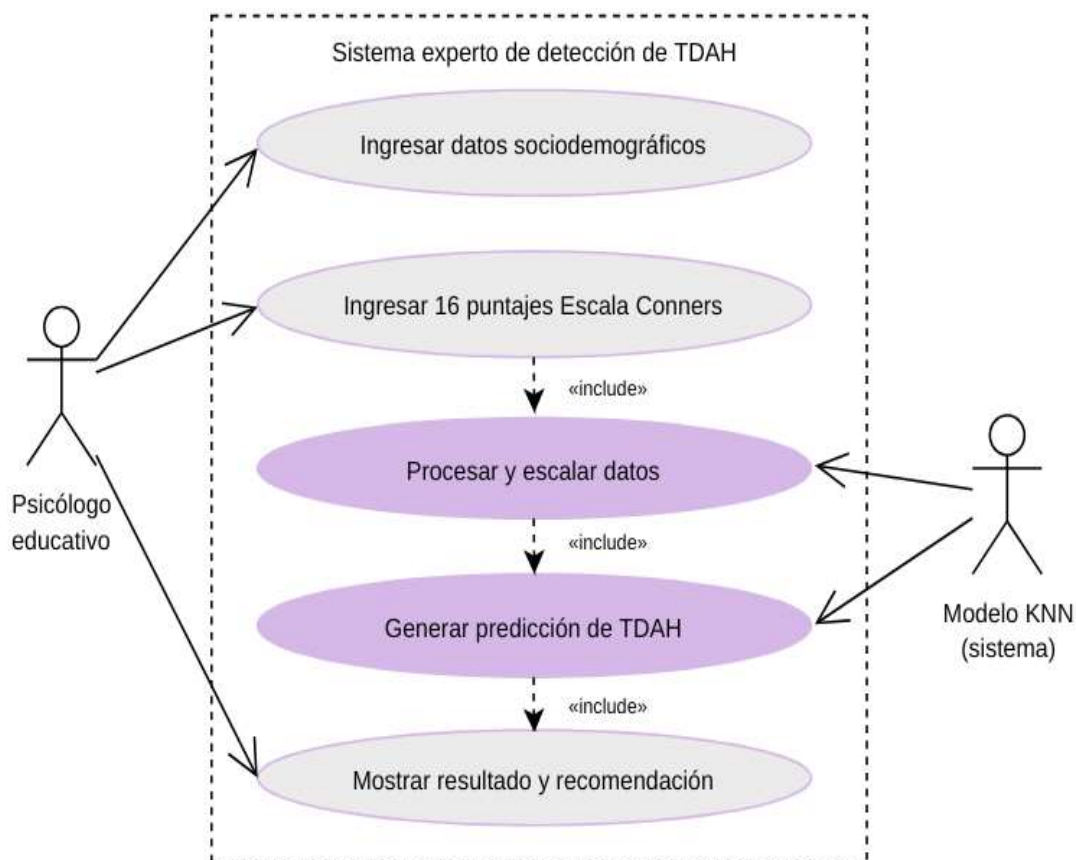
4.5 Despliegue (Análisis, diseño, desarrollo y pruebas)

4.5.1 Análisis

Casos de uso principales:

1. Ingresar Evaluación: El psicólogo accede a la plataforma web e ingresa los 16 puntajes del test de Conners y datos sociodemográficos de un estudiante.
2. Generar Predicción: El sistema recibe los datos, los escala automáticamente y consulta el modelo de IA.

La figura 8 presenta el diagrama de casos de uso del sistema de detección de TDAH.

Figura 8. Diagrama de casos de uso del sistema.

Fuente: Elaboración propia.

Tabla 5. Requisitos funcionales.

Código	Descripción
RF01	El sistema permite ingresar datos sociodemográficos del estudiante
RF02	El sistema permite ingresar los 16 puntajes del test de Conners
RF03	El sistema procesa y escala automáticamente los datos ingresados
RF04	El sistema genera una predicción de probabilidad de TDAH
RF05	El sistema muestra el resultado con un indicador porcentual
RF06	El sistema emite una recomendación de derivación al psicólogo

Fuente: Elaboración propia.

Tabla 6. Requisitos No Funcionales.

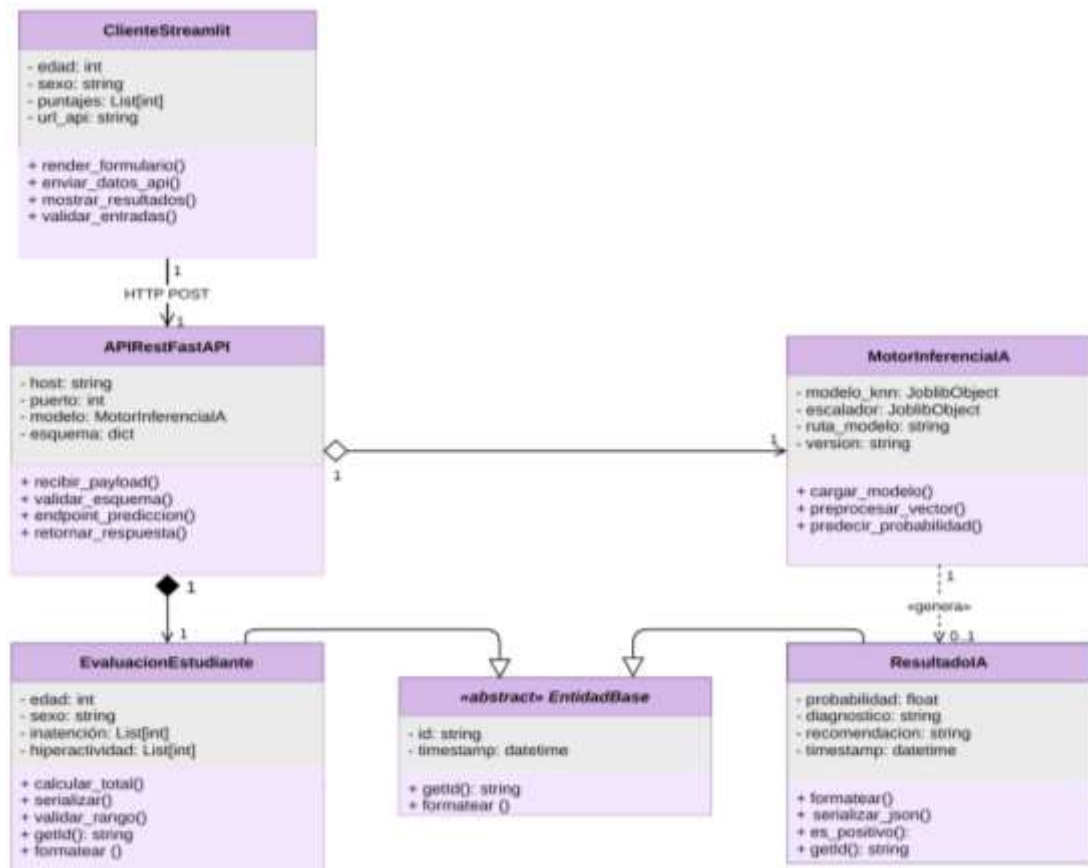
Código	Descripción
RNF01	El sistema debe responder en menos de 3 segundos
RNF02	La interfaz debe ser intuitiva y de fácil uso
RNF03	Los datos ingresados deben ser tratados con confidencialidad
RNF04	El sistema debe estar disponible vía navegador web
RNF05	El modelo debe mantener una exactitud mínima del 75%

Fuente: Elaboración propia.

4.5.2 Diseño de la solución

Diagrama de Clases

El diagrama de clases muestra la estructura del sistema, las clases principales, sus atributos y las relaciones entre ellas, permitiendo representar la organización de los componentes utilizados en la solución propuesta. La figura 9 muestra el diagrama de clases del sistema.

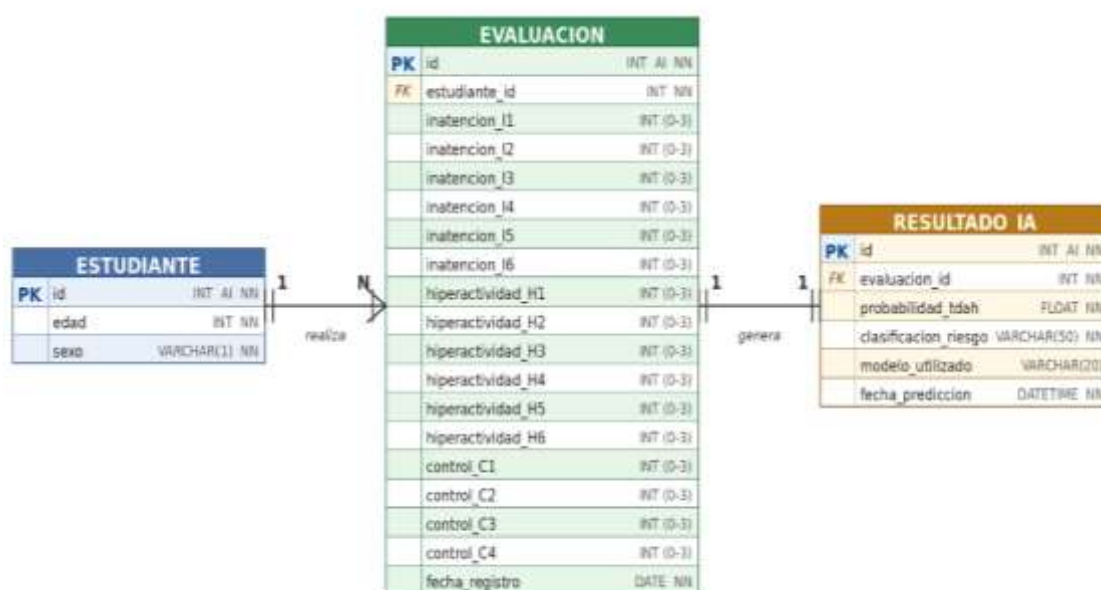
Figura 9. Diagrama de clases del sistema de detección de TDAH.

Fuente: Elaboración propia.

Diagrama de Entidad Relación

El diagrama entidad-relación representa la estructura de los datos del sistema, mostrando las entidades principales, sus atributos y las relaciones existentes entre ellas. Este modelo permitió organizar la información utilizada en el proceso de análisis y clasificación. La figura 10 muestra la relación del conjunto de datos.

Figura 10. Diagrama entidad relación del sistema de detección de TDAH.



Fuente: Elaboración propia.

Diccionario de datos

El diccionario de datos describe la estructura, tipo de dato, restricciones y significado de cada atributo de las entidades que conforman el sistema de detección de TDAH. Las tablas presentadas a continuación corresponden a las tres entidades identificadas en el diagrama entidad relación, donde la tabla 7 muestra datos de la entidad estudiante, la tabla 8 los datos de la entidad evaluación, finalmente la tabla 9 muestra los datos de resultado IA.

Entidad: ESTUDIANTE

Almacena los datos demográficos básicos de cada estudiante evaluado en la I.E.P. María Reina.

Tabla 7. Diccionario de datos de la entidad ESTUDIANTE.

ESTUDIANTE					
Campo	Tipo	Longitud	Nulo	Clave	Descripción
Id	INT	--	NO	PK	Identificador único del estudiante. Se incrementa automáticamente.
edad	INT	--	NO		Edad del estudiante en años. Rango valido: 6 a 12.
sexo	VARCHAR	1	NO		Sexo del estudiante. Valores permitidos: M = Masculino, F = Femenino.

Fuente: Elaboración propia.

Entidad: EVALUACION

Registra los puntajes obtenidos en el Test de Connors adaptado, respondido por padres o docentes. La escala ordinal utilizada es: 0 = Nunca, 1 = A veces, 2 = Frecuentemente, 3 = Muy frecuentemente.

Tabla 8. Diccionario de datos de la entidad EVALUACION.

EVALUACION					
Campo	Tipo	Longitud	Nulo	Clave	Descripción
Id	INT	--	NO	PK	Identificador único de la evaluación. Se incrementa automáticamente.
estudiante_id	INT	--	NO	FK	Referencia al estudiante evaluado. Relacionado con ESTUDIANTE (id).
inatencion_l1	INT	--	NO		Le cuesta mantener la atención en tareas o juegos. Escala 0-3.
inatencion_l2	INT	--	NO		Deja tareas sin terminar o pierde el hilo fácilmente. Escala 0-3.
inatencion_l3	INT	--	NO		Evita tareas que requieren esfuerzo mental sostenido. Escala 0-3.
inatencion_l4	INT	--	NO		Se distrae con facilidad por estímulos externos. Escala 0-3.

inatencion_I5	INT	--	NO	Pierde objetos necesarios para sus actividades. Escala 0-3.
inatencion_I6	INT	--	NO	Parece no escuchar cuando se le habla directamente. Escala 0-3.
hiperactividad_H1	INT	--	NO	Habla en exceso o interrumpe a otros con frecuencia. Escala 0-3.
hiperactividad_H2	INT	--	NO	Actúa sin pensar o responde antes de terminar la pregunta. Escala 0-3.
hiperactividad_H3	INT	--	NO	Le cuesta esperar su turno. Escala 0-3.
hiperactividad_H4	INT	--	NO	Se levanta del asiento cuando debería permanecer sentado. Escala 0-3.
hiperactividad_H5	INT	--	NO	Está en marcha o parece impulsado por un motor. Escala 0-3.
hiperactividad_H6	INT	--	NO	Se entromete en actividades o conversaciones de otros. Escala 0-3.
control_C1	INT	--	NO	Comete errores por responder demasiado rápido. Escala 0-3.
control_C2	INT	--	NO	Tarda mucho en iniciar o completar respuestas sencillas. Escala 0-3.
control_C3	INT	--	NO	Su rendimiento fluctúa mucho de un momento a otro. Escala 0-3.
control_C4	INT	--	NO	Confunde instrucciones o elementos similares. Escala 0-3.
fecha_registro	DATE	--	NO	Fecha en que se aplicó el cuestionario al informante.

Fuente: Elaboración propia.

Entidad: RESULTADO_IA

Almacena la predicción generada por el modelo K-Nearest Neighbors (KNN, $k=3$) para cada evaluación registrada en el sistema.

Tabla 9. Diccionario de datos de la entidad RESULTADO_IA.

RESULTADO_IA					
Campo	Tipo	Longitud	Nulo	Clave	Descripción
Id	INT	--	NO	PK	Identificador único del resultado. Se incrementa automáticamente.
evaluacion_id	INT	--	NO	FK	Referencia a la evaluación analizada. Relacionado con EVALUACION (id).
probabilidad_tdah	FLOAT	--	NO		Probabilidad de TDAH generada por el modelo KNN. Rango: 0.0 a 1.0.
clasificacion_riesgo	VARCHAR	50	NO		Resultado del análisis. Valores: POSIBLE TDAH o SIN INDICADORES.
modelo_utilizado	VARCHAR	20	SI		Nombre del modelo de IA empleado. Valor por defecto: KNN ($k=3$).
fecha_prediccion	DATETIME	--	NO		Fecha y hora exacta en que se generó la predicción por el sistema.

Fuente: Elaboración propia.

4.5.3 Desarrollo de la solución

Herramientas de desarrollo:

- Lenguaje: Python 3.10
- Entrenamiento IA: Scikit-Learn, Imbalanced-Learn (SMOTE), Pandas, Numpy.
- Backend / API: FastAPI
- Frontend / UI: Streamlit.

El código corresponde al archivo `frontend_tesis.py`, desarrollado en Python 3.10 utilizando la biblioteca Streamlit. Este script implementa la interfaz gráfica del sistema experto para la evaluación y detección de posibles casos de TDAH mediante el modelo K-Nearest Neighbors (KNN).

El código fuente completo de la implementación se presenta en el Anexo 9.

Interfaz gráfica del sistema:

La capa de presentación del sistema fue desarrollada utilizando Streamlit, lo que permitió construir una interfaz interactiva y de fácil uso para el proceso de evaluación. La interfaz presenta un diseño organizado en forma de Ficha de Evaluación del Estudiante, en la cual se registran los datos básicos del estudiante y se realiza la evaluación de síntomas asociados al TDAH.

La sección de evaluación se encuentra estructurada en tres dimensiones principales: inatención, hiperactividad y control cognitivo, las cuales corresponden a los principales indicadores conductuales considerados en el cuestionario. Los docentes o profesionales encargados de la evaluación ingresan los valores correspondientes mediante deslizadores interactivos (sliders) con una escala de 0 a 3, donde cada valor representa la frecuencia con la que se observa el comportamiento evaluado.

Una vez completada la información, el usuario puede presionar el botón “Analizar caso con IA”, mediante el cual la interfaz envía los datos al sistema de análisis basado en modelos de aprendizaje automático. Como resultado, el sistema muestra de forma visual el nivel de probabilidad asociado a la presencia de TDAH, presentando un mensaje orientativo acompañado de un indicador porcentual. La figura 11 muestra la interfaz principal del sistema utilizada para el registro de datos y evaluación de síntomas. Por su parte, la figura 12 presenta el resultado generado por el sistema luego del análisis realizado por el modelo de inteligencia artificial.

Figura 11. Interfaz principal del sistema experto para la evaluación de síntomas de TDAH.

Fuente: Elaboración propia.

Figura 12. Visualización del resultado del análisis predictivo generado por el sistema experto para la detección de TDAH.

Fuente: Elaboración propia.

4.5.4 Pruebas

Se realizaron pruebas funcionales al sistema desarrollado con el objetivo de verificar que cada componente cumpliera con los requisitos establecidos. La tabla 10 detalla los casos de prueba.

Tabla 10. Casos de prueba del sistema de detección de TDAH.

Código	Caso de prueba	Datos de entrada	Resultado esperado	Resultado obtenido	Estado
CP01	Ingreso de datos completos	16 puntajes del test de Connors + datos sociodemográficos	Sistema genera predicción correctamente	Predicción generada con probabilidad porcentual	Aprobado
CP02	Ingreso de datos incompletos	Puntajes faltantes	Sistema muestra mensaje de error	Mensaje de error mostrado correctamente	Aprobado
CP03	Generación de predicción positiva	Puntajes elevados en inatención e hiperactividad	Diagnóstico: POSIBLE TDAH	Diagnóstico correcto con probabilidad > 50%	Aprobado
CP04	Generación de predicción negativa	Puntajes bajos en todas las dimensiones	Diagnóstico: SIN INDICADORES	Diagnóstico correcto con probabilidad < 50%	Aprobado
CP05	Tiempo de respuesta	Datos completos ingresados	Respuesta en menos de 3 segundos	Respuesta obtenida en menos de 2 segundos	Aprobado

Fuente: Elaboración propia.

Los resultados obtenidos en las pruebas funcionales confirmaron que el sistema opera correctamente en todos los casos evaluados, respondiendo dentro de los tiempos establecidos y generando predicciones coherentes con los datos ingresados.

4.6 Análisis inferencial

El umbral del 75% de exactitud fue establecido como criterio mínimo de desempeño aceptable en concordancia con investigaciones previas en el campo de la detección del TDAH mediante técnicas de Machine Learning. Chen et al. (2021) y Sánchez (2023) reportaron niveles de exactitud

superiores al 75% en modelos de clasificación aplicados a contextos clínicos y educativos, validando dicho valor como una referencia razonable para evaluar la utilidad predictiva de modelos supervisados orientados a la identificación temprana de indicadores asociados al TDAH.

Asimismo, desde una perspectiva metodológica y estadística, diversos autores sostienen que, en problemas de clasificación binaria con conjuntos de datos pequeños y desbalanceados, un desempeño inferior al 75% refleja una capacidad limitada de generalización y discriminación entre clases, reduciendo la confiabilidad del modelo para fines predictivos (Hastie, Tibshirani y Friedman, 2009). En este sentido, el valor de 0.75 fue utilizado como parámetro de contraste en las pruebas inferenciales del estudio, permitiendo determinar si el desempeño promedio de los algoritmos era significativamente superior al nivel mínimo considerado aceptable.

Finalmente, considerando que el sistema desarrollado fue diseñado como una herramienta de triaje preliminar y apoyo a la detección temprana, mas no como un instrumento de diagnóstico clínico definitivo, el umbral del 75% representa un nivel mínimo de utilidad práctica para respaldar la derivación oportuna de estudiantes a evaluación psicopedagógica especializada en la I.E.P. María Reina.

Dado que el tamaño de muestra para cada modelo fue de 50 observaciones, se evaluó el supuesto de normalidad de la variable Accuracy mediante la prueba de Shapiro-Wilk, adecuada para muestras pequeñas y medianas. Los resultados de las pruebas de normalidad se presentan en la tabla 11.

Supuesto de normalidad

H_0 : Los datos de la variable Accuracy de los modelos de clasificación siguen una distribución normal.

H_1 : Los datos de la variable Accuracy de los modelos de clasificación no siguen una distribución normal.

4.6.1. Nivel de significancia:

Nivel de confianza: 95%

Significancia (α): 0.05

4.6.2. Estimación del p-valor:

Tabla 11. Prueba de Kolmogorov-Smirnov y Shapiro-Wilk para la variable Accuracy.

Modelo	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
KNN	.159	50	.003	.956	50	.062
Random Forest	.278	50	.000	.817	50	.000
SVM	.236	50	.000	.905	50	.001

Fuente: Elaboración propia.

Decisión:

De acuerdo con los resultados de la prueba de Shapiro-Wilk, el modelo KNN presenta un valor de significancia de 0.062, el cual es mayor que el nivel de significancia establecido ($\alpha = 0.05$). Por lo tanto, no se rechaza la hipótesis nula, indicando que los datos de Accuracy para este modelo pueden considerarse normalmente distribuidos.

En contraste, los modelos Random Forest y SVM presentan valores de significancia menores que el nivel de significancia ($p = 0.000$ y $p = 0.001$, respectivamente), por lo que se rechaza la hipótesis nula y se acepta la hipótesis alternativa. Esto indica que los datos de Accuracy para estos modelos no siguen una distribución normal. Por lo tanto, para el modelo KNN se aplicó la prueba paramétrica t de una muestra, mientras que para los modelos Random Forest y SVM se utilizó la prueba no paramétrica de rangos con signo de Wilcoxon.

4.6.3. Prueba de hipótesis específica 1

El modelo K-Nearest Neighbors (KNN) presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por

Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

Prueba t de Student para una muestra

Formulación de hipótesis estadísticas:

$$H_0: \bar{x} \leq 0.75$$

La media de la exactitud (Accuracy) del modelo construido con el algoritmo K-Nearest Neighbors es menor o igual a 0.75.

$$H_1: \bar{x} > 0.75$$

La media de la exactitud (Accuracy) del modelo construido con el algoritmo K-Nearest Neighbors es mayor que 0.75.

Nivel de significancia:

Nivel de Confianza: 95%

Significancia (α): 0.05

Estimación del p-valor:

Para contrastar la hipótesis planteada se aplicó la prueba t de Student para una muestra, con el objetivo de determinar si la media de la exactitud del modelo K-Nearest Neighbors es significativamente mayor que el valor de referencia de 0.75. Los resultados obtenidos se presentan en la siguiente tabla 12.

Tabla 12. Resultados de la prueba t de Student para una muestra aplicada a la exactitud del modelo K-Nearest Neighbors.

Prueba de muestra única						
Valor de prueba = 0.75						
	t	gl	Sig. (bilateral)	Diferencia de medias	95% de intervalo de confianza de la diferencia	
					Inferior	Superior
Accuracy	5.704	49	.000	.06684	.0433	.0904

Fuente: Elaboración propia.

Regla de decisión:

Si el $p - valor \geq 0.05$, se acepta H_0 y se rechaza H_1 .

Si el $p - valor < 0.05$, se rechaza H_0 y se acepta H_1 .

Decisión:

Dado que el p-valor obtenido ($p = 0.000$) es menor que el nivel de significancia establecido ($\alpha = 0.05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

Por lo tanto, existe evidencia estadística suficiente para afirmar que la media de la exactitud del modelo construido con el algoritmo K-Nearest Neighbors es significativamente mayor que 0.75, con un nivel de confianza del 95%.

En consecuencia, se acepta la hipótesis específica 1, la cual establece que: El modelo K-Nearest Neighbors presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

4.6.4. Prueba de hipótesis específica 2

El modelo Random Forest presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

Prueba de rangos con signo de Wilcoxon para una muestra**Formulación de hipótesis estadísticas:**

$$H_0: \tilde{x} \leq 0.75$$

La mediana de la exactitud (Accuracy) del modelo construido con el algoritmo Random Forest es menor o igual a 0.75.

$$H_1: \tilde{x} > 0.75$$

La mediana de la exactitud (Accuracy) del modelo construido con el algoritmo Random Forest es mayor que 0.75.

Nivel de significancia:

Nivel de Confianza: 95%

Significancia (α): 0.05

Estimación del p-valor:

Para contrastar la hipótesis planteada se aplicó la prueba de rangos con signo de Wilcoxon para una muestra, con el objetivo de determinar si la media de la exactitud del modelo Random Forest es significativamente mayor que el valor de referencia de 0.75. Los resultados obtenidos se presentan en la siguiente tabla 13.

Tabla 13. Resultados de la prueba de rangos con signo de Wilcoxon para una muestra aplicada a la exactitud del modelo Random Forest.

Resumen de contrastes de hipótesis			
Hipótesis nula	Prueba	Sig.	Decisión
La mediana de Accuracy es igual a 0.750.	Prueba de Wilcoxon de los rangos con signo para una muestra.	.000	Rechacé la hipótesis nula.

Fuente: Elaboración propia.

Se muestra significaciones asintóticas. El nivel de significancia es .05.

Regla de decisión:

Si el $p - \text{valor} \geq 0.05$, se acepta H_0 y se rechaza H_1 .

Si el $p - \text{valor} < 0.05$, se rechaza H_0 y se acepta H_1 .

Decisión:

Dado que el p-valor obtenido ($p = 0.000$) es menor que el nivel de significancia establecido ($\alpha = 0.05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

Por lo tanto, existe evidencia estadística suficiente para afirmar que la mediana de la exactitud del modelo construido con el algoritmo correspondiente es significativamente mayor que 0.75, con un nivel de confianza del 95%.

En consecuencia, se acepta la hipótesis específica 2 que establece que: El modelo Random Forest presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

4.6.5. Prueba de hipótesis específica 3

El modelo Support Vector Machine (SVM) presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

Prueba de rangos con signo de Wilcoxon para una muestra

Formulación de hipótesis estadísticas:

$$H_0: \tilde{x} \leq 0.75$$

La mediana de la exactitud (Accuracy) del modelo construido con el algoritmo Support Vector Machine (SVM) es menor o igual a 0.75.

$$H_1: \tilde{x} > 0.75$$

La mediana de la exactitud (Accuracy) del modelo construido con el algoritmo Support Vector Machine (SVM) es mayor que 0.75.

Nivel de significancia:

Nivel de Confianza: 95%

Significancia (α): 0.05

Estimación del p-valor:

Para contrastar la hipótesis planteada se aplicó la prueba de rangos con signo de Wilcoxon para una muestra, con el objetivo de determinar si la media de la exactitud del modelo Support Vector Machine (SVM) es significativamente mayor que el valor de referencia de 0.75. Los resultados obtenidos se presentan en la tabla 14.

Tabla 14. Resultados de la prueba de rangos con signo de Wilcoxon para una muestra aplicada a la exactitud del modelo Support Vector Machine (SVM).

Resumen de contrastes de hipótesis			
Hipótesis nula	Prueba	Sig.	Decisión
La mediana de Accuracy 1 es igual a 0.750.	Prueba de Wilconson de los rangos con signo para una muestra.	.000	Rechacé la hipótesis nula.

Fuente: Elaboración propia.

Se muestra significaciones asintóticas. El nivel de significancia es .05.

Regla de decisión:

Si el $p - \text{valor} \geq 0.05$, se acepta H_0 y se rechaza H_1 .

Si el $p - \text{valor} < 0.05$, se rechaza H_0 y se acepta H_1 .

Decisión:

Dado que el p-valor obtenido ($p = 0.000$) es menor que el nivel de significancia establecido ($\alpha = 0.05$), se rechaza la hipótesis nula y se acepta la hipótesis alternativa.

Por lo tanto, existe evidencia estadística suficiente para afirmar que la mediana de la exactitud del modelo construido con el algoritmo correspondiente es significativamente mayor que 0.75, con un nivel de confianza del 95%.

En consecuencia, se acepta la hipótesis específica 3 que establece que: El modelo Support Vector Machine (SVM) presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

4.6.6. Hipótesis general

Los modelos de aprendizaje automático presentan un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por

Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

Decisión

De acuerdo con los resultados obtenidos en las pruebas de hipótesis específicas, los modelos K-Nearest Neighbors, Random Forest y Support Vector Machine presentaron valores de exactitud significativamente superiores al 75%.

En consecuencia, se acepta la hipótesis general, concluyendo que los modelos de aprendizaje automático evaluados presentan un desempeño predictivo superior al umbral establecido en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad.

CONCLUSIONES

Primera: En la presente investigación se aplicaron modelos basados en técnicas de aprendizaje automático con el propósito de apoyar la identificación temprana de posibles casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina de la ciudad de Puerto Maldonado. Asimismo, se desarrolló un sistema para la aplicación y registro del cuestionario basado en la Escala de Conners, lo que permitió recopilar y procesar la información de manera organizada para el análisis de patrones conductuales relacionados con el trastorno. Los resultados mostraron que la aplicación de modelos de Machine Learning constituye una herramienta útil para fortalecer los procesos de detección temprana en el ámbito educativo.

Segunda: Con los datos recolectados y analizados mediante el algoritmo K-Nearest Neighbors se determinó estadísticamente mediante la prueba paramétrica t de Student para una muestra ($p < 0.05$), que el modelo K-Nearest Neighbors presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

Tercera: Con los datos recolectados y analizados mediante el algoritmo Random Forest, se determinó estadísticamente, mediante la prueba no paramétrica de rangos con signo de Wilcoxon para una muestra ($p < 0.05$), que dicho modelo presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

Cuarta: Con los datos recolectados y analizados mediante el algoritmo Support Vector Machine, se determinó estadísticamente, mediante la prueba no paramétrica de rangos con signo de Wilcoxon para una muestra ($p < 0.05$), que dicho modelo presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025.

SUGERENCIAS

Primera: Se recomienda que las instituciones educativas incorporen herramientas tecnológicas basadas en técnicas de aprendizaje automático como apoyo en los procesos de identificación temprana de posibles dificultades relacionadas con el TDAH, con el fin de facilitar intervenciones oportunas que contribuyan al desarrollo académico y personal de los estudiantes.

Segunda: Para futuras investigaciones se sugiere ampliar la cantidad de datos utilizados para el entrenamiento de los modelos, incluyendo información proveniente de otras instituciones educativas. Esto permitiría mejorar el rendimiento de los algoritmos y obtener resultados más representativos y confiables.

Tercera: Se recomienda explorar el uso de otros algoritmos de aprendizaje automático que puedan ofrecer mejores niveles de desempeño en la detección del trastorno, tales como modelos de ensamble o técnicas más avanzadas.

Cuarta: Se sugiere complementar el uso de herramientas tecnológicas con evaluaciones profesionales realizadas por especialistas en neuropsicología, considerando que los sistemas predictivos deben utilizarse únicamente como apoyo en la detección preliminar y no como un diagnóstico clínico definitivo.

Quinta: Finalmente, se recomienda promover programas de capacitación dirigidos a docentes y personal educativo sobre la importancia de la detección temprana del TDAH y el uso adecuado de herramientas tecnológicas que contribuyan a mejorar los procesos de seguimiento y atención a los estudiantes.

REFERENCIAS BIBLIOGRÁFICAS

AGUIRRE, M. et al., 2022. *Trastorno por déficit de atención e hiperactividad y su relación diagnóstica con el Trastorno en línea por estrés postraumático infantil: Una revisión sistemática* [en línea]. Revista de Psicología Clínica con Niños y Adolescentes, vol. 9, núm. 1, págs. 9-19 [consulta: 6 de enero de 2025]. DOI: 10.21134/rpcna.2022.09.1.1. Disponible en: <https://doi.org/10.21134/rpcna.2022.09.1.1>.

AMERICAN PSYCHIATRIC ASSOCIATION, 2022. Diagnostic and Statistical Manual of Mental Disorders [en línea]. 5.^a ed. Washington: APA [consulta: 6 de enero de 2025]. DOI: 10.1176/appi.books.9780890425787. Disponible en: <https://doi.org/10.1176/appi.books.9780890425787>.

BARKLEY, R.A., 2014. Attention-deficit/hyperactivity disorder [en línea]. The Lancet, vol. 383, núm. 9922. DOI: 10.1016/S0140-6736(13)61684-0. Disponible en: [https://doi.org/10.1016/S0140-6736\(13\)61684-0](https://doi.org/10.1016/S0140-6736(13)61684-0).

BREIMAN, L. et al., 2017. Classification and regression trees. Classification and Regression Trees [en línea], [consulta: 18 mayo 2025]. DOI 10.1201/9781315139470/CLASSIFICATION-REGRESSION-TREES-LEO-BREIMAN-JEROME-FRIEDMAN-OLSHEN-CHARLES-STONE/RIGHTS-AND-PERMISSIONS. Disponible en: <https://www.taylorfrancis.com/books/mono/10.1201/9781315139470/classification-regression-trees-leo-breiman-jerome-friedman-olshen-charles-stone>.

CASAS CASTILLO, R. y CELIS GUTIÉRREZ, G., 2023. Aplicación móvil con gamificación para reforzar el aprendizaje de niños con TDAH mediante el método TCT [en línea]. Tesis de grado, Universidad Peruana de Ciencias Aplicadas [consulta: 6 de enero de 2025]. Disponible en: <https://repositorioacademico.upc.edu.pe/handle/10757/668696>.

CASTRO, R. et al., 2018. Detecta kids – estudo de identificação precoce de distúrbios cerebrais relacionados a déficit de atenção e hiperatividade em niños [en línea]. Programa De Iniciação Científica - Pic/Uniceub - Relatórios

De Pesquisa, núm. 3 [consulta: 6 enero 2025]. DOI: 10.5102/pic.n3.2017.5837.

CHEN, Y.-S. et al., 2021. Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Machine Learning. *Applied Artificial Intelligence* [en línea], vol. 35, DOI: 10.1080/08839514.2021.1927722. Disponible en: <https://doi.org/10.1080/08839514.2021.1927722>.

CONNERS, C.K., 2008. Conners 3rd Edition (Conners 3): Technical Manual [en línea]. Toronto: Multi-Health Systems [consulta: 6 de enero de 2025]. Disponible en: <https://www.mhs.com/MHS-Assessment?prodname=conners3>.

CORMEN, T.H. et al., 2022. Introduction to Algorithms [en línea]. 4.^a ed. Cambridge: MIT Press [consulta: 6 de enero de 2025]. Disponible en: <https://mitpress.mit.edu/9780262046305/introduction-to-algorithms/>.

CORTES, C. y VAPNIK, V., 1995. Support-vector networks [en línea]. *Machine Learning*, vol. 20, núm. 3, págs. 273-297 [consulta: 6 de enero de 2025]. DOI: 10.1007/BF00994018. Disponible en: <https://doi.org/10.1007/BF00994018>.

COVER, T. y HART, P., 1967. Nearest neighbor pattern classification [en línea]. *IEEE Transactions on Information Theory*, vol. 13, núm. 1, págs. 21-27 [consulta: 6 de enero de 2025]. DOI: 10.1109/TIT.1967.1053964. Disponible en: <https://doi.org/10.1109/TIT.1967.1053964>.

DAS, W. y KHANNA, S., 2021. A robust machine learning framework for automatic detection of ADHD using pupillometric biomarkers and time series analysis [en línea]. *Scientific Reports*, vol. 11, art. 16370 [consulta: 28 de febrero de 2025]. DOI: 10.1038/s41598-021-95673-5. Disponible en: <https://doi.org/10.1038/s41598-021-95673-5>.

DUEÑAS GUEVARA, M. y MENDOZA MARÍN, A., 2023. Solución tecnológica para niños con TDAH mediante el uso de realidad aumentada en juegos serios [en línea]. Tesis de grado, Universidad Peruana de Ciencias Aplicadas [consulta: 6 de enero de 2025]. Disponible en: <http://hdl.handle.net/10757/668697>.

GRAO-CRUCES, A. et al., 2017. Conners' Rating Scales and their applicability in educational and clinical contexts [en línea]. *Psicothema*, vol. 29, núm. 3, págs. 318-325 [consulta: 6 de enero de 2025]. Disponible en: <https://www.psicothema.com/pdf/4397.pdf>.

HAN, J.Y. et al., 2015. The effects of prenatal exposure to alcohol and environmental tobacco smoke on risk for ADHD: A large population-based study [en línea]. *Psychiatry Research*, vol. 225, núm. 1-2 [consulta: 6 de enero de 2025]. DOI: 10.1016/J.PSYCHRES.2014.11.009. Disponible en: <https://doi.org/10.1016/j.psychres.2014.11.009>.

HASTIE, T., TIBSHIRANI, R. y FRIEDMAN, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* [en línea]. 2.^a ed. New York: Springer [consulta: 6 de enero de 2025]. Disponible en: <https://hastie.su.domains/ElemStatLearn>.

HERNÁNDEZ-NIETO, R.A., 2002. *Contributions to Statistical Analysis*. Mérida: Universidad de Los Andes.

HERNÁNDEZ-SAMPIERI, R. et al., 2014. *Metodología de la investigación* [en línea]. 6.^a ed. México: McGraw-Hill / Interamericana Editores [consulta: 20 de abril de 2025]. Disponible en: <https://drive.google.com/file/d/1Fjufmi0oGY4Zs8EajFiAJYNT2qoecH4k/view>.

LUNA, N. et al., 2023. Percepción de los maestros de educación general básica sobre el trastorno por déficit de atención e hiperactividad en niños [en línea]. *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, núm. 2, págs. 5869-5881 [consulta: 6 de enero de 2025]. DOI: 10.37811/cl_rcm.v7i2.5758. Disponible en: https://doi.org/10.37811/cl_rcm.v7i2.5758.

MANIRUZZAMAN, M. et al., 2023. Optimal channels and features selection based ADHD detection from EEG signal using statistical and machine learning techniques [en línea]. *IEEE Access*, vol. 11, págs. 33570-33583 [consulta: 28 de febrero de 2025]. DOI: 10.1109/ACCESS.2023.3264236. Disponible en: <https://doi.org/10.1109/ACCESS.2023.3264236>.

MESINO, R. et al., 2023. Modelo pedagógico inclusivo para la enseñanza aprendizaje de la matemática a través de la resolución de problemas en niños de grado quinto con TDAH [en línea]. Revista De Gestão E Secretariado, vol. 14, núm. 8, págs. 13561-13588 [consulta: 6 de enero de 2025]. DOI: 10.7769/gesec. v14i8.2488. Disponible en: <https://doi.org/10.7769/gesec.v14i8.2488>.

MINISTERIO DE SALUD, 2020. Más de cinco mil menores con Trastorno por Déficit de Atención e Hiperactividad (TDAH) recibieron servicios de salud integral [en línea]. Lima: Ministerio de Salud, 13 de julio [consulta: 6 de enero de 2025]. Disponible en: <https://www.gob.pe/institucion/minsa/noticias/208627-mas-de-cinco-mil-menores-con-trastorno-por-deficit-de-atencion-e-hipe>.

MITCHELL, T.M., 1997. Machine Learning [en línea]. New York: McGraw-Hill [consulta: 17 de abril de 2025]. Disponible en: <http://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf>.

ORGANIZACIÓN MUNDIAL DE LA SALUD, 2023. Pharmacological and nonpharmacological interventions for children with attention-deficit hyperactivity disorder (ADHD) [en línea]. Ginebra: OMS [consulta: 20 de mayo de 2025]. Disponible en: [https://www.who.int/teams/mental-health-and-substance-use/treatment-care/mental-health-gap-action-programme/evidence-centre/child-and-adolescent-mental-disorders/pharmacological-and-nonpharmacological-interventions-for-children-with-attention-deficit-hyperactivity-disorder-\(adhd\)](https://www.who.int/teams/mental-health-and-substance-use/treatment-care/mental-health-gap-action-programme/evidence-centre/child-and-adolescent-mental-disorders/pharmacological-and-nonpharmacological-interventions-for-children-with-attention-deficit-hyperactivity-disorder-(adhd)).

PARKE, E.M. et al., 2021. Cognición social en niños con TDAH [en línea]. Revista de trastornos de la atención, vol. 25, núm. 4, págs. 519-529 [consulta: 6 de enero de 2025]. DOI: 10.1177/1087054718816157. Disponible en: <https://doi.org/10.1177/1087054718816157>.

RAMOS, D. et al., 2022. El TDAH como reto para la educación inclusiva en las universidades [en línea]. Psicoespacios, vol. 16, núm. 28, págs. 1-12 [consulta: 6 de enero de 2025]. DOI: 10.25057/21452776.1459.

RETAMOZO, B., 2023. Implementación de las TIC en quechua para los estudiantes quechua hablantes del Perú rural [en línea]. Tzhoecoen, vol. 15, núm. 2, págs. 55-67 [consulta: 6 de enero de 2025]. DOI: 10.26495/tzh.v15i2.2590. Disponible en: <https://doi.org/10.26495/tzh.v15i2.2590>.

RICKETTS, E. et al., 2020. Deterioro académico, interpersonal, recreativo y familiar en niños con síndrome de Tourette y trastorno por déficit de atención e hiperactividad [en línea]. Psiquiatría infantil y desarrollo humano, vol. 53, núm. 1, págs. 3-15 [consulta: 6 de enero de 2025]. DOI: 10.1007/s10578-020-01111-4. Disponible en: <https://doi.org/10.1007/s10578-020-01111-4>.

RUSCA-JORDÁN, F. y CORTEZ-VERGARA, C., 2020. Trastorno por déficit de atención con hiperactividad (TDAH) en niños y adolescentes. Una revisión clínica [en línea]. Revista de Neuropsiquiatría, vol. 83, núm. 3, págs. 148-156 [consulta: 2 de abril de 2025]. DOI: 10.20453/rnp.v83i3.3794. Disponible en: http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S0034-85972020000300148.

RUSSELL, S. et al., 2021. Intelligence artificielle: une approche moderne (4e édition) [en línea]. [consulta: 18 de mayo de 2025]. DOI: 10.34894/VQ1DJA. Disponible en: <https://hal.science/hal-04245057>.

SÁNCHEZ, A., 2023. Desarrollo de un modelo de aprendizaje automático para la mejora de la detección temprana de casos de TDAH en niños y adolescentes [en línea]. Tesis de Doctorado, Universitat Politècnica de València [consulta: 6 de enero de 2025]. Disponible en: <https://riunet.upv.es/bitstream/handle/10251/194403/Sanchez%20%20Desarrollo%20de%20un%20modelo%20de%20aprendizaje%20automatico%20para%20la%20mejora%20de%20la%20deteccion%20tempra....pdf?secuencia=1&isAlloed=y>.

SANTARROSA-LÓPEZ, I. et al., 2023. Análisis comparativo de técnicas de aprendizaje automático y aprendizaje profundo para la detección de TDAH [en línea]. Investigación en Ciencias de la Computación, vol. 152, núm. 8 [consulta: 6 enero 2025]. Disponible en: https://rcs.cic.ipn.mx/2023_152_8/Analisis%20comparativo%20de%20tecnic

as%20de%20aprendizaje%20automatico%20y%20aprendizaje%20profundo.pdf.

SANTARROSA-LÓPEZ, I. et al., 2024. DETEC-ADHD: A Data-Driven Web App for Early ADHD Detection Using Machine Learning and Electroencephalography. *Big Data and Cognitive Computing* 2025, Vol. 9, Page 3 [en línea], vol. 9, no. 1, [consulta: 17 abril 2025]. ISSN 2504-2289. DOI 10.3390/BDCC9010003. Disponible en: <https://www.mdpi.com/2504-2289/9/1/3/htm>.

VIZCAÍNO ZÚÑIGA, P.I. et al., 2023. Metodología de la investigación científica: guía práctica [en línea]. *Ciencia Latina Revista Científica Multidisciplinar*, vol. 7, núm. 4, págs. 9723-9762 [consulta: 6 de enero de 2025]. DOI: 10.37811/cl_rcm.v7i4.7658. Disponible en: https://doi.org/10.37811/cl_rcm.v7i4.7658.

WEISS, F. et al., 2023. Triple diagnóstico del trastorno por déficit de atención hiperactividad con coexistencia de trastorno bipolar y trastorno por consumo de alcohol: aspectos clínicos y tratamientos farmacológicos [en línea]. *Neurofarmacología actual*, vol. 21, núm. 7, págs. 1467-1476 [consulta: 6 de enero de 2025]. DOI: 10.2174/1570159X20666220830154002. Disponible en: <https://doi.org/10.2174/1570159X2066622083015400>

ANEXOS

Anexo 1. Matriz de operacionalización de variables.

Variables	Definición Conceptual	Definición Operacional	Dimensiones	Indicadores
Variable Independiente: Técnicas de Machine Learning	Las técnicas de Machine Learning son métodos algorítmicos que permiten a un sistema aprender patrones a partir de datos, mejorando su desempeño en una tarea específica sin necesidad de programación explícita (Mitchell, 1997).	Esta variable fue evaluada mediante la implementación de tres algoritmos de aprendizaje automático: K-Nearest Neighbors (KNN), Random Forest y Support Vector Machine (SVM). Cada modelo fue entrenado utilizando los datos obtenidos del cuestionario aplicado a los estudiantes y posteriormente se comparó su desempeño en la clasificación de posibles casos de TDAH. El rendimiento de los modelos fue medido mediante métricas de evaluación utilizadas en problemas de clasificación.	K-Nearest Neighbors (KNN)	Accuracy, Precisión, Recall, F1-Score, ROC-AUC
			Random Forest	Accuracy, Precisión, Recall, F1-Score, ROC-AUC
			Support Vector Machine (SVM)	Accuracy, Precisión, Recall, F1-Score, ROC-AUC
Variable Dependiente: Detección del Trastorno Por Déficit de Atención e Hiperactividad	El TDAH, Trastorno por Déficit de Atención/Hiperactividad, se define en el DSM-5, como un trastorno del neurodesarrollo caracterizado por un patrón persistente de déficit de atención y/o hiperactividad que es desadaptativo para el paciente, está presente antes de los 12 años e interfiere en sus actividades diarias. En el DSM-5 se informa de una prevalencia del 5% en niños/as en la mayoría de las culturas y de un 2.5% en adultos (Aguirre et al., 2022)	Esta variable fue medida mediante la aplicación de la Escala de Conners dirigida a padres o tutores, instrumento utilizado para identificar comportamientos asociados al TDAH en niños. Las respuestas obtenidas permitieron determinar la presencia de indicadores conductuales relacionados con el trastorno y fueron empleadas como datos de entrada para el modelo de clasificación.	Patrones conductuales asociados al TDAH	Nivel de atención Hiperactividad Impulsividad

Anexo 2. Matriz de consistencia.

Título: Detección del trastorno por déficit de atención e hiperactividad mediante Machine Learning en estudiantes de la Institución María Reina, Puerto Maldonado 2025					
Autores: YABAR YLLA, Fabiola VARGAS FIGUEROA, Álvaro Jesús					
PROBLEMA	OBJETIVOS	HIPÓTESIS	VARIABLES E INDICADORES		METODOLOGÍA
Problema General	Objetivo General	Hipótesis general	Variable (dependiente): Detección del Trastorno Por Déficit de Atención e Hiperactividad		
Problema general ¿Cuál es el desempeño predictivo de los modelos de Machine Learning en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado 2025? Problemas específicos: - ¿Cuál es el desempeño predictivo del modelo K-Nearest Neighbors (KNN) en la detección de casos del trastorno por déficit de atención e hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado 2025? - ¿Cuál es el desempeño predictivo del modelo Random Forest en la detección de casos del trastorno por déficit de atención e hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado 2025? - ¿Cuál es el desempeño predictivo del modelo Support Vector Machine (SVM) en la detección de casos del trastorno por déficit de atención e hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado 2025?	Objetivo general: Evaluar el desempeño predictivo de técnicas de Machine Learning en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la institución María Reina, Puerto Maldonado 2025. Objetivos específicos: -Evaluar el desempeño del modelo K-Nearest Neighbors (KNN) en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado 2025. -Evaluar el desempeño del modelo Random Forest en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado 2025. -Evaluar el desempeño del modelo Support Vector Machine (SVM) en la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa Privada María Reina, Puerto Maldonado 2025.	Hipótesis general: El desempeño predictivo de técnicas de Machine Learning permite la detección de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la institución María Reina, Puerto Maldonado 2025. Hipótesis específicas: - El modelo K-Nearest Neighbors (KNN) presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025. - El modelo Random Forest presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025. - El modelo Support Vector Machine (SVM) presenta un desempeño predictivo superior al 75% en la clasificación de casos de Trastorno por Déficit de Atención e	Dimensiones	Indicadores	Enfoque: cuantitativo Diseño: No-Experimental Tipo: Aplicada Técnicas instrumentales: Test: TEST DE CONNERS ABREVIADO De muestreo: Probabilístico con asignación proporcional De recolección de datos: Test De Connors Abreviado Técnicas De Machine Learning De procesamiento de Datos: análisis inferencial Análisis descriptivo e inferencial De análisis Población: 300 estudiantes Institución Educativa Privada MARIA REINA
			Patrones conductuales asociados al TDAH	Nivel de atención, Hiperactividad, Impulsividad	
			Variable (independiente): Técnicas de Machine Learning		
			Dimensiones	Indicadores	
			K-Nearest Neighbors (KNN)	Accuracy, Precisión, Recall, F1-Score, ROC-AUC	
			Random Forest	Accuracy, Precisión, Recall, F1-Score, ROC-AUC	
Support Vector Machine (SVM)	Accuracy, Precisión, Recall, F1-Score, ROC-AUC				

		Hiperactividad en estudiantes de la Institución Educativa María Reina, Puerto Maldonado, 2025. Justificación La identificación del TDAH en la población estudiantil de la I.E.P. María Reina, ubicada en Puerto Maldonado, se vio seriamente restringida por la brecha de capacitación docente y la carencia de logística especializada. Como resultado, el alto índice de subdiagnóstico repercute de manera perjudicial en el éxito académico y la estabilidad emocional del alumnado. Para revertir esta situación, se postuló el desarrollo de un sistema predictivo fundamentado en Machine Learning. La integración de esta tecnología pretendió reducir la subjetividad inherente a los métodos tradicionales, favoreciendo procesos de detección más precisos que faciliten la derivación oportuna a evaluación especializada. Adicionalmente, la arquitectura replicable de este modelo no solo promovió la innovación pedagógica, sino que también aportó una base metodológica para el desarrollo de herramientas tecnológicas orientadas a la detección temprana de posibles alteraciones del neurodesarrollo en contextos educativos.			Muestra: 169 estudiantes
--	--	--	--	--	--

Anexo 3. Instrumento de recolección de datos.

CUESTIONARIO TIPO CONNERS (ADAPTADO) – ÚNICO INFORMANTE

Edad: ____ años
 Sexo: M / F (círculo) Fecha: ____ / ____ / ____
 Informante: Padre o Madre / Docente (círculo) Contexto: Casa / Escuela (círculo)

INSTRUCCIONES:

Responda pensando en las últimas 6–8 semanas. Marque una opción por ítem.
 Escala: 0 = Nunca, 1 = A veces, 2 = Frecuentemente, 3 = Muy frecuentemente.

ÍTEM	☐ 0	☐ 1	☐ 2	☐ 3
INATENCIÓN				
I1: Le cuesta mantener la atención en tareas o juegos	☐ 0	☐ 1	☐ 2	☐ 3
I2: Deja tareas sin terminar o pierde el hilo fácilmente	☐ 0	☐ 1	☐ 2	☐ 3
I3: Evita tareas que requieren esfuerzo mental sostenido	☐ 0	☐ 1	☐ 2	☐ 3
I4: Se distrae con facilidad por estímulos externos	☐ 0	☐ 1	☐ 2	☐ 3
I5: Pierde cosas necesarias para actividades (útiles, libros)	☐ 0	☐ 1	☐ 2	☐ 3
I6: Parece no escuchar cuando se le habla directamente	☐ 0	☐ 1	☐ 2	☐ 3
IMPULSIVIDAD/HIPERACTIVIDAD				
H1: Habla en exceso o interrumpe a otros con frecuencia	☐ 0	☐ 1	☐ 2	☐ 3
H2: Actúa sin pensar o responde antes de terminar la pregunta	☐ 0	☐ 1	☐ 2	☐ 3
H3: Le cuesta esperar su turno	☐ 0	☐ 1	☐ 2	☐ 3
H4: Se levanta de su asiento cuando debería permanecer sentado	☐ 0	☐ 1	☐ 2	☐ 3

H5: Está "en marcha" o parece impulsado por un motor	☐ 0	☐ 1	☐ 2	☐ 3
--	----------------------------	----------------------------	----------------------------	----------------------------

H6: Se entromete en actividades o conversaciones de otros	☐ 0	☐ 1	☐ 2	☐ 3
---	----------------------------	----------------------------	----------------------------	----------------------------

CONTROL/VARIABILIDAD

C1: Comete errores por responder demasiado rápido	☐ 0	☐ 1	☐ 2	☐ 3
---	----------------------------	----------------------------	----------------------------	----------------------------

C2: Tarda mucho en iniciar o completar respuestas sencillas	☐ 0	☐ 1	☐ 2	☐ 3
---	----------------------------	----------------------------	----------------------------	----------------------------

C3: Su rendimiento fluctúa mucho de un momento a otro	☐ 0	☐ 1	☐ 2	☐ 3
---	----------------------------	----------------------------	----------------------------	----------------------------

C4: Confunde instrucciones o elementos similares	☐ 0	☐ 1	☐ 2	☐ 3
--	----------------------------	----------------------------	----------------------------	----------------------------

TIENE TDAH: SI / NO (círculo)

OBSERVACIONES:

Firma: _____

Instrumento adaptado inspirado en Conners. Uso académico/investigación; no sustituye evaluación clínica.

Anexo 4. Solicitud de autorización para realización de estudio.



UNIVERSIDAD NACIONAL AMAZÓNICA DE MADRE DE DIOS
INGENIERIA DE SISTEMAS E INFORMÁTICA

"Año De La Recuperación Y Consolidación De La Economía Peruana"
"Madre de Dios Capital de la Biodiversidad del Perú"



CARTA N° 003-2025-UNAMAD/EPISI/FYY-AJVF

PARA : PROF. GUILLERMINA EMILIA YATTO DE RIVERO
DIRECTORA DE LA INSTITUCIÓN EDUCATIVA PRIVADA MARIA REINA

DE : Bachiller. VARGAS FIGUEROA, Alvaro Jesus
Bachiller. YABAR YLLA, Fabiola.

ASUNTO : Solicitud de Permiso para Recolectar Datos.

REF. :
a) Resolución de decanatura de la facultad de ingeniería N°169-2025-UNAMAD-DFI
b) Copia de Muestra del cuestionario TIPO CONNERS (ADAPTADO) – UNICO INFORMANTE

FECHA : Puerto Maldonado, 20 de agosto del 2025

Estimada directora,

Nos dirigimos a usted con el debido respeto para solicitar su autorización para realizar entrevistas con los docentes de la Institución Educativa Privada Maria Reina, como parte de nuestro proyecto de investigación titulado "Detección del trastorno por déficit de atención e hiperactividad mediante técnicas de Machine Learning en estudiantes de la institución Maria Reina, Puerto Maldonado 2025." Este estudio, desarrollado en el marco de nuestra tesis para optar por el título de Ingeniero en Sistemas e Informática en la Universidad Nacional Amazónica de Madre de Dios, tiene como objetivo desarrollar un modelo predictivo basado en técnicas de Machine Learning para identificar tempranamente casos de Trastorno por Déficit de Atención e Hiperactividad (TDAH) en estudiantes, mejorando las estrategias pedagógicas y el bienestar educativo.

Las entrevistas se centrarán en recopilar información sobre el comportamiento de los estudiantes, específicamente en aspectos relacionados con atención, hiperactividad e impulsividad, utilizando el Conners' Rating Scale y observaciones conductuales estandarizadas. Estas entrevistas se realizarán en las instalaciones del colegio durante los días 21 – 31 de agosto de 2025, en horarios que se coordinarán previamente para no interferir con las responsabilidades docentes ni con el funcionamiento del colegio. Nos comprometemos a garantizar que:

- La participación de los docentes será completamente voluntaria, con la posibilidad de retirarse en cualquier momento sin ninguna consecuencia.
- Toda la información recopilada será tratada con estricta confidencialidad, asegurando que ni los docentes, ni los estudiantes, ni el colegio sean identificables en los resultados del estudio. Los datos serán codificados y almacenados de manera segura, cumpliendo con las normativas


Fabiola Yabar Ylla.
79382237.


Alvaro Jesus Vargas Figueroa
73621286

Anexo 5. Solicitud de validación de instrumento.

Puerto Maldonado, 10 de agosto de 2025

Señor:
Espinoza Toribio, José Carlos
Director del Centro Psicológico S.O.S.



RECIBIDO
Fecha 12 AGO 2025

Presente. –

De mi consideración:

Yo, Fabiola Yabar Ylla, estudiante de la Universidad Nacional Amazónica de Madre de Dios, me dirijo a usted con el debido respeto para solicitar la validación de contenido del instrumento que será utilizado en mi investigación titulada:

"DETECCIÓN DEL TRASTORNO POR DÉFICIT DE ATENCIÓN E HIPERACTIVIDAD MEDIANTE TÉCNICAS DE MACHINE LEARNING EN ESTUDIANTES DE LA INSTITUCIÓN MARIA REINA, PUERTO MALDONADO 2025"

En este estudio se empleará la Escala de Conners, adaptada para su aplicación en población escolar, con el propósito de evaluar características conductuales y atencionales en estudiantes de educación primaria.

Dado que usted cuenta con experiencia profesional en Psicología Clínica y posee el grado de Maestría en Psicología Clínica, considero que su criterio experto es fundamental para garantizar la pertinencia, claridad, coherencia y validez del instrumento empleado en la investigación.

Adjunto a la presente:

- El instrumento de evaluación,
- La matriz de validación,
- Y la ficha correspondiente para su revisión.

Por lo expuesto, agradeceré pueda brindarme su valiosa evaluación y emitir su juicio de experto respecto al instrumento, lo cual permitirá sustentar académicamente la validez del mismo para su aplicación en el estudio.

Sin otro particular, quedo atenta a su respuesta, reiterándole mi agradecimiento por su colaboración.

Atentamente,



Fabiola Yabar Ylla
Investigadora

Universidad Nacional Amazónica de Madre de Dios
Correo: fyabary@unamad.edu.pe

Anexo 6. Ficha de validación.

VALIDACIÓN DEL INSTRUMENTO**I. DATOS GENERALES**

- 1.1 Apellidos y Nombres del Experto: Espinoza Toribio José Carlos
- 1.2 Grado académico: Maestría en Psicología Clínica
- 1.3 Áreas de experiencia profesional: Psicología Clínica
- 1.4 Cargo e Institución donde labora: Director del Centro Psicologico S.O.S
- 1.5 Nombre del instrumento motivo de Evaluación: ESCALA DE CONNERS
- 1.6 Autor del Instrumento: VARGAS FIGUEROA, Alvaro Jesus y YABAR YLLA, Fabiola.

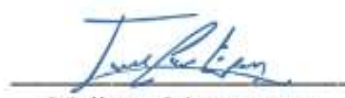
II. OPINIÓN DE APLICABILIDAD:

La Escala es fácil de usar, clara y ayuda a obtener información útil sobre la conducta y la atención. Funciona bien en el trabajo clínico, por lo que se considera adecuada para su aplicación.

III. PROMEDIO DE VALORACIÓN:

INDICADORES	CRITERIOS	Deficiente 0 –20 %	Regular 21– 40 %	Bueno 41– 60 %	Muy bueno 61–80 %	Excelente 81–100 %
1. CLARIDAD	Esta formulada con lenguaje apropiado					X
2. OBJETIVIDAD	Esta expresado en conductas observables				X	
3. ACTUALIDAD						X
4. ORGANIZACIÓN	Existe una organización lógica					X
5. SUFICIENCIA	Comprende los aspectos de cantidad y calidad				X	
6. INTENCIONALIDAD	Adecuado para valorar aspectos del sistema de evaluación y desarrollo de capacidades cognitivas					X
7. CONSISTENCIA	Basados en aspectos Teóricos – científicos de la Tecnología Educativa					X
8. COHERENCIA	Entre los índices, indicadores y las dimensiones					X
9. METODOLOGIA	La estrategia responde al propósito del diagnóstico.				X	

Puerto Maldonado, 15 de agosto de 2025


 Psicólogo - Psicoterapeuta
 Jose Carlos Espinoza Toribio

Anexo 7. Consentimiento informado.



UNIVERSIDAD NACIONAL AMAZÓNICA DE MADRE DE DIOS
INGENIERIA DE SISTEMAS E INFORMÁTICA

"Año De La Recuperación Y Consolidación De La
Economía Peruana" "Madre de Dios Capital de la
Biodiversidad del Perú"



FORMULARIO DE CONSENTIMIENTO INFORMADO PARA PADRES O TUTORES

Institución: Universidad Nacional Amazónica de Madre de Dios

Proyecto de Investigación: Detección del trastorno por déficit de atención e hiperactividad mediante técnicas de Machine Learning en estudiantes de la institución Maria Reina, Puerto Maldonado 2025

Investigadores: Alvaro Jesús Vargas Figueroa y Fabiola Yabar Ylla

Contacto: fyabary@unamad.edu.pe, Teléfono, 997404667 o 921868187

Fecha: 20 de agosto de 2025

ESTIMADO(A) PADRE, MADRE O TUTOR LEGAL,

Nos dirigimos a usted para solicitar su consentimiento informado para que su hijo(a), estudiante de la Institución Educativa Privada Maria Reina, participe en el proyecto de investigación titulado "Detección del trastorno por déficit de atención e hiperactividad mediante técnicas de Machine Learning en estudiantes de la institución Maria Reina, Puerto Maldonado 2025." Este estudio, desarrollado como parte de nuestra tesis para optar por el título de Ingeniero en Sistemas e Informática en la Universidad Nacional Amazónica de Madre de Dios, tiene como objetivo desarrollar un modelo predictivo basado en técnicas de Machine Learning para identificar tempranamente posibles casos de Trastorno por Déficit de Atención e Hiperactividad (TDAH) en estudiantes, contribuyendo a mejorar las estrategias pedagógicas y el bienestar educativo.

Descripción del Estudio

El estudio recopilará datos sobre el comportamiento de los estudiantes, específicamente en áreas como atención, hiperactividad e impulsividad, a través de entrevistas con docentes y la aplicación de instrumentos estandarizados, como la Escala de Conners para padres y maestros. La participación de su hijo(a) implica que los docentes proporcionen observaciones sobre su comportamiento en el aula, sin requerir la intervención directa del estudiante. En algunos casos, se podrá solicitar a los padres completar cuestionarios similares para complementar la información.

Procedimientos

- **Método de recolección de datos:** Entrevistas estructuradas con docentes y, opcionalmente, cuestionarios para padres, realizados en las instalaciones de la Institución Educativa Privada Maria Reina, Puerto Maldonado, durante los días 21-31 de agosto de 2025.
- **Duración:** Las entrevistas con docentes tomarán aproximadamente 20-30 minutos por sesión, y los cuestionarios para padres, si aplica, tomarán unos 15 minutos.
- **Lugar:** Las actividades se realizarán en el colegio, en horarios que no interfieran con las clases.



UNIVERSIDAD NACIONAL AMAZÓNICA DE MADRE DE DIOS
INGENIERIA DE SISTEMAS E INFORMÁTICA

*"Año De La Recuperación Y Consolidación De La
Economía Peruana" "Madre de Dios Capital de la
Biodiversidad del Perú"*



Confidencialidad

- Toda la información recopilada será tratada con estricta confidencialidad. Los datos de su hijo(a) serán codificados con un identificador anónimo para garantizar que ni su nombre, ni el del colegio, ni el de los docentes sean identificables en los resultados del estudio.
- Los datos se almacenarán de manera segura y se utilizarán exclusivamente para fines académicos, conforme a las normativas éticas de investigación aprobadas por el Comité de Ética de la Universidad Nacional Amazónica de Madre de Dios.
- Los resultados se presentarán de forma agregada, sin identificar a ningún estudiante, docente o familia individualmente.

Participación Voluntaria

- La participación en este estudio es completamente voluntaria. Usted puede decidir que su hijo(a) no participe, o retirar su consentimiento en cualquier momento, sin ninguna consecuencia para su hijo(a) o su relación con el colegio.
- No se realizarán evaluaciones directas ni intervenciones con los estudiantes; la información será proporcionada únicamente por los docentes y, en caso de que usted acepte, por usted como padre/madre/tutor.

Beneficios y Riesgos

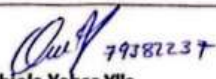
- **Beneficios:** Este estudio busca generar herramientas que apoyen a los docentes en la identificación temprana de TDAH, lo que puede mejorar las estrategias educativas y el apoyo a los estudiantes. Los resultados podrían beneficiar a la comunidad educativa al promover un entorno más inclusivo.
- **Riesgos:** No se anticipan riesgos significativos, ya que la recolección de datos se basa en observaciones docentes y cuestionarios opcionales. Nos comprometemos a proteger la privacidad de todos los participantes.

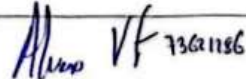
Derechos del Participante

- Usted tiene derecho a recibir una copia de este formulario, a hacer preguntas sobre el estudio en cualquier momento, y a recibir información sobre los resultados generales una vez concluido el proyecto.
- Para cualquier duda o consulta, puede contactar a los investigadores descritos y detallados líneas arriba.

Agradecemos su apoyo y confianza en este proyecto. Si tiene alguna pregunta, no dude en contactarnos.

Atentamente,


Fabiola Yabar Ylla
 Bachiller en Ingeniería de Sistemas e Informática


Alvaro Jesus Vargas Figueroa
 Bachiller en Ingeniería de Sistemas e Informática



UNIVERSIDAD NACIONAL AMAZÓNICA DE MADRE DE DIOS
INGENIERIA DE SISTEMAS E INFORMÁTICA

*"Año De La Recuperación Y Consolidación De La
Economía Peruana" "Madre de Dios Capital de la
Biodiversidad del Perú"*



Declaración de Consentimiento

Yo, _____, he leído y comprendido la información proporcionada sobre el estudio "Detección del trastorno por déficit de atención e hiperactividad mediante técnicas de Machine Learning en estudiantes de la institución Maria Reina, Puerto Maldonado 2025." Entiendo que la participación de mi hijo(a), _____, es voluntaria y que puedo retirar este consentimiento en cualquier momento sin consecuencia alguna. Autorizo la recolección de datos sobre el comportamiento de mi hijo(a) a través de entrevistas con docentes y, si aplica, mi participación mediante cuestionarios, bajo las condiciones de confidencialidad descritas.

Nombre del Padre/Madre/Tutor: _____

Firma: _____

Fecha: _____

Nombre del Estudiante: _____

Agradecemos su apoyo y confianza en este proyecto. Si tiene alguna pregunta, no dude en contactarnos.

[illegible]

Anexo 9. Código fuente del sistema de detección de TDAH.

```
# frontend_tesis.py

import os

import streamlit as st

import requests

st.set_page_config(

    page_title="Tesis TDAH - Maria Reina 2025",

    page_icon="🎓",

    layout="wide"

)

with st.sidebar:

    st.header("🎓 Datos del Proyecto")

    st.info("""

    **Institución Educativa:**

    María Reina

    **Ubicación:**

    Puerto Maldonado, Madre de Dios

    **Año:**

    2025

    """)

    st.markdown("---")

    st.subheader("👥 Autores / Tesistas")

    st.markdown("""
```

```

* Bach. Fabiola Yabar ylla

* Bach. Alvaro Jesus Vargas Figueroa

"""

st.markdown("---")

st.caption("Sistema desarrollado con Python y Machine Learning (KNN).")

st.title("🧠 Sistema Experto para Detección de TDAH")

with st.expander("📄 VER PROPÓSITO DEL PROYECTO", expanded=True):

    st.markdown("""

        **Propósito de la Investigación:**

        Este software es parte de la tesis titulada *"DETECCIÓN DEL TRASTORNO POR DÉFICIT DE ATENCIÓN E HIPERACTIVIDAD MEDIANTE TÉCNICAS DE MACHINE LEARNING"*.

        Su objetivo es servir como una **herramienta de triaje** para el departamento de psicología, permitiendo identificar patrones de riesgo en estudiantes mediante el análisis de datos del Test de Conners modificado.

        > **Nota:** Este sistema no reemplaza el juicio clínico profesional, sino que actúa como una herramienta de apoyo a la toma de decisiones.

    """)

st.markdown("---")

st.subheader("📋 Ficha de Evaluación del Estudiante")

with st.form("ficha_estudiante"):

    c_dat1, c_dat2, c_dat3 = st.columns(3)

    edad = c_dat1.number_input("Edad del Estudiante", min_value=6, max_value=16, value=8)

    sexo = c_dat2.selectbox("Sexo", ["Masculino", "Femenino"])

    grado = c_dat3.text_input("Grado/Sección (Opcional)", "Primaria")

    st.divider()

    st.markdown("#### Evaluación de Síntomas (Escala 0-3)")

    st.caption("0: Nunca | 1: A veces | 2: A menudo | 3: Muy a menudo")

```

```
tab1, tab2, tab3 = st.tabs([" Inatención", " Hiperactividad", " Control Cognitivo"])
```

```
with tab1:
```

```
    c1, c2 = st.columns(2)
```

```
    i1 = c1.slider("I1. Dificultad para prestar atención en tareas", 0, 3, 1)
```

```
    i2 = c2.slider("I2. Dificultad para finalizar tareas iniciadas", 0, 3, 1)
```

```
    i3 = c1.slider("I3. Evita tareas de esfuerzo mental sostenido", 0, 3, 1)
```

```
    i4 = c2.slider("I4. Se distrae fácilmente con estímulos externos", 0, 3, 1)
```

```
    i5 = c1.slider("I5. Pierde objetos necesarios (lápices, libros)", 0, 3, 1)
```

```
    i6 = c2.slider("I6. Parece no escuchar cuando se le habla", 0, 3, 1)
```

```
with tab2:
```

```
    c1, c2 = st.columns(2)
```

```
    h1 = c1.slider("H1. Habla en exceso", 0, 3, 0)
```

```
    h2 = c2.slider("H2. Actúa sin pensar (impulsividad)", 0, 3, 0)
```

```
    h3 = c1.slider("H3. Dificultad para esperar su turno", 0, 3, 0)
```

```
    h4 = c2.slider("H4. Se levanta del asiento en clase", 0, 3, 0)
```

```
    h5 = c1.slider("H5. Actúa como si tuviera un 'motor'", 0, 3, 0)
```

```
    h6 = c2.slider("H6. Interrumpe o se entromete con otros", 0, 3, 0)
```

```
with tab3:
```

```
    c1, c2 = st.columns(2)
```

```
    co1 = c1.slider("C1. Comete errores por descuido o rapidez", 0, 3, 0)
```

```
    co2 = c2.slider("C2. Tarda demasiado en dar respuestas", 0, 3, 0)
```

```
    co3 = c1.slider("C3. Su rendimiento académico fluctúa", 0, 3, 0)
```

```
    co4 = c2.slider("C4. Confunde o malinterpreta instrucciones", 0, 3, 0)
```

```
col_izq, col_cent, col_der = st.columns([1, 2, 1])
```

```
with col_cent:
```

```

submitted = st.form_submit_button("🔍 ANALIZAR CASO CON IA",
use_container_width=True)

if submitted:

    datos_api = {

        "Edad": edad,

        "Sexo_M": 1 if sexo == "Masculino" else 0,

        "Sexo_F": 1 if sexo == "Femenino" else 0,

        "I1_Atencion_tareas": i1, "I2_Terminar_tareas": i2, "I3_Evita_esfuerzo_mental": i3,

        "I4_Distraccion_externa": i4, "I5_Pierde_objetos": i5, "I6_No_escucha": i6,

        "H1_Habla_exceso": h1, "H2_Actua_sin_pensar": h2, "H3_Esperar_turno": h3,

        "H4_Levanta_asiento": h4, "H5_En_marcha_motor": h5, "H6_Se_entromete": h6,

        "C1_Errores_rapidez": co1, "C2_Tarda_respuestas": co2,

        "C3_Rendimiento_fluctua": co3, "C4_Confunde_instrucciones": co4

    }

    URL_API = os.getenv("API_URL", "http://127.0.0.1:8000")

    try:

        with st.spinner('Procesando datos con Modelo KNN (k=3)...'):

            response = requests.post(f"{URL_API}/diagnosticar", json=datos_api)

            if response.status_code == 200:

                resultado = response.json()

                prob = resultado['probabilidad_tdah']

                dx = resultado['diagnostico']

                st.markdown("### 📊 Resultado del Análisis")

                col_res1, col_res2 = st.columns(2)

                if dx == "POSIBLE TDAH":

                    col_res1.error(f"DIAGNÓSTICO: {dx}", icon="⚠️")

                    col_res2.metric("Probabilidad de TDAH", f"{prob}%", delta="Alto Riesgo").

```

```
    else:

        col_res1.success(f"DIAGNÓSTICO: {dx}", icon="✅")

        col_res2.metric("Probabilidad de TDAH", f"{prob}%", delta="Riesgo Bajo")

        st.success(f"**Recomendación:** El perfil del estudiante se encuentra dentro de  
los parámetros neurotípicos esperados.")

    else:

        st.error("Error técnico: La API respondió incorrectamente.")

except Exception as e:

    st.error(f"❌ Error de conexión: Asegúrate de que el archivo 'main.py' se esté  
ejecutando en otra terminal. ({e})")
```

Anexo 10. Panel fotográfico.







