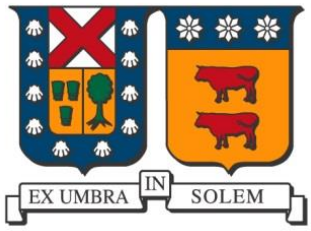
UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
VIÑA DEL MAR - CHILE



"MODELO DE APRENDIZAJE AUTOMÁTICO PROFUNDO PARA LA IDENTIFICACIÓN DE RASGOS DEL ESPECTRO AUTISTA EN ADULTOS"

CHRISTIAN ALEXIS SARABIA NEIRA

TRABAJO DE TITULACIÓN PARA OPTAR AL TÍTULO DE
INGENIERO EN INFORMÁTICA/INGENIERA EN INFORMÁTICA

Profesor Guía: GABRIEL JARA
Profesor Correferente: PATRICIO SANTANDER

04 (del examen) - 2026

CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): [X] Memoria o trabajo de título [ ] Tesis de Postgrado

Título del trabajo: Modelo de aprendizaje automático profundo para la identificación de rasgos del espectro autista en adultos.

Nombre del candidato(a): Christian Alexis Sarabia Neira

Carrera / Grado: Ingeniería en Informática

Campus: Sede José Miguel Carrera **Departamento:** Electrotecnia e Informática

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Gabriel Jara Bulnes, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente **DEJO CONSTANCIA** que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

[X] El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

[ ] El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

[ ] 6 meses [ ] 12 meses [ ] 2 años [ ] 3 años [ ] 5 años [ ] 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 19-05-2026 **Firma:** Gabriel

Estudiante o Candidato(a):

Fecha: 19-05-2026 **Firma:** C.Sarabia

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

DEDICATORIA

Este trabajo esta dedicado, en primer lugar, a mi familia, por su apoyo incondicional, su paciencia y su confianza a lo largo de todo este proceso académico. Su respaldo constante fue fundamental para mantener la motivación y la perseverancia necesaria para culminar esta etapa de formación personal.

Dedico también este trabajo a todas las personas dentro del espectro autista y a quienes, en la adultez, han debido enfrentar el desafío de comprender y explicar sus propias experiencias sin haber contado oportunamente con un diagnóstico. Su realidad inspira la necesidad de generar herramientas que contribuyan a una detección más temprana, accesible y justa.

AGRADECIMIENTOS

Deseo expresar mi más sincero agradecimiento a mi profesor guía, Gabriel Jara, por su orientación, disposición y apoyo durante todo el desarrollo de este trabajo de título. Sus observaciones y sugerencias fueron fundamentales para orientar el proyecto y asegurar un enfoque metodológico riguroso.

Asimismo, agradezco a la Universidad Técnica Federico Santa María, sede José Miguel Carrera, proporcionar el entorno académico, los recursos y la formación necesaria para el desarrollo de este proyecto, así como por fomentar una visión profesional orientada al rigor y la responsabilidad social.

Extiendo mi agradecimiento a todas las personas que, de manera directa o indirecta, aportaron con su apoyo, consejos o motivación durante este proceso, en especial a mi familia y cercanos, cuyo respaldo fue clave para afrontar los desafíos académicos y personales que implicó la realización de este trabajo.

Finalmente, agradezco a la comunidad científica y a los desarrolladores de plataformas de datos abiertos, como Kaggle, que ponen a disposición conjuntos de datos y herramientas que hacen posible la investigación y el desarrollo de soluciones basadas en ciencia de datos y aprendizaje automático.

RESUMEN

El presente trabajo tuvo como objetivo desarrollar y validar un modelo de clasificación basado en técnicas de Machine Learning para apoyar el proceso de screening del Trastorno del Espectro Autista (TEA) en adultos, utilizando el cuestionario estandarizado AQ-10. Para ello, se empleó el conjunto de datos público *Autism Screening on Adults*, el cual fue sometido a un proceso de preprocesamiento que incluyó limpieza de datos, codificación de variables categóricas y estandarización de atributos numéricos, además de la eliminación de una variable que introducía fuga de información.

La solución fue desarrollada siguiendo la metodología CRISP-DM y consistió en la implementación y evaluación comparativa de cuatro modelos de clasificación: Árbol de Decisión, Random Forest, Máquina de Vectores de Soporte (SVM) y una Red Neuronal Artificial. Los modelos fueron validados mediante métricas estándar de clasificación, tales como precisión, sensibilidad (recall), F1-score, matriz de confusión y el área bajo la curva ROC (AUC-ROC), priorizando la reducción de falsos negativos debido a la naturaleza clínica del problema.

Los resultados obtenidos muestran que todos los modelos presentan un desempeño elevado, destacando especialmente el Random Forest y la Red Neuronal Artificial, ambos con valores de AUC cercanos a 0,996 y altos niveles de sensibilidad y precisión. En particular, el Random Forest demostró el mejor equilibrio entre capacidad predictiva, estabilidad y reducción de errores clínicamente relevantes, posicionándose como la alternativa más adecuada para su uso en procesos de screening inicial.

En conclusión, este estudio confirma que las técnicas de Machine Learning pueden constituir una herramienta eficaz de apoyo al tamizaje del TEA en adultos, contribuyendo a una detección más temprana y accesible, siempre como complemento y no como reemplazo del diagnóstico clínico profesional.

Palabras Clave Trastorno del Espectro Autista; Machine Learning; AQ-10; Screening clínico; Clasificación supervisada.

ABASTRACT

The objective of this work is to develop and validate a Machine Learning–based classification model to support the screening process of Autism Spectrum Disorder (ASD) in adults using the standardized AQ-10 questionnaire. For this purpose, the public dataset *Autism Screening on Adults* was used and subjected to a preprocessing stage that included data cleaning, categorical variable encoding, numerical feature standardization, and the removal of a variable that introduced data leakage.

The solution was developed following the CRISP-DM methodology and consisted of the implementation and comparative evaluation of four classification models: Decision Tree, Random Forest, Support Vector Machine (SVM), and an Artificial Neural Network. The models were validated using standard classification metrics, including precision, recall, F1-score, confusion matrix, and the Area Under the ROC Curve (AUC-ROC), with particular emphasis on minimizing false negatives due to the clinical nature of the problem.

The results show that all models achieved high performance, with Random Forest and the Artificial Neural Network standing out, both reaching AUC values close to 0.996 along with high sensitivity and precision. In particular, Random Forest exhibited the best balance between predictive accuracy, stability, and reduction of clinically relevant errors, making it the most suitable alternative for initial screening purposes.

In conclusion, this study demonstrates that Machine Learning techniques can effectively support the screening of ASD in adults, contributing to earlier and more accessible detection, always as a complement to professional clinical diagnosis.

Keywords Autism Spectrum Disorder; Machine Learning; AQ-10; Clinical screening; Supervised classification.

GLOSARIO

- **APA**: American Psychiatric Association. Asociación Estadounidense de Psiquiatría, entidad responsable del DSM-5.
- **ASD**: Autism Spectrum Disorder (equivalente a TEA en español). Trastorno del Espectro Autista.
- **CRISP-DM**: Cross Industry Standard Process for Data Mining. Metodología estándar para proyectos de minería de datos y machine learning.
- **DSM-5**: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Manual diagnóstico y estadístico de los trastornos mentales, quinta edición.
- **DT**: Decision Tree. Árbol de decisión, algoritmo de clasificación basado en reglas de partición del espacio de atributos.
- **EEG**: Electroencephalogram. Electroencefalograma, técnica de registro de la actividad eléctrica cerebral.
- **FN**: False Negative. Falso negativo; caso positivo real clasificado incorrectamente como negativo por el modelo.
- **FP**: False Positive. Falso positivo; caso negativo real clasificado incorrectamente como positivo por el modelo.
- **F1-score**: Métrica que corresponde a la media armónica entre precision y recall, utilizada para evaluar el equilibrio entre ambas.
- **fMRI**: Functional Magnetic Resonance Imaging. Resonancia magnética funcional, técnica de neuroimagen.
- **KDD**: Knowledge Discovery in Databases. Proceso de descubrimiento de conocimiento en bases de datos.
- **ML**: Machine Learning. Rama de la inteligencia artificial que permite a los sistemas aprender a partir de datos.
- **RNA**: Red Neuronal Artificial. Modelo de aprendizaje profundo basado en neuronas artificiales organizadas en capas.
- **ROC**: Receiver Operating Characteristic. Curva que representa la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos.
- **RF**: Random Forest. Algoritmo de aprendizaje basado en un conjunto de árboles de decisión.
- **SVM**: Support Vector Machine. Máquina de Vectores de Soporte, algoritmo de clasificación que maximiza el margen entre clases.
- **TP**: True Positive. Verdadero positivo; caso positivo real correctamente clasificado por el modelo.
- **TN**: True Negative. Verdadero negativo; caso negativo real correctamente clasificado por el modelo.

INDICE DE CONTENIDOS

INDICE DE FIGURAS

INDICE DE TABLAS

INTRODUCCIÓN

El Trastorno del Espectro Autista (TEA) es una condición del neurodesarrollo caracterizada por dificultades en la comunicación social, la interacción interpersonal y la presencia de patrones de comportamiento, intereses o actividades restringidos y repetitivos. Aunque históricamente la detección del TEA se ha enfocado en la infancia, una proporción significativa de la población adulta permanece sin diagnóstico o lo recibe de forma tardía, lo que puede generar consecuencias relevantes a nivel personal, emocional, social y laboral, afectando la calidad de vida y la integración social.

En la población adulta, la detección clínica del TEA se basa principalmente en entrevistas especializadas y en el uso de instrumentos estandarizados, como el cuestionario Autism-Spectrum Quotient (AQ-10). Si bien estos métodos son fundamentales para el diagnóstico, su aplicación suele ser prolongada, costosa y dependiente de la disponibilidad de profesionales especializados, lo que limita su accesibilidad y uso masivo. En este contexto, surge la necesidad de desarrollar herramientas tecnológicas complementarias que permitan apoyar el proceso de screening inicial de manera eficiente, objetiva y escalable, sin reemplazar el juicio clínico. El screening es un proceso de evaluación preliminar que funciona como un filtro inicial de alta sensibilidad.

El presente Trabajo de Título aborda esta problemática desde el ámbito de la Ingeniería en Informática, proponiendo el desarrollo de un modelo de clasificación del TEA en adultos mediante técnicas de Machine Learning, junto con la implementación de un prototipo funcional que actúe como herramienta de apoyo al screening clínico. La solución busca facilitar una evaluación preliminar rápida y confiable, optimizando la identificación temprana de posibles casos y apoyando la toma de decisiones de los profesionales de la salud.

El desarrollo del proyecto sigue una metodología propia de la ciencia de datos. En primer lugar, se realiza una revisión de la literatura científica relacionada con la detección del TEA en adultos. Posteriormente, se efectúa el preprocesamiento y preparación del conjunto de datos utilizado, correspondiente al dataset público *Autism Screening on Adults* disponible en Kaggle, incluyendo la limpieza de datos, la codificación de variables y el análisis exploratorio. Finalmente, se diseñan, entrenan y evalúan distintos modelos de Machine Learning —Árbol de Decisión, Máquina de Soporte Vectorial (SVM) y Red Neuronal—, los cuales son validados mediante métricas estándar de clasificación y el análisis del área bajo la curva ROC (AUC-ROC), permitiendo evaluar su desempeño y capacidad predictiva de manera integral.

CAPÍTULO 1: DEFINICIÓN DEL PROBLEMA

El Trastorno del Espectro Autista (TEA) en adultos representa un desafío significativo para los sistemas de salud y para la comunidad científica, debido a que su identificación suele ser tardía, incompleta o, en muchos casos, inexistente. A diferencia de la población infantil donde existen protocolos de evaluación más estructurados y una mayor visibilidad del trastorno los adultos frecuentemente llegan a la etapa diagnóstica después de años de dificultades acumuladas en áreas como la interacción social, la organización personal, la estabilidad emocional y el desempeño académico o laboral. Esta realidad deriva en un problema de subdiagnóstico ampliamente documentado en la literatura.

La detección del TEA históricamente se ha basado en evaluaciones clínicas, entrevistas estructuradas y cuestionarios aplicados por especialistas. Sin embargo, estos procesos suelen ser extensos, costosos y subjetivos, lo que limita su accesibilidad.

En este contexto, surge la necesidad de investigar y desarrollar un modelo basado en técnicas de machine learning que pueda apoyar el proceso de detección del autismo en adultos. Para ello, se trabajará con un dataset público como el autism-screening-on-adults de kaggle que serán preprocesados y optimizados para el entrenamiento de modelos predictivos.

Es importante destacar que este trabajo se enmarca exclusivamente en el ámbito académico y se define estrictamente como una **herramienta tecnológica de tamizaje (*screening*) o de apoyo a la decisión clínica**. Bajo ningún concepto los resultados emitidos por los modelos de aprendizaje automático deben interpretarse como un diagnóstico clínico definitivo de Trastorno del Espectro Autista (TEA). El objetivo es contribuir desde la ingeniería informática con un **modelo computacional** que asista en la identificación temprana de rasgos conductuales para optimizar la derivación a especialistas, pero **no sustituye la evaluación profesional** realizada por médicos o psicólogos acreditados. Esta herramienta servirá como base sólida para investigaciones futuras y como soporte técnico en el área de la salud mental.

Objetivos

Objetivo general

- Desarrollar un modelo de clasificación del trastorno del espectro autista (TEA) en adultos utilizando técnicas de machine learning, e implementar un prototipo funcional que facilite su uso como herramienta de apoyo al proceso diagnóstico clínico.

Objetivos Específicos

1. Revisar la literatura científica existente sobre la detección del TEA en adultos, así como recopilar, preparar y preprocesar un conjunto de datos relevante.

2. Diseñar y entrenar diferentes modelos de machine learning, evaluando algoritmos como, máquina de soporte vectorial (SVM), árbol de decisión vs redes neuronales, con el propósito de identificar la técnica con mejor desempeño.

3. Evaluar, analizar y justificar el desempeño de los modelos generados, seleccionando métricas de evaluación adecuadas al contexto del *screening* del TEA en adultos, tales como *precision*, *recall*, *F1-score*, matriz de confusión y Área Bajo la Curva ROC (AUC-ROC), asegurando que la validación refleje de manera confiable la capacidad de generalización y la relevancia clínica del modelo.

4. Implementar un prototipo pipeline de la herramienta de detección, que integre el modelo seleccionado y facilite su uso como apoyo al proceso de diagnóstico clínico en adultos.

Justificación

La justificación de este proyecto radica en la necesidad de abordar el vacío de detección en la población adulta. La literatura epidemiológica global estima que una porción significativa (entre el 1.2% y el 1.5% de la población según revisiones como la de Bölte & Girdler, 2020) podría tener TEA sin un diagnóstico formal. Esta omisión se ve exacerbada por el fenómeno del camuflaje social (Lai et al., 2011) y la limitada disponibilidad de especialistas. Por lo tanto, la inversión de tiempo y recursos en este modelo de Machine Learning se justifica al ofrecer una solución objetiva y escalable que puede servir como un primer filtro masivo, garantizando que los recursos clínicos se concentren en los casos de mayores riesgos.

CAPÍTULO 2: MARCO CONCEPTUAL Y ESTADO DEL ARTE

El marco conceptual proporciona los fundamentos teóricos y técnicos que sustentan el desarrollo del modelo de predictivo. En esta sección se definen los conceptos clave relacionados con la problemática del TEA, se profundiza en las herramientas y técnicas de la ciencia de datos y el machine learning para la solución.

2.1 Fundamentos del TEA

2.1.1 definición y desafíos en la adultez

El TEA es una condición del neurodesarrollo caracterizada por déficits persistentes en la comunicación e interacción social, junto con patrones de comportamiento, intereses o actividades restringidos y repetitivos (American Psychiatric Association [APA], 2013). La clasificación como "espectro" subraya la amplia variabilidad en la presentación de los síntomas, lo que complica el diagnóstico, especialmente en la adultez.

- **Camuflaje (*Masking*):** En la edad adulta, los individuos a menudo desarrollan mecanismos de compensación social o camuflaje (*masking*), lo que dificulta la identificación clínica basada en la observación superficial (Lai et al., 2011). Este fenómeno hace que la detección oportuna dependa en gran medida de la autoevaluación y la información estructurada.

2.1.2 EL CUESTIONARIO AUTISM-SPECTRUM QUOTIENT (AQ-10)

El **AQ-10** es una herramienta de screening breve y estandarizada, derivada del cuestionario original de 50 ítems desarrollado por Baron-Cohen et al. (2001). Consiste en 10 preguntas binarias (respuestas 0 a 1) diseñadas para medir rasgos autistas en adultos. Este cuestionario es el **Fundamento de los datos de entrada** para el modelo propuesto, donde el análisis de las diez respuestas individuales (A1_Score a A10_Score) constituye el núcleo de los atributos conductuales.

AQ-10

Cociente de Espectro Autista

Nombre: _____

Fecha: _____

Una guía rápida para remitir a adultos con sospecha de autismo qué no cuentan con dificultades de aprendizaje.

Marque únicamente una opción por pregunta:		**Definitivamente de acuerdo**	**Ligeramente de acuerdo**	**Ligeramente en desacuerdo**	**Definitivamente en desacuerdo**
1	Con frecuencia percibo pequeños sonidos cuando los demás no lo hacen				
2	Usualmente me concentro en toda la película en lugar de pequeños detalles				
3	Se me facilita hacer más de una cosa a la vez				
4	Si hay una interrupción, puedo volver inmediatamente a donde estaba				
5	Se me facilita "leer entre líneas" cuando alguien me habla				
6	Puedo decir cuando alguien me está escuchando o cuando se está aburriendo				
7	Cuando estoy leyendo una historia, se me dificulta identificar las intenciones de los personajes				
8	Me gusta coleccionar información acerca de categorías de cosas (ejemplo: tipos de autos, de aves, de trenes, tipos de plantas, etc.)				
9	Se me facilita saber lo que alguien está pensando o sintiendo simplemente mirándole a la cara				
10	Se me dificulta distinguir las intenciones de la gente				

PUNTAJE: Únicamente anote un punto por cada pregunta. Anote un punto por *Definitivamente o Ligeramente de acuerdo en los ítems 1,7,8 y 10*. Anote un punto por *Definitivamente o Ligeramente en desacuerdo en los ítems 2,3,4,5,6 y 9*. Si el individuo puntúa más de 6 ítems de los 10 que son en total, se considera remitirle a un diagnóstico con un especialista.

Esta prueba se recomienda para: reconocimiento, remisión, diagnóstico y manejo de autismo, para adultos en el espectro autista (NICE clinical guide line CG142). www.nice.org.uk/CG142 .

Referencia: Allison C, Auyeung B, and Baron-Cohen S, (2012) Journal of the American Academy of Child and Adolescent Psychiatry 51(2):202-12

Desarrollada en: Universidad de Cambridge , Autism Research Centre.

Figura 1 Test AQ-10 español. Fuente: University of Florida

2.2 Estado del arte

El uso de algoritmos de Machine learning para el diagnóstico y screening del TEA ha sido un campo de intensa investigación en la última década. El objetivo común es aumentar la accesibilidad y reducir los tiempos de espera al automatizar el análisis de cuestionarios

La revisión sistemática de Brain Sci. (2020) sobre el uso de la clasificación supervisada para el TEA establece un claro precedente para la metodología propuesta. Los autores identificaron consistentemente que los modelos basados en la Máquina de Soporte Vectorial (SVM) y los Árboles de Decisión (Decision Tree) son los más utilizados y muestran un rendimiento superior en comparación con otras técnicas.

Un aspecto fundamental abordado en esta revisión corresponde al tipo de datos empleados en los estudios analizados. De manera predominante, la literatura se basa en cuestionarios conductuales estandarizados, utilizados ampliamente en contextos clínicos y de investigación. Entre los instrumentos más frecuentes se encuentran el *Autism Spectrum Quotient (AQ)*, en sus distintas versiones (AQ-50 y AQ-10), el *Social Communication Questionnaire (SCQ)* y el *Modified Checklist for Autism in Toddlers (M-CHAT)*. Estos tests evalúan rasgos relacionados con la comunicación social, la interacción interpersonal y los patrones de comportamiento, permitiendo representar las respuestas mediante variables categóricas o binarias, lo que facilita su procesamiento mediante algoritmos de *Machine Learning*.

Asimismo, la revisión destaca el uso de datasets tabulares públicos, como los disponibles en el *UCI Machine Learning Repository*, que incluyen versiones para población adulta, adolescente e infantil. Estos conjuntos de datos contienen atributos conductuales y demográficos derivados directamente de cuestionarios de screening, manteniendo una estructura similar a la del instrumento AQ-10. Este enfoque ha demostrado ser especialmente adecuado para el desarrollo de modelos predictivos orientados al tamizaje inicial del TEA, debido a su bajo costo, facilidad de recolección y alta interpretabilidad.

Si bien la literatura también reporta investigaciones basadas en datos biomédicos complejos, como electroencefalogramas (EEG), resonancia magnética funcional (fMRI) o conjuntos de datos de neuroimagen como ABIDE (*Autism Brain Imaging Data Exchange*), la revisión de *Brain Sciences* (2020) señala que estos enfoques se utilizan principalmente en contextos de investigación clínica avanzada. Su alto costo, complejidad técnica y limitada escalabilidad reducen su aplicabilidad práctica en procesos de screening masivo o evaluación inicial, razón por la cual no constituyen el enfoque predominante en la literatura revisada.

En consecuencia, la evidencia científica respalda de forma consistente el uso de cuestionarios conductuales estandarizados como base para modelos de clasificación supervisada del TEA, particularmente en contextos de screening. Este planteamiento se alinea directamente con el presente trabajo, que utiliza el dataset público *Autism Screening on Adults*, disponible en la plataforma Kaggle, construido a partir de respuestas al cuestionario AQ-10. De este modo, tanto el tipo de datos como los algoritmos seleccionados

se encuentran plenamente alineados con el estado del arte, reforzando la validez metodológica y la pertinencia clínica de la solución propuesta.

2.3 SET DE DATOS A UTILIZAR

El proyecto se base en el conjunto de datos **"Autism Screening on Adults"** obtenido a través de la plataforma de Kaggle. Este dataset es una compilación de datos de screening estructurados diseñados para la investigación de la detección automatizada de TEA.

El conjunto de datos comprende más de 704 registros y 21 variables, clasificadas de la siguiente manera:

- Variables de Puntuación (10 Features): A1_Score a A10_Score. Estas son las respuestas binarias (0 a 1) al cuestionario AQ-10.

- Variables Demograficas y Contextuales (10 Features): Incluyen la edad(age), el género(gender), la etnicidad(ethnicity), el país de residencia(contry_of_res), si él tiene un familiar con autismo(autism), si usó la aplicación antes(used_app_before) de los 704 registro solo 12 personas es decir el 1.7% utilizaron la aplicación, mientras que 689 personas es decir el 98.3% no utilizaron la aplicación, el descriptor de edad(age_desc), y la relación del encuestado con el sujeto(relation).

- Variable Objetivo (1 Feature): Class/ASD. Es la etiqueta binaria que indica la presencia (YES) o ausencia (NO) de un diagnóstico de TEA, utilizada para el entrenamiento supervisado.

Durante la etapa de preparación de los datos, se identificó la presencia de una variable adicional denominada result, la cual es generada a partir de la variable Class/ASD mediante una codificación binaria numérica. Dicha variable presenta una relación de colinealidad directa con la clase objetivo, ya que representa la misma información semántica expresada en un formato distinto.

Debido a esta colinealidad, la inclusión simultánea de *result* y *Class/ASD* dentro del conjunto de datos podría inducir filtración de información (*data leakage*), afectando artificialmente el desempeño de los modelos y comprometiendo la validez experimental de los resultados. En consecuencia, la variable result es eliminada del conjunto de variables predictoras y se conserva únicamente la etiqueta Class/ASD.

2.4 Bases de la Ciencia de Datos y el Machine Learning

El **Machine Learning** es una rama de la inteligencia artificial que permite a los sistemas aprender automáticamente a partir de datos, identificar patrones complejos y hacer predicciones sin ser programados explícitamente (Géron 2019).

- **Clasificación Supervisada:** Este es el paradigma utilizado en el proyecto. El algoritmo se entrena con un conjunto de datos etiquetado (donde la variable objetivo, Class/ASD, ya está definida como *YES/NO*). El objetivo es que el modelo aprenda a mapear las features de entrada (respuestas AQ-10, edad, género, etc.) a la etiqueta de salida con la mayor precisión posible.

2.4.1 Algoritmos de Clasificación a Evaluar

Se seleccionarán tres familias de algoritmos para una evaluación comparativa, buscando el mejor desempeño predictivo para la clasificación binaria de riesgo de TEA:

2.4.1.1 Máquina de Vectores de Soporte (Support Vector Machine, SVM)

Algoritmo de clasificación que busca el **hiperplano óptimo** que maximiza el margen entre las clases. Es robusto en la gestión de alta dimensionalidad y utiliza funciones *kernel* para manejar la separación de datos no lineales (Cortes & Vapnik, 1995).

- **Principales Hiperparámetros:**
    - **Kernel:** Define la función de transformación que se aplica a los datos. Los más comunes son: Lineal (para datos linealmente separables), RBF (Radial Basis Function o Gaussiano) y Polinomial (para separación no lineal)

    - **C (Parámetro de Regularización):** Controla el trade-off entre tener un margen suave (amplio) y clasificar correctamente los puntos de entrenamiento. Un valor bajo de C tolera más errores de clasificación (margen más amplio); un valor alto exige una clasificación perfecta (margen más estrecho).

    - **Gamma (para kernels RBF/polinomial):** Determina la influencia que tiene un solo ejemplo de entrenamiento. Un valor alto Gamma da a los puntos cercanos más influencia (límite de decisión más complejo); un valor bajo da una influencia más amplia (límite más suave).

2.4.2.2 Árbol de Decisión (Decision Tree, DT)

Modelo predictivo no paramétrico y altamente interpretable que utiliza un conjunto de reglas de decisión. Opera mediante la división iterativa del espacio de *features* en subregiones, minimizando la impureza en cada nodo para lograr la mejor clasificación (Quinlan, 1986).

- **Principales HIPER PARÁMETROS:**
  - **Criterio de Partición (Criterion):** La función utilizada para medir la calidad de una división (la reducción de impureza). Los criterios más comunes son la Impureza Gini y la Ganancia de Información (basada en la Entropía).
  - **Profundidad Máxima (*Max Depth*):** Limita el número máximo de niveles o divisiones que puede tener el árbol. Es fundamental para prevenir el **sobreajuste (*overfitting*)** al restringir la complejidad del modelo.
  - **Mínimo de Muestras para Dividir (*Min Samples Split*):** El número mínimo de instancias o registros que debe tener un nodo para que se considere una división adicional.
  - **Mínimo de Muestras en Hoja (*Min Samples Leaf*):** El número mínimo de instancias requeridas para que un nodo se considere una hoja final (nodo terminal).

2.4.2.3 Random Forest (RF)

El random forest (RF) es una técnica popular de aprendizaje automático en el campo de la minería de datos (). Opera bajo la supervisión de un grupo y ha recibido un reconocimiento significativo. La minería de datos se puede clasificar en dos tipos principales: descriptiva y predictiva. La minería de datos descriptiva se centra en proporcionar descripciones detalladas y resúmenes de datos. Por otro lado, la minería de datos predictiva implica el estudio de datos históricos para identificar patrones y tendencias que puedan utilizarse para realizar predicciones sobre el futuro. La minería de metadatos es el proceso de describir y resumir datos, descubrir patrones y relaciones dentro de ellos y utilizar datos históricos para realizar predicciones sobre tendencias futuras. Los modelos predictivos se construyen analizando las características de los factores predictivos para proporcionar hipótesis que ayuden a tomar decisiones futuras.

- **Principales Hiper Parámetros:**
    - **Numero de árboles (n_estimators):** Corresponde a la cantidad de árboles de decisión que componen el bosque. Un mayor número de árboles generalmente mejora la estabilidad y el desempeño del modelo, reduciendo la varianza, aunque incrementa el costo computacional.
    - **Numero de máximo de variables por división (max_features):** Define cuántas características se seleccionan aleatoriamente en cada nodo para evaluar las posibles divisiones. Valores más bajos introducen mayor aleatoriedad y reducen la correlación entre árboles, lo que mejora la capacidad de generalización del modelo.
    - **Profundidad máxima del árbol (max_depth):** Limita la profundidad de cada árbol individual. Controlar este parámetro es fundamental para evitar el sobreajuste, especialmente cuando el número de muestras es limitado.
    - **Mínimo de muestras por hojas (min_samples_leaf):** Determina el número mínimo de muestras que debe contener un nodo hoja. Este parámetro ayuda a suavizar el modelo y a evitar que se ajusten patrones espurios.
    - **Criterio de impureza (criterion):** Define la métrica utilizada para evaluar la calidad de una división, siendo las más comunes la Impureza Gini y la Entropía.

2.4.2.4 Redes Neuronales Artificiales (RNA)

Estructuras de *Deep Learning* compuestas por capas de neuronas interconectadas. Su capacidad para modelar relaciones complejas y no lineales es esencial para capturar patrones sutiles en los datos de salud conductual (Goodfellow et al., 2016).

- **Principales Hiper Parámetros:**
    - **Arquitectura (Capas y Neuronas):** El número de capas ocultas y la cantidad de neuronas en cada una. Una arquitectura más profunda (más capas) o ancha (más neuronas) aumenta la capacidad de modelado, pero también el riesgo de sobreajuste.

    - **Tasa de Aprendizaje (*Learning Rate*):** Controla el tamaño de los pasos que toma el optimizador (por ejemplo, ADAM u SGD) al actualizar los pesos de la red. Un valor demasiado alto puede hacer que el modelo diverja; uno demasiado bajo ralentiza la convergencia.

    - **Función de Activación:** Determina la salida de cada neurona en función de su entrada. Las más comunes para las capas ocultas son **ReLU** (*Rectified Linear Unit*) y **Sigmoide** o **Softmax** (para la capa de salida de clasificación).

    - **Regularización (*Dropout*):** Una técnica para prevenir el sobreajuste que consiste en ignorar aleatoriamente un porcentaje de neuronas en cada paso de entrenamiento.

2.5 Herramientas tecnológicas y Plataformas

Esta sección describe el conjunto de herramientas de software, lenguajes de programación y bibliotecas de *Machine Learning* que se utilizarán para la implementación y validación del modelo de clasificación.

Python es el lenguaje de programación principal seleccionado debido a su robustez, su sintaxis legible y, crucialmente, su vasto ecosistema de bibliotecas orientadas a la Ciencia de Datos y la inteligencia Artificial (Géron, 2019).

Estas bibliotecas constituyen el framework principal para manipular y modelar los datos:

- **Pandas y NumPy:** Para la manipulación y el procesamiento de datos.
- **Matplotlib / Seaborn:** Para el análisis exploratorio y la visualización de resultados.
- **Scikit-learn y TensorFlow/Keras:** Para la implementación, entrenamiento y evaluación de los modelos de *Machine Learning*.
- **Plotly:** Utilizada para visualizaciones interactivas, facilitando un análisis exploratorio más dinámico de los datos, permitiendo explorar, analizar y presentar datos con gráficos de líneas, barras, dispersión, 3D.
- **Tabulate:** Empleada para la presentación estructurada de resultados en formato tabular dentro del entorno de ejecución, mejorando la legibilidad de métricas y comparaciones entre modelos.
- **Warnings:** Utilizada para la gestión y supresión de advertencias durante la ejecución del código, contribuyendo a una salida más limpia y controlada durante los experimentos.

Plataformas que permiten el desarrollo de código Python de manera interactiva y la mezcla de código ejecutable con narración y documentación (markdown), como google colab/ jupyter notebook.

2.6 Preprocesamiento de Datos

El preprocesamiento de datos es una etapa esencial del proceso de descubrimiento de información o KDD (Knowledge Discovery in Databases, en inglés) (Han, Kamber y Pei, 2011; Zaki y Meira, 2014). Esta etapa se encarga de la limpieza de datos, su integración, transformación y reducción para la siguiente fase de minería de datos (García, Luengo y Herrera, 2015).

Se realizarán tareas de limpieza, *Codificación One-Hot* para variables categóricas nominales (como **ethnicity y gender**), y *Estandarización* para la variable *age*, con el fin de optimizar el rendimiento de los algoritmos y prepararlos para la fase de modelado.

Adicionalmente, durante el análisis exploratorio y la validación inicial de los datos, se identificó la variable *result* como un atributo derivado directamente de la variable objetivo *Class/ASD*, presentando una alta colinealidad con esta. La inclusión de dicha variable habría introducido una filtración de información (*data leakage*), comprometiendo la validez de los resultados del modelo. Por este motivo, la variable *result* fue excluida del conjunto de datos previo al entrenamiento, asegurando que el proceso de aprendizaje se basara únicamente en información disponible de forma legítima en un escenario real de *screening*.

Un aspecto crítico en el control de sesgos fue la identificación de la variable 'result', la cual representaba la sumatoria de las respuestas del cuestionario AQ-10. Se determino que incluir esta variable en el entramiento constituiría un error de fuga de datos (data leakage), ya que el algoritmo aprendería una regla aritmética determinista en lugar de identificar patrones conductuales complejos. Al eliminar esta variable, se garantizó que los modelos(especialmente Random forest y Redes neuronales) fueran validados sobre la capacidad de generalización real de las respuestas individuales, asegurando un proceso de aprendizaje libre de sesgos metodológicos directos.

2.7 Métricas de Evaluación

Se utilizarán métricas sensibles a los errores de clasificación, cruciales en el ámbito de la salud, ya que se busca minimizar los Falsos Negativos (riesgo de no diagnosticar un caso real).

La elección de estas métricas no es solo estadística, sino que responde a una estrategia de mitigación de sesgos metodológicos. Al haber eliminado variables deterministas como 'result', las métricas Recall y F1-Score adquieren una validez real, pues obligan al modelo a demostrar su capacidad de generalización sobre los rasgos conductuales individuales y no sobre sumatorias aritméticas predecibles.

2.7.1 Matriz de confusión

La Matriz de Confusión es una herramienta esencial en el campo del *Machine Learning* que permite visualizar y cuantificar el desempeño de un algoritmo de clasificación supervisada. En términos prácticos, esta matriz nos muestra qué tipos de aciertos y errores está teniendo nuestro modelo al contrastar las predicciones con los valores reales (*ground truth*). Cada columna de la matriz representa el número de predicciones para cada clase, mientras que cada fila representa a las instancias reales.

- **Recall:** Mide la capacidad del modelo para identificar correctamente los casos positivos reales de TEA. Se define como la proporción de Verdaderos Positivos (TP) respecto al total de casos positivos reales (TP + FN). Su maximización es directamente proporcional a la minización de los Falsos Negativos (FN).
  Un recall elevado permite reducir los falsos negativos, lo que es clave en contextos de *screening* o apoyo diagnóstico para evitar la omisión de casos de riesgo. Aunque priorizar esta métrica puede aumentar los falsos positivos, esta implicancia es aceptable en el estudio, ya que el modelo se utiliza como herramienta de apoyo y no como diagnóstico definitivo, privilegiando la detección temprana del TEA en adultos.

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

- **Precision:** Se refiere a la dispersión del conjunto de valores obtenidos a partir de mediciones repetidas de una magnitud. Cuanto menor es la dispersión mayor la precisión. Se representa por la proporción de verdaderos positivos dividido entre todos los resultados positivos (tanto verdaderos positivos, como falsos positivos).

$$\text{Precisión} = \frac{TP}{TP + FP}$$

- **F1-score:** Esta métrica es la media armónica de la Sensibilidad (*Recall*) y la Precisión (*Precision*). Es crucial para balancear la Sensibilidad alta con la necesidad de evitar Falsas Alarmas excesivas. El F1-score se utiliza como el criterio final de selección del modelo, ya que garantiza el mejor equilibrio entre la seguridad clínica (alto *Recall*) y la eficiencia de recursos (alta *Precision*) en un entorno de datos desbalanceado (Fawcett, 2006).
  Dado que el dataset presenta una distribución de clases donde los casos positivos representan aproximadamente el 26.8% de la muestra (189 de 704 registros), el uso del F1-score y la Curva ROC es mandatorio para evitar el sesgo de la 'paradoja de la exactitud', donde un modelo podría parecer exitoso simplemente prediciendo siempre la clase mayoritaria.

$$F1 = \frac{2 \cdot (\text{Precisión} \cdot \text{Recall})}{\text{Precisión} + \text{Recall}}$$

2.7.2 Curva ROC

La **Curva ROC** es una herramienta gráfica fundamental que evalúa el rendimiento de un clasificador a través de todos los posibles umbrales de clasificación. Se grafica la **Tasa de Verdaderos Positivos (Sensibilidad)** contra la **Tasa de Falsos Positivos (1 - Especificidad)**.

- **Tasa de verdaderos positivos**
  - La tasa de verdaderos positivos, también conocida como sensibilidad o recuerdo, refleja la capacidad de un modelo para identificar correctamente los casos positivos. Mide la proporción de casos positivos reales que el modelo identifica con éxito. Matemáticamente, esto puede expresarse mediante la siguiente ecuación:

    - **Eje Y (TPR / Sensibilidad / Recall):**

$$\boldsymbol{TPR = \frac{TP}{TP + FN}}$$

- **Tasa de falsos positivos**
  - FPR representa la frecuencia con la que nuestro modelo clasifica incorrectamente como positivas las instancias de clase negativas. Mide la proporción de instancias negativas reales que el modelo identifica incorrectamente como positivas, lo que indica la tasa de falsas alarmas. Matemáticamente esto se puede expresar de la siguiente manera:

    - **Eje X (FPR / Tasa de Falsos Positivos):**

$$\boldsymbol{FPR = \frac{FP}{FP+TN}}$$

Al variar progresivamente el umbral de clasificación, el modelo genera distintas combinaciones de TPR y FPR, las cuales se representan como puntos en el plano ROC. La unión de estos puntos da lugar a la curva ROC, que describe el comportamiento del modelo frente a distintos criterios de decisión. Un modelo adecuado para tareas de *screening* de TEA adulto tenderá a presentar curvas cercanas al vértice superior izquierdo del gráfico, reflejando una alta sensibilidad junto con una tasa controlada de falsos positivos.

El desempeño global del modelo se resume mediante el Área Bajo la Curva ROC (AUC, *Area Under the Curve*), la cual cuantifica la capacidad del clasificador para discriminar correctamente entre adultos con y sin TEA, independientemente del umbral seleccionado. Un valor de AUC cercano a 1 indica una alta capacidad discriminativa, mientras que un valor cercano a 0,5 indica que el clasificador tiene capacidad discriminativa, pero está *prediciendo de forma inversa.* Es decir, la probabilidad de que el modelo asigne una puntuación más alta a un caso positivo que a uno negativo es menor a la del azar. En el contexto de este trabajo, un AUC elevado resulta especialmente relevante, ya que respalda el uso del modelo como una herramienta de apoyo al *screening* de TEA en adultos, orientada a la detección temprana y no al diagnóstico clínico definitivo.

CAPÍTULO 3: PROPUESTA DE SOLUCION

En este capítulo se presenta la estrategia global diseñada para abordar el desafío del modelo de clasificación del TEA en la población adulta, utilizando datos de bajo costo y clasificadores de machine learning de alta eficiencia.

3.1 Propuesta de Solución: Modelo de clasificación comparativo

La solución propuesta consiste en el desarrollo y la evaluación comparativa de tres modelos de **Clasificación Binaria Supervisada** basados en un dataset estructurado (AQ-10). El objetivo principal es crear una herramienta de screening de riesgo inicial para adultos, que pueda operar de manera escalable y accesible para si poder ayudar a los especialistas aumentar la clasificación en TEA.

3.1.1 Ejes del Aporte Creativo

El valor de esta solución reside en un enfoque que integra la robustez del *Machine Learning* con las necesidades de la práctica clínica. Este diseño cumple con la necesidad de demostrar la creación, diseño y/o ejecución de una solución profesional.

3.1.1.1 Priorización de la Detección

Dada la alta implicación de un diagnóstico fallido (Falso Negativo) en el ámbito de la salud, la solución no busca simplemente la mayor **Exactitud (*Accuracy*).** La propuesta se enfoca en maximizar la **Sensibilidad (*Recall*)** mediante la técnica de Ponderación de **Clases (*Class Weighting*)** en la fase de entrenamiento. Esto es un diseño intencional para un sistema de *screening* inicial.

Esta decisión de diseño asume explícitamente el costo asociado a la clasificación errónea de algunos individuos sin TEA como casos positivos (Falsos Positivos), los cuales podrán ser posteriormente descartados mediante evaluaciones clínicas más específicas. Dicho costo se considera aceptable en el contexto de un sistema de screening inicial, ya que resulta preferible identificar un mayor número de posibles casos, aun a costa de generar falsas alarmas, antes que omitir individuos que efectivamente presentan TEA y que podrían beneficiarse de una evaluación diagnóstica oportuna.

De este modo, el enfoque adoptado prioriza la reducción de Falsos Negativos por sobre la optimización global de la exactitud, alineándose con el objetivo principal del sistema como herramienta de apoyo al proceso diagnóstico y no como un mecanismo de diagnóstico clínico definitivo.

3.1.1.2 Evaluación con Terna de Algoritmos

Se seleccionaron tres algoritmos que cubren un amplio espectro de estrategias matemáticas, lo que garantiza que la solución óptima no dependa de un único principio de modelado:

- **Árbol de Decisión (DT):** Este modelo destaca por su alta interpretabilidad, especialmente relevante en contextos de salud, ya que permite identificar de forma explícita la influencia y el orden de las variables en la toma de decisiones. En este estudio, esta característica fue clave para detectar tempranamente la variable *result* como una fuente de *data leakage*, al presentarse como un atributo dominante directamente derivado de la variable objetivo. Este hallazgo permitió depurar el conjunto de datos antes del entrenamiento definitivo, evidenciando la utilidad de los modelos interpretables tanto para la validación clínica como para la detección de problemas metodológicos.

- **Random Forest (RF):** Se utiliza como un método de ensamblado que combina múltiples árboles de decisión, mejorando la capacidad de generalización y reduciendo el sobreajuste. Además, permite analizar la importancia relativa de las variables, lo que resulta útil para evaluar la contribución de los distintos atributos en la predicción del TEA.

- **Máquina de Soporte Vectorial (SVM):** Es especialmente eficaz en la separación de clases en espacios de alta dimensionalidad. Mediante el uso de funciones *kernel*, la SVM permite modelar fronteras de decisión no lineales, lo que la convierte en una alternativa sólida para evaluar la capacidad de separación entre casos positivos y negativos de TEA en función de los atributos disponibles.

- **Red Neuronal Artificial (RNA):** Se incorporan debido a su capacidad para modelar relaciones no lineales y patrones sutiles inherentes a datos conductuales y de cuestionarios psicométricos. Este tipo de modelo permite capturar combinaciones complejas de variables que no siempre son evidentes mediante enfoques más lineales o basados en reglas, aunque a costa de una menor interpretabilidad directa.

3.2 Metodología: CRISP - DM

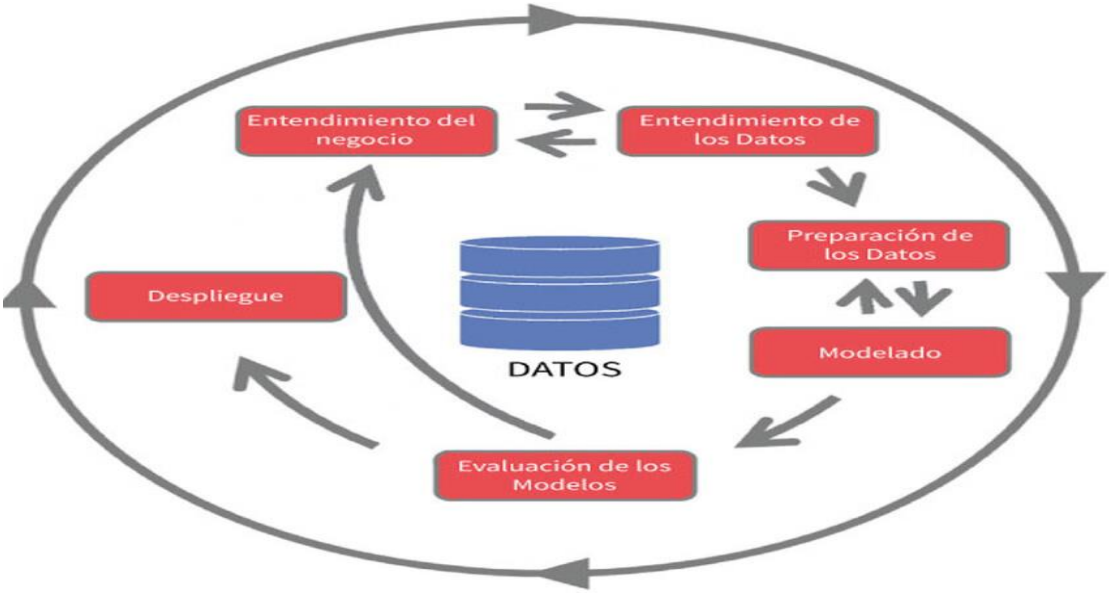


Figura 2 Diagrama CRISP-DM. Fuente: ResearchGate

Se ha seleccionado la metodología *CRISP-DM,* un estándar consolidado en la industria para proyectos de minería de datos y aprendizaje de datos. La elección de este marco de trabajo responde a un enfoque cíclico e iterativo, lo que permite una revisión constante de los resultados obtenidos en cada etapa frente a los objetivos de negocio iniciales. Al segmentar el proceso en fases interconectadas desde la comprensión del problema clínico hasta la evaluación técnica de los algoritmos, se garantiza un rigor metodológico que permite identificar el modelo más robusto para el apoyo en la detección de rasgos del espectro autista en adultos.

A continuación, se describen las actividades realizadas en cada una de las fases aplicadas a este estudio:

- **Comprensión del Negocio:** Definición del problema del subdiagnóstico de TEA en adultos y establecimiento de los requisitos de un modelo de *screening* eficiente, priorizando la reducción de falsos negativos.
- **Comprensión de los Datos:** Análisis exploratorio del dataset seleccionado para comprender la distribución de las respuestas al AQ-10 y las características demográficas de la muestra.

- **Preparación de los Datos:** Limpieza de datos, tratamiento de nulos y aplicación de técnicas de preprocesamiento como *One-Hot Encoding* y normalización de variables. Se eliminó la variable result para evitar el sesgo en el entrenamiento.
- **Modelado:** Selección y configuración de cuatro algoritmos de clasificación (Árbol de Decisión, Random Forest, SVM y Redes Neuronales) para comparar su capacidad predictiva.
- **Evaluación:** Análisis exhaustivo de los resultados mediante métricas de desempeño, matrices de confusión y curvas ROC, determinando que el modelo de **Random Forest** es el más apto por su equilibrio entre precisión y generalización.
- **Despliegue:** En esta etapa, el proyecto culmina con la entrega de *las conclusiones técnicas y recomendaciones* para una futura implementación de los modelos evaluados en entornos clínicos o de auto-consulta.

CAPÍTULO 4: VALIDACIÓN DE LA SOLUCIÓN

En este capítulo se presentan los métodos utilizados para evaluar la efectividad y robustez de los modelos de predicción del Trastorno del Espectro Autista basados en el cuestionario AQ-10. Debido a que el enfoque de este trabajo es la determinación del algoritmo más preciso para el tamizaje, la validación se centra en el rendimiento estadístico, la capacidad de generalización y la relevancia clínica de los resultados obtenidos.

Con el fin de asegurar una validación integral, este capítulo considera tres dimensiones clave:

1. **Validez técnica del modelo predictivo:** Evaluación exhaustiva mediante métricas estándar de clasificación como precisión, sensibilidad (recall), puntaje F1 y exactitud.

2. **Capacidad de discriminación:** Análisis mediante curvas ROC y el área bajo la curva (AUC) para determinar la eficacia de los modelos (especialmente Random Forest y Redes Neuronales) en la distinción de casos positivos.

3. **Análisis de Error y Pertinencia Clínica:** Evaluación de las matrices de confusión con especial énfasis en la minimización de falsos negativos, factor crítico para que el modelo sea considerado una herramienta de apoyo confiable en un contexto de screening inicial.

A continuación, se describe detalladamente cada uno de los procesos y resultados obtenidos.

4.1 Validación Técnica del Modelo Predictivo

La validación técnica corresponde a la verificación del desempeño del sistema apartir de métricas cuantitativas obtenidas sobre un conjunto de prueba independiente. Esto permite evaluar la capacidad del modelo para generalizar nuevos casos y garantizar que las predicciones realizadas no dependen exclusivamente de los datos de entrenamiento.

Para este propósito, el conjunto de datos *Autism Screening on Adults*, obtenido desde la plataforma Kaggle, fue dividido en dos subconjuntos disjuntos: un conjunto de entrenamiento, utilizado para el ajuste de los parámetros de los modelos, y un conjunto de prueba, reservado exclusivamente para la evaluación final del desempeño. Esta separación es fundamental para mitigar el riesgo de sobreajuste (*overfitting*), ya que impide que los modelos sean evaluados sobre datos previamente vistos durante el aprendizaje.

Para la validación se entrenaron tres modelos: **Árbol de Decisión, Máquina de Vectores de Soporte y Red Neuronal Artificial.** Cada modelo fue evaluado en términos de precisión (precision), sensibilidad (recall), f1-score y acurracy.

Asimismo, se utilizaron la **Matriz de Confusión** y la **Curva ROC-AUC** como herramientas principales para visualizar el desempeño.

4.1.1 Árbol de decisión

4.1.1.1 Matriz de clasificación Árbol de Decisión

	precision	recall	f1-score	support
0	0.94	0.92	0.93	98
1	0.82	0.86	0.84	43
accuracy			0.90	141
macro avg	0.88	0.89	0.88	141
weighted avg	0.90	0.90	0.90	141

Tabla 1 Matriz de Clasificación Árbol de decisión. Fuente: Elaboración Propia

Durante una evaluación preliminar, el modelo de Árbol de Decisión presentó un desempeño perfecto, alcanzando valores de *accuracy*, *precision*, *recall* y *F1-score* iguales a 1.00 en ambas clases. Si bien estos resultados podrían interpretarse inicialmente como altamente positivos, un análisis crítico permitió identificar que un rendimiento perfecto en un problema clínico realista resulta estadísticamente inusual y metodológicamente sospechoso.

Dado que la evaluación se realizó sobre un conjunto de prueba independiente, este comportamiento no podía atribuirse a sobreajuste. En consecuencia, se revisó el proceso de preparación de los datos, identificándose la presencia de fuga de información (data leakage) provocada por la inclusión de la variable *result* como característica de entrada, a pesar de su relación directa con la etiqueta objetivo. Esta situación permitió que el modelo accediera implícitamente a la respuesta correcta, invalidando la evaluación inicial.

Tras corregir el conjunto de datos y eliminar la variable problemática, el modelo fue nuevamente entrenado y evaluado, obteniendo métricas más realistas (accuracy = 0.90, precision = 0.82, recall = 0.86 y F1-score = 0.84 para la clase positiva). Estos resultados reflejan un comportamiento coherente con el problema abordado y confirman que el desempeño perfecto inicial no correspondía a una capacidad predictiva genuina.

Este proceso de validación evidencia la importancia del criterio del investigador en la interpretación de métricas de desempeño, así como el rol de la validación técnica como una herramienta no solo cuantitativa, sino también diagnóstica, orientada a detectar errores metodológicos y fortalecer la validez científica del estudio.

4.1.1.2 Matriz de confusión Árbol de Decisión

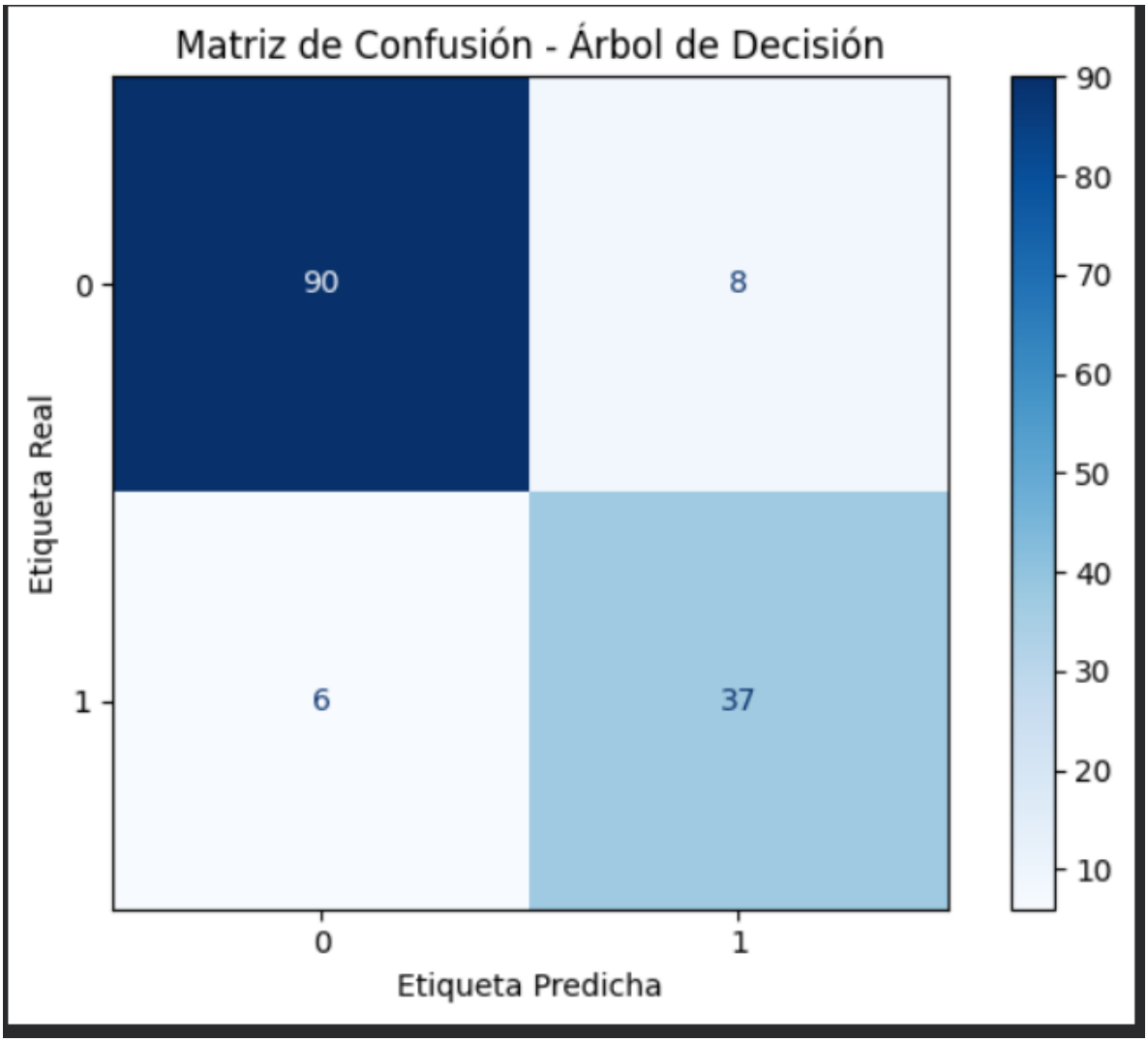


Figura 3 Matriz de confusión Árbol de Decisión. Fuente: Elaboración Propia

La matriz de confusión del árbol de decisión muestra 90 verdaderos negativos y 37 verdaderos positivos, junto con 8 falsos positivos y 6 falsos negativos, lo que evidencia que el modelo no logra separación perfecta entre ambas clases. Si bien presenta un desempeño aceptable, la existencia de falsos negativos implica la omisión de algunos casos reales de TEA, mientras que los falsos positivos generan alertas innecesarias. Estos resultados reflejan un comportamiento realista del modelo y confirman que, aunque es interpretable y sencillo, su desempeño es inferior al de modelos más avanzados como random forest o redes neuronales para un proceso de screening inicial.

4.1.1.3 Curva ROC-AUC Árbol de Decisión

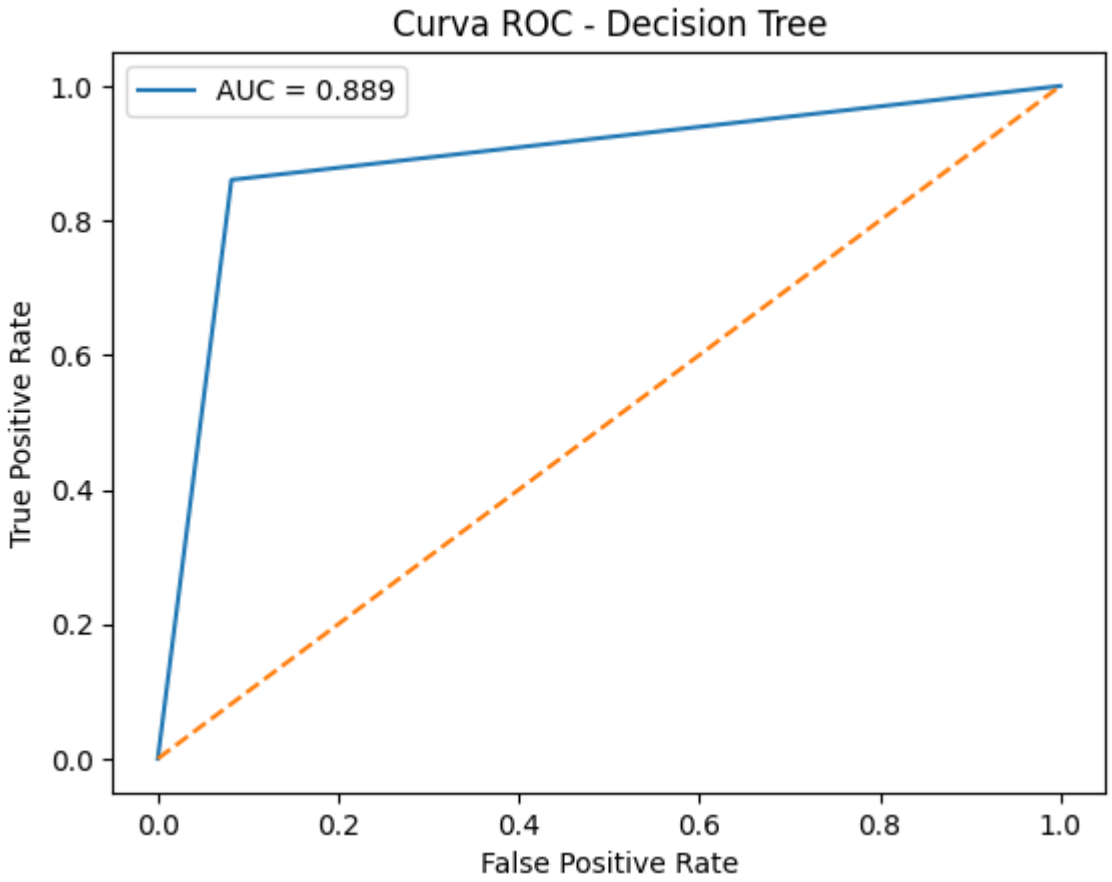


Figura 4 Curva ROC-AUC Árbol de Decisión. Fuente: Elaboración Propia

El Árbol de Decisión obtuvo un AUC = 0.889, lo que indica una buena capacidad de discriminación entre individuos con y sin indicios de TEA. Este valor refleja que, en la mayoría de los umbrales de decisión, el modelo asigna probabilidades más altas a los casos positivos que a los negativos, aunque existe cierto traslape entre ambas distribuciones. En consecuencia, el clasificador no logra una separación perfecta entre las clases, lo que es consistente con la presencia de falsos positivos y falsos negativos observados en la matriz de confusión. Este desempeño, aunque adecuado, confirma que el Árbol de Decisión presenta limitaciones frente a modelos más avanzados en tareas de *screening* clínico.

4.1.2 Random Forest

4.1.2.1 Matriz de clasificación de Random Forest

	precision	recall	f1-score	support
0	0.96	0.97	0.96	98
1	0.93	0.91	0.92	43
accuracy			0.95	141
macro avg	0.94	0.94	0.94	141
weighted avg	0.95	0.95	0.95	141

Tabla 2 Matriz de clasificación de Random Forest. Fuente: Elaboración Propia

El modelo Random Forest evidencia un desempeño robusto y consistente, con una accuracy de 0.95, acompañada de valores elevados de precision (0.93), recall (0.91) y F1-score (0.92) para la clase positiva. A diferencia del Árbol de Decisión, el modelo presenta una reducción significativa de errores, manteniendo un equilibrio adecuado entre falsos positivos y falsos negativos.

El análisis de la matriz muestra que la mayoría de los individuos con indicios de TEA fueron correctamente clasificados, lo que se traduce en una alta sensibilidad del modelo. Al mismo tiempo, la tasa de falsos positivos se mantiene acotada, evitando una sobreestimación de casos que podría derivar en derivaciones clínicas innecesarias. Este balance es especialmente relevante en un contexto de *screening*, donde resulta fundamental minimizar la omisión de casos reales sin generar un exceso de alertas.

Desde una perspectiva metodológica, el desempeño observado en Random Forest se considera **realista y confiable**, ya que no presenta resultados artificialmente perfectos, lo que sugiere una adecuada capacidad de generalización sobre el conjunto de prueba. La naturaleza de ensamble del modelo contribuye a una mayor estabilidad frente a la variabilidad de los datos, reduciendo el impacto de decisiones individuales y mejorando la separación entre clases.

En consecuencia, Random Forest se posiciona como un modelo sólido y bien balanceado, ofreciendo un compromiso adecuado entre rendimiento predictivo y confiabilidad clínica, lo que lo convierte en un candidato relevante para su uso como herramienta de apoyo en procesos de *screening* inicial

4.1.2.2 Matriz de Confusión Random Forest

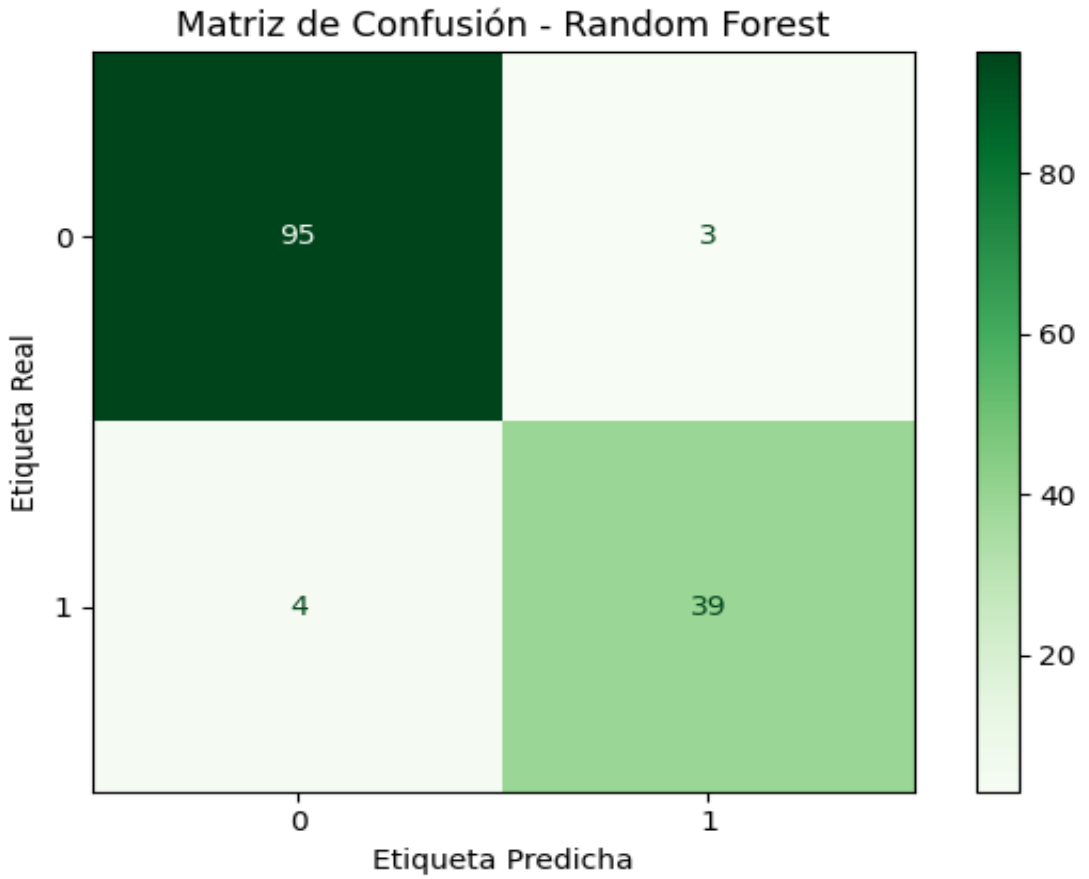


Figura 5 Matriz de Confusión Random Forest. Fuente: Elaboración Propia

La matriz de confusión del random forest muestra 95 verdaderos negativos y 39 positivos, junto con solo 3 falsos positivos y 4 falsos negativos, lo que evidencia una capacidad superior para separar ambas clases en comparación con el árbol de decisión. La baja cantidad de falsos negativos indica que la mayoría de los casos reales de TEA son correctamente identificados, reduciendo el riesgo de omisiones en procesos de derivación clínica. Al mismo tiempo, el reducido número de falsos positivos limita la generación de alertas innecesarias. Este comportamiento balanceado y consistente confirma que random forest ofrece un desempeño más confiable y clínicamente adecuado para un proceso de screening inicial.

4.1.2.3 Curva ROC – AUC Random Forest

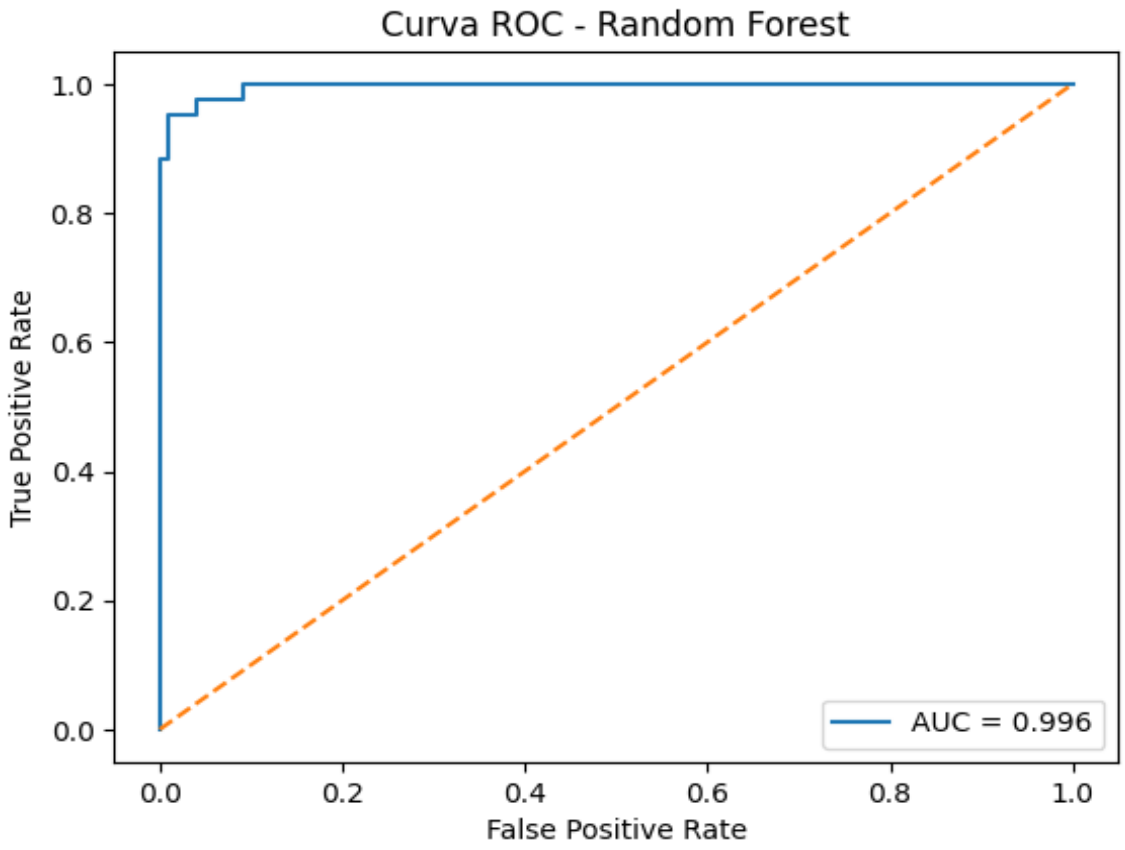


Figura 6 Curva ROC - AUC Random Forest.Fuente Elaboración Propia

Random Forest obtuvo un AUC = 0.996, lo que representa un desempeño excepcional en la curva ROC y una capacidad casi perfecta para discriminar entre individuos con y sin indicios de TEA. Este valor indica que existe una separación muy clara entre ambas clases a lo largo de prácticamente todos los umbrales de decisión, con un traslape mínimo entre las probabilidades asignadas a casos positivos y negativos. En términos clínicos, este comportamiento implica que el modelo es altamente confiable para apoyar procesos de *screening*, ya que maximiza la detección de casos reales y minimiza tanto las omisiones como las alertas innecesarias, consolidándose como la alternativa más robusta entre los modelos evaluados.

4.1.3 Maquina de vectores de soporte (SVM)

4.1.3.1 Matriz de clasificación SVM

	precision	recall	f1-score	support
0	0.95	1.00	0.98	98
1	1.00	0.88	0.94	43
accuracy			0.96	141
macro avg	0.98	0.94	0.96	141
weighted avg	0.97	0.96	0.96	141

Tabla 3 Matriz de Clasificación SVM. Fuente: Elaboración Propia

la Máquina de Vectores de Soporte (SVM) muestra un desempeño global elevado, alcanzando una accuracy de 0.96, junto con una precision perfecta (1.00) para la clase positiva y un F1-score de 0.94. Estos resultados indican que el modelo no generó falsos positivos, clasificando correctamente a todos los individuos identificados como casos con indicios de TEA.

No obstante, el análisis detallado de la matriz revela la presencia de falsos negativos, reflejada en un recall de 0.88, lo que implica que algunos casos reales no fueron detectados por el modelo. En un contexto de *screening* clínico, esta característica resulta especialmente relevante, ya que la omisión de casos puede retrasar procesos de evaluación y derivación profesional.

Desde el punto de vista metodológico, el desempeño del modelo se considera consistente y realista, ya que, a diferencia de evaluaciones preliminares con resultados perfectos, las métricas obtenidas reflejan un compromiso natural entre precisión y sensibilidad. El comportamiento observado es coherente con la naturaleza de SVM, que prioriza la maximización del margen de separación entre clases, favoreciendo la reducción de falsos positivos a costa de una menor sensibilidad.

En consecuencia, aunque SVM destaca por su alta capacidad de discriminación y su precisión absoluta en la identificación de casos positivos, su menor recall respecto a otros modelos sugiere que podría no ser la opción más adecuada cuando el objetivo principal es minimizar la omisión de casos en un proceso de *screening* inicial. Sin embargo, su desempeño lo posiciona como una alternativa sólida en escenarios donde se prioriza la confiabilidad de las clasificaciones positivas.

4.1.3.2 Matriz de confusión SVM

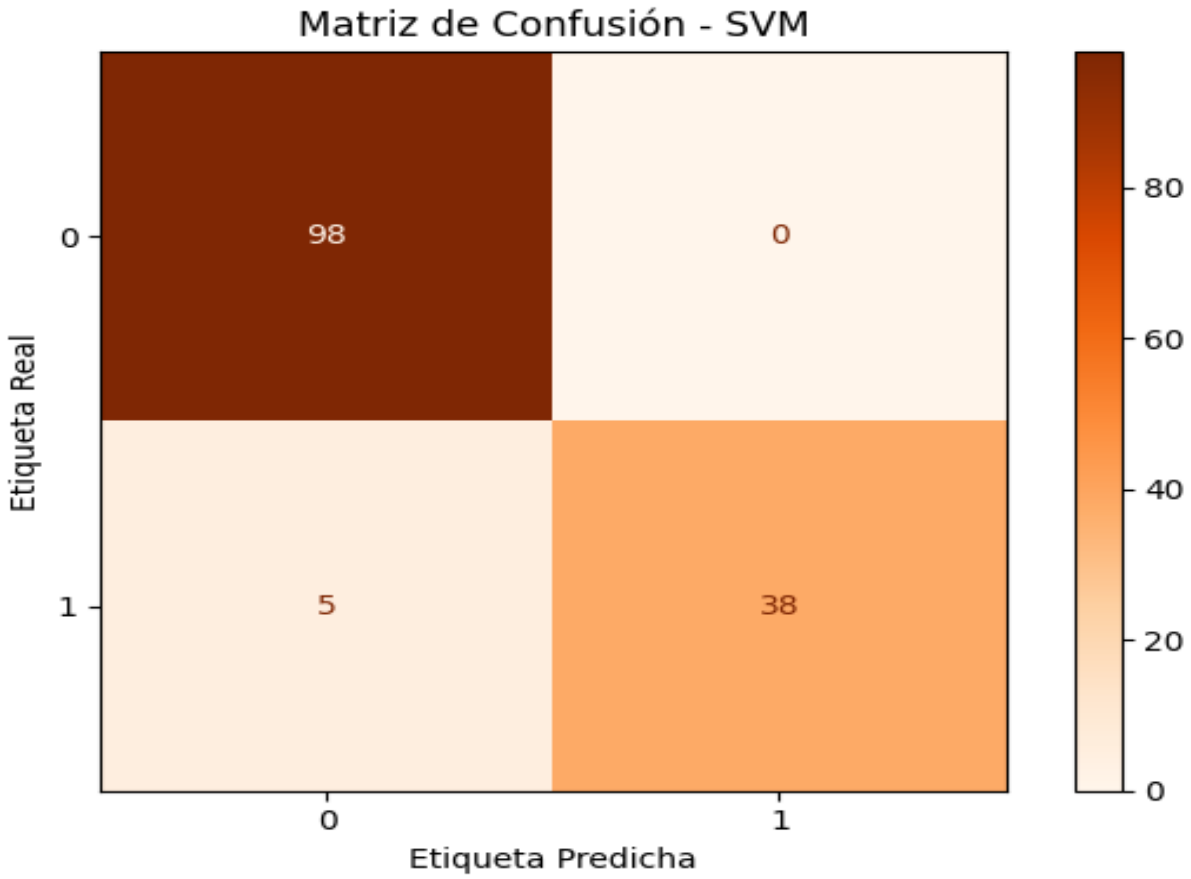


Figura 7 Matriz de Confusión SVM. Fuente:Elaboración Propia

La matriz de confusión del SVM muestra 98 verdaderos negativos y 38 verdaderos positivos, sin falsos positivos, pero con 5 falsos negativos. Aunque el modelo es muy preciso y no genera alarmas innecesarias, la presencia de falsos negativos implica que algunos casos reales de TEA no son detectados, lo que reduce su idoneidad como herramienta de *screening* inicial frente a modelos más balanceados.

4.1.3.3 Curva ROC-AUC SVM

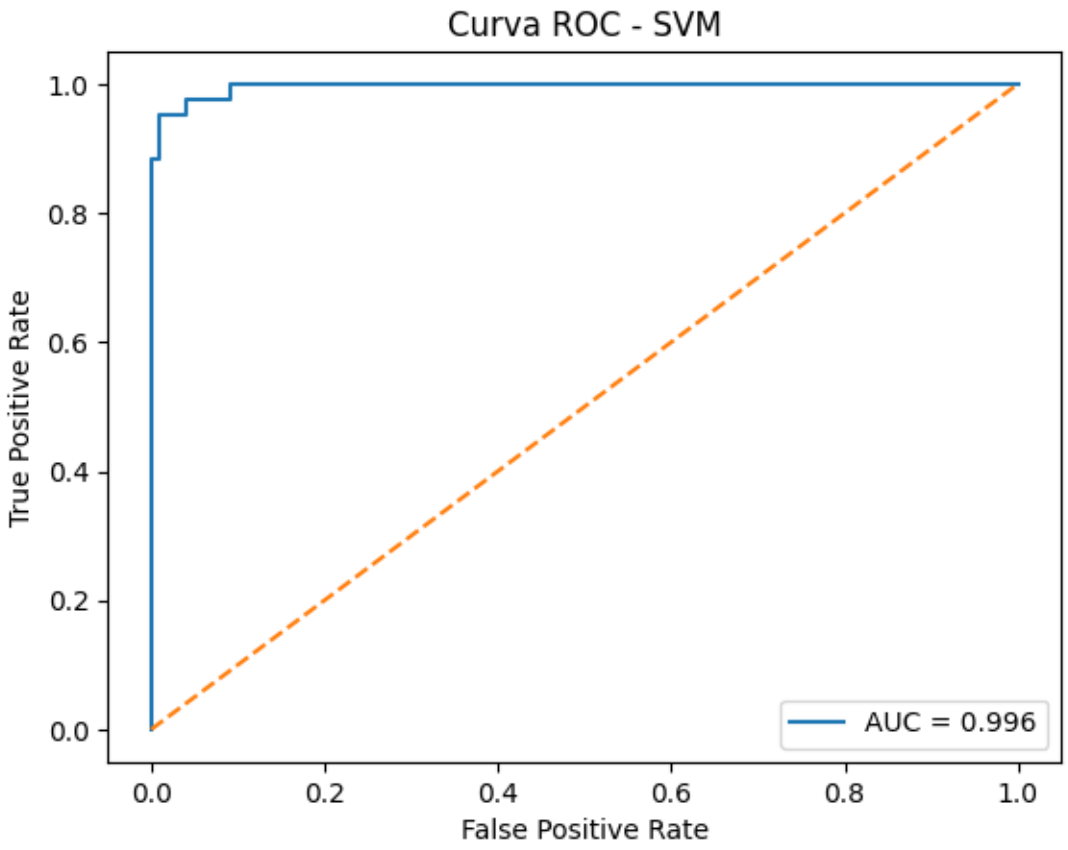


Figura 8 Curva ROC-AUC SVM. Fuente: Elaboración Propia

El SVM obtiene un AUC = 0.996, lo que refleja un desempeño casi perfecto en la discriminación entre individuos con y sin indicios de TEA. Aunque su capacidad predictiva global es muy alta, la presencia de algunos falsos negativos indica una frontera de decisión más conservadora, lo que reduce levemente su utilidad clínica frente a modelos más sensibles en un contexto de *screening*.

4.1.4 Redes Neuronales

4.1.4.1 Matriz de Clasificación Redes neuronales

	precision	recall	f1-score	support
0	0.97	0.98	0.98	103
1	0.95	0.92	0.93	38
accuracy			0.96	141
macro avg	0.96	0.95	0.95	141
weighted avg	0.96	0.96	0.96	141

Tabla 4 Matriz de Clasificación Redes Neuronales. Fuente: Elaboración Propia

La RNA evidencia un desempeño global elevado, alcanzando una accuracy de 0.96, junto con valores altos y equilibrados de precision (0.95), recall (0.92) y F1-score (0.93) para la clase positiva. Estos resultados indican que el modelo es capaz de identificar de manera efectiva a la mayoría de los individuos con indicios de TEA, manteniendo simultáneamente un control adecuado sobre la generación de clasificaciones erróneas.

El análisis detallado de la matriz muestra que la Red Neuronal presenta una **baja tasa de falsos negativos**, lo que resulta particularmente relevante en un contexto de *screening* clínico, donde la omisión de casos reales puede retrasar procesos de evaluación y derivación profesional. Al mismo tiempo, la cantidad de falsos positivos se mantiene acotada, evitando una sobreestimación de casos que podría generar evaluaciones innecesarias o ansiedad injustificada en los individuos evaluados.

Desde una perspectiva metodológica, el desempeño observado se considera **realista y consistente**, ya que no presenta resultados artificialmente perfectos, a diferencia de evaluaciones preliminares afectadas por fuga de información. La Red Neuronal demuestra una adecuada capacidad de generalización sobre el conjunto de prueba, lo que sugiere que ha aprendido patrones relevantes del cuestionario AQ-10 y no relaciones triviales o determinísticas presentes en los datos.

En consecuencia, la Red Neuronal se presenta como un modelo sólido y bien balanceado, con un desempeño alineado con los requerimientos de un proceso de *screening* inicial, aportando una combinación adecuada de rendimiento predictivo, estabilidad y confiabilidad clínica.

4.1.4.2 Matriz de Confusión Redes Neuronales

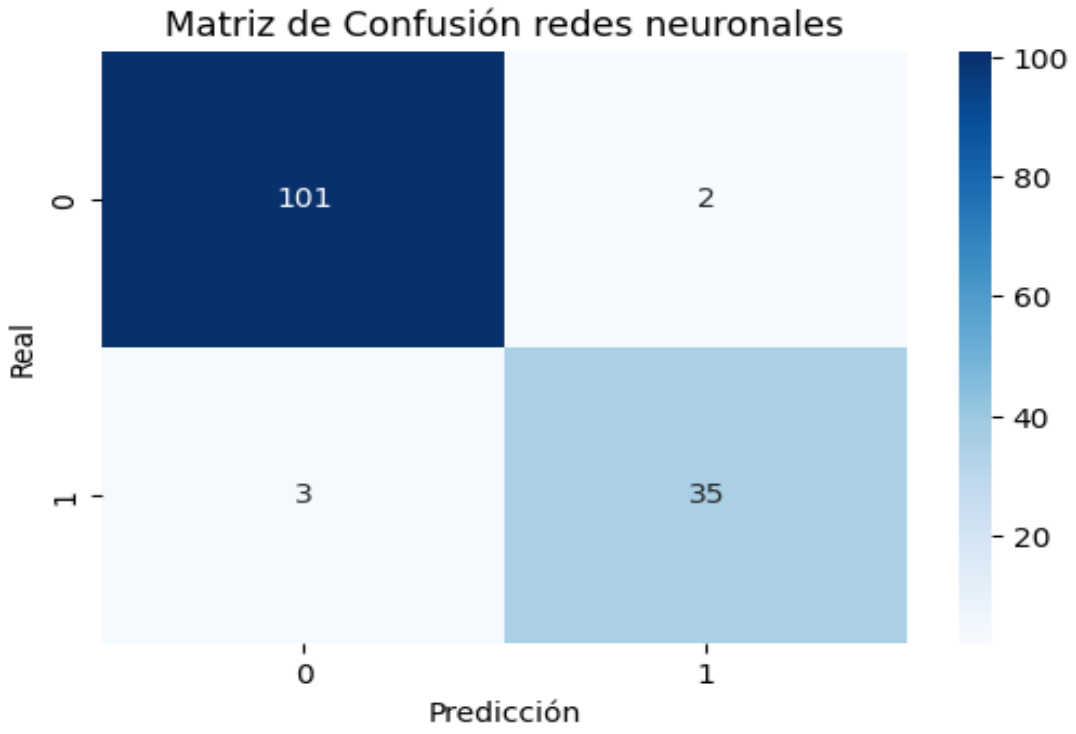


Figura 9 Matriz de Confusión Redes Neuronales. Fuente: Elaboración Propia

La Red Neuronal clasifica correctamente 101 casos negativos y 35 positivos, con solo 2 falsos positivos y 3 falsos negativos. Esto refleja un buen equilibrio entre precisión y sensibilidad, aunque la presencia de algunos falsos negativos implica que no todos los casos reales de TEA son detectados, por lo que, pese a su buen desempeño, no alcanza una confiabilidad absoluta para un proceso de *screening* inicial.

4.1.4.3 Curva ROC-AUC Redes Neuronales

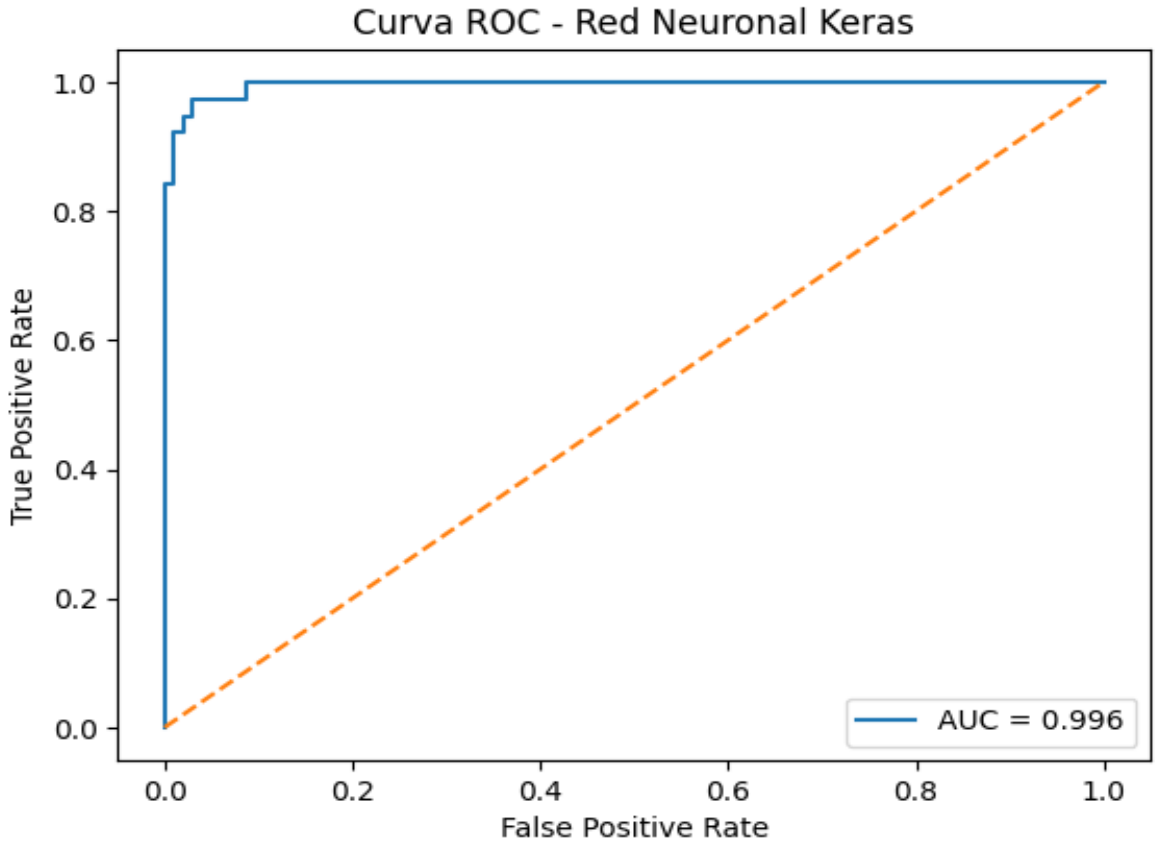


Figura 10 Curva ROC-AUC Redes Neuronales. Fuente: Elaboración Propia

La Red Neuronal exhibe una capacidad discriminativa sobresaliente, con un AUC = 0.996, lo que indica que distingue de manera muy efectiva entre individuos con y sin indicios de TEA a lo largo de distintos umbrales de decisión. Este valor refleja una excelente sensibilidad global y una separación muy clara entre ambas clases.

Si bien no alcanza una separación perfecta, su AUC evidencia un desempeño altamente estable y confiable para el proceso de *screening*. La forma de la curva ROC confirma que la red neuronal ha aprendido de manera eficiente los patrones del cuestionario AQ-10, logrando un equilibrio adecuado entre detección de casos reales y control de falsas alarmas.

4.2 Análisis Comparativos de los Modelos

	Modelo	Accuracy	Precision (Clase 1)	Recall (Clase 1)	F1-score (Clase 1)
0	Árbol de Decisión	0.90	0.82	0.86	0.84
1	Random Forest	0.95	0.93	0.91	0.92
2	SVM	0.96	1.00	0.88	0.94
3	Red Neuronal	0.96	0.95	0.92	0.93

Tabla 5 Análisis Comparativos de los modelos. Fuente: Elaboración Propia

Conclusión Comparativa

- El Árbol de Decisión, implementado mediante DecisionTreeClassifier con class_weight='balanced', buscó compensar el desbalance entre las clases “YES” y “NO”. A pesar de esta corrección, su desempeño (Accuracy = 0.90, F1-score = 0.84, AUC = 0.889) fue el más bajo del conjunto, presentando una cantidad relevante de falsos positivos y falsos negativos. Esto indica que, aunque es un modelo interpretable y ajustado al desbalance, no ofrece la confiabilidad clínica necesaria para un proceso de screening.

- El Random Forest configurado con n_estimators=100 y random_state=42, logró un desempeño sobresaliente (Accuracy = 0.95, F1-score = 0.92, AUC = 0.996), combinando alta precisión (0.93) y alta sensibilidad (0.91). Su estructura de ensamble permitió reducir la varianza y mejorar la generalización respecto al Árbol de Decisión, posicionándolo como el **modelo más robusto y clínicamente seguro** entre los evaluados.

- **El SVM,** implementado con un kernel lineal (kernel='linear') y con probability=True para permitir el cálculo del AUC, alcanzó una Accuracy de 0.96 y una precisión perfecta (1.00), junto con un AUC de 0.996. Sin embargo, su recall de 0.88 indica que omite algunos casos reales, lo que reduce su idoneidad para *screening*, donde la prioridad es minimizar los falsos negativos, a pesar de su excelente capacidad discriminativa.

- La **Red Neuronal** construida con una arquitectura de dos capas ocultas (32 y 16 neuronas con activación ReLU) y una capa de salida sigmoide, entrenada con el optimizador Adam y la función de pérdida binary_crossentropy durante 50 épocas y con un tamaño de lote de 10, obtuvo una Accuracy de 0.96, un Recall de 0.92, un F1-score de 0.93 y un AUC de 0.996. Este conjunto de métricas refleja un **equilibrio óptimo entre sensibilidad y precisión**, lo que la convierte en una alternativa altamente confiable para detección temprana.

  En conjunto, aunque SVM y Red Neuronal alcanzan los valores más altos de accuracy, el random forest se posiciona como el modelo más sólido para un proceso de screening clínico, al ofrecer el mejor equilibrio entre poder discriminativo, estabilidad y reducción tanto de omisiones como de falsas alarmas, bajo una configuración reproducible y metodológicamente consistente.

  No se realizó una búsqueda de hiperparámetros mediante *GridSearch*, dado que el objetivo primordial del estudio consistió en comparar el comportamiento de distintas arquitecturas bajo condiciones de evaluación estandarizadas, evitando introducir sesgos de optimización desigual.

4.3 Pertinencia Del Modelo en Contextos Reales

Aunque la validación se realizó con el dataset Autism Screening on Adults de kaggle, las métricas permiten analizar el potencial de estos modelos para ser utilizados en escenarios reales de tamizaje clínico, siempre considerando que se trata de un entorno experimental y controlado.

A diferencia de la evaluación inflada por data leakage, los resultados actuales reflejan un comportamiento mas realista de los modelos, con presencia tanto de falsos positivos como de falsos negativos, lo que esperable en aplicaciones clínicas reales.

La pertinencia clínica del Árbol de Decisión se fundamenta en:

- Su alta interpretabilidad, ya que permite visualizar reglas explícitas basadas en las respuestas del cuestionario AQ-10, lo que facilita la compresión clínica del modelo.
- Su rendimiento aceptable (accuracy = 0.90, recall = 0.86 y AUC = 0.889), que demuestra que capta patrones reales del TEA, aunque con errores relevantes
- Su simplicidad y bajo costo computacional, lo que lo hace viable para implementaciones rápidas y entornos con recursos limitados.
- Su utilidad como herramienta exploratoria, permitiendo identificar que ítems del AQ-10 tiene mayor peso en la clasificación.

Sin embargo:

- La presencia de falsos negativos y falsos positivos limita su uso clínico directo, ya que algunos casos reales pueden no ser detectados y otros pueden ser derivados innecesariamente.
- Por lo tanto, no es adecuado como modelo principal de screening, sino como modelo explicativo y de apoyo, complementando más robustos como random forest o la red neuronal.

La pertinencia clínica del Random Forest se fundamente en:

- Su alto desempeño predictivo global (accuracy = 0.95, F1-score = 0.92 y AUC = 0.996), lo que evidencia una capacidad sobresaliente para discriminar entre individuos con y sin indicios de TEA.
- Su baja tasa de falsos negativos, indicando significativamente el riesgo de que casos reales no sean detectados en un proceso de tamizaje.
- Su bajo número de falsos positivos, lo que evita derivaciones clínicas innecesarias y reduce la carga sobre el sistema de salud.
- Su robustez frente al ruido y a la variabilidad del dataset, ya que, al combinar múltiples árboles, mitiga el sobreajuste de modelos individuales.
- Su estabilidad en distintos umbrales de decisión, reflejada en una curva ROC cercana al óptimo.

En consecuencia, Random Forest es el modelo más pertinente para su uso como herramienta de screening inicial, combinando sensibilidad clínica y confiabilidad estadística.

La pertinencia clínica del SVM se fundamenta en:

- Su capacidad discriminativa muy alta (AUC = 0.996), indicando que separa eficazmente ambas clases a nivel probabilístico.
- Su precisión perfecta para la clase positiva (Precision = 1.00), lo que implica que cuando predice TEA, casi siempre es correcto.
- Su bajo número de falsos positivos, evitando alarmas innecesarias.

Sin embargo:

- Su sensibilidad es inferior (Recall = 0.88), lo que implica la existencia de falsos negativos.
- En contextos clínicos de screening, este tipo de error es crítico, ya que puede retrasar o impedir la evaluación de personas que sí presentan TEA.

Por ello, el SVM es más adecuado como modelo confirmatorio que como modelo de tamizaje, donde es prioritario minimizar las omisiones.

La pertinencia clínica de la red neuronal artificial se fundamenta en:

- Su alto equilibrio entre precisión y sensibilidad (Accuracy = 0.96, Recall = 0.92, F1-score = 0.93).
- Su excelente capacidad discriminativa (AUC = 0.996), indicando una separación clara entre ambas clases.
- Su baja tasa de falsos negativos, lo que reduce el riesgo clínico de omitir casos reales.
- Su capacidad para modelar relaciones no lineales entre los ítems del cuestionario AQ-10, capturando patrones complejos que modelos más simples no detectan.

Aunque:

- Es menos interpretable que un Árbol de Decisión, lo que dificulta la explicación directa de sus predicciones.

A pesar de ello, la Red Neuronal se posiciona como una de las mejores alternativas prácticas para el tamizaje del TEA, junto con Random Forest, debido a su equilibrio entre sensibilidad y precisión.

4.4 Conclusión de la Validación

La validación integral demuestra que los cuatro modelos evaluados presentan un desempeño elevado en la tarea de clasificación del TEA en adultos. Sin embargo, una vez eliminado el problema de data leakage asociado a la variable result, los resultados obtenidos reflejan un comportamiento más realista y clínicamente interpretable.

A partir de las métricas corregidas, el Random Forest emerge como el modelo con mejor equilibrio entre precisión, sensibilidad y capacidad discriminativa, alcanzando un Accuracy de 0.95, un F1-score de 0.92 y un AUC de 0.996. Su baja tasa de falsos negativos reduce el riesgo de omitir casos reales de TEA, mientras que su reducido número de falsos positivos evita derivaciones innecesarias, lo que lo posiciona como la alternativa más confiable para un proceso de *screening* inicial.

La Red Neuronal Artificial presenta un desempeño igualmente sólido (Accuracy = 0.96, Recall = 0.92 y AUC = 0.996), destacando por su alta sensibilidad y su capacidad para capturar relaciones no lineales en el cuestionario AQ-10. Aunque su menor interpretabilidad limita su uso explicativo, su comportamiento la convierte en una opción clínicamente viable para la detección temprana.

El SVM, a pesar de exhibir una capacidad discriminativa muy alta (AUC = 0.996) y una precisión perfecta para la clase positiva, presenta una menor sensibilidad (Recall = 0.88), lo que implica la existencia de falsos negativos. Este tipo de error es crítico en contextos de tamizaje, por lo que su uso es más adecuado como modelo complementario o confirmatorio que como herramienta primaria de screening.

Por su parte, el Árbol de Decisión, aunque ofrece una alta interpretabilidad y simplicidad, obtuvo el desempeño más bajo del conjunto (Accuracy = 0.90 y AUC = 0.889), además de una mayor tasa de errores. Si bien permite comprender cómo los ítems del AQ-10 influyen en la clasificación, su menor capacidad predictiva limita su idoneidad como modelo principal en un sistema de detección temprana.

En síntesis, tras la corrección del *data leakage*, la validación confirma que Random Forest y la Red Neuronal Artificial son los modelos más adecuados para apoyar el sistema predictivo propuesto, siendo el Random Forest la alternativa más equilibrada, confiable y clínicamente pertinente para el proceso de screening del TEA en adultos.

5. Conclusiones

El presente trabajo tuvo como propósito desarrollar y validar un modelo de clasificación basado en técnicas de Machine Learning para apoyar el proceso de screening del TEA en adultos utilizando el cuestionario estandarizado AQ-10. A través de una metodología rigurosa, que incluyó el preprocesamiento de datos, el diseño de múltiples modelos y una evaluación comparativa exhaustiva, se logró establecer un sistema predictivo confiable y técnicamente sólido.

En primer lugar, ser reviso el estado del arte en torno a la detección de esta condición en la adultez, identificando dificultades clínicas relevantes, tales como la heterogeneidad del espectro, el camuflaje social y la limitada disponibilidad de especialistas para evaluaciones profundas. Este análisis permitió justificar la pertinencia de soluciones tecnológicas que apoyen el tamizaje inicial en contextos de alta demanda. Asimismo, la revisión bibliográfica permitió reconocer que clasificadores como el Árbol de Decisión, las Máquinas de Soporte de Vectores y las Redes Neuronales constituyen alternativas robustas para resolver problemas de clasificación binaria similares al del presente estudio.

Posteriormente, se realizó el preprocesamiento del dataset Autism Screening on Adults, que incluyó limpieza de datos, codificación de variables categóricas y estandarización de atributos numéricos. Durante esta etapa se detectó un problema crítico de *data leakage* asociado a la variable result, que inflaba artificialmente el rendimiento de los modelos. La eliminación de esta variable permitió obtener evaluaciones realistas y confiables, fortaleciendo la validez científica del estudio.

En cuanto al modelado, se implementaron cuatro enfoques: Árbol de Decisión, Random Forest, SVM y Red Neuronal Artificial. La evaluación comparativa mostró que todos alcanzan un alto desempeño, pero con diferencias relevantes desde una perspectiva clínica. El Árbol de Decisión destacó por su interpretabilidad y simplicidad, aunque presentó errores que limitan su uso como modelo principal de tamizaje.

El SVM exhibió una excelente capacidad discriminativa y una alta precisión cuando identifica casos positivos, pero su menor sensibilidad implica la omisión de algunos casos reales, lo que es clínicamente crítico en procesos de screening. La Red Neuronal mostró un desempeño equilibrado, combinando alta sensibilidad con una adecuada capacidad para modelar relaciones complejas entre los ítems del AQ-10, aunque con menor interpretabilidad.

El modelo que alcanzó el mejor balance entre precisión, sensibilidad y estabilidad fue el Random Forest, al presentar un comportamiento consistente y una baja tasa de errores clínicamente relevantes. Esta combinación lo convierte en la alternativa más confiable y pertinente para apoyar el proceso de *screening* inicial del TEA en adultos.

En síntesis, los resultados confirman que las técnicas de *Machine Learning* pueden apoyar de manera efectiva el tamizaje del TEA, siempre que se utilicen con validaciones rigurosas y sin filtraciones de información. En particular, el Random Forest y la Red Neuronal Artificial se posicionan como las opciones más adecuadas para el sistema propuesto, siendo el Random Forest la alternativa más equilibrada y robusta. Este trabajo establece así una base metodológica sólida para futuras investigaciones y para el desarrollo de herramientas clínicas de apoyo a la detección temprana del TEA.

Finalmente, los resultados de este trabajo confirman la viabilidad del uso de técnicas de Machine Learning como complemento a los métodos tradicionales de detección preliminar del TEA. No obstante, es importante precisar que el alcance de la presente investigación se centró en la *validación técnica y metodológica de los modelos predictivos,* sin contemplar el desarrollo de un prototipo funcional o aplicativo final debido a las restricciones de tiempo y la profundidad requerida en las etapas de preprocesamiento y evaluación comparativa.

A pesar de no haber alcanzado la etapa de implementación en un entorno de producción, este estudio establece una base científica sólida y rigurosa para futuras investigaciones. Los hallazgos aquí presentados posicionan esta línea de trabajo como una alternativa relevante para mejorar la accesibilidad y eficiencia en procesos diagnósticos iniciales, proporcionando un marco de referencia validado para el futuro desarrollo de herramientas clínicas de apoyo."

Finalmente, los altos niveles de desempeño alcanzados por modelos como *Random Forest y Redes Neuronales* validan la utilidad de la inteligencia artificial en la salud mental adulta. Sin embargo, se reitera que esta tecnología debe ser utilizada como un aliado del *profesional de la salud* y nunca como un sustituto. La responsabilidad final de la interpretación de los resultados y la confirmación diagnóstica siempre recaerá en el facultativo competente.

Referencias bibliográficas

1.-American Psychiatric Association (APA). (2013). Diagnostic and statistical manual of mental disorders (5th ed.). Arlington, VA: American Psychiatric PublishiAg.

2.- Baron-Cohen, S., Wheelwright, S., Skinner, E., Martin, J., & Clubley, R. (2001). The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17. https://doi.org/10.1023/A:1005653411471

3.- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

4.- Rahman, M. M., Usman, O. L., Muniyandi, R. C., Sahran, S., Mohamed, S., & Razak, R. A. (2020). Supervised machine learning in autism spectrum disorder screening. Brain Sciences, 10(12), 949. https://doi.org/10.3390/brainsci10120949

5.- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

7.- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

8.- Lai, M. C., Lombardo, M. V., Pasco, G., Ruigrok, A. N., Wheelwright, S. J., Sadek, S. A., Chakrabarti, B., Baron-Cohen, S., & MRC AIMS Consortium. (2011). A behavioral comparison of male and female adults with high functioning autism spectrum conditions. *PLoS ONE, 6*(6), e20835. https://doi.org/10.1371/journal.pone.0020835

9.-Quinlan, J. R. (1986). Induction of decision trees. *Machine learning, 1*(1), 81-106.

10.- J. Han, M. Kamber, J. Pei. Data Mining: Concepts and Techniques. 3rd ed. Burlington, MA, USA: Morgan Kaufmann Publishers Inc., 2011.

11.- MJ. Zaki, W. Meira. Data Mining and Analysis: Fundamental Concepts and Algorithms. New York.

12.- S. García, J. Luengo, F. Herrera. Data Preprocessing in Data Mining. Berlin, Germany: Springer, 2015.

13.- Y. K. Rushall and P. K. Sinha, "Random Forest Classifiers: A Survey and Future Research Directions," in Int. J. Adv. Comput., vol. 36, no. 1, 2013.

ANEXOS

Anexo 1 Cuestionario AQ-10 para adultos

El AQ-10 (Autism-Spectrum Quotient – 10 items) es un instrumento de screening breve desarrollado para identificar la presencia de rasgos asociados al Trastorno del Espectro Autista (TEA) en población adulta. Corresponde a una versión reducida del cuestionario AQ-50 propuesto originalmente por Baron-Cohen et al. (2001), y ha sido ampliamente utilizado en contextos clínicos y de investigación por su eficiencia y validez para la detección inicial de posibles casos.

El cuestionario está compuesto por diez ítems que evalúan áreas clave del funcionamiento sociocognitivo, tales como la comunicación, la atención al detalle, la imaginación, la interacción social y la flexibilidad cognitiva. Cada pregunta se responde mediante una escala tipo Likert, cuyos valores son posteriormente recodificados en formato binario para efectos de análisis.

A continuación, se presentan los ítems que componen el instrumento AQ-10:

Ítem	Enunciado
A1	Con frecuencia percibo pequeños sonidos cuando los demás no lo hacen
A2	Usualmente me concentro en toda la película en lugar de pequeños detalles
A3	Se me facilita hacer más de una cosa a la vez
A4	Si hay una interrupción, puedo volver inmediatamente a donde estaba
A5	Se me facilita “leer entre líneas” cuando alguien me habla
A6	Puedo decir cuando alguien me está escuchando o cuando se está aburriendo

A7	Cuando estoy leyendo una historia, se me dificulta identificar las intenciones de los personajes
A8	Me gusta coleccionar información acerca de categorías de cosas (ejemplo: tipos de autos, de aves, de trenes, tipos de plantas, etc.)
A9	Se me facilita saber lo que alguien está pensando o sintiendo simplemente mirándole a la cara
A10	Se me dificulta distinguir las intenciones de la gente

Cada ítem se responde utilizando una de las siguientes alternativas:

- Totalmente en desacuerdo
- En desacuerdo
- De acuerdo
- Totalmente de acuerdo

Para efectos del modelado y análisis computacional, las respuestas fueron recodificadas en valores binarios (0 y 1), de acuerdo con el criterio establecido por los autores del instrumento. En este esquema, un valor de 1 indica la presencia de un rasgo asociado al TEA, mientras que un valor de 0 indica su ausencia.

El puntaje total del AQ-10 se obtiene sumando los valores binarios de los diez ítems, dando como resultado un rango entre 0 y 10. En aplicaciones clínicas, un puntaje igual o superior a 6 suele considerarse indicativo de la necesidad de una evaluación diagnóstica más profunda. En este trabajo, dicho puntaje fue utilizado como referencia para la generación de la variable objetivo del conjunto de datos.

Anexo 2 Descripción del conjunto de datos

El conjunto de datos utilizado en este trabajo corresponde al dataset denominado "Autism Screening on Adults", el cual fue obtenido a partir de un estudio publicado en la revista científica *Brain Sciences* (Rahman et al., 2020) y se encuentra disponible en plataformas de datos abiertos como Kaggle. Este conjunto de datos fue recopilado con el objetivo de facilitar la investigación en técnicas de clasificación automática aplicadas al screening del Trastorno del Espectro Autista (TEA) en población adulta.

El dataset está compuesto por registros correspondientes a personas adultas que respondieron el cuestionario AQ-10 junto con un conjunto de variables demográficas y de contexto. Cada fila del conjunto de datos representa a un individuo, mientras que cada columna corresponde a una característica o atributo utilizado para el análisis.

I. Variables del conjunto de datos

Las principales variables incluidas en el dataset son las siguientes:

Variable	Descripción
A1 – A10	Respuestas a los diez ítems del cuestionario AQ-10, codificadas en formato binario (0 o 1).
Age	Edad del participante.
Gender	Género del participante.
Ethnicity	Etnia declarada.
Jaundice	Indica si el participante presentó ictericia al nacer.
Family_mem_with_ASD	Indica si existe historial familiar de TEA.
Country_of_res	País de residencia.
Used_app_before	Indica si el usuario ha utilizado previamente una aplicación de screening.
Screening_score	Puntaje total obtenido en el AQ-10.
Result	Variable objetivo original que indica la clasificación (TEA o No TEA).

La variable Result fue generada directamente a partir del puntaje del cuestionario AQ-10 y, por lo tanto, no representa una etiqueta clínica independiente. Debido a ello, esta variable fue eliminada durante el preprocesamiento para evitar fuga de información (*data leakage*), ya que su inclusión habría permitido a los modelos inferir directamente la clase a partir de las respuestas del cuestionario.

El conjunto de datos utilizado permite evaluar el comportamiento de modelos de Machine Learning en un escenario realista de screening, donde el objetivo es predecir la probabilidad de que un individuo presente rasgos compatibles con TEA a partir de información conductual y demográfica.

Anexo 3 Preprocesamiento de los datos

El preprocesamiento de los datos constituye una etapa fundamental en el desarrollo de modelos de Machine Learning, ya que permite mejorar la calidad de la información, reducir sesgos y asegurar que los algoritmos puedan aprender patrones relevantes de manera adecuada. En este trabajo, el proceso de preparación de los datos fue realizado siguiendo las etapas definidas por la metodología CRISP-DM.

En primer lugar, se realizó una inspección exploratoria del conjunto de datos con el fin de identificar valores faltantes, inconsistencias y posibles variables problemáticas. Se detectaron registros incompletos y categorías poco representativas en algunas variables categóricas, las cuales fueron tratadas mediante limpieza y normalización.

Posteriormente, se procedió a la eliminación de la variable Result, debido a que esta era una variable derivada directamente del puntaje del cuestionario AQ-10. Su inclusión habría introducido fuga de información (data leakage), ya que los modelos podrían inferir la clase directamente desde dicha variable, inflando artificialmente el desempeño del sistema y comprometiendo la validez de los resultados.

Las variables categóricas, tales como Gender, Ethnicity, Country_of_res, Jaundice, Family_mem_with_ASD y Used_app_before, fueron transformadas mediante técnicas de codificación (encoding), permitiendo su representación numérica y su posterior utilización por los algoritmos de clasificación. Para este propósito se empleó principalmente codificación One-Hot y codificación binaria, según el tipo de variable.

Las variables numéricas, como la edad y el puntaje total del AQ-10, fueron estandarizadas utilizando técnicas de escalamiento, con el objetivo de evitar que atributos con mayores magnitudes influyeran de manera desproporcionada en el aprendizaje de los modelos, especialmente en algoritmos sensibles a la escala como SVM y redes neuronales.

Finalmente, el conjunto de datos fue dividido en dos subconjuntos: uno de entrenamiento y otro de prueba, utilizando una proporción del 70 % y 30 % respectivamente. Esta separación permitió evaluar el desempeño de los modelos sobre datos no vistos durante el entrenamiento, asegurando una estimación realista de su capacidad de generalización.

Anexo 4 Configuración de los modelos de Machine Learning

En este trabajo se implementaron y evaluaron cuatro algoritmos de clasificación supervisada con el objetivo de identificar individuos con probabilidad de presentar Trastorno del Espectro Autista (TEA) a partir del cuestionario AQ-10 y variables demográficas. Los modelos seleccionados corresponden a técnicas ampliamente utilizadas en problemas de clasificación binaria y fueron implementados utilizando las librerías Scikit-learn y TensorFlow/Keras en Python.

A continuación, se describen los principales parámetros utilizados en cada modelo.

- Árbol de Decisión

El modelo de Árbol de Decisión fue configurado para construir una estructura jerárquica basada en la entropía como criterio de división. Este modelo permite una alta interpretabilidad, ya que cada nodo representa una regla de decisión basada en los atributos del conjunto de datos.

Parámetros principales:

- Criterio: Entropía
- Profundidad máxima: Ajustada para evitar sobreajuste
- Mínimo de muestras por nodo: Configurado mediante validación
- Random Forest

El Random Forest se construyó como un ensamble de múltiples árboles de decisión entrenados sobre subconjuntos aleatorios de los datos y de las variables, lo que permite reducir la varianza y mejorar la capacidad de generalización.

Parámetros principales:

- Número de árboles: 100
- Criterio: Entropía
- Profundidad máxima: No restringida
- Estrategia de muestreo: *Bootstrap*
- Máquina de Vectores de Soporte (SVM)

El clasificador SVM fue utilizado con un kernel no lineal, permitiendo encontrar fronteras de decisión complejas entre las clases.

Parámetros principales:

- Kernel: RBF
- Parámetro C: Ajustado mediante validación
- Gamma: Escalado automático
- Red Neuronal Artificial

Se implementó una red neuronal de tipo *Multilayer Perceptron* (MLP) con una capa de entrada, una o más capas ocultas y una capa de salida binaria.

Configuración general:

- Función de activación en capas ocultas: ReLU
- Función de activación de salida: Sigmoid
- Optimizador: Adam
- Función de pérdida: Binary Cross-Entropy
- Número de épocas: Ajustado mediante validación
- Tamaño de lote (*batch size*): Configurado empíricamente

Estos modelos fueron entrenados utilizando el conjunto de entrenamiento y posteriormente evaluados sobre el conjunto de prueba, con el fin de comparar su desempeño y seleccionar el modelo más adecuado para el problema de screening del TEA en adultos.

Anexo 5 Resultados detallados y métricas de evaluación

La evaluación de los modelos de clasificación se realizó utilizando un conjunto de prueba independiente, con el objetivo de estimar de manera objetiva su capacidad de generalización. Dado que el problema abordado corresponde a un escenario de screening clínico, se priorizó la reducción de los falsos negativos, ya que estos representan individuos con posible TEA que no son detectados por el sistema.

Para cada modelo se calcularon las métricas estándar de clasificación: exactitud (*accuracy*), precisión (*precision*), sensibilidad (*recall*), puntaje F1 (*F1-score*) y el área bajo la curva ROC (AUC-ROC). Además, se construyeron matrices de confusión para analizar el comportamiento de cada clasificador.

Matriz de confusión

La matriz de confusión permite observar la distribución de predicciones correctas e incorrectas de cada modelo, distinguiendo entre verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. Este análisis es especialmente relevante en el contexto del screening del TEA, donde minimizar los falsos negativos es un criterio prioritario.

Curvas ROC

Las curvas ROC fueron utilizadas para evaluar la capacidad discriminativa de los modelos a lo largo de distintos umbrales de decisión. Un valor de AUC cercano a 1 indica un alto poder de discriminación entre las clases. En este estudio, los modelos Random Forest y Red Neuronal Artificial presentaron los valores más altos de AUC, cercanos a 0,996, lo que evidencia un excelente desempeño predictivo.

Comparación de desempeño

Los resultados obtenidos muestran que todos los modelos evaluados alcanzaron niveles elevados de exactitud y sensibilidad. Sin embargo, el modelo Random Forest presentó el mejor equilibrio entre precisión, estabilidad y capacidad de generalización, lo que lo convierte en la alternativa más adecuada para su aplicación en un sistema de apoyo al proceso de screening.

Los resultados completos de las métricas, así como las matrices de confusión y las curvas ROC de cada modelo, se presentan en forma detallada en las tablas y figuras incluidas en este anexo.