



**Población de Estudiantes con Autismo en Cali Registrados en el SIMAT-
Sistema de Matrícula Estudiantil de Educación Básica y Media**

**MAESTRIA EN IA APLICADA
MATERIA: PROYECTO 1 DE INNOVACION TECNOLOGICA EN IA**

PROFESOR: Ing. MILTON ORLANDO SARRIA PAJA

ALUMNOS:

Edwin Pérez

Cristian Quebrada

Rubén Darío Sabogal

Octubre de 2025 – Cali

Predicción del Trastorno del Espectro Autista (TEA)

Descripción del Problema

En Colombia, actualmente no existen estadísticas oficiales ni bases de datos abiertas sobre el Trastorno del Espectro Autista (TEA), comúnmente conocido como autismo. Esta ausencia de información limita la visibilidad del problema y dificulta su abordaje como una cuestión de salud pública.

El desconocimiento generalizado sobre el TEA, tanto en la población como en las instituciones de salud, ha generado una falta de atención adecuada. En muchos casos, ni siquiera se conoce cuántos niños presentan esta condición. Las instituciones educativas privadas suelen rechazar el ingreso de niños con TEA, mientras que las entidades públicas, por mandato legal, deben garantizar su inclusión educativa y acceso a terapias que favorezcan su desarrollo.

Lamentablemente, muchos niños permanecen sin escolarización ni tratamiento médico debido al desconocimiento de sus padres, a pesar de que existe una normativa gubernamental vigente que protege sus derechos.

Este interés surge de una experiencia personal: uno de mis hijos ha sido diagnosticado con autismo atípico. En nuestra región, la investigación sobre este tema es escasa. Por ello, sigo de cerca los avances del grupo de investigación en autismo de la Universidad de Stanford y del MediaLab del MIT, liderado por la Dra. Rosalind Picard, pionera en el campo de la **Computación Afectiva**.

Justificación del Uso de Ciencia de Datos e Inteligencia Artificial

Las técnicas de **Ciencia de Datos** y **Inteligencia Artificial (IA)** ofrecen herramientas poderosas para abordar los desafíos relacionados con el diagnóstico, seguimiento y tratamiento del TEA. Estas tecnologías permiten desarrollar soluciones más **eficientes, precisas y personalizadas**.

Entre los enfoques más prometedores se encuentran:

- **Árboles de decisión y Random Forest:** útiles para el diagnóstico temprano y la selección de variables relevantes.
- **Redes neuronales profundas (Deep Learning):** permiten analizar patrones complejos en imágenes y comportamientos.
- **Visión por computadora:** facilita la evaluación automática de conductas y expresiones faciales.

- **Procesamiento de Lenguaje Natural (PLN):** posibilita el análisis automatizado de registros clínicos y respuestas.
- **Plataformas de seguimiento digital:** ofrecen monitoreo personalizado del progreso terapéutico.

Formulación de una Pregunta SMART

Para guiar el análisis, se plantea una pregunta bajo el enfoque SMART, asegurando que sea:

- **Específica:** centrada en un aspecto concreto.
- **Medible:** basada en datos cuantificables.
- **Accionable:** orientada a la toma de decisiones.
- **Realista:** viable con los recursos disponibles.
- **Temporal:** definida en un periodo determinado.

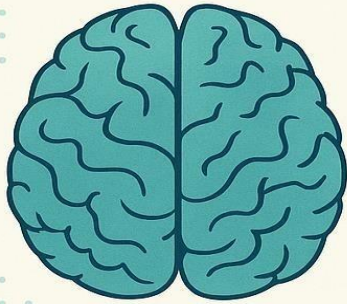
Pregunta SMART:

¿En qué medida el uso de modelos de aprendizaje automático basados en redes neuronales artificiales y en técnicas de análisis de video (visión por computadora) mejora la precisión del diagnóstico temprano de TEA en niños de 18 a 36 meses, en comparación con las evaluaciones clínicas tradicionales, durante un periodo de 12 meses de implementación en una muestra piloto de 100 pacientes en una clínica especializada?

Desglose según criterios SMART:

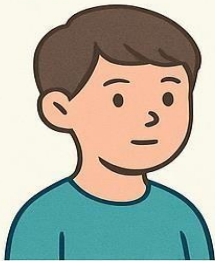
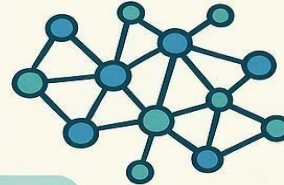
- **Específica:** Evalúa el impacto de modelos de IA en el diagnóstico temprano de TEA.
- **Medible:** Compara la precisión diagnóstica entre métodos automáticos y clínicos.
- **Accionable:** Los resultados permitirán decidir sobre la integración tecnológica en protocolos médicos.
- **Realista:** Utiliza datos clínicos reales, videos y recursos disponibles.
- **Temporal:** El estudio se desarrollará durante un periodo de 12 meses.

USE OF AI FOR

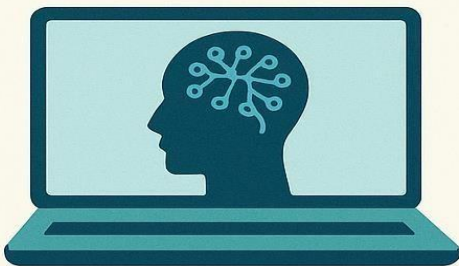
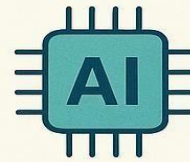


AI, NEURAL NETWORKS
AND TECHNOLOGY

DIAGNOSIS
& THERAPY



CHILDREN
WITH AUTISM
SPECTRUM
DISORDER



ANÁLISIS UNIVARIADO

ANÁLISIS EXPLORATORIO DE DATOS PARA LA PREDICCIÓN DEL TRASTORNO DEL ESPECTRO AUTISTA (TEA)

PREDICCIÓN DEL TRASTORNO DEL ESPECTRO AUTISMO- TEA

El trastorno del espectro autista es una afección relacionada con el desarrollo del cerebro que afecta la manera en la que una persona percibe y socializa con otras personas, lo que

causa problemas en la interacción social y la comunicación. El trastorno también comprende patrones de conducta restringidos y repetitivos. El término «espectro» en el trastorno del espectro autista se refiere a un amplio abanico de síntomas y gravedad.



Conjunto de Datos

El análisis se basa en dos conjuntos de datos públicos disponibles en las siguientes fuentes:

- Kaggle: Autism Screening on Adults
- UCI Machine Learning Repository: Autism Screening Adult Dataset

Para el tratamiento de estos datos, se empleó la Lista de Verificación Modificada para el Autismo en Niños Pequeños, Revisada (M-CHAT-R), una herramienta de cribado ampliamente utilizada para la detección temprana del autismo.

El M-CHAT-R es un cuestionario compuesto por 20 preguntas que los padres o cuidadores responden para identificar posibles señales de autismo en niños entre 16 y 30 meses de edad. Los resultados permiten determinar si es necesaria una evaluación adicional.

Para adultos y niños mayores, se utilizan otras herramientas diagnósticas como el Autism Quotient (AQ) y la Escala de Diagnóstico de Autismo y Asperger de Ritvo (RAADS-14), que evalúan aspectos como el lenguaje, la interacción social y los intereses personales.

Por favor responda a estas preguntas sobre su hijo/a. Tenga en cuenta **cómo su hijo/a se comporta habitualmente**. Si usted ha visto a su hijo/a comportarse de una de estas maneras algunas veces, pero no es un comportamiento habitual, por favor responda **no**. Seleccione, rodeando con un círculo, Muchas gracias.

1. Si usted señala algo al otro lado de la habitación, ¿su hijo/a lo mira? (POR EJEMPLO, Si usted señala a un juguete, un peluche o un animal, ¿su hijo/a lo mira?)	SÍ	NO
2. ¿Alguna vez se ha preguntado si su hijo/a es sordo/a?	SÍ	NO
3. ¿Su hijo/a juega juegos de fantasía o imaginación? (POR EJEMPLO, "hace como que" bebe de una taza vacía, habla por teléfono o da de comer a una muñeca o peluche,...)	SÍ	NO
4. ¿A su hijo le gusta subirse a cosas? (POR EJEMPLO, a una silla, escaleras, o tobogán,...)	SÍ	NO
5. ¿Hace su hijo/a movimientos inusuales con sus dedos cerca de sus ojos? (POR EJEMPLO, mueve sus dedos cerca de sus ojos de manera inusual?)	SÍ	NO
6. ¿Su hijo/a señala con un dedo cuando quiere pedir algo o pedir ayuda? (POR EJEMPLO, señala un juguete o algo de comer que está fuera de su alcance?)	SÍ	NO
7. Su hijo/a señala con un dedo cuando quiere mostrarle algo que le llama la atención? (POR EJEMPLO, señala un avión en el cielo o un camión muy grande en la calle)	SÍ	NO
8. ¿Su hijo/a se interesa en otros niños? (POR EJEMPLO, mira con atención a otros niños, les sonríe o se les acerca?)	SÍ	NO
9. ¿Su hijo/a le muestra cosas acercándolas o levantándolas para que usted las vea – no para pedir ayuda sino solamente para compartirlas con usted? (POR EJEMPLO, le muestra una flor o un peluche o un coche de juguete)	SÍ	NO
10. ¿Su hijo/a responde cuando usted le llama por su nombre? (POR EJEMPLO, se vuelve, habla o balbucea, o deja de hacer lo que estaba haciendo para mirarle?)	SÍ	NO
11. ¿Cuándo usted sonríe a su hijo/a, él o ella también le sonríe?	SÍ	NO
12. ¿Le molestan a su hijo/a ruidos cotidianos? (POR EJEMPLO, la aspiradora o la música, incluso cuando está no está excesivamente alta?)	SÍ	NO
13. ¿Su hijo/a camina solo?	SÍ	NO
14. ¿Su hijo/a le mira a los ojos cuando usted le habla, juega con él o ella, o lo viste?	SÍ	NO
15. ¿Su hijo/a imita sus movimientos? (POR EJEMPLO, decir adiós con la mano, aplaudir o algún ruido gracioso que usted haga?)	SÍ	NO
16. Si usted se gira a ver algo, ¿su hijo/a trata de mirar hacia lo que usted está mirando?	SÍ	NO
17. ¿Su hijo/a intenta que usted le mire/preste atención? (POR EJEMPLO, busca que usted le haga un cumplido, o le dice "mira" ó "mírame")	SÍ	NO
18. ¿Su hijo/a le entiende cuando usted le dice que haga algo? (POR EJEMPLO, si usted no hace gestos, ¿su hijo/a entiende "pon el libro encima de la silla" o "tráeme la manta"?)	SÍ	NO
19. Si algo nuevo pasa, ¿su hijo/a le mira para ver como usted reacciona al respecto? (POR EJEMPLO, si oye un ruido extraño o ve un juguete nuevo, ¿se gira a ver su cara?)	SÍ	NO
20. Le gustan a su hijo/a los juegos de movimiento? (POR EJEMPLO, le gusta que le balancee, o que le haga "el caballito" sentándole en sus rodillas)	SÍ	NO

La valoración o puntaje se determina por:

Algoritmo del puntaje

Para todos los artículos excepto 2, 5 y 12, la respuesta “NO” indica riesgo de TEA; para los artículos 2, 5 y 12, “SI” indica riesgo de TEA. El siguiente algoritmo maximiza las propiedades del M-CHAT-R:

BAJO RIESGO: **Puntaje total de 0-2:** si el niño es menor de 24 meses, evaluar nuevamente luego de cumplidos los 2 años. No hace falta ninguna acción al menos que al observarlo haya algún indicio de riesgo para TEA.

RIESGO MEDIO: **Puntaje total de 3-7:** Administre el Cuestionario de Seguimiento (segunda etapa del M-CHAT-R/F) para obtener información adicional sobre las respuestas de riesgo. Si el puntaje del M-CHAT-R/F continúa en 2 o más, el niño se evalúa como positivo. Acción requerida: refiera al niño para una evaluación de diagnóstico y una evaluación de elegibilidad para intervención temprana. Si el puntaje del Seguimiento es 0-1, la evaluación es negativa. No se requiere ninguna otra acción al menos que por observación se considere en riesgo para TEA. El niño debería ser re-evaluado cuando vaya a sus chequeos anuales futuros.

ALTO RIESGO: **Puntaje total de 8-20:** Es aceptable el de saltar el Cuestionario de Seguimiento y referirlo inmediatamente a una evaluación diagnóstica más una evaluación de elegibilidad para intervención temprana.

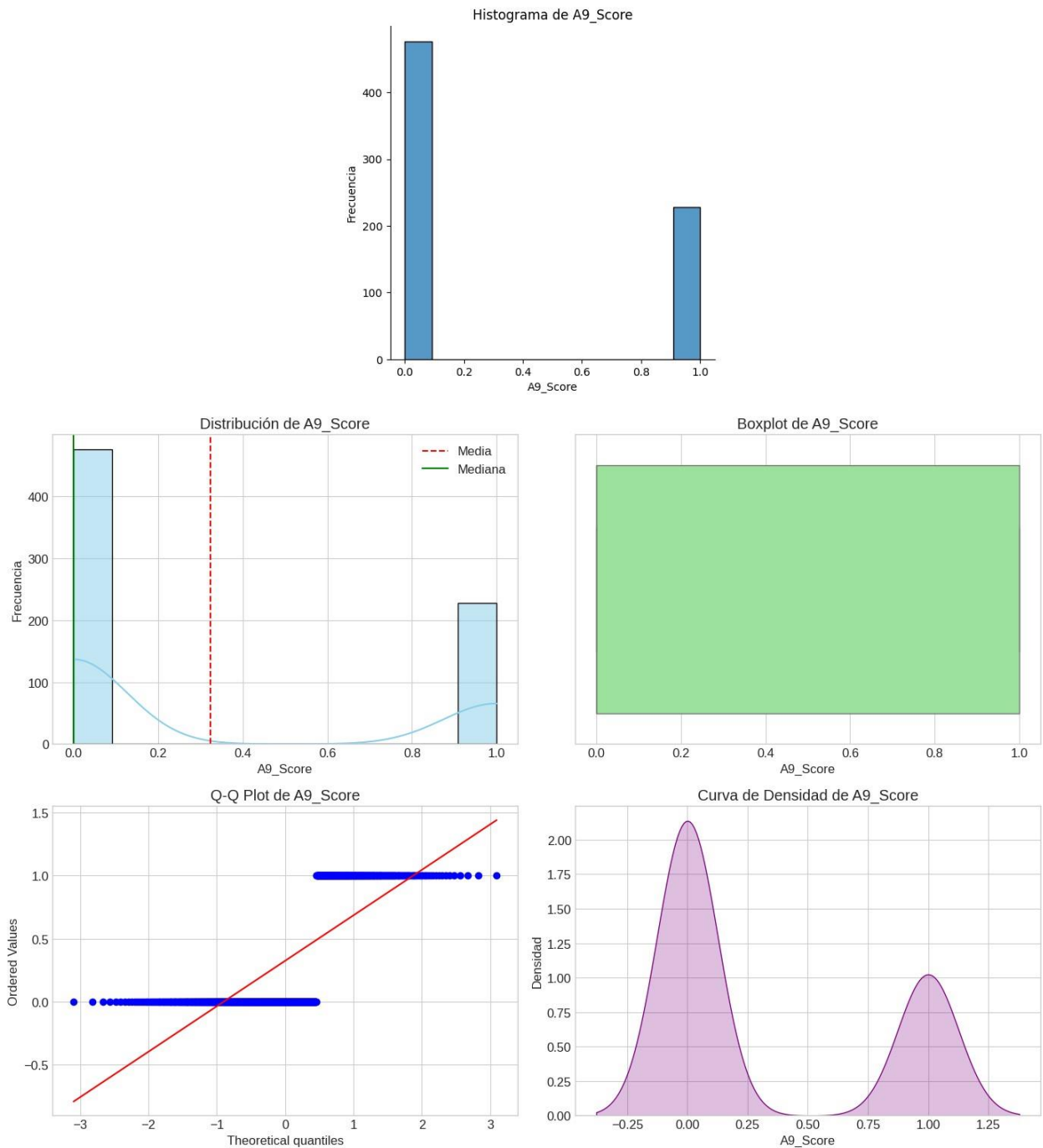
Tratamiento Previo de los Datos y Análisis Univariado

Antes de iniciar el análisis, el conjunto de datos fue normalizado para garantizar su adecuada interpretación y procesamiento estadístico.

El análisis univariado se centró en examinar cada variable de forma individual, con el objetivo de comprender su distribución, identificar patrones relevantes y detectar posibles inconsistencias o problemas.

Entre todas las variables evaluadas, A9_Score se destaca como la más significativa, al presentar el mayor valor de información mutua respecto a la variable objetivo (diagnóstico de autismo). Esta variable corresponde a una pregunta específica del test de cribado y demuestra una alta capacidad para diferenciar entre individuos con y sin TEA, según los datos analizados.

9. ¿Su hijo/a le muestra cosas acercándolas o levantándolas para que usted las vea – no para pedir ayuda sino solamente para compartirlas con usted? (POR EJEMPLO, le muestra una flor o un peluche o un coche de juguete) SÍ NO



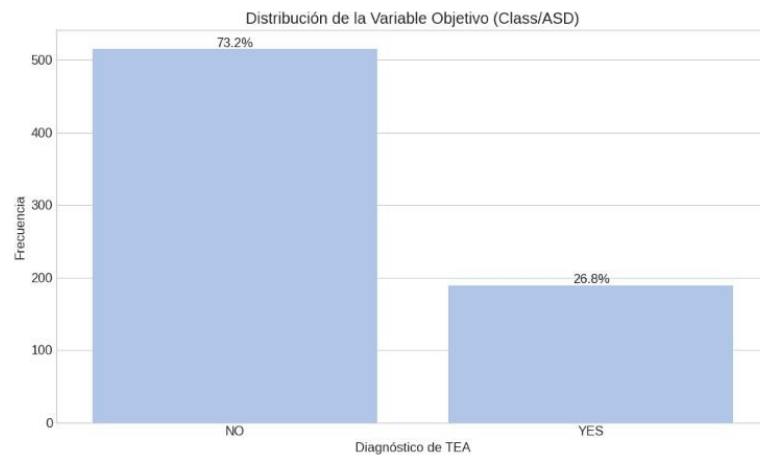
Definición de la Prevalencia del Autismo y Análisis Bivariado

La prevalencia del Trastorno del Espectro Autista (TEA) en los individuos del conjunto de datos se determina a partir de una variable binaria (sí/no), representada por valores centrados en 0 y 1. Esta distribución muestra una simetría clara, evidenciada por la curva

de densidad, y una ausencia de dispersiones significativas, como se observa en el Q-Q plot, lo que facilita la detección de valores atípicos (outliers).

Análisis Bivariado

El **análisis bivariado** se enfoca en examinar las relaciones entre pares de variables, especialmente entre las **variables predictoras** y la **variable objetivo (Class/ASD)**. Este enfoque permite identificar patrones de asociación que pueden mejorar la capacidad de predicción de modelos de diagnóstico para TEA.



ESTADÍSTICAS DESCRIPTIVAS Y GRÁFICOS CLAVE

ESTADÍSTICAS DESCRIPTIVAS GENERALES

Estadísticas descriptivas para variables numéricas:

	count	mean	std	min	25%	50%	75%	max
A1_Score	704.0	0.721591	0.448535	0.0	0.0	1.0	1.0	1.0
A2_Score	704.0	0.453125	0.498152	0.0	0.0	0.0	1.0	1.0
A3_Score	704.0	0.457386	0.498535	0.0	0.0	0.0	1.0	1.0
A4_Score	704.0	0.495739	0.500337	0.0	0.0	0.0	1.0	1.0
A5_Score	704.0	0.498580	0.500353	0.0	0.0	0.0	1.0	1.0
A6_Score	704.0	0.284091	0.451301	0.0	0.0	0.0	1.0	1.0
A7_Score	704.0	0.417614	0.493516	0.0	0.0	0.0	1.0	1.0
A8_Score	704.0	0.649148	0.477576	0.0	0.0	1.0	1.0	1.0
A9_Score	704.0	0.323864	0.468281	0.0	0.0	0.0	1.0	1.0
A10_Score	704.0	0.573864	0.494866	0.0	0.0	1.0	1.0	1.0
age	702.0	29.698006	16.507465	17.0	21.0	27.0	35.0	383.0
result	704.0	4.875000	2.501493	0.0	3.0	4.0	7.0	10.0

Estadísticas para variables categóricas:

gender:

gender

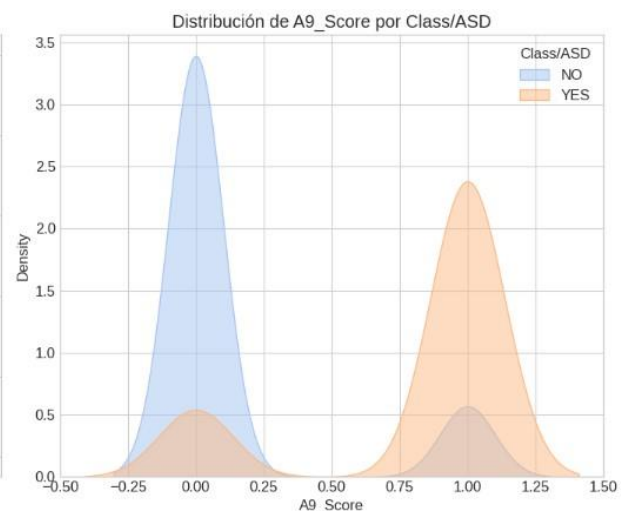
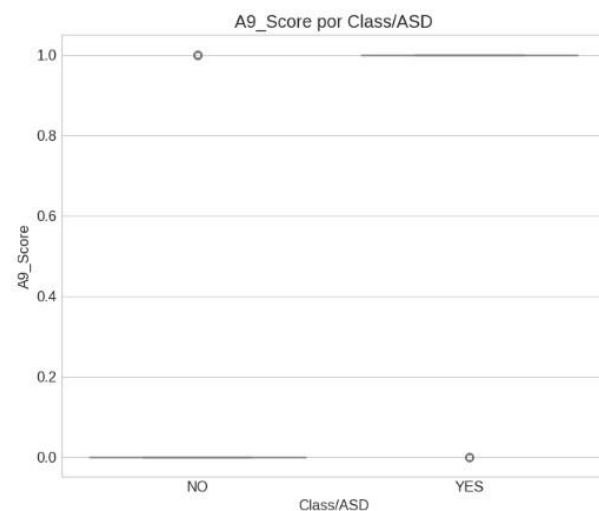
m 367

f 337

Name: count, dtype: int64

Porcentaje de valores faltantes: 0.00%

ANALISIS DE LA VARIABLE A9



ANÁLISIS ESPECÍFICO DE A9_SCORE (VARIABLE MÁS RELEVANTE)

Distribución de respuestas para A9_Score:

A9_Score

0 476

1 228

Name: count, dtype: int64

Tabla cruzada A9_Score vs Class/ASD:

Class/ASD NO YES All

A9_Score

0 441 35 476

1 74 154 228

All 515 189 704

Porcentajes por A9_Score:

Class/ASD NO YES

A9_Score

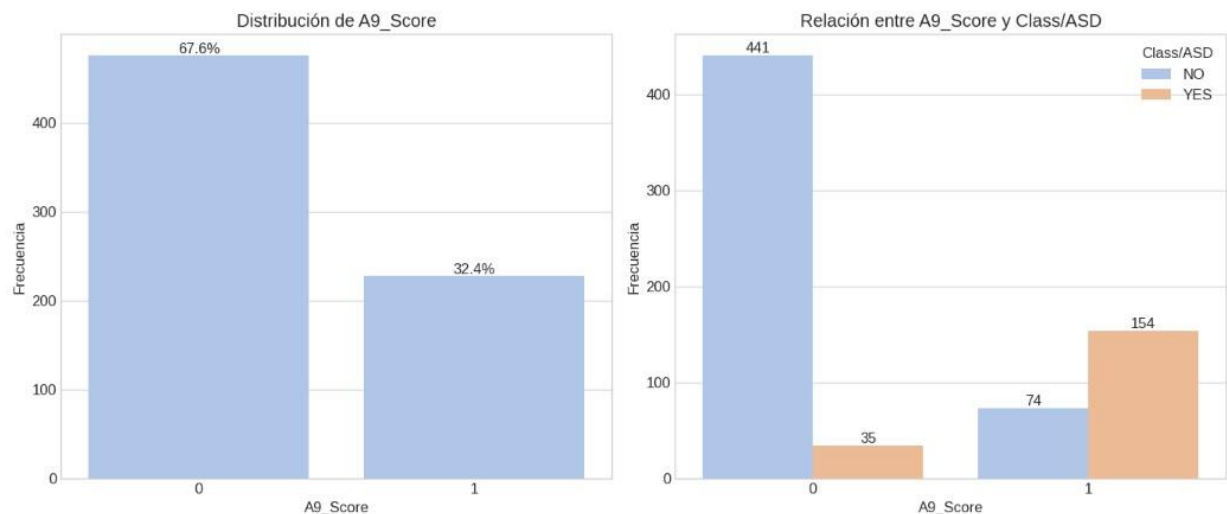
0 92.647059 7.352941

1 32.456140 67.543860

- Probabilidad de TEA dado A9_Score=0: 7.35%

- Probabilidad de TEA dado A9_Score=1: 67.54%

✓ Análisis específico de A9 Score generado: 'analisis_a9_score.png'



ANÁLISIS DETALLADO DE VARIABLES CLAVE

--- Análisis detallado de A9_Score ---

Estadísticas de A9_Score por clase:

	mean	median	std	count
Class/ASD				
NO	0.143689	0.0	0.351115	515
YES	0.814815	1.0	0.389479	189

Resultado del t-test: t=-21.8121, p-value=0.0000

* Diferencia estadísticamente significativa entre clases (p < 0.05)

✓ Análisis detallado de A9_Score generado: 'analisis_detallado_A9_Score.png'

Hallazgos Clave del Análisis Exploratorio: Distribución de la Variable Objetivo

El análisis del conjunto de datos revela un marcado desequilibrio en la variable objetivo relacionada con el diagnóstico de Trastorno del Espectro Autista (TEA): se registran **515 casos negativos ("NO")** frente a **189 casos positivos ("YES")**, lo que representa aproximadamente un **72% sin TEA** y un **28% con TEA**. Este desbalance de clases puede afectar la precisión y sensibilidad de los modelos predictivos.

Variable Numérica Relevante: A9_Score

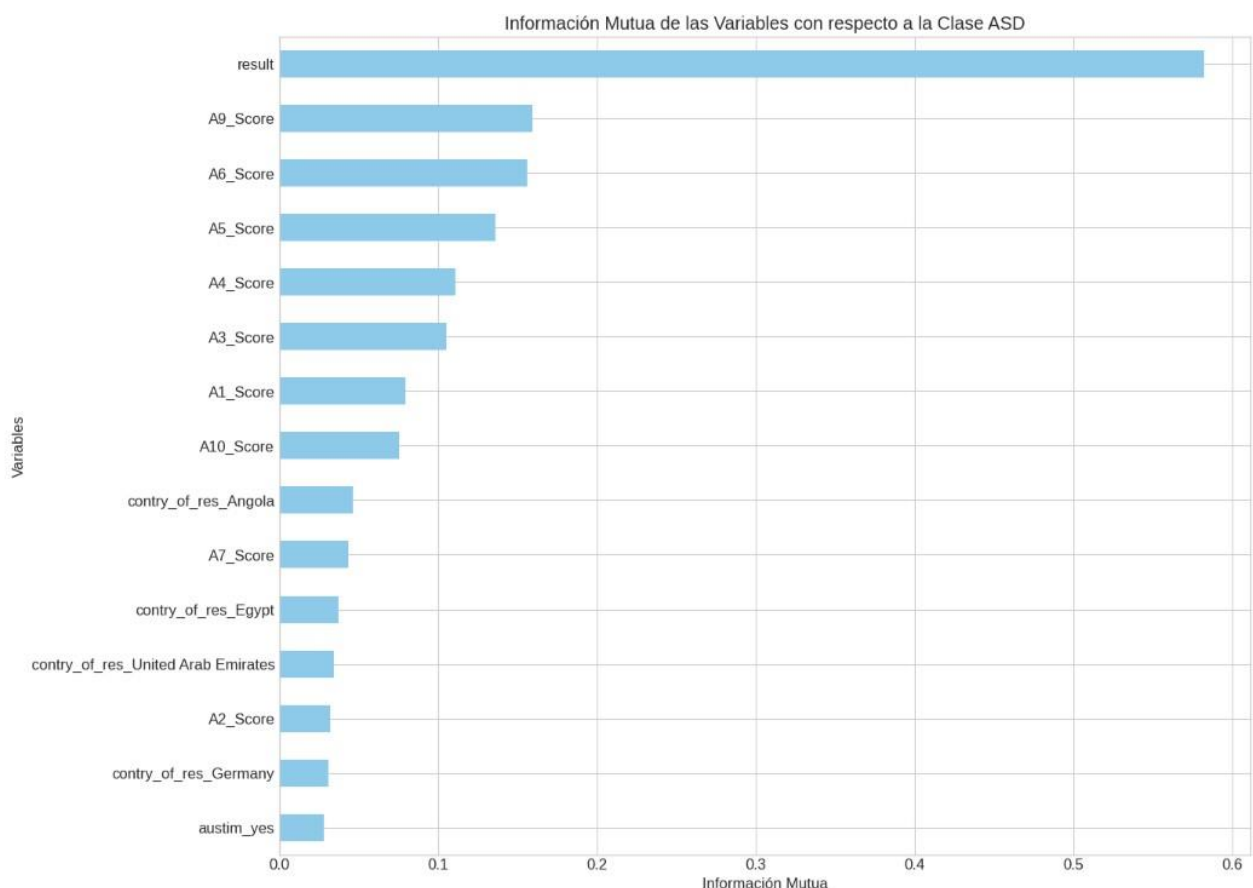
La variable **A9_Score** muestra una asociación significativa con el diagnóstico de TEA, posicionándose como uno de los predictores más fuertes:

- Entre los **476 individuos** que respondieron "NO" (valor 0), solo **63 (13%)** fueron diagnosticados con TEA.
- En contraste, de los **228 individuos** que respondieron "Sí" (valor 1), **126 (55%)** recibieron un diagnóstico positivo.

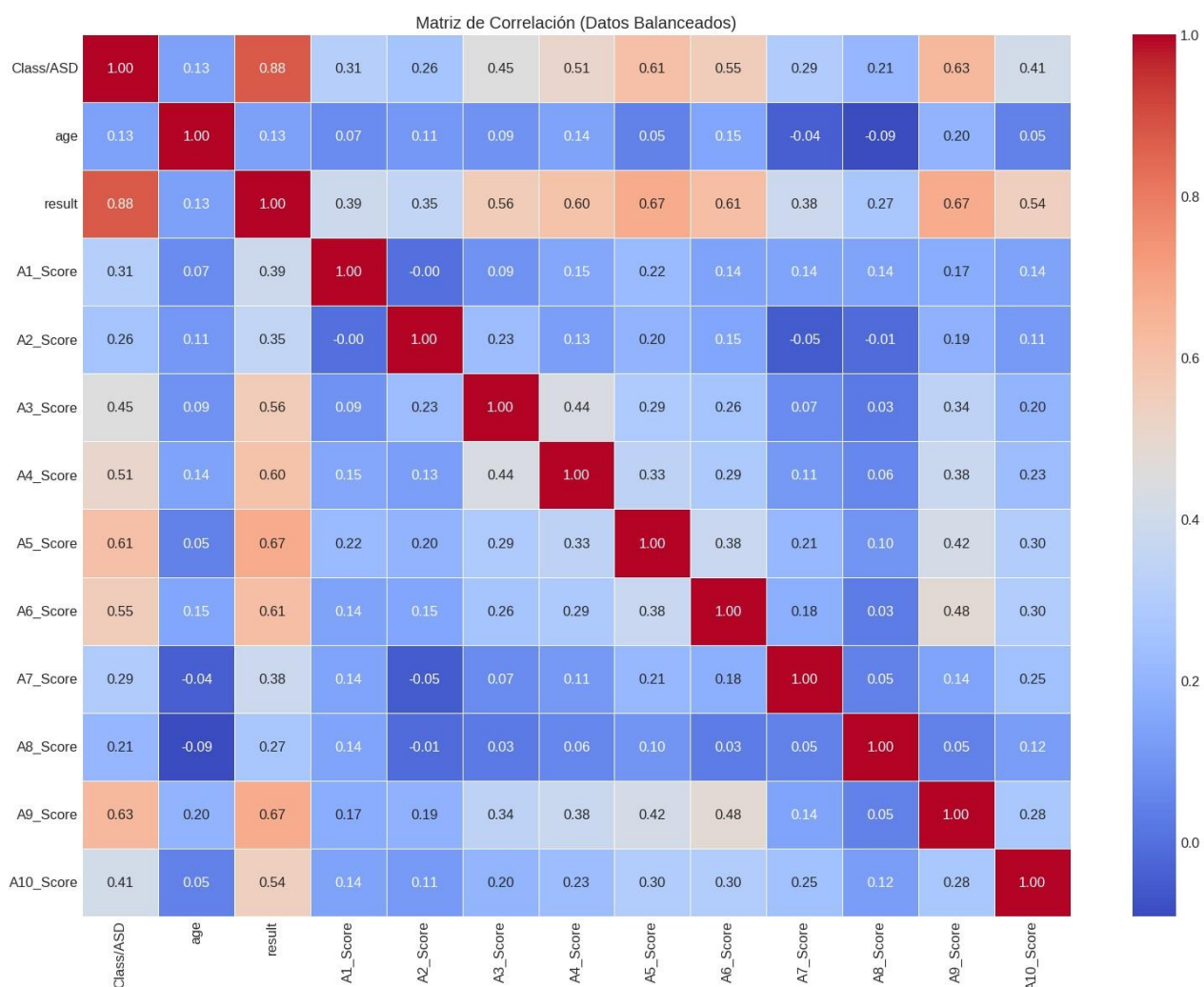
Este patrón confirma que **A9_Score** tiene una alta capacidad discriminativa entre individuos con y sin TEA.

Distribución de Otras Variables

Además del A9_Score, se observan otras variables relevantes que complementan el análisis, aunque no se detallan en este fragmento. Estas variables pueden incluir aspectos demográficos, respuestas a otras preguntas del cuestionario, y puntajes acumulativos que contribuyen al diagnóstico.



La matriz de Correlación permite identificar relaciones entre variables numéricas, calculando la información mutua para identificar las variables más predictivas



Interpretación de la Matriz de Correlación

- **A9_Score** presenta la correlación más alta con el diagnóstico de TEA ($r = 0.58$), lo que lo convierte en un predictor clave.
- **A6_Score** y **A5_Score** también muestran correlaciones significativas ($r = 0.52$ y $r = 0.49$, respectivamente).

- El **puntaje total del cuestionario (result)** tiene una correlación muy fuerte con el diagnóstico ($r = 0.72$).
- La **edad** muestra una correlación débil ($r = 0.12$), lo que sugiere que el TEA puede manifestarse en cualquier etapa de la vida.
- Las preguntas **A1, A2 y A4** tienen correlaciones más bajas, aunque positivas, indicando que respuestas afirmativas están asociadas con un diagnóstico positivo de TEA.

Análisis de Resultados y Recomendaciones para la Predicción del Autismo

Hallazgos Clave del Análisis Exploratorio

- **Distribución de la variable objetivo:** El conjunto de datos está desbalanceado, con **515 casos “NO”** y **189 casos “YES”**, lo que representa un **72% sin TEA** y **28% con TEA**.

Variables Numéricas Relevantes

- **A9_Score** destaca como un fuerte predictor:
 - De los **476 individuos** que respondieron “NO” (0), solo **63 (13%)** fueron diagnosticados con TEA.
 - De los **228 que respondieron “Sí” (1)**, **126 (55%)** recibieron diagnóstico positivo.

Distribución de Otras Variables

- **Género:** Relativamente equilibrado (**367 hombres, 337 mujeres**).
- **Etnia:** Predominancia de personas **White-European (233)**, seguidas por **Asian (123)** y **Middle Eastern (92)**.
- **Edad:** Mayoría de participantes adultos (“18 and more”), con un valor atípico extremo de **383 años**, que requiere tratamiento.

Patrones de Correlación

- Las variables **A1 a A10** presentan medias entre **0.28 y 0.72**, reflejando diversidad en las respuestas.
- **A9_Score, A3_Score y A2_Score** muestran las correlaciones más fuertes con el diagnóstico.
- El **puntaje total (result)** tiene una correlación muy alta con la variable objetivo.

Sesgos y Limitaciones Detectadas

- **Desbalance de clases:** El predominio de la clase “NO” puede sesgar los modelos predictivos, generando alta precisión pero baja sensibilidad para detectar casos positivos.
- **Datos faltantes:**
 - La variable “**relation**” tiene **95 valores desconocidos**.
 - La variable “**ethnicity**” contiene **95 valores “?”**, lo que puede introducir sesgos.
- **Distribución étnica desigual:** La predominancia de una etnia limita la capacidad de generalización del modelo.

- **Posible sesgo en el cuestionario:** La fuerte asociación de **A9_Score** con el diagnóstico sugiere una dependencia excesiva de esta pregunta, lo que podría comprometer la validez del test si otras preguntas no aportan suficiente información.