



Descripción del Dataset Original

Origen del Dataset

Este conjunto de datos proviene de un **test de tamizaje** basado en las preguntas del AQ-10-Child y otras variables sociodemográficas y médicas, diseñado para detectar señales tempranas del **Trastorno del Espectro Autista (TEA)** en niños. Fue recopilado por **Fadi Fayez Thabtah** y está disponible públicamente en el repositorio de aprendizaje automático de la **UCI Machine Learning Repository**.

Aclaraciones:

¿Qué es AQ-10-Child? El **test AQ-10-Child** es un cuestionario breve que evalúa **rasgos de comportamiento asociados al TEA**. Cada respuesta se registra en el dataset como un valor binario (**0 = No, 1 = Sí**), lo que permite analizar patrones y su posible relación con el diagnóstico de TEA.

¿Qué es un test de tamizaje o cribado? Ambos términos son equivalentes. En este contexto, se refieren a cuestionarios estructurados que **buscan detectar señales tempranas de TEA**, pero **sin proporcionar un diagnóstico definitivo**.

Información del Dataset

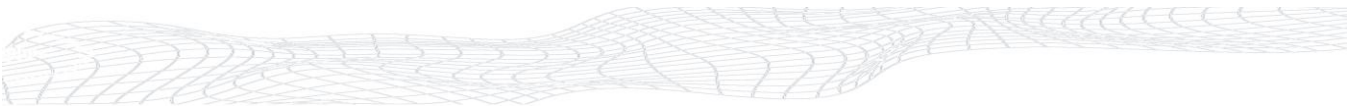
- **Nombre:** *Autistic Spectrum Disorder Screening Data for Children*
- **Fuente oficial:** [UCI Machine Learning Repository](https://doi.org/10.24432/C5659W)
- **Cita recomendada:**

Thabtah, F. (2017). *Autistic Spectrum Disorder Screening Data for Children* [Dataset]. UCI Machine Learning Repository. <https://doi.org/10.24432/C5659W>

- **Archivos disponibles:**
 - Autism-Child-Data.arff
 - Autismo-Detección-Niño-Datos Description.docx
- **Ubicación en el repositorio local:**
 - data/raw/Autism-Child-Data.arff (dataset).
 - data/raw/Autismo-Detección-Niño-Datos Description.docx (documento descriptivo)
- **Formato original:** .arff
- **Método de carga:** `scipy.io.arff + pandas.DataFrame`
- **Fecha de descarga:** 25/05/2025

Este dataset servirá como base para la construcción de modelos predictivos con fines de salud pública, particularmente en **Tierra del Fuego**.

Descripción General del Dataset

- ✓ **Observaciones:** 292 registros (cada registro representa un niño/a)
 - ✓ **Variables:** 21 columnas
 - ✓ **Valores nulos:** age (4 nulos)
 - ✓ **Errores de codificación:** nombres como `jundice`, `austim`, `contry_of_res`
 - ✓ **Errores de formato:** columnas binarias y numéricas mal tipadas (de `object` a `int`)
- 

Descripción de Variables

- Variables Demográficas y Clínicas

Las siguientes variables describen características generales de los participantes en el estudio:

Variable	Tipo	Descripción y propósito
age	Numérica	Edad del niño en años. Permite analizar si la edad influye en la probabilidad de presentar signos de TEA.
gender	Categórica	Género del niño (m = masculino, f = femenino). Se investiga si hay diferencias de prevalencia según el género.
ethnicity	Categórica	Grupo étnico del niño. Puede ayudar a explorar variaciones culturales o genéticas en el diagnóstico.
jaundice	Binaria	Historial de ictericia neonatal (yes = Sí, no = No). Se evalúa su posible correlación con TEA. (*1)
autism	Binaria	Antecedentes familiares de TEA (yes = Sí, no = No). Indica si un familiar cercano ha sido diagnosticado con TEA.
age_desc	Categorica	Descripción textual de la edad ('4-11 years').
relation	Categórica	Relación del encuestado con el niño (padre, madre, cuidador, médico, etc.). Permite identificar quién proporcionó la información del test.
country_of_res	Categórica	País de residencia del niño. Se analiza si hay factores ambientales o culturales asociados.
used_app_before	Binaria	Si el usuario ha usado una aplicación de cribado de TEA previamente (yes = Sí, no = No). (*2)
result	Numérica	Puntaje final del test de cribado, calculado automáticamente según las respuestas. Indica si hay signos consistentes con TEA.

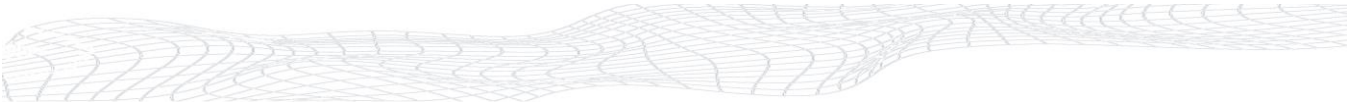
(*1) jaundice es ictericia, ¿qué es? <https://www.topdoctors.es/diccionario-medico/ictericia/> refiere a la bilirrubina.

(*2) La documentación del dataset no especifica qué aplicación en particular se utilizó, pero hace referencia a herramientas digitales de cribado para detectar signos tempranos de TEA.

- Preguntas del Cuestionario AQ-10-Child

Las siguientes variables representan preguntas del **cuestionario AQ-10-Child**, diseñadas para evaluar rasgos de comportamiento relacionados con el TEA. Cada respuesta es binaria (**0 = No, 1 = Sí**):

Pregunta	Descripción
A1_Score	¿El niño evita el contacto visual?
A2_Score	¿Responde cuando lo llaman por su nombre?



Pregunta	Descripción
A3_Score	¿Comparte intereses con los demás?
A4_Score	¿Juega de forma imaginativa?
A5_Score	¿Señala para mostrar cosas que le interesan?
A6_Score	¿Responde a expresiones emocionales?
A7_Score	¿Sigue instrucciones simples?
A8_Score	¿Muestra afecto hacia otros?
A9_Score	¿Inicia interacciones sociales?
A10_Score	¿Utiliza gestos como saludar?

🔗 Fuente del test: [Autism Research Centre](#)

- **Variable Objetivo**

Variable	Descripción
Class/ASD	Diagnóstico estimado: YES (posible TEA), NO (descartado)

