# **Descripción del Dataset Procesado**

## 

## **Resumen General**

Este archivo describe el **dataset final preprocesado** resultante del proyecto de análisis predictivo del Trastorno del Espectro Autista (TEA) en niños. Tras un exhaustivo proceso de limpieza, transformación y depuración, se obtuvo un conjunto de datos adaptado para el modelado con algoritmos de aprendizaje automático.

## **Cambios Realizados al Dataset Original**

### **Renombrado de Columnas**

Se tradujeron y normalizaron los nombres de las variables para facilitar la interpretación:

| Original | Renombrada |
| --- | --- |
| A1\_Score | Evita\_Contacto\_Visual (A1) |
| A2\_Score | Responde\_a\_Su\_Nombre (A2) |
| … | … |
| gender | Género |
| ethnicity | Etnia |
| jundice | Ictericia\_Neonatal |
| austim | Antecedente\_Familiar\_TEA |
| used\_app\_before | Uso\_App\_Anterior |
| Class/ASD | Clase\_TEA |

### **Eliminación de Variables Irrelevantes o Problemáticas**

Se eliminaron las siguientes variables por razones de redundancia, alta cardinalidad o riesgo de sesgo:

* Resultado\_Diagnóstico
* Grupo\_Edad
* Relación\_con\_el\_Niño
* País\_de\_Residencia
* Uso\_App\_Anterior
* Etnia

**Justificación** completa incluida en el análisis exploratorio de la notebook.

### **Tratamiento de Valores Nulos**

* La variable Edad contenía 4 valores nulos.
* Se imputaron mediante la **mediana** de la columna (Edad = 6).

### **Conversión y Codificación de Variables**

* Edad: de float64 a int.
* Variables binarias (Género, Ictericia\_Neonatal, Antecedente\_Familiar\_TEA, Clase\_TEA) fueron convertidas a valores numéricos:
  + m/‘f’ → 0/1
  + no/yes → 0/1
  + NO/YES → 0/1

## **Descripción de las Variables Finales**

Aquí tienes la tabla con formato listo para pegar en Word. Te recomiendo usar una tabla con bordes visibles para una presentación más clara, pero si vas a pegarlo como texto plano, puedes usar esta versión con tabulaciones o espacios ajustados:

|  |  |  |
| --- | --- | --- |
| Variable | Tipo | Descripción |
| Evita\_Contacto\_Visual (A1) | Binaria | ¿Evita el niño el contacto visual? |
| Responde\_a\_Su\_Nombre (A2) | Binaria | ¿Responde cuando se lo llama por su nombre? |
| Comparte\_Intereses (A3) | Binaria | ¿Comparte intereses con los demás? |
| Juega\_Imaginativo (A4) | Binaria | ¿Participa en juegos imaginativos? |
| Señala\_Para\_Mostrar (A5) | Binaria | ¿Señala objetos para compartir interés? |
| Responde\_Emociones (A6) | Binaria | ¿Responde a expresiones emocionales? |
| Sigue\_Instrucciones\_Simples (A7) | Binaria | ¿Sigue instrucciones simples? |
| Muestra\_Afecto (A8) | Binaria | ¿Muestra afecto hacia otras personas? |
| Inicia\_Interacciones (A9) | Binaria | ¿Inicia espontáneamente interacciones sociales? |
| Utiliza\_Gestos (A10) | Binaria | ¿Utiliza gestos comunicativos como saludar o señalar? |
| Edad | Entera | Edad del niño en años (entre 4 y 11) |
| Género | Binaria | 0 = Masculino, 1 = Femenino |
| Ictericia\_Neonatal | Binaria | Antecedentes de ictericia neonatal (0 = No, 1 = Sí) |
| Antecedente\_Familiar\_TEA | Binaria | Historia familiar de TEA (0 = No, 1 = Sí) |
| Clase\_TEA | Binaria | Variable objetivo: 0 = No indicios de TEA, 1 = Posibles indicios de TEA |

Referencia:

|  |
| --- |
| 0 = masculino, 1 = femenino |
| 0 = no, 1 = sí |

## **Formato Final del Dataset**

* **Observaciones:** 292
* **Variables:** 15 (todas numéricas)
* **Sin valores nulos**
* **Formato de guardado:** data/processed/dataset\_procesado.csv