

Predicción del abandono escolar en nivel secundario

Tierra del Fuego

Entrega 2 - Descripción del Dataset

Dataset

-
- **Nombre del archivo:** DatosAbandonoTDF (consolidado a partir de múltiples fuentes públicas)
 - **Instancias:** 21
 - **Columnas:** 55
 - **Tipo de datos:** Numéricos (tasas porcentuales y cantidades) y categóricos (años)
 - **Variables destacadas:**
 - rep_total_pct: porcentaje de repitencia total
 - sobr_total_pct: tasa de sobreedad total
 - prom_total_pct: tasa de promoción efectiva total
 - mat_total_pct: matrícula general
 - aband_total_pct: tasa de abandono general
 - abandono_binario: variable objetivo binaria
 - También se incluyen tasas y cantidades por cursos, tipo de gestión y ciudades (Río Grande/Ushuaia)
 - **Fuente de los datos:** <https://ipiec.tierradelfuego.gob.ar/>
 - **Datasets utilizados:**
 - Tasa de repitencia por año (Excel)
 - Promoción efectiva, sobreedad y abandono por año
 - Matrícula por gestión y localidad
 - Fecha de recolección: 05 de Mayo 2025

Preprocesamiento aplicado

1. Limpieza de cada archivo:

- Eliminación de filas vacías
- Eliminación de encabezados
- Estandarización de columnas

2. Unificación:

- Se utilizó merge() sucesivamente entre los distintos archivos, unificando por columnas compartidas como 'año' y estructuras comunes de cursos.3.

3. Renombrado:

- Uso de prefijos por indicador: rep_, prom_, sobr_, mat_, aband_

4. Creación de variable objetivo:

```
df_merged['abandono_binario'] = df_merged['aband_total_pct'].apply(lambda x: 1 if pd.notna(x) and x > 5 else 0)
```