

Problema elegido: Predicción de abandono escolar.

1. Tipo de aprendizaje: Supervisado (tenemos datos históricos con la variable objetivo: abandono / no abandono).

2. Tipo de tarea: Clasificación (la salida es binaria: “abandona” o “no abandona”).

3. Datos necesarios y posibles fuentes:

- Datos académicos (notas, asistencia, participación en actividades).
- Información socioeconómica (ingresos familiares, nivel educativo de los padres).
- Datos de comportamiento (uso de plataformas de estudio, tiempos de conexión).
- Fuente: registros internos del colegio/universidad y encuestas a estudiantes.

4. Etapas a seguir:

- **EDA:** analizar distribución de variables, detectar valores faltantes y outliers.
- **División de datos:** entrenamiento (70%), validación (15%), prueba (15%).
- **Entrenamiento:** aplicar distintos modelos de clasificación.
- **Evaluación:** métricas como accuracy, F1-score y matriz de confusión.

5. Algoritmos candidatos:

- Regresión logística.
- Random Forest.
- Gradient Boosting.