

Actividad práctica: Análisis de precisión vs. recall en la detección de fraude

Objetivo:

Los estudiantes deben aprender a diferenciar entre **precisión** y **recall**, y entender por qué **recall** es más importante que precisión en situaciones como la **detección de fraude**, donde los **falsos negativos** pueden ser muy costosos.

Descripción del experimento:

1. **Cargar y preprocesar los datos:** Utilizaremos un conjunto de datos simulado o el conjunto de datos de **detección de fraude** (puedes usar cualquier conjunto de datos donde la clase positiva esté muy desbalanceada).
2. **Entrenar un modelo de clasificación:** Los estudiantes entrenarán un modelo (como un **Random Forest** o un **Árbol de Decisión**) para predecir si una transacción es fraudulenta o no.
3. **Evaluar el modelo con precisión y recall:** Los estudiantes evaluarán el modelo utilizando tanto **precisión** como **recall**, observando las diferencias en estas métricas cuando las clases están desbalanceadas.
4. **Analizar el impacto de un alto recall:** Los estudiantes observarán los efectos de priorizar **recall** sobre **precisión**, viendo cómo el modelo puede tener una mayor tasa de falsos positivos pero aún así capturar más fraudes.

1. Cargar y preprocesar los datos:

Usaremos un conjunto de datos simulado de **fraude** donde la clase positiva (fraude) está muy desbalanceada.

2. Entrenar el modelo:

Usamos un **Random Forest** para entrenar el modelo en el conjunto de datos.

3. Evaluar el modelo con precisión y recall:

Los estudiantes evaluarán el modelo utilizando las métricas de **precisión** y **recall**.

4. Analizar los resultados:

Los estudiantes deben observar los siguientes puntos:

- **Precisión** (accuracy): ¿Qué tan exactas son las predicciones del modelo, especialmente en las predicciones positivas?
- **Recall** (sensibilidad): ¿Cuántos de los casos realmente positivos (fraudes) fueron capturados?
- **Matriz de confusión**: Observar el número de **falsos negativos** (FN) y **falsos positivos** (FP) es crucial. Los estudiantes deben enfocarse en los **falsos negativos** porque son los que más afectan cuando se está priorizando el **recall**.