

Resultados del Proyecto de Clasificación

1. Introducción y Objetivos del Proyecto

El objetivo principal de este proyecto fue evaluar y comparar el rendimiento de varios modelos de clasificación utilizando un conjunto de datos sobre montos totales.

Se implementaron diferentes modelos, y se realizó una validación cruzada para seleccionar el mejor según métricas clave como la precisión, recall, F1-Score y AUC de la curva ROC.

2. Descripción del Conjunto de Datos

Características: Datos de transacciones que incluyen características como 'Total Amount', categorizadas en valores binarios (mayor que la mediana o no).

Objetivo: Predecir si un valor de 'Total Amount' es mayor o menor que la mediana.

Transformaciones: Se aplicó un preprocesamiento de datos que incluyó la codificación one-hot para variables categóricas y escalado de características numéricas.

3. Principales Análisis y Hallazgos

Se evaluaron seis modelos: Logistic Regression, KNN, Decision Tree, Random Forest, XGBoost, y LightGBM.

Los modelos fueron entrenados con validación cruzada para asegurar su robustez.

Se calculó el AUC de la curva ROC para cada modelo, lo que permitió comparar la capacidad de cada uno para diferenciar entre las dos clases (mayor que la mediana o no).

Resultados del Proyecto de Clasificación

4. Visualización de la Curva ROC

Se utilizó la curva ROC para comparar el rendimiento de los modelos. Este gráfico muestra la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR).

El AUC indica la capacidad del modelo para discriminar entre las clases.

5. Conclusiones y Recomendaciones

