

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 varios modelos: Logistic Regression, Random Forest, XGBoost, y LightGBM.

Cada modelo fue entrenado con validación cruzada y se compararon sus curvas ROC.

El AUC de cada modelo permitió evaluar su capacidad para diferenciar entre las dos clases.

4. Visualización de la Comparación de Curvas ROC

Resultados del Proyecto de Clasificación

Se compararon las curvas ROC de varios modelos de clasificación: Logistic Regression, Random Forest, XGBoost y LightGBM.

Cada curva muestra la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR).

El área bajo la curva (AUC) es un indicador de la capacidad del modelo para discriminar entre las dos clases.

5. Conclusiones y Recomendaciones

Modelo más destacado: El modelo XGBoost mostró el mejor rendimiento en la comparación de curvas ROC.

Recomendación: Se recomienda utilizar XGBoost para la clasificación debido a su mayor capacidad de discriminación.

También se sugiere continuar con la experimentación de los hiperparámetros de otros modelos.

