Informe Ejecutivo – Predicción de Propinas Generosas

Proyecto: Modelado predictivo para la Comisión de Taxis y Limusinas de Nueva York

Realizado por: Juan (Analista de Datos)

OVERVIEW

Objetivo
Desarrollar un modelo de aprendizaje automático que prediga si un pasajero dejará una propina generosa (≥20%) antes del viaje.

📊 Datos Utilizados
- Datos históricos de viajes (TLC NYC)
- Variables: tarifa, propina, tipo de pago, distancia, peajes, fecha y hora
- Variable objetivo: `generoso` (1 si la propina fue ≥20%)
- Registros analizados: +22.000

PROJECT STATUS

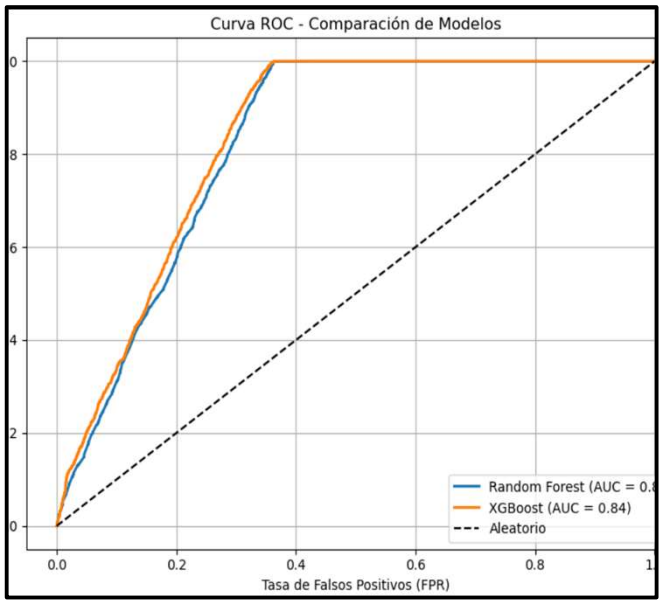
⚙️ Modelos Construidos

Random Forest Classifier
- Precisión: 78%
- Recall (Generoso): 86%
- AUC: 0.83

XGBoost Classifier
- Precisión: 81%
- Recall (Generoso): 100%
- AUC: 0.84

KEY INSIGHTS

✅ Recomendaciones
- Usar XGBoost como predictor base
- Probar integración en app TLC
- Agregar variables externas (clima, tráfico)
- Automatizar retroalimentación al conductor



NEXT STEPS

🧠 Hallazgos Principales
- Forma de pago: principal indicador
- Viajes largos/tarifas altas → más propinas generosas
- XGBoost identificó todos los casos generosos (recall 100%), aunque con más falsos positivos

📘 Guía de Conceptos Clave en Evaluación de Modelos

- 🔹 Precisión (Accuracy)Porcentaje total de predicciones correctas.Fórmula: (VP + VN) / Total

- 🔹 Recall o SensibilidadQué tan bien identifica los casos positivos.Fórmula: VP / (VP + FN)

- 🔹 Precisión (Precision)Qué proporción de casos predichos como positivos realmente lo son.Fórmula: VP / (VP + FP)

- 🔹 F1-ScorePromedio armónico entre precisión y recall. Útil si hay desbalance.Fórmula: 2 * (Precision * Recall) / (Precision + Recall)

- 🔹 AUC – Área bajo la curva ROCCuánto discrimina el modelo entre clases. Más cerca de 1 = mejor.

- 🔹 Importancia de variablesMide qué tan influyente es cada variable para la predicción final de modelo.

- 🔹 Comparación entre modelos

Un modelo puede tener mejor precisión, otro mejor recall. Elegí según lo que más te importe (e detectar generosidad sin dejar pasar casos).