Predicting Generous Tippers for the NYC Taxi & Limousine Commission

Using Machine Learning Models to Identify Patterns in Customer Generosity

Overview

Automatidata partnered with the **New York City Taxi & Limousine Commission (TLC)** to explore how different trip characteristics influence passenger tipping behavior. The goal was to use predictive modeling to determine whether a customer is likely to be a *generous tipper* (tip ≥ 20%).
This work supports TLC's goal of improving **driver satisfaction** and **service quality** through data-driven insights.

Objective

- Develop and evaluate machine learning models to predict whether a rider will give a generous tip.
  Identify the **key factors** influencing tipping behavior.
  Recommend actionable insights for TLC to improve driver earnings and customer experience.
  Two models were tested: **Random Forest Classifier,XGBoost Classifier.**
  Both models were optimized using **F1 score**, which balances precision and recall — ideal since both false positives and false negatives carry similar costs.

Results

The **XGBoost model** slightly outperformed Random Forest in accuracy and precision.

Both models showed consistent generalization with balanced precision and recall.

Random Forest (Test)	0.6697	0.7698	0.7163	0.6790
XGBoost (Test)	**0.6765**	**0.7586**	**0.7152**	**0.6820**

Next Steps

Deploy the XGBoost model in pilot programs to predict generous tippers and inform driver allocation strategies.

Collect more contextual data such as:
Weather conditions
Traffic or time-of-day congestion
Passenger and driver profiles

Retrain and monitor model performance over time to ensure reliability.

Share insights with TLC stakeholders to improve incentive programs and driver engagement.