Machine Learning Model Outcomes

Executive summary report for the New York City Taxi and Limousine Commission
Prepared by Automatidata

Overview

The New York City Taxi & Limousine Commission partnered with Automatidata to build a machine learning model that predicts whether a passenger is likely to leave a generous tip. The goal is to give drivers better insights into rider behavior.

Problem

At first, the team considered predicting non-tippers, but ethical concerns led them to rethink that approach. Instead, they chose to focus on identifying generous tippers those who leave tips of 20% or more. This shift aims to fairly balance the needs of both drivers and passengers.

Solution

Two different modeling techniques were tested and compared. Both performed well, but the Random Forest model showed slightly better results. The next step is to run pilot tests with actual taxi drivers to gather feedback and refine the system.

Details

Behind the data

- The data team began with the idea that the trip route, estimated fare amount, and time of day could directly impact the tip value. The expectation was that these elements would be significant enough to allow reliable predictions of more generous tipping.

- After building the models and running the tests, the results confirmed this assumption. These factors do contribute to predicting tip amounts. The model reached an F1 score of 0.7235, which reflects consistent and promising performance.

	model	precision	recall	F1	accuracy
0	RF CV	0.674919	0.757312	0.713601	0.680233
0	RF test	0.675297	0.779091	0.723490	0.686538
0	XGB CV	0.673074	0.724487	0.697756	0.669669
0	XGB test	0.675660	0.747978	0.709982	0.678349

Image Alt-Text: F1 scores for random forest and XGboost models

Future model suggestions

- Expand the collection of detailed driver and user-level data, including historical tipping behavior.

- Use clustering techniques such as the K-means algorithm to segment the data and uncover meaningful insights from the identified patterns.

Results Summary

The developed algorithm makes it possible to identify riders who are likely to give generous tips, showing solid precision, recall, F1 score, and overall accuracy. For suggestions on how to enhance this model, please refer to the "Next Steps" section.

Next Steps

The Automatidata data team may reach out to the New York City Taxi and Limousine Commission to share the model's results and propose its use as an indicator of tip amounts. However, to achieve meaningful improvements in the model's performance, additional data would be required.