New York City TLC Project – Machine Learning Model

Executive Summary Report V

Commission Prepared by **Automatidata**

Project Overview

The New York City Taxi & Limousine Commission wants a machine learning model that can predict if a passenger is likely to give a generous tip. The initial goal of predicting non-tippers was dismissed due to ethical concerns. Instead, the focus shifted to identifying "generous" tippers, defined as those giving tips of 20% or more. This adjustment aims to balance the interests of both drivers and passengers.

Key Insights

- The developed model is capable of identifying potential generous tippers with good precision, recall, F1 score and overall accuracy.
- The model can undergo beta testing with taxi drivers for additional feedback.
- Trip details, such as route, estimated fare and time of travel, are significantly related to tipping behavior. In fact, these factors are indeed effective in predicting tipping behavior.
- Analyzing clusters that are created using K-means clustering could provide deeper insights into the data.
- Additional features, such as trip distance categorization (short, medium, long) and the ratio of fare values relative to multiples of $5 or $10, could potentially enhance model accuracy. These features could capture customer tendencies to round their tips.
- Having access to historical tipping behavior, accurate data for cash tips and a larger dataset would likely boost model performance.

Details

- The team tested two different model types and compared their effectiveness: Random Forest and XGBoost. Both models delivered acceptable results.
- It was assumed that trip details, such as route, estimated fare and time of travel, would be significantly related to tipping behavior. After building and testing the models, it became evident that these factors were indeed effective in predicting tipping behavior. The model achieved an F1 score of 72%.

model	precision	recall	F1	accuracy
Random Forest Train	0.680002	0.762601	0.718856	0.686128
Random Forest Test	0.670784	0.777225	0.720092	0.681952
XGBoost Train	0.679115	0.742065	0.709077	0.679576
XGBoost Test	0.681391	0.767890	0.722060	0.688831

Negative (0) = Ungenerous tipper

Positive (1) = Generous tipper

Next Steps

1. Present the model's findings to the New York City Taxi & Limousine Commission, recommending that the model be used as a predictor of tipping tendencies.
2. Collect more data at both the driver and customer levels to significantly enhance the model's performance.