Classification Modeling for Taxi Tips

Executive Summary Report for New York City Taxi and Limousine Commission

Prepared By: Rebecca Iglesias-Flores

Project Overview

One of Automitadata's clients, the NYC Taxi & Limousine Commission, has asked us to build a model that predicts whether a customer will leave a tip. Due to the ethical concerns of this objective, we are re-formulating the goal to predict how generous a tipper will be.

Key Insights

Things to consider

- The model predicts twice as many false positives (predicted to tip >= 20%, but don't) than false negatives (predicted to tip poorly, but tip well >= 20%).
- The efficacy of the model then really depends on the end goal and prioritized end user.
- It is not as simple to say false positives are worse for cab drivers or false negatives are worse for customers.
- Given a class imbalance of roughly 10% difference in the population favoring generous tippers, rejecting generous tippers due to false negatives leads to greater lost earnings than accepting non-tippers due to false positives and therefore recommendation is to prioritize recall as our evaluation metric..

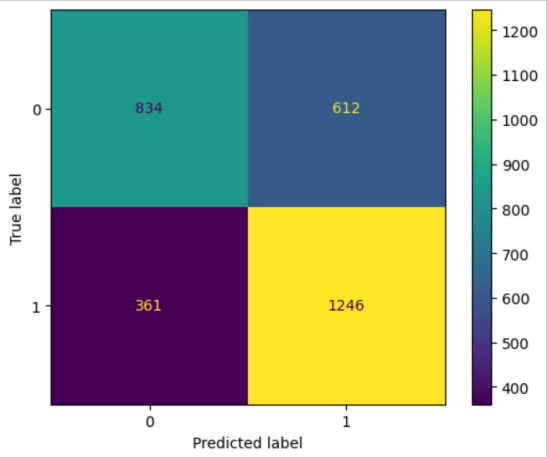

Figure 2. Confusion Matrix of top performing model

Details

model	precision	recall	F1	accuracy
RandomForestClassifier Test	0.673139	0.776602	0.721179	0.683917
XGBoost Test	0.680556	0.762290	0.719108	0.686538
RandomForestClassifier Val	0.674701	0.757934	0.713778	0.680233
XGBoost Val	0.670590	0.717797	0.693246	0.665739

Figure 2. Results from model experiments

- A random forest and an XGBoost classifier were constructed using a 4-fold cross validation process and evaluated on validation and test sets.
- This model's performance is acceptable. The top performing model had an F1 score of 0.72119 and it had an overall accuracy of 0.6839.
- It correctly identifies ~78% of actual generous tippers in the test set, which is 48% better than a random guess.
- It may be worthwhile to test the model with a select group of taxi drivers to get feedback.
- It may be worthwhile to prioritize recall, given key insights into how rejecting false negatives would result in higher lost earnings (recall aligned with F1 this time, but this might not be the case in future experiments and improvements)

Next Steps

- This model performs well and we can recommend beta-testing with cab drivers as it stands
- There is room for improvement and more feature engineering is needed
- potentially use k-means clustering to add predictive signal in features
- Recommend collect/adding more granular driver and user-level data, including past tipping behavior and cash value tips