Machine Learning Model Outcomes

Overview

Automatidata was contracted by the New York City TLC to construct a machine learning model to predict whether patrons of NYC TLC taxi drivers will be generous tippers.

Problem

The initial model goal of predicting non-tipping and tipping customers was decided to be ethically problematic. Instead, goals shifted to predict whether customers would be "generous" tippers instead, defined by those predicted to tip greater than 20% of their fare. This would continue to work toward the original business goals while limiting bias toward customers by TLC drivers.

Solution

Different modeling methodologies were utilized and compared, with both performing at an adequate level. Ultimately, random forest modeling yielded the most favorable results, and is recommended for initial testing by drivers before large-scale utilization.

Details

Data observations

- Initial assumptions about relevant trip attributes were proven to have a strong relationship with generous tipping. These factors included predicted fare amounts, time of day, and travel duration/distance.

- A random forest model produced an F1 score of 0.748, considered the best indicator of model quality based on a desired balance between false positive and negative results.

model	precision	recall	F1	accuracy
RF CV	0.693438	0.812539	0.748210	0.712168
RF test	0.684375	0.817673	0.745109	0.705536
XGB CV	0.673074	0.724487	0.697756	0.669669
XGB test	0.675660	0.747978	0.709982	0.678349

F1 scores and additional metrics for random forest and XGboost models

Future model suggestions

- Promote collection of tipping behaviors, including cash-based payment methods and tipping patterns for individual customers.
- Utilize clustering techniques such as K-means to identify novel patterns and insights from the data.

Results Summary

The champion algorithm is a usable predictor by NYC TLC drivers for riders who may be generous tippers. The model is reasonably strong based on calculated metrics, but may still have room for improvement to increase its usability.

Next Steps

Going forward, Automatidata is able to recommend this model for use by the New York City Taxi & Limousine Commission. It may become a component of helping their drivers increase income through predictions of customer tipping behaviors, but may also require additional efforts to improve the model's predictive capabilities.