Machine Learning Model Outcomes

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

Prepared by Automatidata

Overview

Previously, a project proposal was developed, and a Multiple Linear Regression (MLR) model was created to predict customer fare amounts. Now, the New York City Taxi and Limousine Commission (TLC) has requested us to develop another model that can predict the amount of gratuity a driver receives.

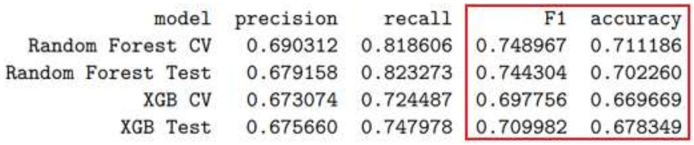
Objective

The objective is to develop a model that can accurately predict the gratuities received by drivers. This model must be constructed in a manner that is both ethical and morally sound. The initial modeling goal was rejected due to ethical concerns related to predicting non-tippers. Consequently, a different modeling objective was implemented: predicting generous tippers.

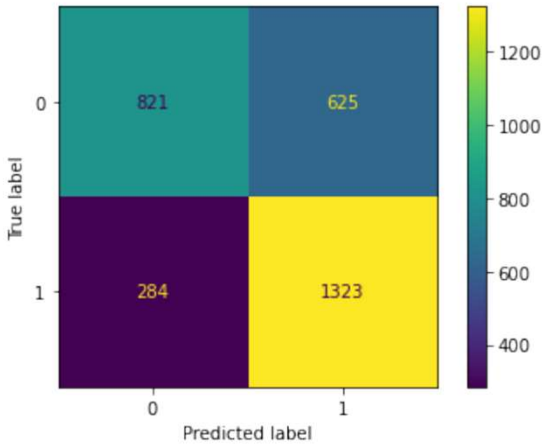
Results

The data teams' assumptions that predicted fare amount, time of day, and a trip's itinerary were confirmed to be valid, as relationships between these variables and the target variable were identified.

Two modeling algorithms were tested and compared, with the Random Forest model performing better. Accuracy and F1 score were used as performance metrics due to the relatively balanced target variable. Although the model performed well on the test set, it exhibited a potentially problematic Type I error rate. Nevertheless, the model overall delivers satisfactory results, and therefore we can proceed with implementing the model in a test environment.



```
             model  precision    recall        F1  accuracy
  Random Forest CV   0.690312  0.818606  0.748967  0.711186
Random Forest Test   0.679158  0.823273  0.744304  0.702260
            XGB CV   0.673074  0.724487  0.697756  0.669669
          XGB Test   0.675660  0.747978  0.709982  0.678349
```



Next Steps

Given the positive results, the model can be further refined if needed. More data and clustering algorithms could provide additional insights. To evaluate real-world performance, test on a small subset of taxi drivers. Share the results with the New York City TLC and recommend its use as a tip amount predictor. Significant improvements may require more data.