

Machine Learning Model Outcomes

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

Prepared by Automatidata

Overview

New York City Taxi & Limousine Commission has contracted with us to build a machine learning model to predict whether a rider will be a generous tipper.

Problem

The challenge we encountered was defining a value to assign to the word “generous.” We also considered ethical consequences of building this model, particularly regarding the errors that we will encounter when the model is built.

Solution

We built two different models and compared results and parameters. Despite hypertuning and boosting, the outcome does not deliver as strong of predictions as anticipated. We would recommend using this model only with the understanding that its predictions succeed at roughly a 0.65 rate.

Details

Behind the data

- Our assumption was that location and time of day may have a strong enough relationship with tip amount that we could accurately predict generous tipping from just these variables.
- As we built our models and performed the testing, it became clear that there was not as strong a correlation as anticipated; but rather a passable one.

	model	precision	recall	f1	accuracy
0	random forest: f1	0.661	0.309	0.421	0.714

	model	precision	recall	f1	accuracy
0	XGBoost 1: f1	0.659	0.484	0.558	0.742

Future model suggestions

- Remove “last_evaluation” as a feature
- Predict performance and evaluation scores
- Cluster with K-means and analyze the clusters to derive insights from the data

Results Summary

The resulting algorithm is usable to predict riders who might be generous tippers, but it is not a model with an overly high degree of correct predictions.

Next Steps

- Contact New York City TLC and provide them with a summation of our work.
- Share the results and recommendation that the model could be used as an indicator, as long as there is an understanding of its limits on correct prediction.