New York TLC Machine Learning Model

Executive Summary Report

Overview

The Automatidata team has been requested to build a machine learning model on whether a customer will be a generous tipper.

Problem

The original task was to predict whether a customer will LEAVE A TIP or not. This is unethical and it will be a name-calling model. Instead, to avoid this, the generous tipper (leaving 20% tip or more) model is built.

Solution

The data team used two different supervised classification model types: random forest and XGBoost. Both models have good results, with the random forest having a slightly better one.

Details

	model	precision	recall	F1	accuracy
0	RF CV	0.691244	0.813783	0.747515	0.710613
0	RF test	0.690489	0.817673	0.748718	0.711104
0	XGB CV	0.683153	0.745177	0.712786	0.683999
0	XGB test	0.683509	0.766024	0.722418	0.690141

All four metrics were evaluated and the validation score F1 shows that the random forest slightly edged the XGBoost model.

Next Steps

The data result is to be shared with the New York TLC as the final model results and be recommended that it will be used as an indicator of generous tipping.