

# 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 the Automatidata data team to build a machine learning model to predict whether a NYC TLC taxi cab rider will be a generous tipper.

## Problem

One of the concerns for the drivers were if the client was going to tip after the trip (> 20%). Therefore, we propose a model that could predict which clients will leave a generous tip.

## Solution

The team developed two different modeling architecture and compared their results. Both of the models perform great, but random forest was the model with better metrics

## Details

### Behind the data:

- The first assumption made from the team was that the trip's itinerary, predicted fare, and time were the variables that determine the tip situation
- After creating our model and performed the testing. The team discovered that some that the variables that affect the tipping situation were predicted fare, mean duration, and mean distance. The f1 model's score was 0.71941

### Results Summary

The resulting algorithm is usable to predict riders who might be generous tipper, according to the metrics from the table.

	model	precision	recall	F1	accuracy
0	RF CV	0.675733	0.759023	0.714830	0.681461
0	RF test	0.672069	0.774113	0.719491	0.682280
0	XGB CV	0.669775	0.716086	0.692084	0.664592
0	XGB test	0.672241	0.750467	0.709203	0.676056

Image Alt-Text Here F1 Scores for random forest and XGB models.

### Future model suggestions:

- Collect more granular driver and user-level data.
- Use the K'means to cluster the information.

## Next Steps

As a next step, the Automatidata data team can consult the New York City Taxi and Limousine commission to share the model results and recommend that the model could be used as an indicator of tip amount. However, additional data would be needed to realize significant improvement to the model.