Generous Tip Prediction ML Model Outcomes

Executive summary report for the New York TLC

Prepared by Automatidata

OVERVIEW

To identify which variables or factors influence the amount of gratuity a rider gives a driver while predicting If rider will be a generous tipper. This in turn will help make informed business decisions that will increase gratuities and subsequently improve driver satisfaction.

PROJECT STATUS

- The initial objective (predicting non-tippers) was rejected out of ethical concern that cash tippers' data wasn't available.
- It was decided to predict "generous" tippers—those who tip ≥ 20%.
- This decision was made to balance the sometimes competing interests of taxi drivers and potential passengers.
- The model gives a decent F1 score of 0.77. and thus is usable to predict riders who might be generous tippers.

NEXT STEPS

- The model could be used as an indicator of tip amount.
- It can be improved to a limit by including features like (actual time taken / mean time for those location pairs), tax percentage in total amount, etc.
- But still there are some really important untracked features like behaviour of the driver, temperature inside the car, previous history of rider, etc which are needed to get proper prediction.

KEY INSIGHTS

- A lot of false negatives are predicted i.e. the driver will think that a customer won't give generous tips when the customer actually would have given a generous tip. The driver might not even pick them up.
- But if the model says a customer will give a generous tip then its highly likely that they'll actually give a generous tip.
- The tip amounts are highly dependent on mean duration and distance and passenger count.
- The current best recommendation is to try to invent ways for reducing trip duration, maybe use better navigation.
- Using the dataset and similar model, we can also predict whether a driver would pick a customer or not.