NY taxi time prediction

By: RISHIT DHOLAKIA

Report

NY taxi time prediction

# Synopsis of the code

This was a code that started of with , first making the generalized table for my test and training then rearranging the features like splitting the time and the rest of the other feature conversion like scaling etc.

Then it was time to check on the anomalies, addition of simplified columns and addition of extra information like weather that is not provided in the dataset.

ANN is the algorithm used to train this dataset, reason being the dataset is very large.

# What more could be done ……..

If I had extra time I would have wanted to explore the date feature more in depth by adding features of week day, weekend, which of the days there were more cabs running.

Using the shortest distance from going from one place to another and then comparing with the observed distance and analyzing the delay that can give an idea on the delay feature for that period of time.

Another algorithm related ensemble learning called xgboost was what I would want to experiment on if I had time and compare it with the ANN.

# If given a DGX

Would use stratified cross validation on my dataset as this is a complex process and cannot be handled by then machine I have.

Increase the layers I am currently working on to get a better result at a faster rate.