NEW YORK CITY TAXI TRIP DURATION PREDICTION

WEILING DENG

Contents

1. Introduction 2
2. Process 2
3. Conclusions 2

Date: (None).
2020 *Mathematics Subject Classification.* Artificial Intelligence.
Key words and phrases. New York City Taxi Trip Duration Prediction.

1. Introduction

This project required to build a model that predicts the total ride duration of taxi trips in New York City. The primary dataset is one released by the NYC Taxi and Limousine Commission, which includes pickup time, geo-coordinates, number of passengers, picktime and dropoff time, and several other variables.
In this section, The train data which we have 1458644 Rows and 11 columns. The test data which we have 625134 Rows and 9 columns.
We built the linearregression model, using distance to predict the trip duration and test.

2. Process

There are steps as fellows:
Step 1, Data ETL, inclouding read data, cleaning data, recheck NAs and decrible data types.
Step 2, Knowledge Discovery, Calculate the distance by latitude and longitude. Distribution distance and trip duration, and then visualization then, in order to find the characteristics about distance and trip duration, prepare for select featuers.
Step 3, Model Built and Prediction, set distance and trip duration as attribuates. In this section we first to prepare data, which including Select features and gropby data to train data and test data, from 0 to 150 as train data, 150 to 180 as test data, named as train and test.
And then, Built LinearRegresion Model and find best parameter to predict, we use distance to predict the trip duration and test. The result presents that the LinearRegression RMSLE is 4.616529350404572.
Last step which is visualization the prediction result.

3. Conclusions

The Propose of this project which is to find the data features, select the attruibuates to built the linearregression model, find the best parameter using distance data to predict the trip duration and test.But , from the chart we can see the model is not very good, maybe we an explore other model to improve the accuracy.

(A. 1) Business School,, Deakin University, Burwood 3125, Australia
Email address, A. 1: ddengpp@gmail.com