Uber Predictions in New York City

Pritam Borate, Puneeth Devabhaktuni, Aditi Helekar, Lily Naoudom

Abstract

Uber Technologies, Inc. is a transportation network company and a pioneer in the sharing economy. The Uber software application is used by drivers and clients during a transaction, and documents pertinent data including the date and time of pick-up locations, longitude and latitude coordinates of pick-up locations, and even base company dispatchers. Using this data, we can apply machine learning algorithms to predict Uber pick-up locations along the boroughs in New York City.

Objectives

- Identify and collect relevant training and test data sets for Uber transactions in New York City.
- Apply machine learning algorithms to classify and predict likelihood of Uber pick-up based on longitude and latitude coordinates and hour of the day provided in the data sets.
- Compare results of Uber predictions with Lyft, one of Uber's business competitors.

Materials & Methods

Data Sets

Uber data (April 2014 – Sept 2014) from Kaggle.com

Attributes:

- Date / time
- Longitude and latitude of pick-up locations
- Base company dispatcher

Machine Learning Algorithms

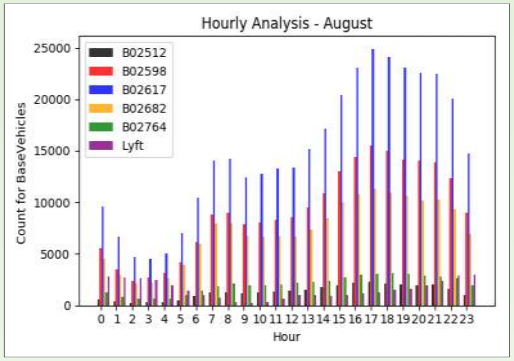
- Decision Trees
- K-Nearest Neighbor
- Neural Networks
- K Means Clustering
- Naïve-Bayes
- Ensemble (Adaboost)

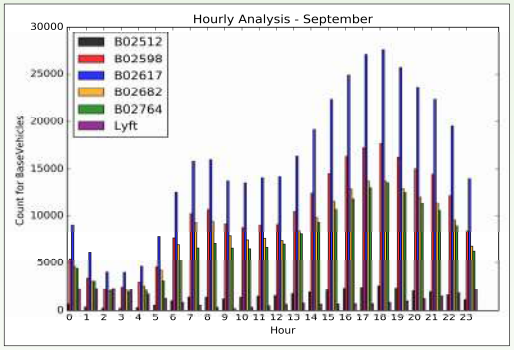
Language & Libraries

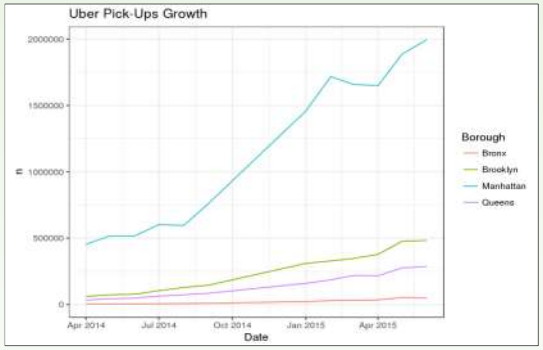
python

scikit learn

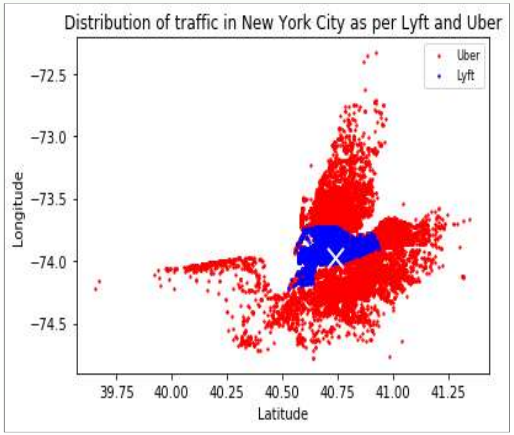
Results







Analysis & Conclusions



As seen from the scatter plot on the left, Uber pick-ups are spread out in New York City as compared to that of Lyft. Lyft pick-ups are clustered in the central part of the city.

Using K Nearest Neighbors and distance as weight we were able to predict Uber pick-ups with 88% accuracy.

Combining our research methods along with the increasing use of Uber Technologies, Uber drivers and base-company dispatchers can use this data to optimize their client transactions based on geographical location.