Uber Fare Prediction

Present By – Preeti Sahani(46)

What Is Uber?

- A transportation company with an app.
- Founded by Travis Kalanick and Garrett Camp, in 2009.
- It had 118 million monthly active users worldwide

UBER

Overview

- This is a basic regression model to predict the fare of the uber ride trained on top of Sklearn RandomForestRegressor model. The model would take the different trip parameters (number of passengers, pick up and drop off geographical coordinates, date and time of the trip) as the input and predict the fare amount as output. The model has been trained with 2 lakh records and acquired an score of 0.968.

About Dataset

2000 Rows & 8 Columns

- key
- fare_amount
- pickup_datetime
- pickup_longitude/latitude
- dropoff_longitude/latitude
- passenger_count

UBER

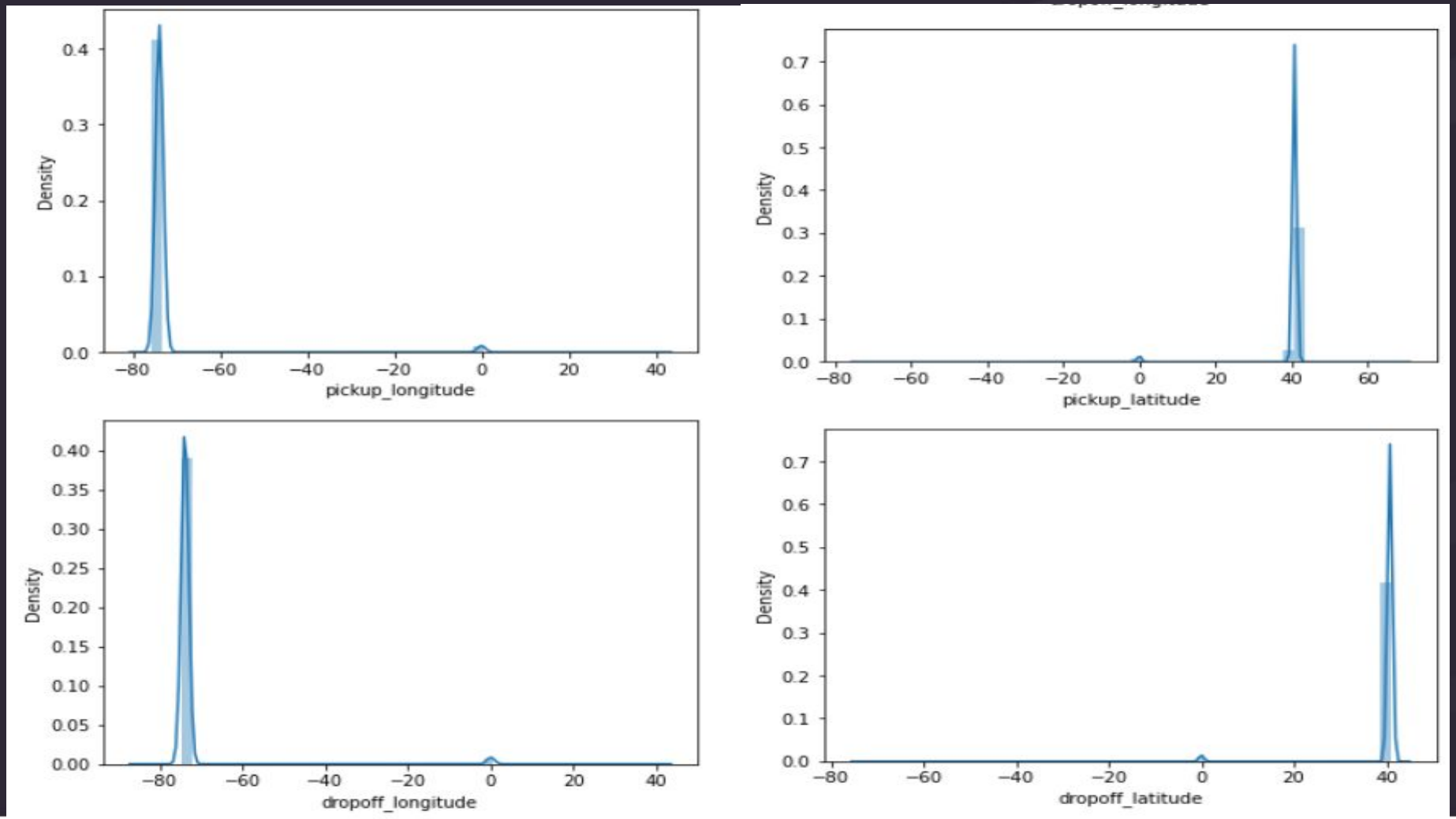
Data Preprocessing

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model.

Step:

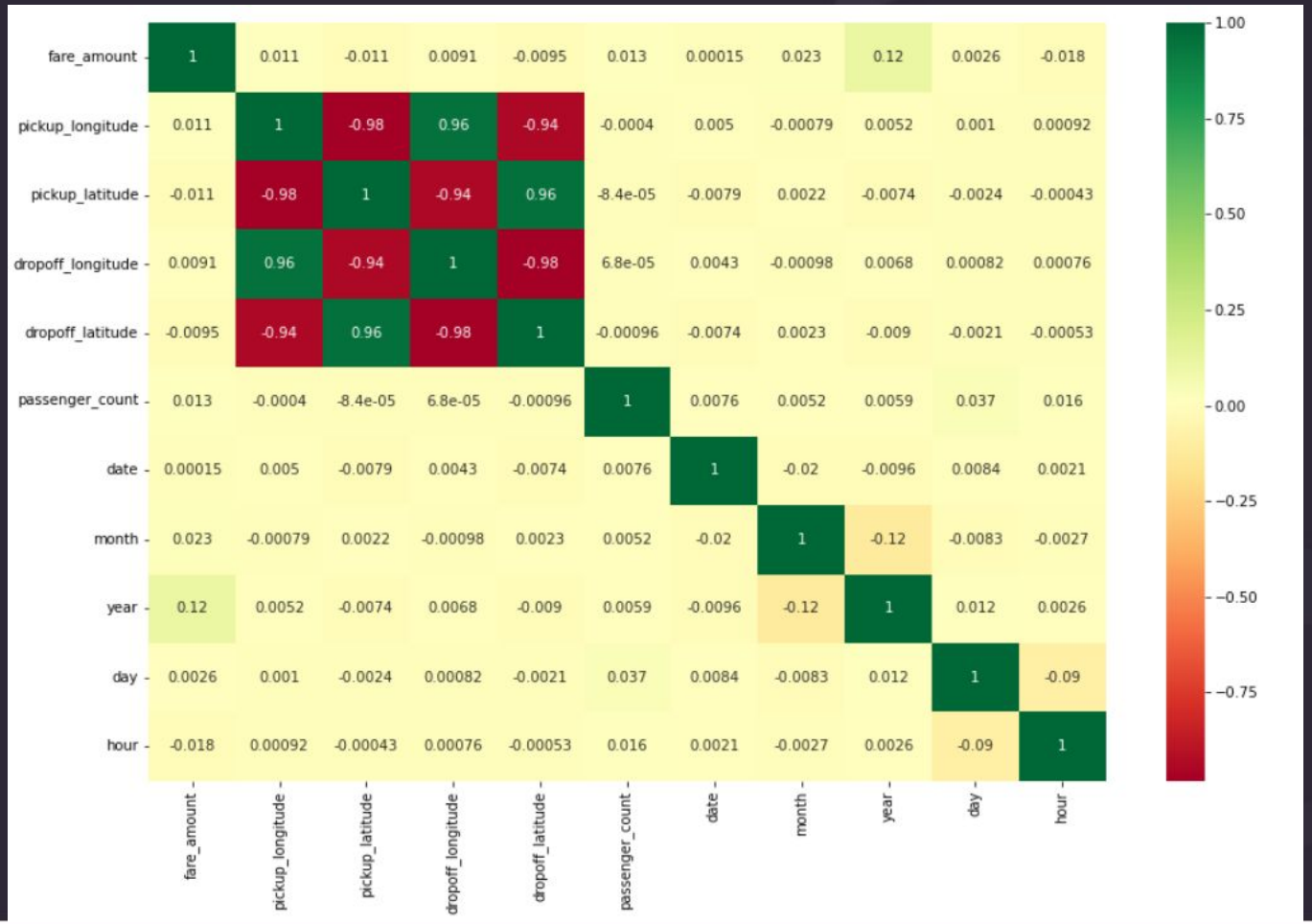
- Getting the dataset
- Importing libraries
- Importing datasets
- Finding Missing Data
- Encoding Categorical Data
- Splitting dataset into training and test set
- Feature scaling

Distribution Plot



UBER

Correlation



UBER

Random Forest

- A random forest is a meta-estimator.
- A supervised learning algorithm.
- It is a bagging technique.
- It operates by multiple decision trees at training time and give output as a individual tress.

Benefits & Challenges

- Reduced risk of overfitting
- Easy to determine feature importance
- Provides flexibility

- Time-consuming process
- Requires more resources
- More complex

Result

- Accuracy score = 96.87%
- RMSE = 5.22

CONCLUSION

- The Random Forest Regression Model for prediction of dynamic price of trips is providing an efficiency of 96.87%. It is better suited for the prediction of target variable which is trip fare, and it performs very well. Further this work can be carried out using different machine learning algorithms and techniques in order to get higher efficiency and lower errors

Thank You