

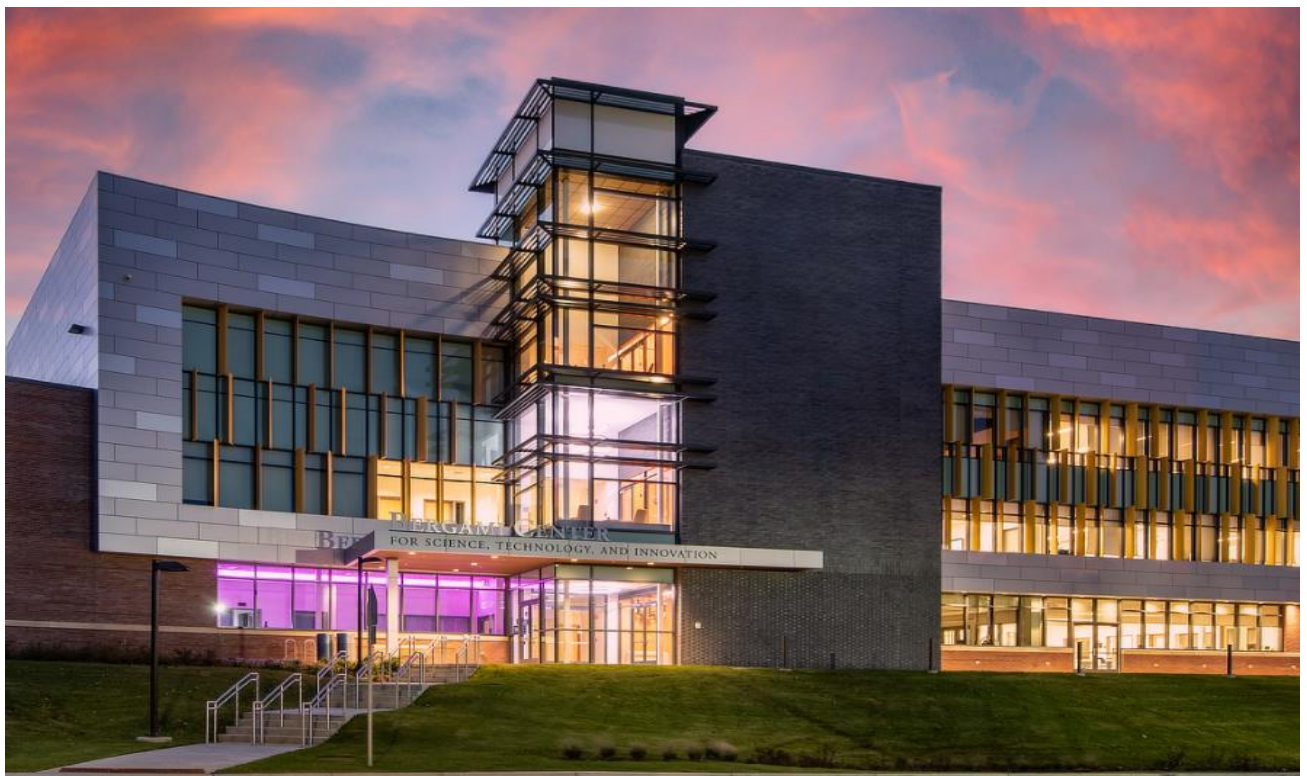


University of New Haven

TAGLIATELA COLLEGE OF ENGINEERING

Electrical & Computer Engineering and Computer Science

Electrical & Computer Engineering & Computer Science (ECECS)



CONTENTS

Project Name	2
Summary	2
Abstract	3
Highlights	3
Methodology	4
CRISP DM	5
Conclusion	5
References	6

Price Prediction by analyzing Uber and lyft data

Summary

The main idea of the project is to know the effect of each field on price with every other field of the dataset. By applying different machine learning models to complete the analysis. The data is collected from the Kaggle. Then data is processed and analyzed by using various libraries in python like Pandas, Skit learn. A question is framed based on the collected data and is processed through various operations to analyze the situation and obtain an answer. After the data is read, cleaned, analyzed/visualized, and represented, this analyzed data helps in getting the desired output which is the objective of the project. Based on the output conclusion for the project is determined

Team Members:

Name 1 Harshavardhan Reddy

Name 2 Sai Mohan Majoju

Name 3 Sharath P

Name 4 Raghavendra Reddy

Submitted on:

10-19-2022

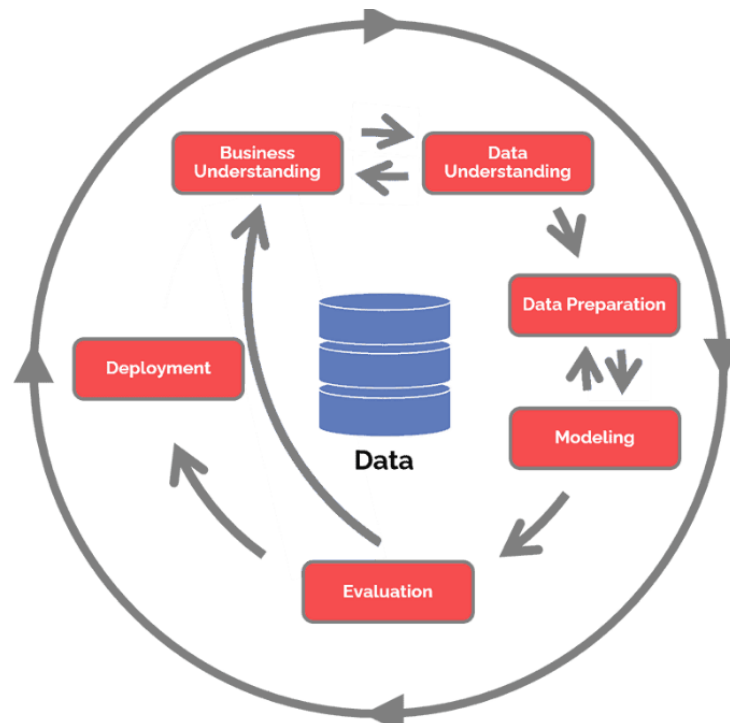
Abstract

The business concept of using platforms that are competitive, like Uber and Lyft which provide rides. Therefore, the goal of a superior establishment is to give customers a better deal than the competition. The pricing for our business model will become clearer by analyzing the price trends of uber and lyft. It's crucial to think about whether we will do this first or whether competitors have already personalized prices using data analysis. and "Can we offer lower prices than rivals?"

Highlights of the project

The steps involved in the project are:

1. Data Collection.
2. Analyzing
3. Visualization.
4. Evaluation.
5. Deployment.



CRISP DM

Data Collection:

We need data for understanding the effect of other fields on the price column. The required data can be gathered by a known technique called as Web Scraping. First of all the data required for the project need to be selected from various sources like Kaggle, data sets or any other platform. In our instance, we selected from the kaggle. After selecting the data we used the data gathering techniques and collected the data into a CSV file. The process of gathering data from different platforms for the study of data is called Web Scraping.

Analyzing:

The data gathered from different sources by using various techniques is to be analyzed with data analyzing tools like Pandas, NumPy and matplotlib, etc. Analyzing/cleaning is to read the data which is gathered. During this process, the unwanted data is cleaned and made precise for data visualization. Based on this analysis the data study is performed on the gathered data and relevant output or outcomes is/are determined.

Visualization:

The analyzed data based on the need may be required to be visualized and represented with visualization. We normally use libraries like Matplotlib and Bokeh for data visualization. Visualization is one kind of technique that converts raw data into an easily understandable format. This may be the representation of the data in the form of bar graphs, pie charts, scatter plots, etc. This process helps in understanding the data and representing the data in a simpler format that can be used to draw conclusions quite easily.

Evaluation:

Based on the results of the analysis and the visualization techniques, the results that are obtained will be evaluated to arrive at a model for prediction.

CRISP DM

Business understanding: What is the primary requirement in this project providing a fair price for the rides to improve business

Data understanding: What Type of data we are dealing with and how to understand the data plays very important role. In this project we are dealing with structured data which contains previous ride information and price.

Data preparation: Finding the relation between the features and cleaning the data is part of this step. We use various techniques to prepare the data to train a particular model

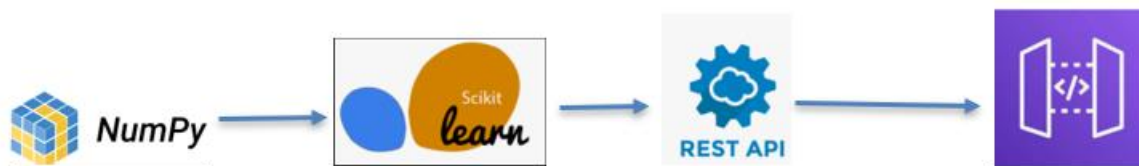
Modeling: After Analysis the model is selected for training. Training and test data splits are made to train and test the model.

Evaluation: The trained model is evaluated based depending upon the accuracy the model is deployed.

Deployment: The trained model is deployed in server and ready for consumption

Conclusion:

Depending upon the accuracy of the model. The model is retrained and deployed.



Reference:

<https://www.kaggle.com/datasets/fivethirtyeight/uber-pickups-in-new-york-city>

<https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma>

<https://github.com/majojusaimohan/DSCI-6007-02->