Proposal

Graduation Project

Project Include: <u>Data Science Project</u>

1) Data Analysis   **Business Cases:**

2) Data Cleaning   * Uber Trip Price Prediction

3) Data Preprocessing

4) Machine Learning

5) Deployment   Dataset: from Kaggle

Link:

1) https://www.kaggle.com/datasets/brllrb/uber-and-lyft-dataset-boston-ma
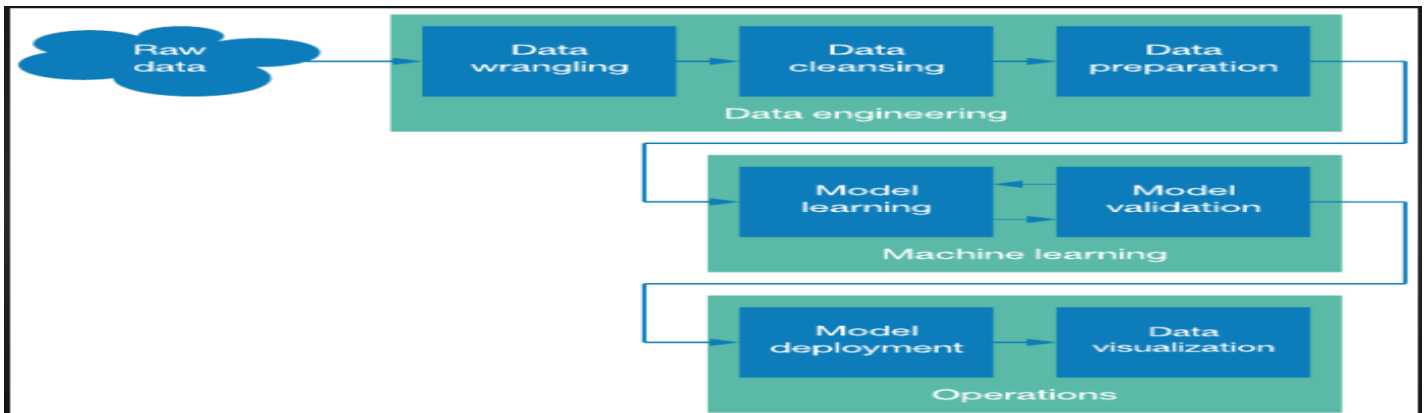
This dataset consists of (693071 rows × 57 columns)

- Problem Understanding:

- This is data sets collection from uber and lyft at United States in specific at Boston

- Information of the Boston in 24th-most populous city in the country at united states

- The goal of this data set is predicting the price with specific feature

* Machine Learning Problem: Regression

Project Pipeline

Project Steps

- We will store the dataset that we brought from the Kaggle website in our own database

- We will apply the stages of data cleaning

      1- Handel a missing data   2- Detect and handle Outliers

      3- Drop the duplicates   4- work with inconsistent data

- And then we enter the stage of data analysis (Descriptive Model)

    At this stage, we will try to get insights from the data that can benefit business.

– We will apply the stages of data preprocessing

      1- Split data into input and output

      2- Feature Encoding

      3- Split data into train and test

      4- Feature Scaling

–Model evaluation

-We will see which technology has given us more accuracy and we will use it

- We are making a small website using streamlit to display the results of the models

- We will deploy the model on website using cloud service