Low Level Design (LLD)

Petrol Price Forecasting (Machine Learning)

By

Kavitha Narsapur

Abstract

The ONGCF is a organization dedicated to the exploration and production of oil and natural gas. Price information is supplied on a weekly basis. It seeks to forecast crude oil prices for the following 16 months, from January 1, 2019 to April 1, 2020. The main goal is to predict and forecast the prices based upon the best model. Different models like Auto regressive (AR), Moving average (MA), Auto regressive moving average (ARMA), Auto regressive integrated moving average (ARIMA) and Seasonal auto regressive integrated moving average (SARIMA) models were built. Out of all the models, ARIMA best predicted the petrol prices with least RMSE of 0.21 and MAPE of 3.03%. Hence ARIMA model can be best utilized to forecast the petrol prices.

Introduction

Why this Low-Level Design Document?

This Low-Level Design (LLD) Document summarizes the data used for forecasting petrol prices and also provides an overview of steps used in this project as Data preprocessing, Exploratory data analysis, Model building and Model deployment.

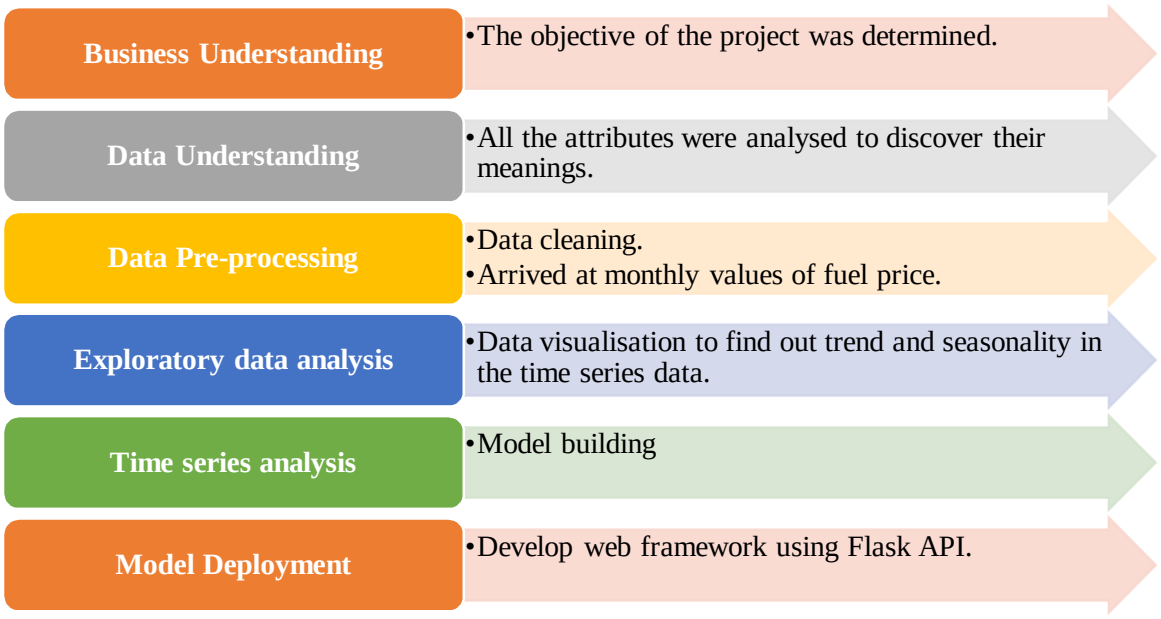
Scope

The LLD documentation presents the structure of the system, such as the application architecture (layers), application flow (Navigation), and technology architecture. The LLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system. This software system will be a Web application. This system will be designed to forecast the petrol prices.

Dataset overview

The data contains petrol prices (Petrol (USD)) and Date column in train_data.csv. The test data has the dates for which the predictions are to be made (column names: Date and Prediction), corresponding to which Prediction is blank.

Architecture



Architecture Description

1. Business objective:

The aim of this project is to develop a solution using data science and machine learning to forecast the petrol price for next 16 months.

2. Data Understanding:

An attempt was made to understand the meanings of all the variables present in the data. The data type of each variable was also determined.

3. Data pre-processing:

Data cleaning: Data was checked for the presence of missing values.
Data was rolled up to arrive at the monthly values of price.

4. Exploratory data analysis:

By means of graphical analysis, trend and seasonality in time series were visualized.

5. Time series forecasting:

Train and Test data creation: Train and Test data were created by splitting the data and 70% of the data was used as train data and 30% was used as test data.
Model building: Different models like Auto regressive (AR), Moving average (MA), Auto regressive moving average (ARMA), Auto regressive integrated moving average (ARIMA) and Seasonal auto regressive integrated moving average (SARIMA) models were built.
Model evaluation: The built models were tested on the test data and evaluation metrics used were RMSE and MAPE.

6. Model Deployment:

Best performing model was saved in Pickle format and the model was tested using Flask API on local system.

Conclusion

The web framework developed in this project can be used by the users to forecast the petrol price.