High 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 High-Level Design Document?

This High-level Design (HLD) Document indicates all the necessary steps that were carried out prior to building time series model as Data pre-processing and Exploratory data analysis. After data preparation, the document indicates which models were built and tested on test data and also describes model deployment.

Scope

The HLD documentation presents the structure of the system, such as the application architecture (layers), application flow (Navigation), and technology architecture. The HLD 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 petrol price.

General Description

Introduction

Thus this project aims to forecast petrol price using machine learning models.

Problem Statement

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.

Approach

This project involves data pre-processing like Data cleaning, Exploratory data analysis, Model building and Model deployment. 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 and then tested on to test data.

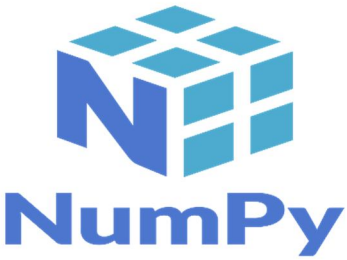
Dataset overview

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.

Further Improvements

The performance of machine learning models can be further enhanced by tuning various hyperparameters of the models.

Tools used

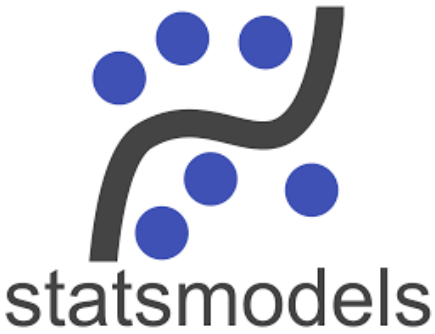



scikit
learn

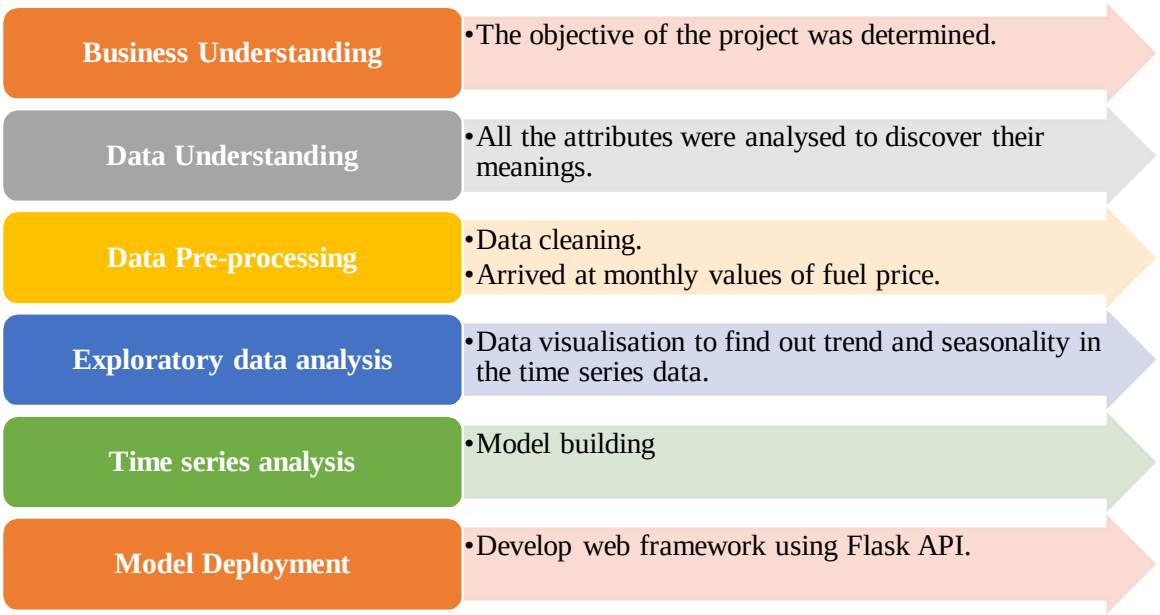

Pickle


Flask


python


statsmodels

Architecture



Conclusion

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.