

Project Report on

**“Forecasting the price of crude oil”**

Submitted in fulfilment of the requirement for the award of the internship certificate

as

## “Data Scientist Intern”

Submitted by

### Mr. THILAK RAJU S

### Mrs. RUCHI HARADHARA

### Mr. RISHI GUPTHA

Under the guidance of

## Mrs. Aishwarya Mate

## Ai Variant – Mentor

## Certificate Code – AIV/23-24/Q3/09/13432

## #49, Ground Floor, 27th Main Near IQRA International School, opposite to WIF Hospital, 1st Stage, BTM Layout, Bengaluru, Karnataka 560068

## ABSTRACT

Develop a forecasting model to accurately predict the future prices of crude oil based on historical data, market trends, and other relevant variables. The model should consider the impact of various factors such as geopolitical events, economic indicators, supply and demand dynamics, and changes in government policies. The goal of this model is to provide reliable and accurate forecasts of crude oil prices, enabling stakeholders in the energy industry to make informed decisions regarding production, investment, trading, and risk management

**INTRODUCTION**

Crude oil is the mixture of petroleum liquids and gases that is extracted from the ground by oil wells. It is an important source of fuel and is used in the production of several products. Given the important role price of the crude oil plays, it becomes extremely important for managers to predict future oil price while making operational decisions such as:

* when to purchase material
* how much to produce
* what modes of transportation to use?

The **goal** is to develop a forecasting model to predict the oil prices that aid management to reduce operational costs, increase profit and enhance competitive advantage.

**Effects of price change in oil**

 The changes in crude oil prices are often great indicators as to changes in the overall economy and global markets.

Cases:

* during the US Recession of 2008, the housing crisis caused the price of crude oil to drop, rendering many forecasting models ineffective in predicting future changes
* More recently, the onset of the COVID-19 pandemic in early March 2020 led to a significant decrease in demand for crude oil.
* When the representatives of OPEC + (the informal alliance of OPEC and non-OPEC countries) met to discuss oil production cuts, Russia, a non-OPEC member of the summit refused the demands to cut production, leading to an oil price war against OPEC countries, and in particular Saudi Arabia, causing the prices of crude oil to plummet dramatically.

To develop a model for forecasting crude oil price, we follow below processes.

**Process involved.**

1. Understanding the Business objective.
2. Gathering data from various sources.
3. Data Preparation.
4. Performing Exploratory Data Analysis.
5. Model Building.
6. Model Evaluation.
7. Model Deployment.

* **Dataset summarization​**

Basic information of the data is gathered such as no of rows and columns and datatype of the dataset. Mean, minimum and maximum data is observed.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

* **Exploratory Data Analysis**

1. Dataset summarization
2. Check for duplicate values
3. Conversion of data types to datetime
4. Setting Index
5. Check for missing values and Imputations
6. Outlier detection and removal using capping method
7. Visualization of the data

* **Check for duplicate values**

Duplicated records are checked in the dataset.

A screenshot of a computer code

Description automatically generated

* **Conversion of datatypes to datetime**

Date Column is having object/string data type. We will convert it to Date-Time type

A screenshot of a computer

Description automatically generated

* **Setting Index**

As this is a Time Series Data set, we will set Date column as index

A screenshot of a computer

Description automatically generated

* **Check for missing values and Imputations**

Missing values and missing dates are checked, and corrections is performed using forward fill.

A computer code with text

Description automatically generated

A screen shot of a computer

Description automatically generated

* **Outlier detection and removal using capping method**

A screenshot of a computer code

Description automatically generatedBoxplot is used for outlier detection. Capping method is used to deal with the outliers present in the dataset

A screenshot of a diagram

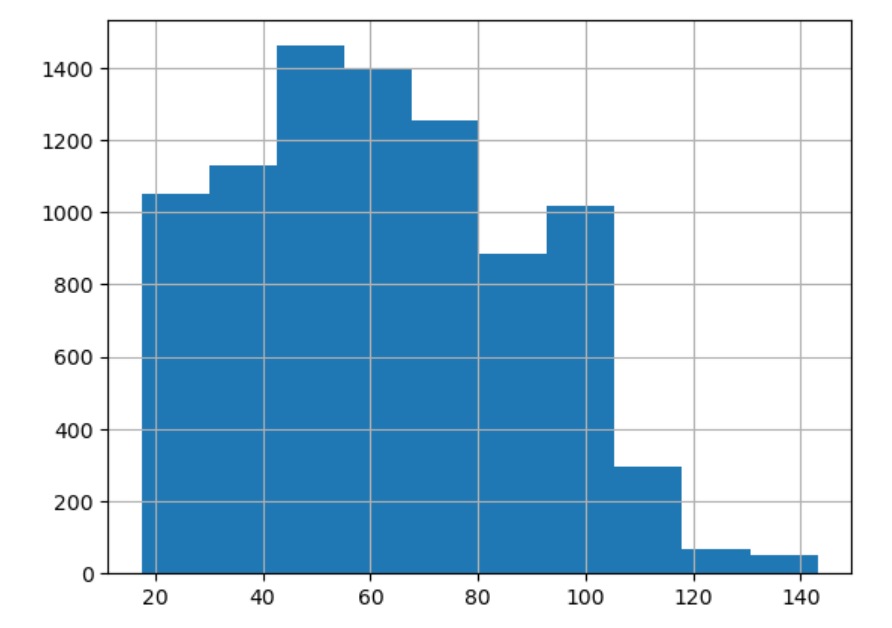
Description automatically generated

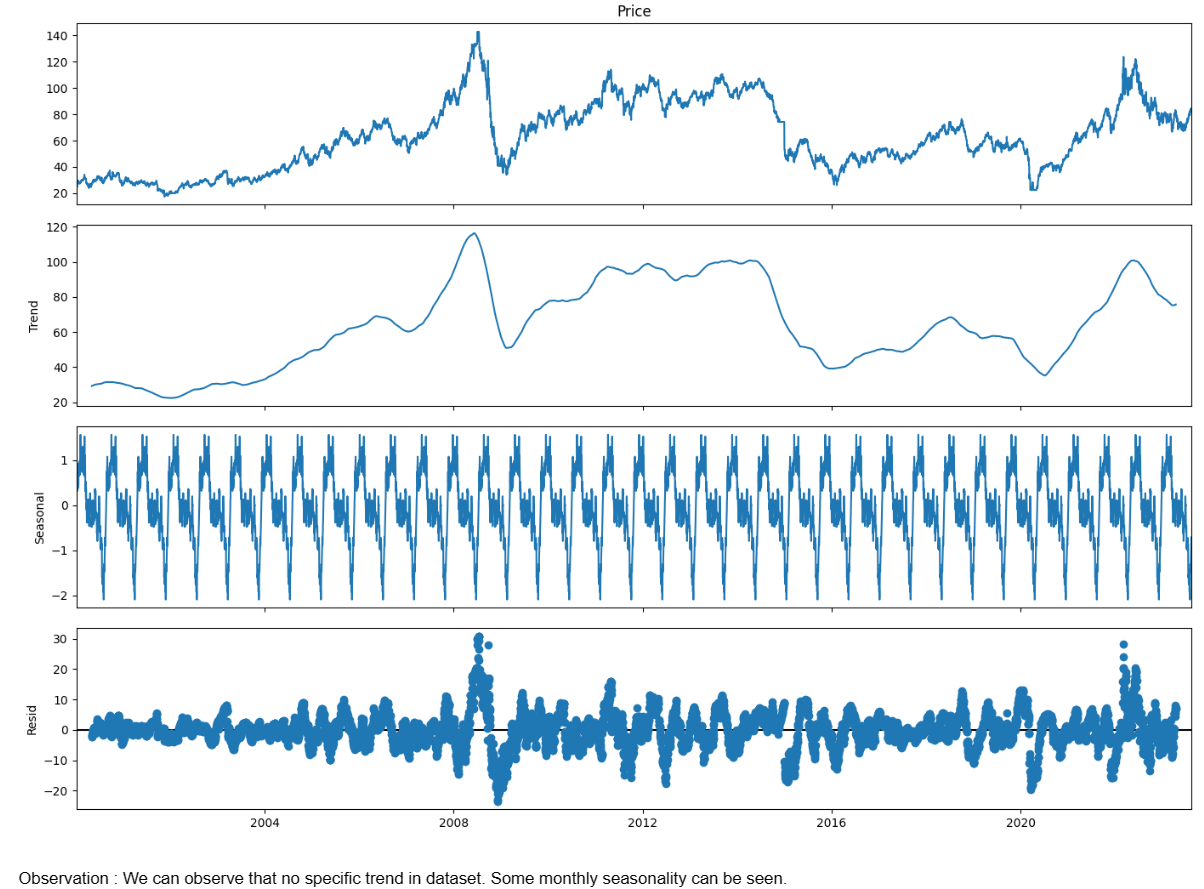
* **Visualizations performed on the dataset**

1. Line Plot
2. Histogram Plot
3. Density Plot
4. Box Plot
5. Lag Plot
6. Moving Averages Plot
7. Time Series Decomposition Plot
8. Auto correlation Plot

A graph showing the growth of the stock market

Description automatically generated



 A blue line graph with numbers

Description automatically generated

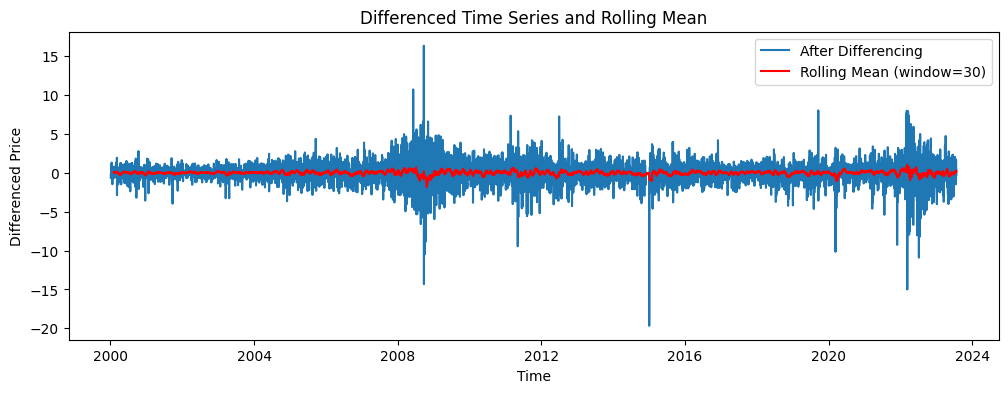
A blue line graph with numbers

Description automatically generated

* **MODEL BUILDING**

By Differencing Method, the p-value of data becomes less than 0.05 and when it gets less than the same, our data becomes stationary, and we checked the same by Ad-fuller Test. Then by sequential method, data is split into train & test.

Where we will get the minimum RMSE value we will build the final model on the same



**RMSE - Root Mean Squared Error**

def RMSE(pred, org):

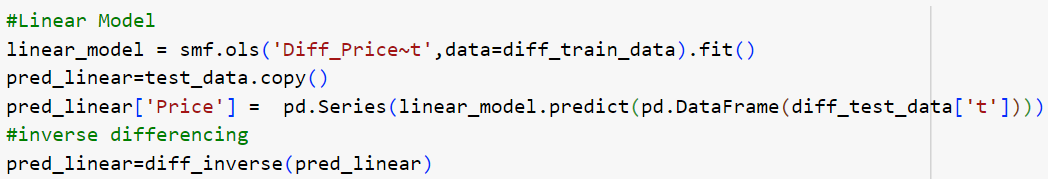
MSE = np.square(np.subtract(org, pred)).mean()

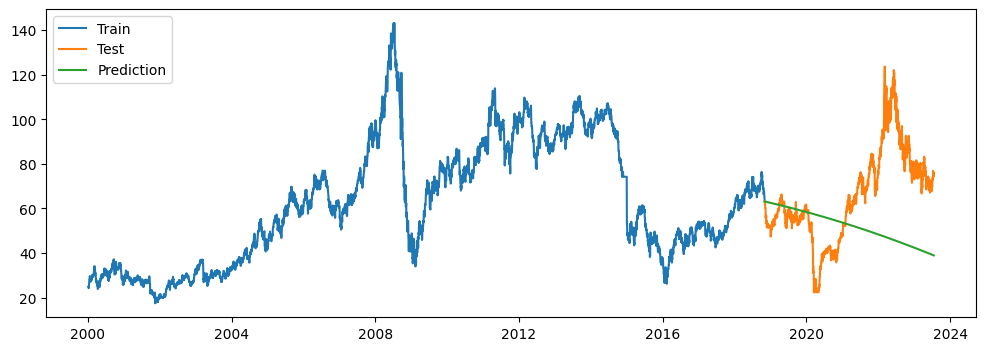
RMSE = np.sqrt(MSE)

return rmse

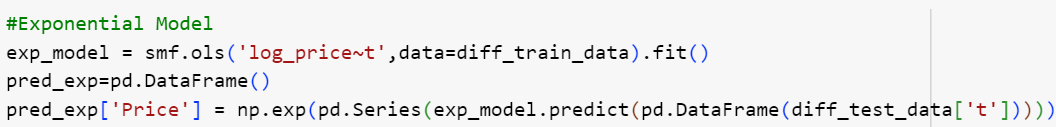
* **MODEL BASED FORECASTING**

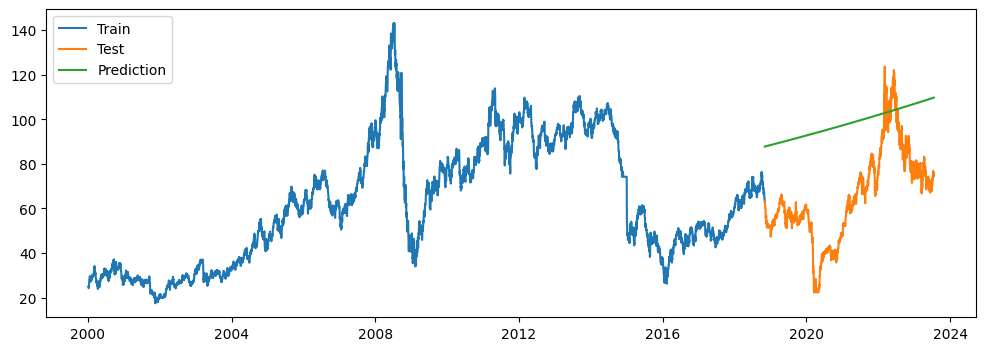
**Linear Model**



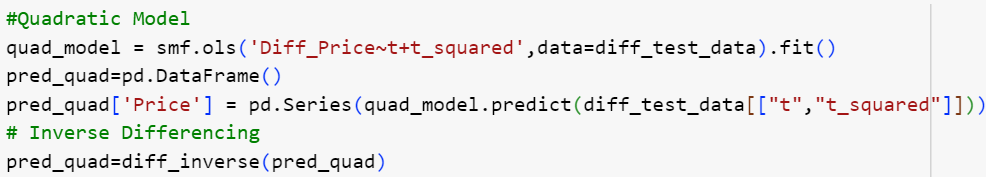


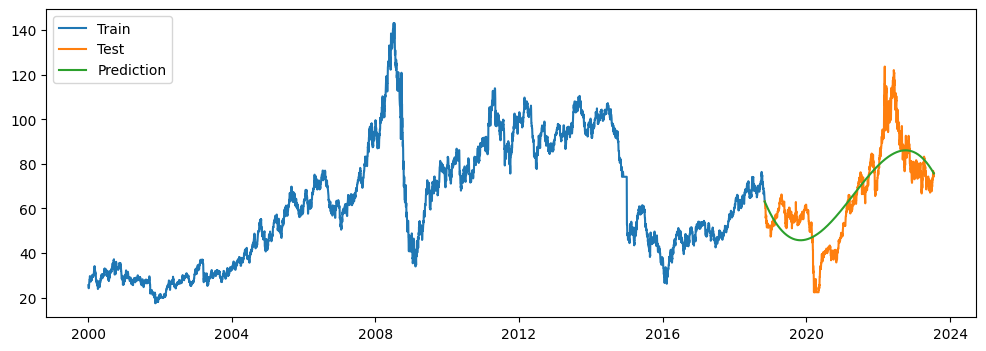
**Exponential Model**





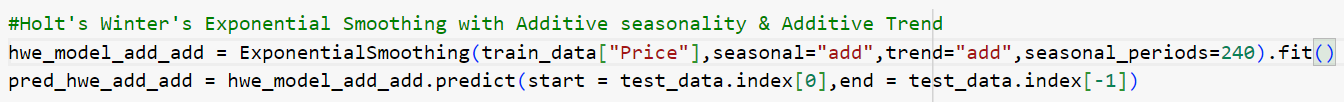
**Quadratic Model**

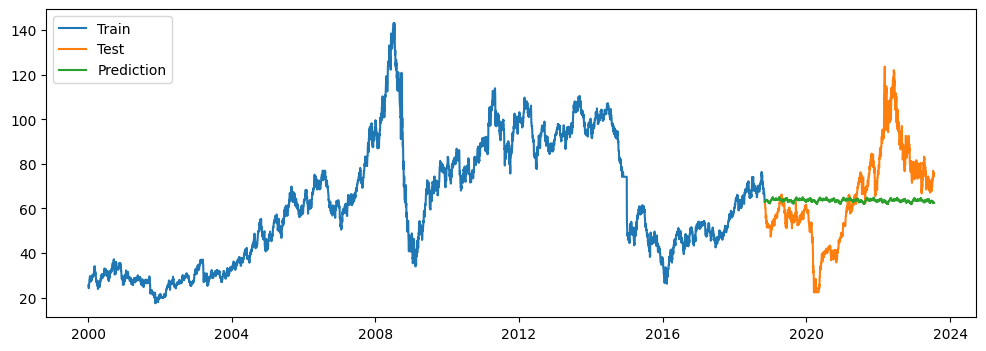




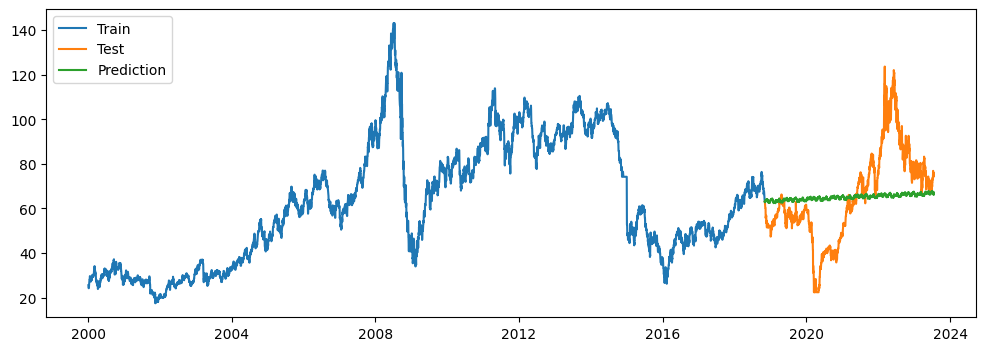
* **DATA DRIVEN MODELS**

**Holt-Winter’s Exponential Smoothing-Additive Seasonality & Additive Trend**

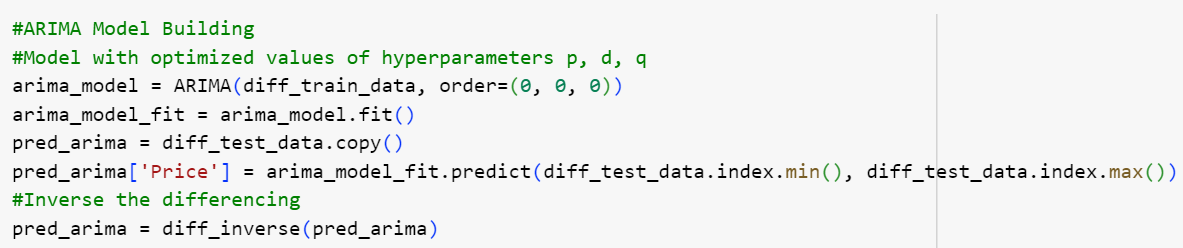


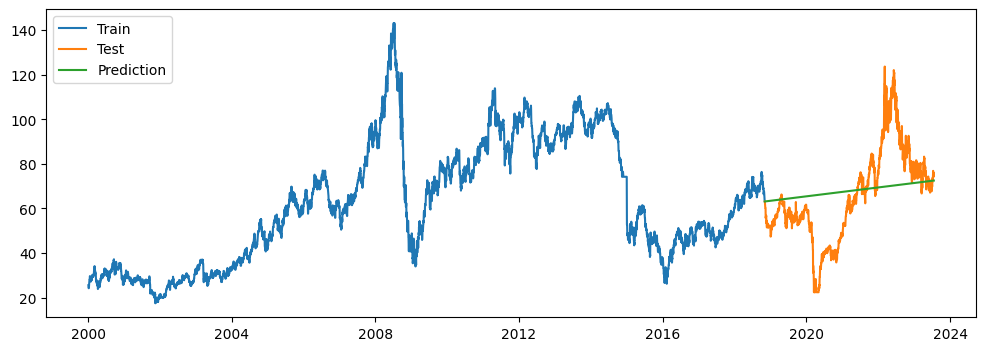


**Holt-Winter’s Exponential Smoothing-Multiplicative Seasonality & Additive Trend**

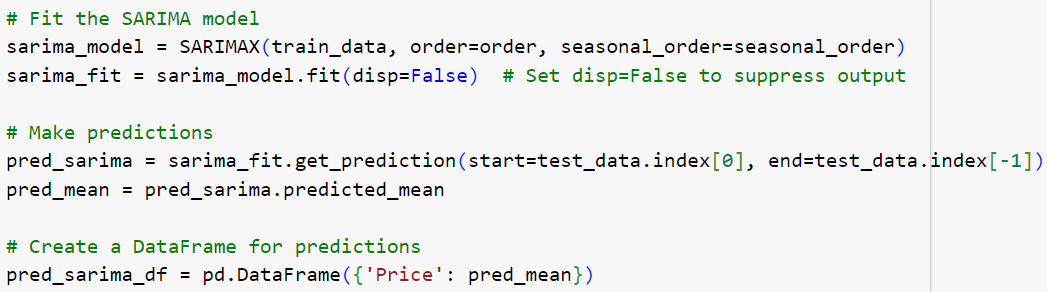
****

**Autoregressive Integrated Moving Average (ARIMA) Model**

****

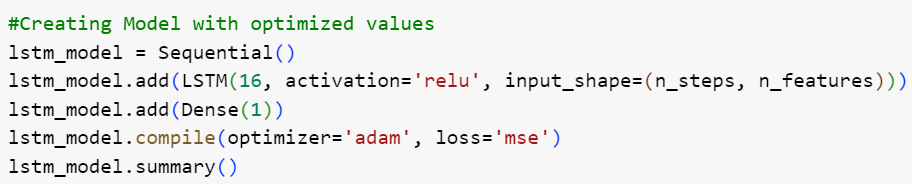


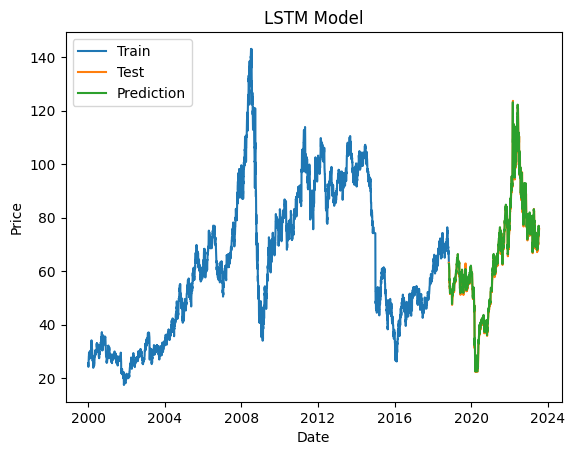
**Seasonal Autoregressive Integrated Moving Average (SARIMA) Model**





**Long Short-Term Memory (LSTM) Model**

****



**Model Comparison**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. No.** | **Method** | **MAPE** | **RMSE** |
| 1 | LSTM Model | 2.036772 | 1.998515 |
| 2 | Quadratic Model | 17.764719 | 11.911926 |
| 3 | FB Prophet Model | 27.708343 | 18.084728 |
| 4 | Seasonal ARIMA Model | 26.510467 | 18.887542 |
| 5 | ARIMA Model | 27.843029 | 18.961463 |
| 6 | Holt's Winter Exponential Smoothing with Multiplacative seasonality & Additive Trend | 28.234913 | 19.859294 |
| 7 | Simple Exponential Smoothing | 29.731985 | 20.491819 |
| 8 | Holt's Winter's Exponential Smoothing with Additive seasonality & Additive Trend | 29.014476 | 20.647820 |
| 9 | Linear Model | 34.228209 | 28.991709 |
| 10 | Exponential Model | 66.613889 | 36.879873 |
| 11 | Holt's Method | 721.161062 | 566.208259 |

**Final Model Building**

* LSTM Model is selected as final model as it has the lowest RMSE value.
* Trained the model with whole data
* Saved the model in Json file
* **Deployment**

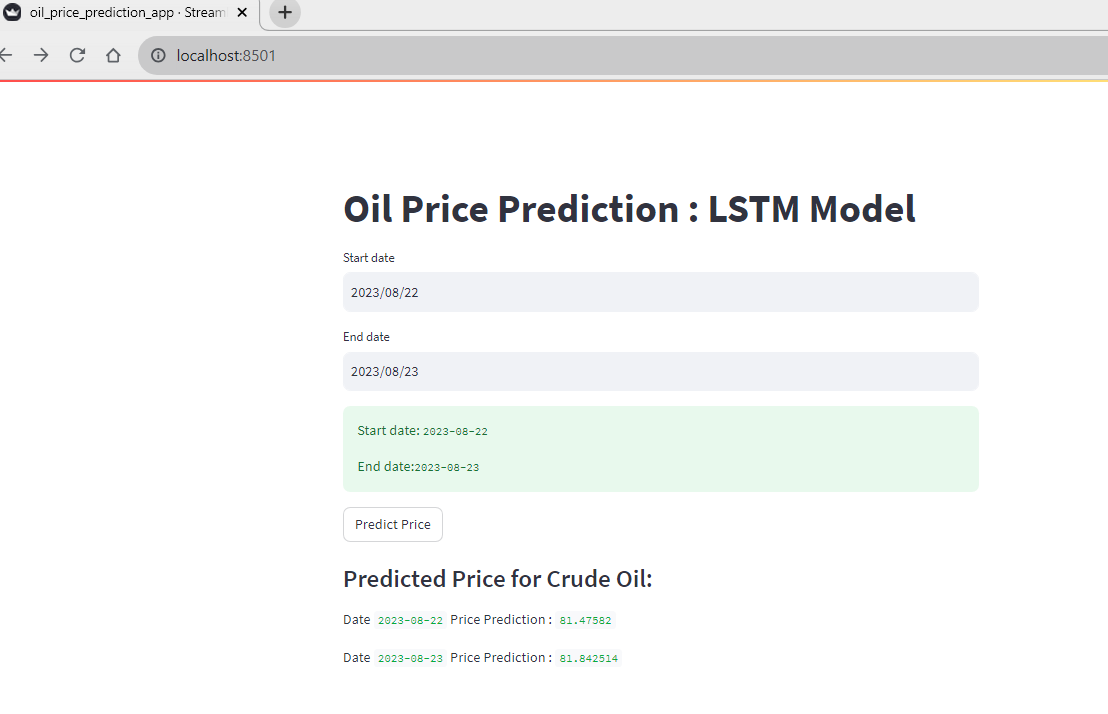
-Created .py file for deployment

-Loaded model from Json file

-Streamlet Library for UI elements



* **Oil Price Prediction Application**



**Conclusion**

Out of 11 models built, LSTM model is selected as it has the lowest Root mean squared error compared to the other models and performed well. Whole data is trained by using LSTM model and the model is saved in Json file. Trained model is deployed using Stream-lit Library for UI elements and an application is created for forecasting the price of the data. The application contains start data and end data as inputs. Once input is given, the predicted price from the start date to end date is forecasted by the model flawlessly.

**References and Supporting File Links**

Introduction - <https://link.springer.com/article/10.1007/s40622-021-00279-5>

GitHub Link - <https://github.com/Thilakraju/Ai-Varient-Intern-Forecasting-Project>