Financial Time Series Analysis Of Crude Oil

Department of Computer Science and Engineering
Amrita Vishwa Vidyapeetham
Amritapuri Campus, India
Kollam 690525
Kerala

Lakshmi G Pillai
AM.EN.U4AIE19074
lakshmigpillai@am.students.amrita.edu

Nithin Sylesh
AM.EN.U4AIE19044
nithinsylesh@am.students.amrita.edu

Ritika R Prasad
AM.EN.U4AIE19053
ritikarprasad@am.students.amrita.edu

Vysakh S Nair
AM.EN.U4AIE19072
vysakhsnair@am.studuents.amrita.edu

Abstract- *Forecasting the future price of crude oil, which has a significant influence in the global economy, is a popular topic among investment firms and governments. However, due to the nonlinear dynamics of the crude oil time series, including chaotic behavior and inherent fractality, forecasting the price of crude oil with high precision is a difficult endeavor. In this paer we are using Brent oil prices to create a forecasting model on how the oil prices varied during the counted years.To anticipate the price of oil, we employed time series and neural network models such as LSTM,ARIMA, and Prophet.*

Keywords : LSTM,ARIMA,Prophet

I.Introduction

Oil is an essential energy resource for economic growth and development, and it is seen as a vital economic engine in both established and developing nations. Demand and supply levels dictate oil prices, but they are also influenced by natural causes of volatility such as business cycles, speculative activities, and political factors.Due to a variety of circumstances, including wars and political instability, economic and financial slowdowns, terrorist attacks, and natural disasters, Crude oil prices have experienced tremendous volatility in recent decades. This is the first study to look at the relationship between spot and future oil prices during four distinct periods of turmoil marked by significant changes in oil prices: the Gulf War, the Asian Crisis, the 9/11 terrorist attack in the United States, and the Global Financial Crisis.

Forecasting of crude oil prices is a popular study area.Conventional statistical and machine learning approaches, such as VAR, error correction models and support vector machines, were used to solve this problem.Lanza et al. (2005) [5] investigated crude oil prices using ECMs. Consequently, due to the highly erratic behavior of the crude oil market, these strategies have frequently proven fruitless.. As per Weigend et al. (1994) [7], if just standard statistical and economic models are applied, the forecasts may perform poorly. Nonetheless, as a machine Progress in ml algorithms, deep learning, and adaptive ARIMA models are currently being utilized to forecast crude oil prices.prediction.

II.Related Work

1.Miroslava Zavadska, Lucía Morales demonstrates how variable oil prices may be amid major worldwide disasters such as 9/11 and the impact on the global economy. The study's findings have major significance for the time horizon that market players should take into account when constructing hedging strategies.

2.Hooman Abdollahi and Seyed Babak Ebrahim[2]proposed a hybrid model provided to estimate daily Brent crude oil prices. The research began with a data collection procedure in which the authors obtained daily time series of Brent oil prices from December 2009 to June 2017 and ended with pairwise comparisons of the forecast results. The aim of this study is to provide a hybrid model with the lowest forecast error and,more importantly, to introduce a hybridization that can better capture the main properties of oil price time series such as non-linearity, lag and market relationships.

3.Benjamin M and Tabak, Daniel [3] The efficiency of crude oil markets (Brent and West Texas Intermediate) is examined in this research by evaluating the fractal structure of these time series. They employ the Rescaled Range Hurst analysis to search for time-varying degrees of long-term reliance and uncover evidence that this market has been more efficient over time. When adjusting for short-term autocorrelation with a shuffling technique, these results are robust

III.Models Approached

1.ARIMA

The Arima model is a modification of the ARMA model. ARIMA is used because it predicts future patterns based on past data. The distinction is the presence of I, or Integrity, in the ARIMA model. I is the differencing factor, which converts non-stationary models to stationary ones.ARIMA models may be used to represent any non-seasonal time series that has patterns and is not random white noise.An ARIMA model is represented as ARIMA(p,d,q)p denotes the order of the AR phrase, d denotes the the order of the moving average and q denotes number of differencing necessary to make the time series

B.LSTM

LSTM stands for long term short memory , is a type of recurrent neural network in the area of artificial intelligence and deep learning.The RNN, particularly the Long Short Term Memory (LSTM), is a network that addresses the problem of vanishing and exploding gradients.The LSTM solves the problem of vanishing gradient descent.LSTM are specifically designed for to neglect the long term depepndency problem.

The four main categories of LSTM model are:

- Memory Cell
- Forget Gate
- Input Gate
- Output Gate

The RNN, particularly the Long Short Term Memory(LSTM), is a network that addresses the problem of vanishing and exploding gradients.The LSTM contains three types of gates: input gate, output gate, and forget gate.

1.*Forget Gate*:The forget gate deletes data from the lstm unit that is no longer useful. It has two inputs, x t and h t-1, which are fed into the gates and multiplied using weight matrices.The result is sent into an activation function, which produces a binary output.

2.*Input Gate*:First, the information is controlled via a sigmoid function, and the stored values are filtered through the inputs h_t1 and x_t, similar to the forgetting gate. The tan function then creates a vector that returns an output from 1 to +1 containing all possible values for h_t1 and xt. Finally, multiply the value of the vector by the adjusted value to get useful information

3.*Output Gate*:The output gate is in charge of collecting valuable information from the current cell state supplied as output. To begin, use the tanh function to construct a vector from the cell. A sigmoid function is then used to drive the information, which is subsequently filtered by the inputs h_t1 and x_t based on the stored values. Finally, the vector value is multiplied by the control value and delivered as actual output towards the next unit.

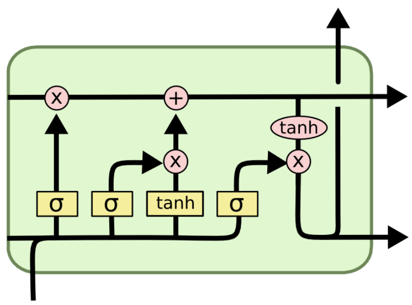

Fig 1 show the representation of LSTM

3.PROPHET

Prophet is a method for predicting time series data.Its a open source library developed by the facebook.Its implementation takes into account trends, seasonality, and vacations.Prophet detects the optimal spots in the learning algorithm for trend-changes to best suit the input while model fitting the trend is considered to be straight at any particular time, but the gradient of the trend might vary. As it could be observed in such samples, the more trend-changes discovered and the wider the gap between neighbouring elevations, the greater the ambiguity in forecasting.

V.Evaluation

a.DATASET

Crude oil price changes are influenced by a variety of factors. This information was obtained from the United States Energy Information Administration: Europe Brent Spot Price FOB (Dollars per Barrel).The goal of this collection and study is to forecast future Crude Oil Prices using historical data from the data-set.The data set includes daily Brent oil prices from May 17, 1987 to February 25, 2020.

The following data-set is available from the US Energy Information Administration:Europe Brent Spot Price FOB (Dollars per Barrel), which is updated weekly. The high level of competition in the Data Science field, as well as the availability of the new Prophet method, made it easier to predict future prices, which is what you might find when predicting oil prices with this data-set.

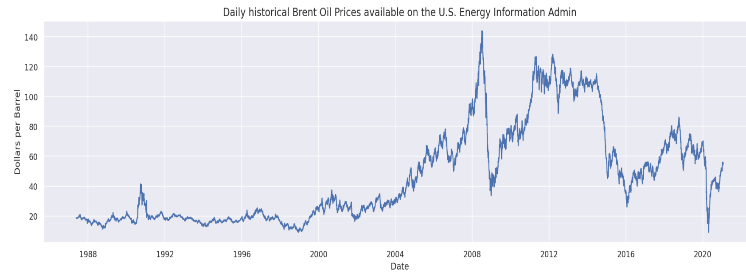

Fig 1.1show the information about the dataset

VI.Results

We attempted to divide the dataset into training and testing data using a machine learning algorithm. Moving prediction indicator analysis methods such as moving average and exponential moving average were also used.To test and confirm the presence of stationary time series, we used Augmented Dickey-Fuller to look for a unit root.In addition, ACF and PACF were introduced to determine the order and trend.ARIMA and the Prophet models were later added.

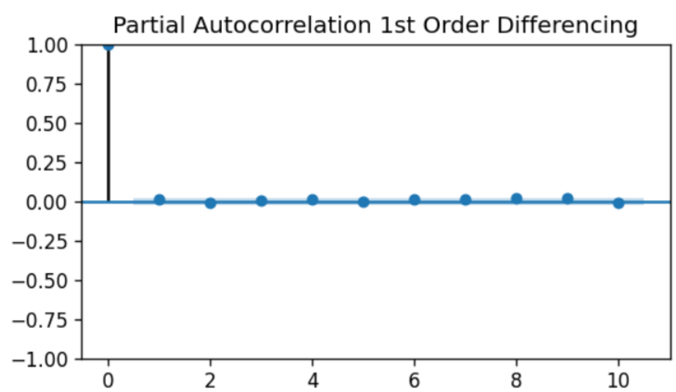

Fig 2 shows the analysis plot of PACF

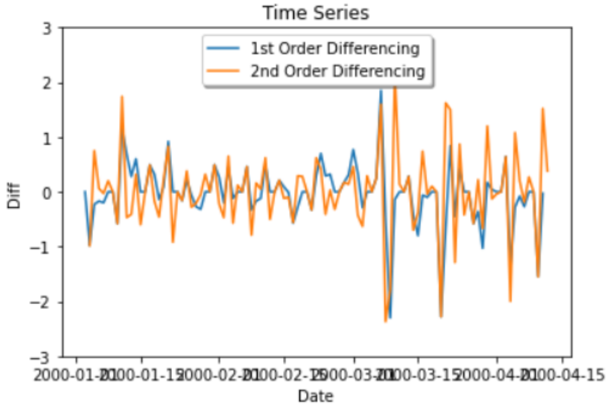

Fig 2.1 shows the result obtained from ARIMA model

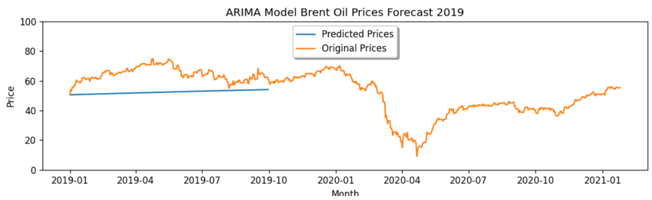

Fig 2.2 shows the Statistical plot of Prediction analysis

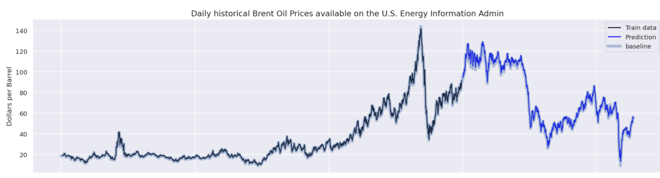

Fig 2.3 shows the Analysis plot of training data,prediction data and baseline

VII.Conclusion

We were able to broaden our knowledge of forecasting and analysis through this project.Our dataset ranges from 1980 to 2021, and has all the ups and downs that have happened due to various global conflicts like the Gulf Wars, The infamous 9/11, etc. All the price variance of the Crude from those periods have been stored in our data set and are taken for execution on the basis of the conventional machine learning split of 70 to 30 for the training to test ratio. We have put our Brent Oil Dataset through three different models, namely ARIMA, LSTM and Prophet. We investigated various possibilities in order to find the best model that can provide the most reliable and efficient model. We later performed root mean square error and mean square error calculations and mean average error to produce computational accuracy.We came to make use of LSTM model.We later used the ARIMA and Prophet models for forecasting.And arrived at the conclusion that the best model for prediction and analysis is the prophet model.There are also other models that are available that can outperform Prophet. In future a lot more better models can be brought in for smoother forecasting and prediction of crude oil prices.

VII.Reference

1.Miroslava Zavadska, Lucía Morales, Joseph Coughlan,Brent crude oil prices volatility during major crises,Finance Research Letters,Volume 32,2020,101078,ISSN1544-6123,https://doi.org/10.1016/j.frl.2018.12.026.

2.Hooman Abdollahi, Seyed Babak Ebrahimi,A new hybrid model for forecasting Brent crude oil price,Energy,Volume200,2020,117520,ISSN 0360-5442,https://doi.org/10.1016/j.energy.2020.117520

3.Benjamin M. Tabak, Daniel O. Cajueiro,Are the crude oil markets becoming weakly efficient over time? A test for time-varying long-range dependence in prices and volatility,Energy Economics,Volume 29, Issue 1,2007,Pages28-36,ISS0140-9883,https://doi.org/10.1016/j.eneco.2006.06.007.