WTI Crude Oil Price Time series Forecasting

A Thesis Submitted to the Faculty of

St. Joseph's College (Autonomous)

Bangalore University

By

Nancy Prakash Karadi

In partial fulfilment of the

requirements of the degree

of

Master of Science

in

Big Data Analytics

July 2022

This thesis paper is dedicated to Almighty God, my family, and the Project Guides.

Acknowledgment

I invested a lot of time and energy into this research. This research, however would not have been possible without the invaluable advice, assistance, and encouragement from my supervisors. I take this opportunity to express my sincere appreciation to each and every one of them. I want to start by expressing my gratitude to my mentor Mrs. Asha K Department of Big Data Analytics for always encouraging, directing, and supervising me as needed. She gave me the unwavering support and encouragement I needed most, beyond all else, she has been a continual source of ideas thanks to her genuine intuition, which has greatly enriched and inspired my development as a student. She will always have a place in my heart and soul.

To my mentor **Mr. Alok Pandey -C3A** at **TATA CONSULTANCY SERVICES** I would like to express my appreciation for his guidance, oversight, and significant contribution.

I am grateful for the assistance I received from all of the faculty at St. Joseph's College in Bangalore. I sincerely thank **Dr. Srinivas Bhogle**, a guest lecturer from St. Joseph's College in Bangalore, for his excellent presentation on a number of subjects that I will think about throughout the rest of my career. I am particularly appreciative of the assistance I received from my classmates, who also contributed their time and ideas during discussions about the progress of my internship. I want to express my gratitude to the Almighty God, my family, and everyone who has ever lived inside of me for their blessings, unwavering love, and support.

CERTIFICATE

ST. JOSEPH'S COLLEGE (AUTONOMOUS)

36, Lalbagh Road, Bangalore, 560027

DEPARTMENT OF BIG DATA ANALYTICS

CERTIFICATE

This is to certify that Master of Science thesis entitled **" WTI Crude Oil Price Time series Forecasting"** is a bon-a-fide work done by Nancy Prakash Karadi bearing Roll No: 20BDA03 in the 4th semester in St. Joseph's College during the year 2020-22 in the partial fulfilment of the requirement for the award of Master of Science in Big Data Analytics from St. Joseph's College (Autonomous).

Project Guide
Mrs. Asha K
Assistant Professor
Department of Big Data Analytics
St. Joseph's College (Autonomous)

Project Mentor
Mr. Alok Pandey
C3A at TCS
Bangalore

Declaration of the Candidate

I do hereby declare that this work entitled, "WTI Crude Oil Price Time Series Forecasting'' has been originally carried out by me under the guidance and supervision of Mr. Alok Pandey C3A at TCS Bangalore. This workhas not been submitted elsewhere for the award of any other degree or diploma certificate.

Bangalore

April 2022

Nancy Prakash Karadi

Reg. No.: - 20BDA03

Master of Science in Big DataAnalytics

Internship Offer Letter

Ref: TCSL/AIP 2021-22/Winter/DT20218671884

Date: 25-Jan-2022

Nancy Karadi
St. Joseph College (Autonomous), Bangalore
prakashnancy08@gmail.com

Dear **Nancy Karadi,**

Sub: Internship Offer

We are pleased to offer you internship in Tata Consultancy Services (TCS) with the following terms and conditions:

1. The tentative start date is 15-Feb-2022 and end date is 14-June-2022. These dates can be changed in discussion with the Project Guide
2. You will be assigned a Project Guide under whose supervision you will work on the project assigned to you.
3. You shall complete your project in accordance with the requirements and guidance of the TCS Project Guide, and maintain qualitative standards as required. You will maintain the discipline, dignity, honor and goodwill of TCS.
4. The arrangement is not that of an employer and an employee and as such you shall not be eligible to any allowances or other benefits as may be available to the employees of TCS.
5. You will observe the rules & regulations and discipline of TCS, and also maintain complete confidentiality and secrecy of the matters pertaining to TCS and/or any data that has been provided to you in the course of your project work. The detailed terms of Confidentiality, Data and Intellectual Property Protection are enclosed as Annexure A. You will be permitted to attend any classes in the college / university at the discretion of the Project Guide if so called for during the period of your project assignment.
6. On completion of your internship you will be required to submit a copy of your project report, which will be the sole property of TCS.
7. You shall not undertake any internship in parallel with this internship
8. In the event of any misconduct or breach of terms of this internship on the part of the Intern during the internship period, TCS reserves the right to terminate internship without any notice.
9. This offer of Internship will be governed as per the Laws of India.

You are required to sign and return a copy of this Internship Offer letter and the Annexure towards your acceptance of the terms and conditions stated therein.

For Tata Consultancy Services

Chandra Koduru
Head, Academic Interface Programme

Accepted,

Name of the Intern: Nancy Karadi

Date:

TATA CONSULTANCY SERVICES
Tata Consultancy Services Limited
Yantra Park, Opp. Voltas HRD Trg. Center, Subhash Nagar, Pokhran Road No. 2, Thane (West) 400 601 India
Tel +91 22 6778 2000/2222 Fax +91 40 6778 2190 website: www.tcs.com
Registered Office Nirmal Building, 9th Floor, Nariman Point, Mumbai 400 021

Table of Contents

Table Of Figures

1. About The Organisation

TATA consultancy Services

Tata Consultancy Services (TCS) is an Indian multinational information technology (IT) services and consulting company with its headquarters in Mumbai. It is a part of the Tata Group and operates in 149 locations across 46 countries.

Tata Consultancy Services (TCS) is an Indian multinational information technology (IT) services and consulting company with its headquarters in Mumbai. It is a part of the Tata Group and operates in 149 locations across 55 countries.

Tata Consultancy Services is an IT services, consulting, and business solutions organization that has been partnering with many of the world's largest businesses in their transformation journeys for over 50 years. TCS is a software and services provider in India. It is part of the Tata Group, which oversees operations for over 100 companies in seven business sectors: communications and information technology, engineering, materials, services, energy, consumer products and chemicals.

TCS offers a consulting-led, cognitive-powered, integrated portfolio of business, technology, and engineering services and solutions. This is delivered through its unique Location Independent Agile delivery model, recognized as a benchmark of excellence in software development.

A part of the Tata group, India's largest multinational business group, TCS has over 592,000 of the world's best-trained consultants in 55 countries. TCS' proactive stance on climate change and award-winning work with communities across the world have earned it a place in leading sustainability indices such as the MSCI Global Sustainability Index and the FTSE4Good Emerging Index.

2. About The Department

TCS' Analytics & Insights (A&I) service line helps organizations view, understand, and reimagine their businesses through an intelligent data-centric approach. They design innovative solutions for superior business outcomes and help customers execute effective data-driven strategies.

The A&I unit researches key emerging trends, including Artificial Intelligence, immersive analytics, mobility, cloud computing and social networking to develop innovative, practical, and powerful applications to deliver business results.

TCS A&I unit help organizations analyze their operational, business, and external data to gain insights, enabling them to be more agile and responsive to the market. This covers customer analytics, AML and fraud risk analytics, sales and marketing analytics, regulatory and

compliance analytics, supply chain analytics, social media analytics, and web analytics. We provide end-to-end services across the entire analytics value chain—data management, reporting, descriptive analytics, predictive modeling, and optimization.

3. Training Period

A minimum of 7–10 hours of work had to be done each week during the course, which ran for 30 days. The course is created using a self-paced online structure. And each module has a time requirement of 1-2 hours. Tata Consultancy Services (TCS), a preeminent provider of IT services, has introduced "TCS iON Career Edge." The company's website describes it as a free 15-day job preparation programmed with certain required courses in an effort to give today's children the fundamental employable skills they will need in the future.

4. Problem Statement

WTI Crude Oil Price Forecasting

Crude Oil Price Prediction oil prices dictating interest rates across regions

5. Project Details

Introduction

5.1 Background

Crude oil, one of the most significant commodities on the international market, is crucial to the worlds of economics, politics, and technology. Forecasting patterns in crude oil prices would be very valuable in gaining an understanding of the volatility of the energy resource market. This would assist the government and energy agencies in formulating policies and provide direction for economic growth. Therefore, making precise crude oil price predictions will be useful.

The majority of studies employ 2 typical forecasting techniques. One approach is from a qualitative perspective, while the other is from a quantitative perspective using tools like statistical and econometric models. And the majority of academics chose the latter.

However, non-linearity exists in crude oil market and it brings difficulty to predict the market movement. Neural network technique provides a favorable tool for nonlinear time series forecasting.

5.2 History of Crude Oil Usage

Crude oil is an organic substance made up of hydrocarbon deposits and other organic elements that is a naturally occurring, unprocessed petroleum product. Crude oil, a type of fossil fuel, can be refined to create useful products like gasoline, diesel, and different kinds of petrochemicals. It is a limited resource since it is nonrenewable, which means that it cannot be replenished naturally at the pace at which humans use it.

Typically, crude oil is collected through drilling, where it is frequently discovered alongside other resources like saline water and natural gas (which is lighter and so rests above the crude oil) (which is denser, and sinks below). After being refined and transformed into a range of products, including gasoline, kerosene, and asphalt, it is then offered for sale to end users.

Although fossil fuels such as coal have been mined in some form for centuries, crude oil was found and exploited during the Industrial Revolution, and its industrial applications were first established in the nineteenth century. Crude oil pricing and profitability are significantly influenced by supply and demand, much like any other sector. The top three oil producers in the world are the United States, Saudi Arabia, and Russia.

6. Purpose Statement

In this project, I will use the deep learning knowledge of recurrent neural network (e.g. LSTM, GRU, etc.) and feed forward neural network (dense layer), to perform time series forecasting for WTI crude oil prices. And from the prediction given by the models, to provide business insights for the crude oil industry.

Project Goal

- Crude oil, being one of the most important commodities on the global market, plays an important role in the worlds of economics, politics, and technology.
- Forecasting crude oil price movements would be extremely useful in gaining an understanding of the energy resource market's unpredictability.
- Benefits in not just the government and energy departments by facilitating policymaking, but it would also provide economic development advice.
- Result, being able to accurately estimate crude oil prices will be beneficial.

7. Methodology

Below are the two APIs that I have used to gather data

https://data.nasdaq.com/search?query=%20wti%20crude%20oil

https://fred.stlouisfed.org/searchresults?st=crude%20oil&t=crude&ob=sr&od=desc

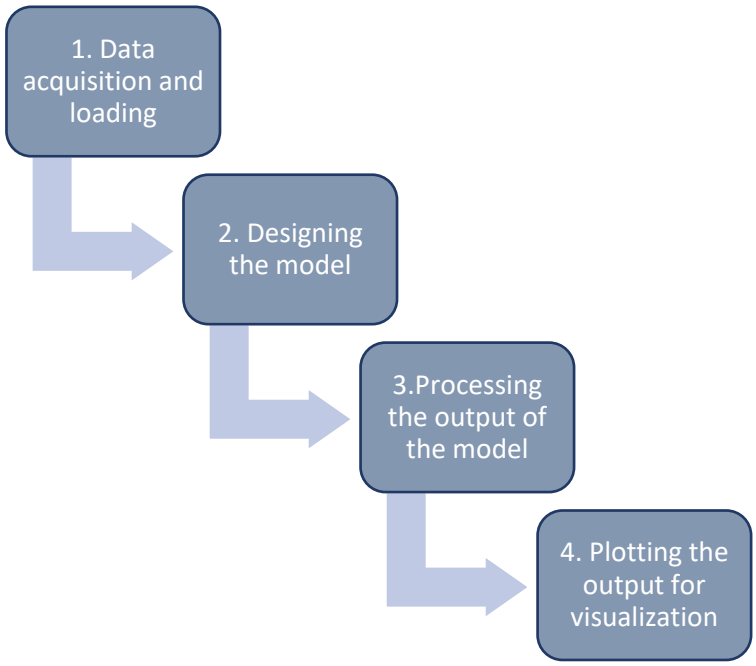


Figure 1 Project Structure

8. Data Description

Numerous influencing elements have an impact on how the price of crude oil changes. These components interact dynamically and intricately, which would explain the crude oil price movement's puzzling character. [6] I am considering the prices of relevant energy resources and oil-sensitive stocks in the analysis because I think that these two factors have a significant impact on how crude oil prices change.

I created my own data and collected data from several sources

1. **Federal Reserve Economic Data**
2. **Energy Sources in the United States**

8.1 Price of Energy Resource Background

1)Crude Oil:
Prices for crude oil I obtained the WTI (Western Texas Intermediate) crude oil price daily historical data for the current prices from a U.S. Energy Information Administration API (EIA). The 1980s through 2020 are covered by the dataset on oil prices.

2) Natural gas and propane:
In addition to crude oil, popular and essential energy sources include propane and natural gas.

Although propane is a byproduct of the processing of crude oil and natural gas, the majority of the propane generated comes from crude oil refineries. The price of crude oil has a greater impact on the price of propane than it does on natural gas since propane is in competition with other fuels based on crude oil.
According to EIA correlation research between crude oil and other commodities, there is a 0.25 correlation coefficient between crude oil and natural gas, which means that changes in the price of oil might cause a 25% shift in the price of natural gas (on average, throughout the study period).
This demonstrates the strong relationship between the price of crude oil and the price of natural gas.

In order to forecast the price of crude oil, I also looked for daily pricing data for propane and natural gas. These two energy sources' data, which span the years 1990 to 2020, are also obtained from the EIA.

3) Stocks sensitive to oil:
Stock prices reflect current market information because they are not subject to modification. The availability of stock price time series, along with its potential to be a useful predictor, can assist increase the precision of forecasts for the price of crude oil. Consequently, I also included oil-sensitive stock values in the predictors.

4) Stock prices of oil firms:
Stock prices of oil businesses are a key indicator of how well the oil industry is doing. As

indicators of the volatility of the oil market, I used the stock prices from Yahoo Finance for the following oil companies:

British Petroleum (BP) is a global oil and gas corporation. the sixth-largest oil and gas firm in the world.
ExxonMobil (XOM) is a Texas-based, multinational American oil company. one of the largest oil businesses in the world.
One of the major oil firms that dominates the world's petroleum market is Chevron (CVR). The company produced 791M net barrels of oil-equivalent per day on average in the United States in 2018.
In the stock prices datasets, I took the "adjclose" as "stock price" I used in predictors, since it's the closing price after adjustments for all applicable splits and dividend distributions. The stock price history is covering from 1997 to 2020.

5) Prices of Solar Company Stock:
The share values of solar companies have a strong relationship with the price of crude oil. This most likely reflects the opinion of the finance markets that the demand for solar electricity is mostly determined by the cost of oil (as a proxy for the price energy). In order to include its stock values in the final dataset, I chose NextExtra Energy (NEE), a U.S. solar firm, and its stock prices from 1997 to 2022.

8.2 Time Series Modification

Following the collection of all the data sources, I modified the dataset's time range and created a time series spanning from Jan. 8, 1997 to Nov. 3, 2022.

8.3 Final Dataset

- Date
- BP Stock Price
- XOM Stock Price
- CVR Stock Price
- NEE Stock Price
- Propane Price
- Natural Gas Price

Below a time-series plot is produced for each continuous feature in the data.

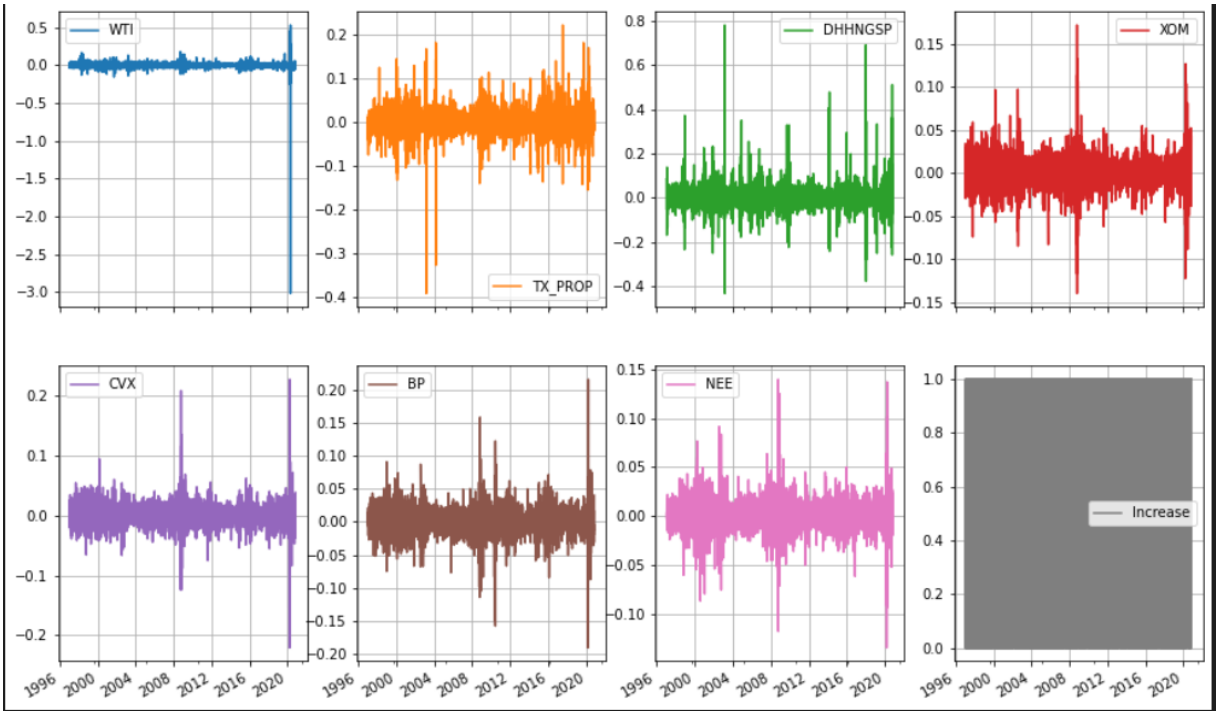


Figure 2 Combining the three datasets together assigning proper column name.

- After combining the data, it was saved as csv file in my google drive.
- The final dataset created has 5951 rows × 8 columns.

8.4 The target:

- Crude Oil Price
- Visualization of Data
- Correlation analysis as text variables is not present in correlation.

Figure 3 Confusion Matrix

9. Designing The Model

9.1 A Recurrent Neural Network

RNN involves sequential processing of the data for learning. This sequential process is justified by its ability to retain a memory of what came before the current sequence being processed. It is called recurrent because the output at each time step is utilized in the next time step as input. This is done by remembering the output of the previous time step. This in turn allows us to learn long-term dependencies in the training data. Like other neural networks, it contains an input layer, hidden layers and an output layer. In our proposed model we have one input Layer, three hidden layers and an output layer. Our model will take in the input sequence, process it through a hidden layer of 500 units, and produce a single valued output. The model is compiled with the Adam optimizer and a fully-connected dense layer with sigmoid activation. This adds additional representational capacity to the network. A dropout layer to prevent overfitting to the training data.

Basic Formula of RNN:

$$h^{(t)} = f\left(h^{(t-1)}, x^{(t)}; \Theta\right)$$

Basic Equations

$$a^{(t)} = b + Wh^{(t-1)} + Ux^{(t)}$$
$$h^{(t)} = \tanh(a^{(t)})$$
$$o^{(t)} = c + Vh^{(t)}$$

9.2 LSTM

LSTM is an extension of RNN. RNN's are networks with loops in them, allowing information to persist. The LSTM contains a forget gate that can be used to train individual neurons on what is important and how long it remains important. An ordinary LSTM unit consists of a block input z_t, an input gate i_t, a forget gate f_t, an output gate o_t, and a memory cell c_t . The forget gate f_t is used to remove information that is no longer useful in the cell state using Equation. The input at a given time x_t and the previous cell output h_{t-1} are fed to the gate and multiplied by weight matrices, followed by the addition of the bias. The result is passed through a sigmoid function that returns a number between 0 and 1. If the output is 0, the information is forgotten for a given cell state; if the output is 1, the information is retained for future use. Adding useful information to the cell state is performed by the input gate i_t using Equation. First, the information is controlled by the sigmoid function, which filters the values to be stored, similar to the forget gate. Then, a vector of new candidate values of h_{t-1} and x_t is generated with the block gate z_t using Equation, which outputs from -1 to $+1$. The vector values and the controlled values are multiplied to obtain useful information using Equation. The output gate o_t decides which information in the cell is used to calculate the output of the LSTM unit using Equation. In the model we build, we passed sigmoid function in the input gate with a dropout of 0.2 for regularization. All other layers other than output layer has dense 50. For compiling LSTM, ADAM optimizer is used. The loss is calculated using the mean square.

$f_t = \sigma_g(W_f * x_t + U_f * h_{t-1} + b_f)$ - - - - - -

$i_t = \sigma_g(W_i * x_t + U_i * h_{t-1} + b_i)$ - - - - - - -

$o_t = \sigma_g(W_o * x_t + U_o * h_{t-1} + b_o)$ - - - - - - -

$\acute{c}_t = \sigma_c(W_c * x_t + U_c * h_{t-1} + b_c)$ - - - - - - -

$c_t = f_t . c_{t-1} + i_t . \acute{c}_t$ - - - - - - - - -

$h_t = o_t . \sigma_c(c_t)$ - - - - - - - - - -

σ_g: *sigmoid*

σ_c: *tanh*

f_t *is the forget gate*

i_t *is the input gate*

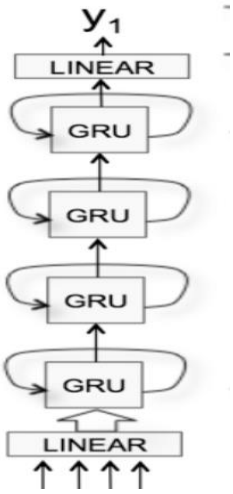
o_t *is the output gate*

c_t *is the cell state*

h_t *is the hidden state*

9.3 Gradient Recurrent Unit (GRU)

GRU belongs to the family of Recurrent Neural Networks. A Gradient Recurrent Unit (GRU) is one of the proposed models and we have one input layer and three hidden layers and an output layer. Our model will take in the input sequence through 500 units, processed through a hidden layer of 500 units with dropout being 0.2, finally single valued output is produced. The model is compiled with the Adam optimizer and loss function being the mean squared error.



10.Processing The Output of The Model

In the proposed model I have One input Layer Three hidden layers and a output layer.

- model will take in the input sequence, process it through a hidden layer of 500 units, and produce a single valued output

- The model is compiled with the Adam optimizer A fully-connected Dense layer with Sigmoid activation. This adds additional representational capacity to the network.

- A Dropout layer to prevent overfitting to the training data.
  - I'am using the Keras Sequential API which means I built the network up one layer at a time.
  - I Trained The model for 500 epoch.

11.Shrinkage Of Crude Oil

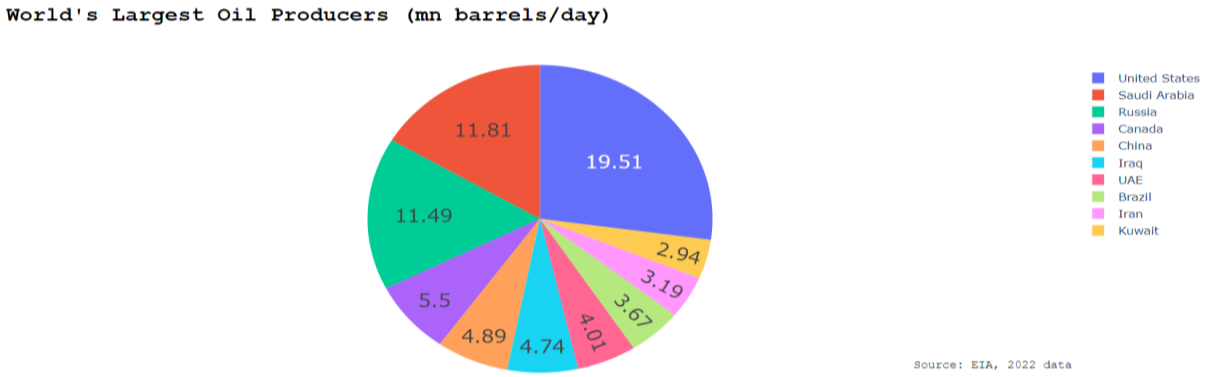


Figure 4 World's largest oil producers

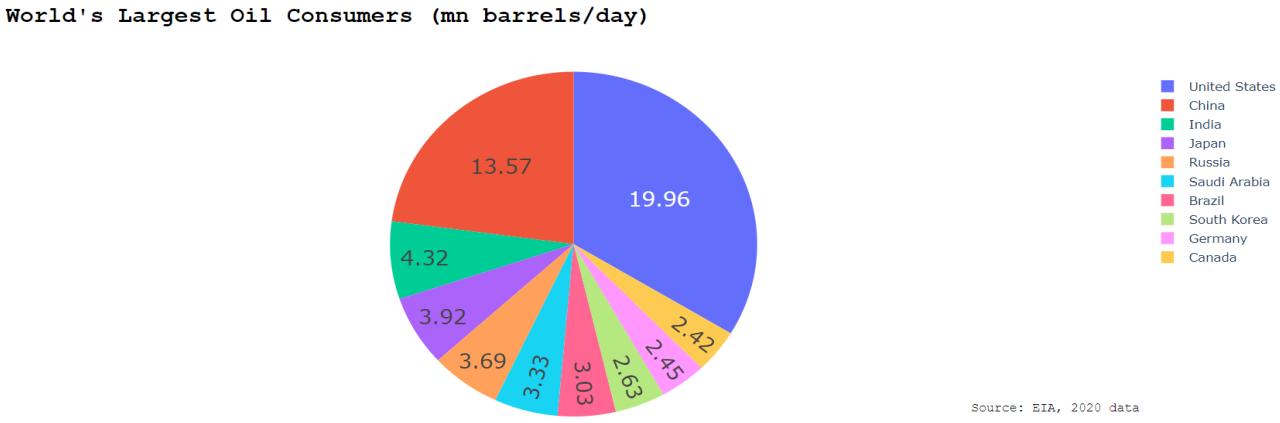


Figure 5 World's largest oil customers

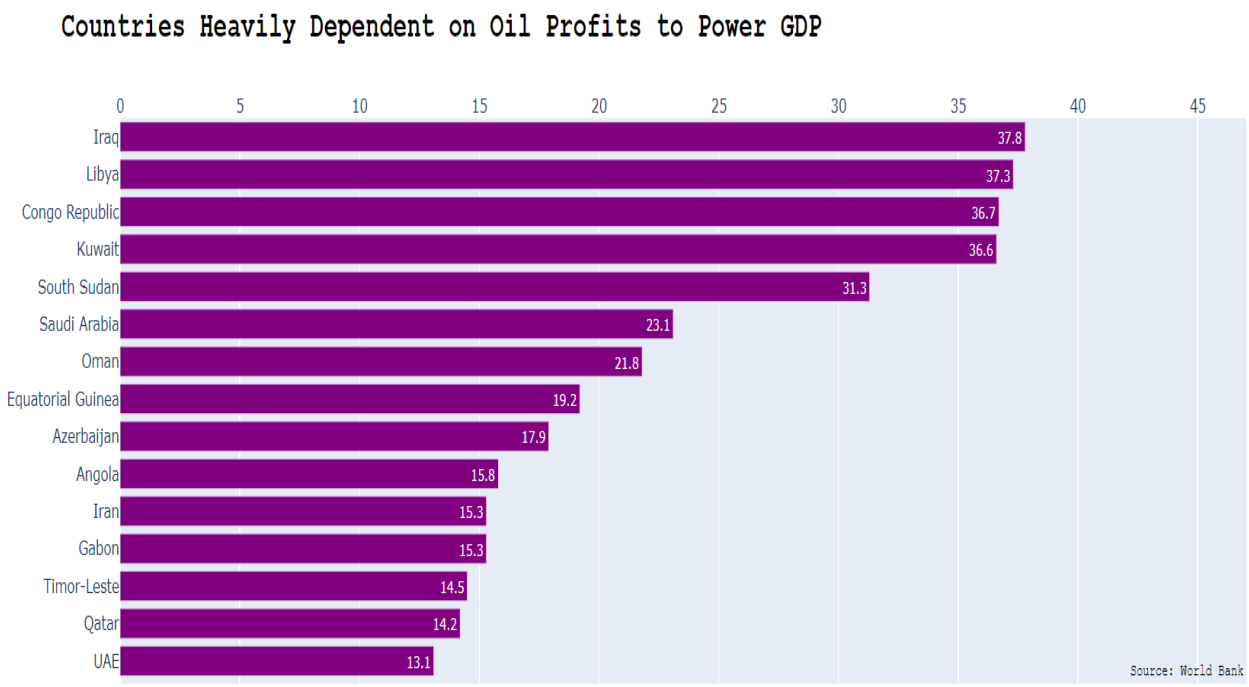


Figure 6 Countries Heavily Dependent on oil profit

12.Result

Model Number(details) Test Accuracy

1. SimpleRNN (Model 1) 98%
2. One Layer LSTM(Model 2) 99%
3. LSTM with drop out layers (Model 3) 99%
4. Mix bag monstor model (Model 4) 95%

The commercially advisable design is model 3 that is LSTM with drop out layers .If I was effective in predicting whether the price of crude oil would be rising or falling, I would use this. Countries that purchase crude oil can put in place economic safeguards against future inflations caused by oil prices. These are only two of the numerous advantages that this approach may offer to businesses, the government, and people

13.Model Summary

- I started with collecting energy industry data to predict change in oil price. After collectin g data from different sources, I combined the data did some cleanup and saved a backup c opy for cleaned data on shared drive.
- Next, I performed data preprocessing by getting rid on NAs and did percentage change o n each column and creating a new column that marks gain or loss. I then shifted newly cr eated column by one so that each row tends to predict increase of next day.
- I then partitioned the data and created look back period of 10 so that a multi variate times eries model can be trained. Now, ready to perform modeling.
- First model was the baseline model with just one simpleRNN layer, the model was decent and gave an accuracy of 98% on test data.
- Second model that I trained was one-
  layer LSTM which performed more or less the same as single layer simpleRNN with acc uracy of 99% on test data.
- Third model is little more complex model it has multiple LSTM layers with slow learning via dropout layers this gave even better results with test accuracy of 99%.
- Fourth model is a mix bag that has convolutional, LSTM and GRU layer, this model perf ormed the worst of all with test accuracy of 95%.

14.References

[1] Li, Xuerong, Shang, Wei, & Wang, Shouyang. (2019). Text-based crude oil price forecasting: A deep learning approach. International Journal of Forecasting, 35(4), 1548–1560.

https://doi.org/10.1016/j.ijforecast.2018.07.006

[2] Benjamin Hunt, Peter Isard, & Douglas Laxton. (2002). THE MACROECONOMIC EFFECTS OF HIGHER OIL PRICES. National Institute Economic Review, 179(179), 87–103. https://doi.org/10.1177/002795010217900111

[3] Y. Fan, K. Wang, Y. J. Zhang et al., "International crude oil market analysis and price forecast in 2009," Bulletin of Chinese Academy of Sciences, vol. 4, no. 1, pp. 42–45, 2009.

[4] Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. Energy Economics, 23(3), 325–338. https://doi.org/10.1016/s0140-9883(00)00075-x

[5] S. Mirmirani and H. Cheng Li, "A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil," Advances in Econometrics, vol. 19, pp. 203–223, 2004.

[6] Chen, Yanhui, He, Kaijian, & Tso, Geoffrey K.F. (2017). Forecasting Crude Oil Prices: a Deep Learning based Model. Procedia Computer Science, 122, 300–307. https://doi.org/10.1016/j.procs.2017.11.373

[7] Factors That Affect Propane Prices

https://lingas.com/hot-topics/factors-that-affect-propane-prices/#:~:text=Propane%20competes%20with%20other%20crude,than%20that%20of%20natural%20gas.&text=You%20remember%20learning%20the%20supply,t%20changed%20over%20the%20years.

[8] U.S. Energy Information Administration. “What drives crude oil prices?” Page 21. Accessed June 3, 2020.

https://www.eia.gov/finance/markets/crudeoil/reports_presentations/crude.pdf

[9] Chen, S. (2014). FORECASTING CRUDE OIL PRICE MOVEMENTS WITH OIL-SENSITIVE STOCKS. Economic Inquiry, 52(2), 830–844. https://doi.org/10.1111/ecin.12053

[10] Impact of Oil Price on Solar Industry

http://www.greenrhinoenergy.com/solar/market/solar_oil.php#:~:text=Conclusion,proxy%20for%20price%20of%20energy).