Project Report

Date	**01 November 2022**
Team ID	**PNT2022TMID41179**
Project name	**Crude Oil Price Prediction**
Team member	**Sathish.M** **Sathishkumar.R** **Soodappan.K** **Vijaysankar.G**

CRUDE OIL PRICE PREDICTION

1.INTRODUCTION

Since crude oil is one of the main items sold worldwide and hence incorporates global measurements, it plays a crucial role in the world. Errors in crude oil forecasting have a composite supply-demand structure at their root. Crude oil volatility has a significant impact on economic variables such as the country's economic growth, unemployment, and exchange rate, which also depends on the export and import of crude oil. Machine learning techniques can now be used in a variety of applications. Powerful computational tools and algorithms that can learn for themselves and anticipate data with a lengthy short-term memory are provided by machine learning. Recurrent neural networks built on LSTM are used in this paper to predict the price of crude oil. The most effective and powerful models for processing time-series-based sequential data are recurrent neural networks (RNNs). In addition to prediction, LSTM variations can be applied to tasks including voice, handwriting, and polyphonic modelling. Through the use of random search and variance framework analysis, the hyperparameters of the variations were evaluated. ANNs are made up of a group of nodes that resemble the neurons in a biological brain.

1.1Project overview

Oil demandis inelastic,therefore therise inprice isgood newsfor producers because they will see an increase in their revenue. Oil importers,however, will experience increased costs of purchasing oil. Because oil is the largest traded commodity, the effects are quite significant. A rising oilprice caneven shifteconomic/political power from oil importers to oilexporters. Thecrude oilprice movementsare subjectto diverseinfluencingfactors.

This GuidedProject mainlyfocuses onapplying NeuralNetworks topredict the Crude Oil Price. This decision helps us to buy crude oil at the propertime. Time series analysis is the best option for this kind of prediction because we are using the Previous history of crude oil prices to predictfuture crudeoil. Sowe wouldbe implementingRNN(Recurrent NeuralNetwork) withLSTM(Long ShortTerm Memory)to achievethe task.

Crudeoilpricefluctuationshaveafarreachingimpactonglobaleconomiesand thus price forecasting can assist in minimising the risks associated with volatility in oil prices. Price forecasts are very important to variousstakeholders: governments, public and private enterprises, policymakers, andinvestors.

1.2Purpose

Crudeoil is amongst the most important resources intoday's world, it is the chief fuel and its cost has a direct effect on the global habitat,our economy and oil exploration, exploitation and other activities. Prediction of oil priceshas become the need of the hour, it is a boonto many large and smallindustries, individuals, the government.The evaporative nature of crudeoil, its price prediction becomes extremely difficult and it is hard to be precise with the same. Several different factorsthat affect crude oil prices. We propose a contemporary and innovativemethod of predicting crude oil prices using the artificial neural network(ANN).

The main advantage of this approach of ANN is that it continuously captures the unstable pattern of the crudeoil prices which have been in corporated by finding out the optimal lag and number of the delay effect that controls the prices of crudeoil.

Variation of lag in a period of time has been done for the most optimumand close results, we then have validated our results by evaluating the root means quareerror and the results obtained using the proposed model have significantly out performed.

2. LITERATURE SURVEY

We divide crude oil price forecasting approaches into three categories: (1) heuristic approaches; (2) econometric models; and (3) machine learning techniques.

Heuristic approaches for oil price prediction include professional and survey forecasts, which are mainly based on professional knowledge, judgments, opinion and intuition. Another heuristic approach, the so-called no-change forecast, uses the current price of oil as the best prediction of future oil prices. Despite its simplicity, the no-change forecast appeared to be a good baseline approach for oil price prediction and was better than other heuristic judgmental approaches (Alquist et al., 2013).

Econometric models are the most widely used approaches for oil price prediction, which include autoregressive moving average (ARMA) models and vector autoregressive (VAR) models, with possibly different input variables (Pindyck, 1999, Frey et al., 2009). These econometric models provide more accurate prediction than the no-change model at least at some horizons (Alquist et al., 2013, Baumeister and Kilian, 2015). Recently, a forecast combination approach was proposed by Baumeister and Kilian (2015), which combines 6 different oil price prediction models including both econometric models (such as the VAR model) and the no-change model. It should be noted that most of the econometric models are linear models and are not be able to capture the nonlinearity of oil prices.

Several machine learning techniques were proposed for oil price prediction, such as artificial neural networks (ANN) (Yu et al., 2008, Kulkarni and Haidar, 2009), and support vector machine (SVM) (Xie et al., 2006). These are nonlinear models which may produce more accurate predictions if the oil price data are strongly nonlinear (Behmiri and Pires Manso, 2013). However, these machine learning techniques, like other traditional machine learning techniques, rely on a fixed set of training data to train a machine learning model and then apply the model to a set.

2.1 Existing problem

Crude oilis oneof themost importantcommodities inthe world,accounting for one-third of global energy consumption. It is a starting material for most of the products that we use in everyday life, ranging fromtransportation fuels to plastics. Crude oil price fluctuations have a far reachingimpact onglobal economiesand thusprice forecastingcan assistin minimisingthe risksassociated withvolatility inoil prices.

Price forecasts are very important to various stakeholders: governments, public and private enterprises, policymakers, and investors. According toeconomic theory, the price of crude oil should be easily predictable from the equilibrium between demand and supply, wherein demand forecasts areusually made from GDP, exchangerates anddomestic prices,and supplyis predictedfrom pastproduction dataand reservedata.

Predicting demand for oil is usually straightforward, however supply is heavilyaffected bypolitical activitysuch ascartelisation byOPEC toregulate prices, technological advances leading to the extraction of higher amountsof oil,and warsand otherconflicts whichcan affectsupplyunpredictably.

Models incorporating economic parameters such as supply and demandand their determinants are known as structural models (see Equation 1). Even though structural models are found to be the most logical ways ofmodelling theprices ofindustrial products,the priceof crudeoil isaffected by many other factors. One of these factors is that the price of crude oil isdetermined inthe futuresmarket whichenables thepurchase ofa predefinedamount ofoil ata particularprice inthe future.

2.2 Reference

1.KaufmannR.K.,UllmanB.

Oilprices, speculation, and fundamentals: Interpreting causal relations among spot and futures prices Energy Economics,31(4)(2009),pp.550-558.

2.ShobhitNigam

Chapter84 Single Multiplicative Neuron Modelin Reinforcement Learning, Springer Science and Business Media LLC(2019).

3."Harmony Search and Nature Inspired Optimization Algorithms", Springer Science and Business Media LLC,2019.

4.ShuangGao,YalinLei.

"Anew approach for crude oil price prediction based on stream learning", Geoscience Frontiers,2017.

5.RamakantaMohanty.

"Software Reliability Prediction Using Group Method of Data Handling", Lecture Notes in ComputerScience,2009.

6.KulkarniS.,HaidarI.

Forecasting model for crude oil price using artificial neural networks andcommodity future prices.International Journal of Computer Science andInformationSecurity,2(1)(2009).

7.HaykinS.

Neural Networks: A Comprehensive Foundation (2nd edition), Prentice Hall(1999),p.842.

8.OnurDursun

3Methodology,WalterdeGruyterGmbH(2014)

9.Lakshmanan, Indhurani, and Subburaj Ramasamy.
"An Artificial Neural-Network Approach to Software Reliability Growth Modeling", Procedia ComputerScience,2015.

10.HaykinS.
Neural Networks and Learning Machines (3rd edition), Pearson (2009), p.938.

11.N.RajKiran,V.Ravi.
"Software reliability prediction by soft computing techniques",
Journal of Systems and Software,2008.

12.LeanYu.
“AnEMD-Based Neural Network Ensemble Learning Model for World Crude Oil Spot Price Forecasting”, Studiesin Fuzziness and Soft Computing,2008.

2.3 Problem Statement Definition

Crude oil price prediction is a wide area of research that has been on for a very long time in history and numerous approaches have been proposed inpredicting crude oil price. Research works on crude oil price prediction approachescan bebroadly categorizedin three:statistical andeconometric models,AI andhybrid modelingtechniques.

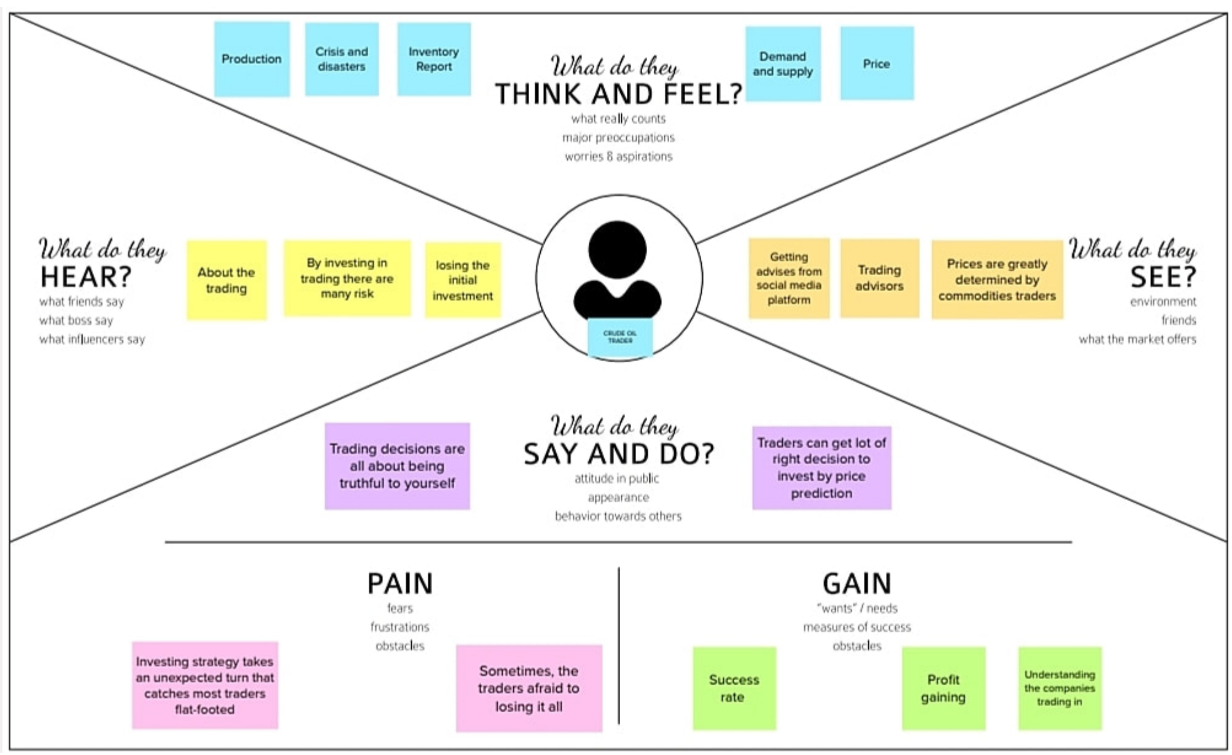
Oil demand is insensitive to changes in price or income., therefore the rise in price is good news for producers because they will see an increase intheir revenue. However some people will see increase in the prices like Oil importers. Because oil is the largest traded commodity, the effects arequite significant.A risingoil price can even shift economic or politicalpower from oil importers to oil exporters. The crude oil price changes are theinfluencing factors.

This Project mainly focuses on applying Neural Networks to predict the Crude Oil Price. This decision helps us to buy crude oil at the proper time.Time series analysis is the best option for this kind of prediction because we are using the Previous history of crude oil prices to predict future crude oil. So we would be implementing RNN(Recurrent Neural Network) withLSTM(Long ShortTerm Memory)to achievethe task.

3.IDEATION & PROPOSED SOLUTION

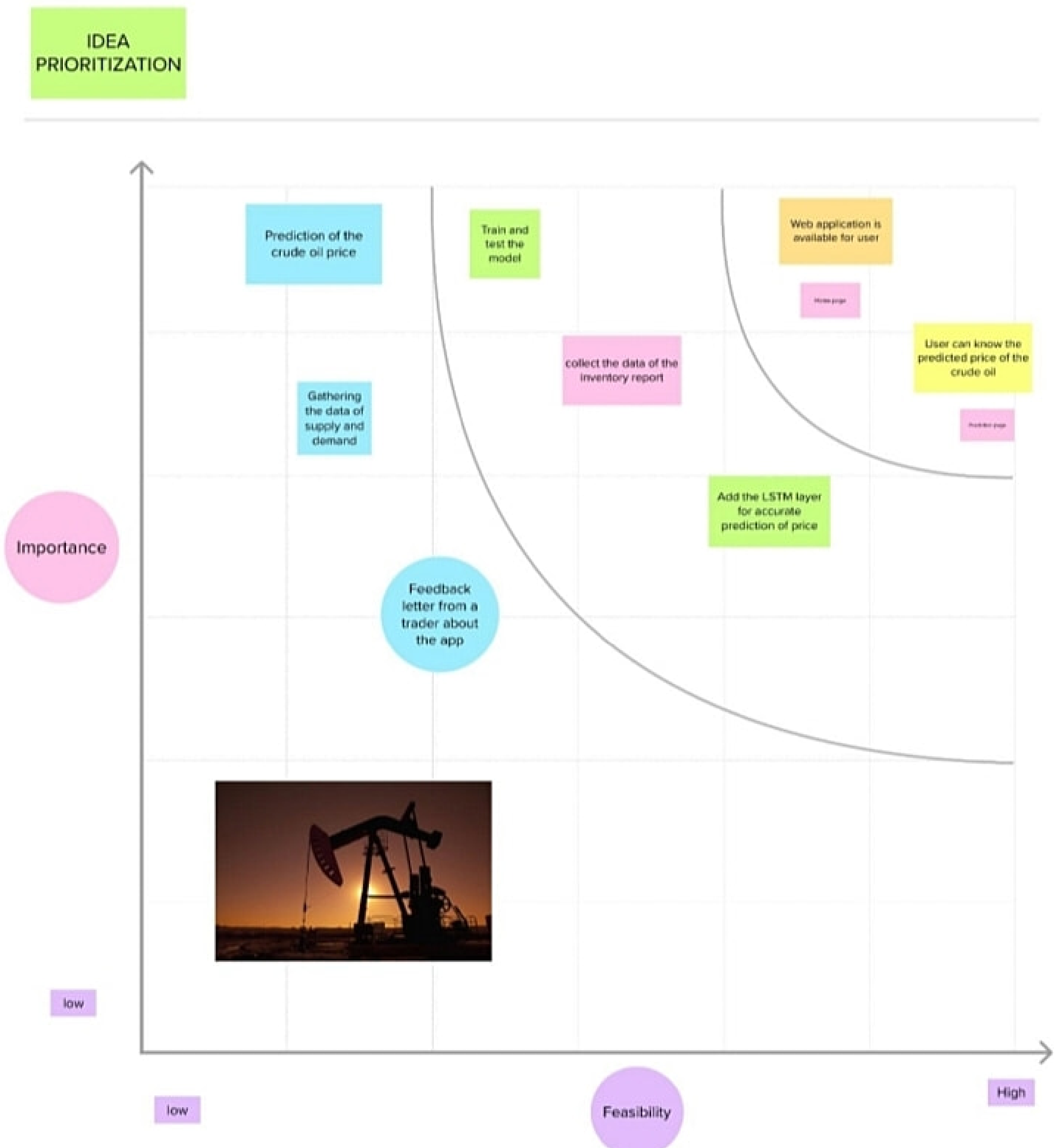
3.1Empathy Map Canvas

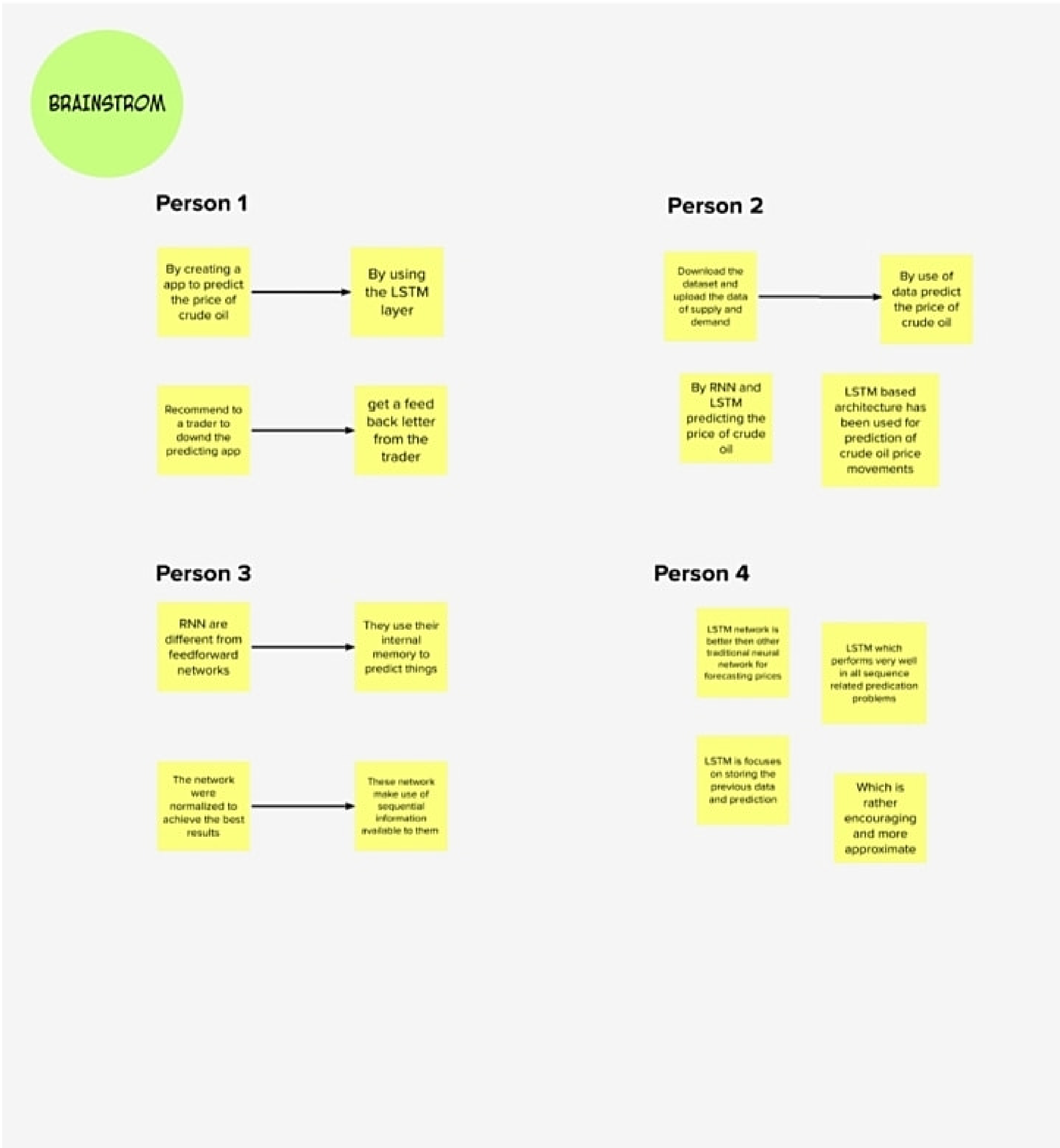
Anempathymapisasimple,easy-to-digestvisual thatcaptures knowledgeabout a user's behaviours and attitudes. It is a useful tool to helps teams betterunderstand theirusers. Creatingan effectivesolutionrequiresunderstanding the true problemand the person who is experiencing it. The exercise of creating the map helps participants consider things from theuser's perspectivealong withhisor hergoals andchallenges.



3.2 Ideation & Brainstorming

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process thatleads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged tocollaborate, helping each other develop arich amountof creative solutions.





3.3 Proposed Solution

oil demand is inelastic, therefore the rise in price is good news for producers because they will see an increase in their revenue. Oil importers, however, will experience increased costs of purchasing oil. Because oil is the largest traded commodity, the effects are quite significant. A rising oil price can even shift economic/political power from oil importers to oil exporters. The crude oil price movements are subject to diverse influencing factors. This Guided Project mainly focuses on applying Neural Networks to predict the Crude Oil Price. This decision helps us to buy crude oil at the proper time. Time series analysis is the best option for this kind of prediction because we are using the Previous history of crude oil prices to predict future crude oil. So, we would be implementing RNN (Recurrent Neural Network) with LSTM (Long Short-Term Memory) to achieve the task.

SOLUTION DESCRIPTION:

Recurrent Neural Networks:

RNN are different from feedforward networks. They use their internal memory to predict things. They are very good at tasks at which humans are not good at such as handwriting recognition and speech recognition. They were initially developed in 1980. These networks make use of sequential information available to them. Traditionally, we assumed that inputs do not depend on each other. But that was not a valid assumption. As if we want to predict the next words in a sentence, we must know the previous words. They can be thought of having a memory which stores the information for future use. There exist various extensions of RNN. One of them is the Bidirectional RNN. In these networks, output at time t may depend on future inputs as well. The other popular variant of the RNN is the deep RNN. In these recurrent networks, there exist multiple layers per time step.

LSTM Networks:

The most popular and widely used type of RNN is the LSTM and these types of recurrent networks have been used for this study. These networks learn order dependence in sequence prediction problem. The LSTM networks are able to solve two major issues encountered in RNN i.e., vanishing gradients and exploding gradients. The key to the solution of these problems were the internal structure that has been used in LSTM. In this, there exists one input layer, one hidden layer and one output layer. This most simple architecture of LSTM networks is known as vanilla LSTM which performs very well in all sequence related prediction problems.

Data retrieval and pre-processing:

In data retrieval, datasets can be fetched such of news data, black gold price data and market data. Dataset from news can be retrieved through headlines as it is easier to obtain and justifies in one line. Factors that affect the prediction are export business, stock market and later business.

Methodology:

LSTM based architecture has been used for prediction of crude oil price movements. The proposed architecture consists of four layers of LSTM layers followed by a dense layer with ten neurons and at the end dense layer with only one neuron. All the inputs to the proposed network were normalized to achieve the best results.

Libraries Required:

Make sure that the following libraries are installed on your working machine before proceeding further

- Keras
- Tensorflow
- Numpy
- Pandas

NOVELTY:

In the era of big data, deep learning for predicting crude oil price has become even more popular than before. We collected 2 years of data from world global data and proposed a comprehensive customization of feature engineering and deep learning-based model for predicting price of crude oil. The proposed solution is comprehensive as it includes pre-processing of the crude oil dataset, utilization of multiple feature engineering techniques, combined with a customized deep learning-based system for crude price prediction. We conducted comprehensive evaluations on frequently used machine learning models and conclude that our proposed solution outperforms due to the comprehensive feature engineering that we built. The system achieves overall high accuracy for crude oil price prediction. With the detailed design and evaluation of prediction term lengths, feature engineering, and data pre-processing methods, this work contributes to the stock analysis research community both in the financial and technical domains.

SOCIAL IMPACT:

Crude oil price prediction has long been the subject of research because of the importance of accuracy of prediction and the difficulty in forecasting. Traditionally, forecasting has involved linear models such as LSTM and RNN using standardized numerical data such as corporate financial data and crude oil price data. However, we know little about which characteristics of crude oil price affect the accuracy of predictions and to what extent. The purpose is to analyse the effects of crude oil price characteristics on crude oil price prediction via RNNs. To this end, we define the characteristics of crude oil price and identify significant differences in prediction the implications of making predictions only with which are unstructured data, without using large amounts of standardized data. Finally, we identify issues for future research.

BUSINESS MODEL:

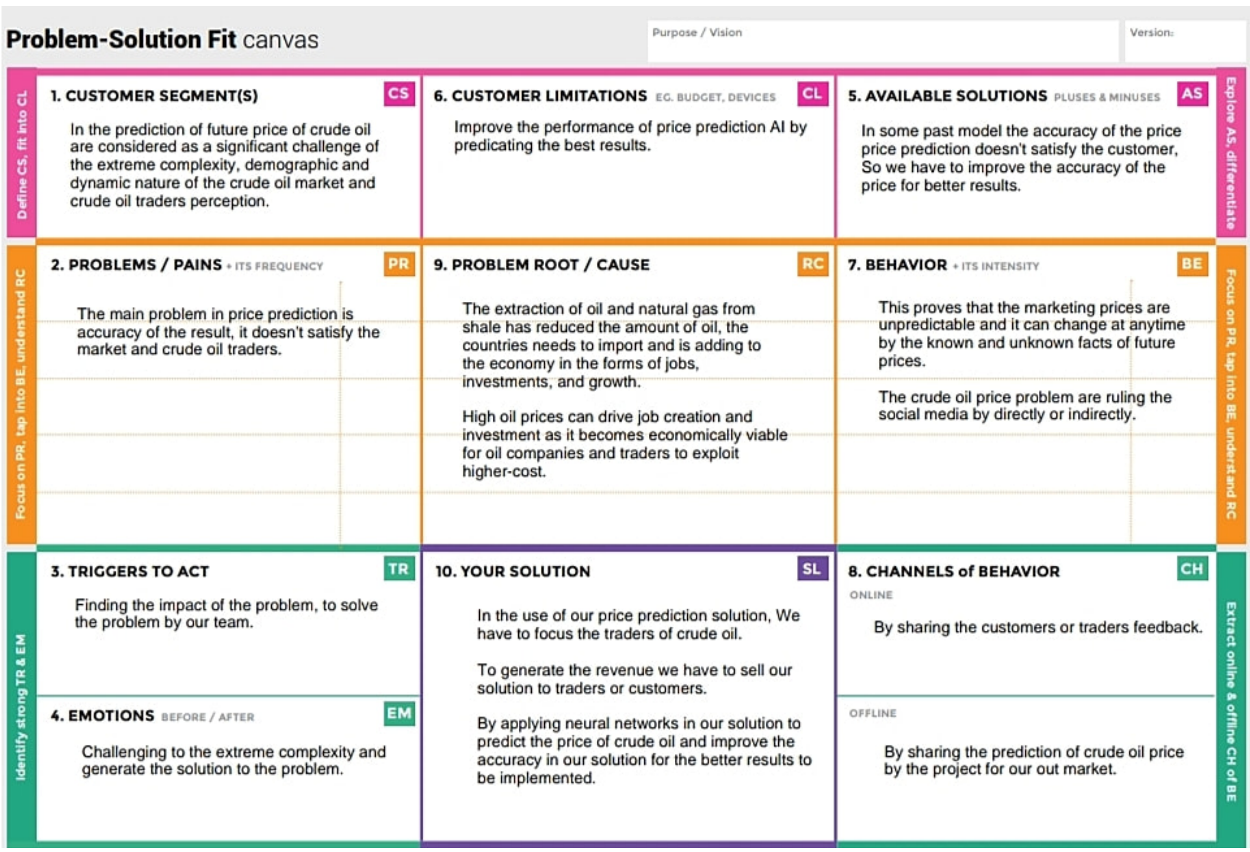
Crude oil price prediction is of course based on data, but when using AI, you are delving into the world of big data. That means that you have more data and more detail to your data. You are then able to take into account each customer's specific individual behaviour and therefore have a more precise price prediction. Learning from mistakes is one of the most valuable things you can do as a human. The thing is that, when it comes to predicting, AI is much better at learning and adjusting than we are. This is not only due to the speed with which a computer can understand and re-evaluate data but also because it is unbiased. Model management (which is minimizing the gap between reality and predicating) is a key element to a good AI-powered crude oil price prediction. When we analyse the prediction and reality we might compare and adjust a tenfold of combinations. With AI the number of combinations that can be made between result data and predicting data. It is therefore concluded that AI creates more and better improvements than we could do manually.

SCALABILITY OF SOLUTION:

Crude oil price forecasting plays a significant role in world economy and its accurate prediction has significant benefits for the economic conditions of a country. In this direction, an effort has been in this paper. This paper has proposed an LSTM based network for better prediction of crude oil prices. The results obtained from the work are quite encouraging. The results indicate that large lookups do not necessarily improve the accuracy of the predictions of crude oil prices. It has been found that lookups up to the value of 10 are ideal for crude oil price prediction purposes. It has also been found that just increasing the number of LSTM layers do not have much impact on the accuracy of the results. Here it can be 90% accurate in price prediction. In future work, current market and political conditions can also be taken into consideration in crude oil price forecasting for even better results.

3.4 Problem Solution fit

The Problem-SolutionFit simply means that you have found a problem with your customer and that the solution you have realized for it actually solves the customer's problem.It helps entrepreneurs, marketers and corporate innovators identify behaviora lpatterns and recognize what would work.



4. REQUIREMENT ANALYSIS

4.1 Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Registration	Registration through Form Registration through Gmail
FR-2	User Confirmation	Confirmation via Email Confirmation via OTP
FR-3	Graph	Obtaining the data from the dataset to show the graph
FR-4	Customer Support	Providing answers for the queries asked by users.
FR-5	Database	User's information will be stored
FR-6	Price information	Information of the oil prices will be updated by admin
FR-7	Notification	Price alert are sends to user by notification

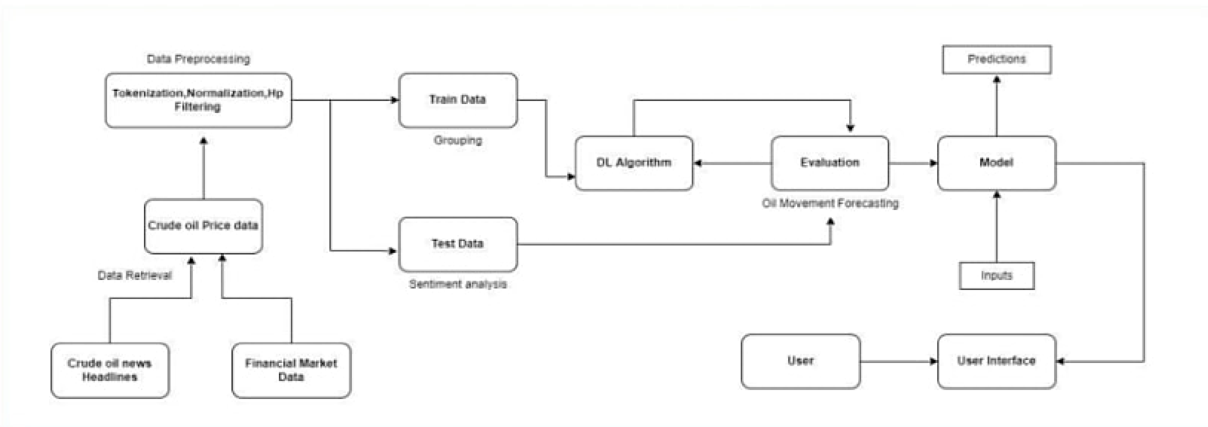
4.2 Non Functional requirement

FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	It can use by wide variety of user, it is very simple to learn and not complex to proceed.
NFR-2	**Security**	In this model we are using login for the user and the information will be secured, so that it will be very secure to use
NFR-3	**Reliability**	This model will be reliable, that it can update with very time period, so that the accuracy will be good.
NFR-4	**Performance**	This model will be performed fast and secure even at the lower bandwidth.
NFR-5	**Availability**	Prediction will be available for every user not only for premium user. News, database and price alert will be alert to ever user.
NFR-6	**Scalability**	It is scalable, that we are going to use data in kb, so that it's quite amount of storage is satisfied to user.

5. PROJECT DESIGN

5.1 Data Flow Diagrams

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict therightamount of the system requirement graphically. It shows how data entersandleaves thesystem, whatchanges theinformation, andwheredata isstored.



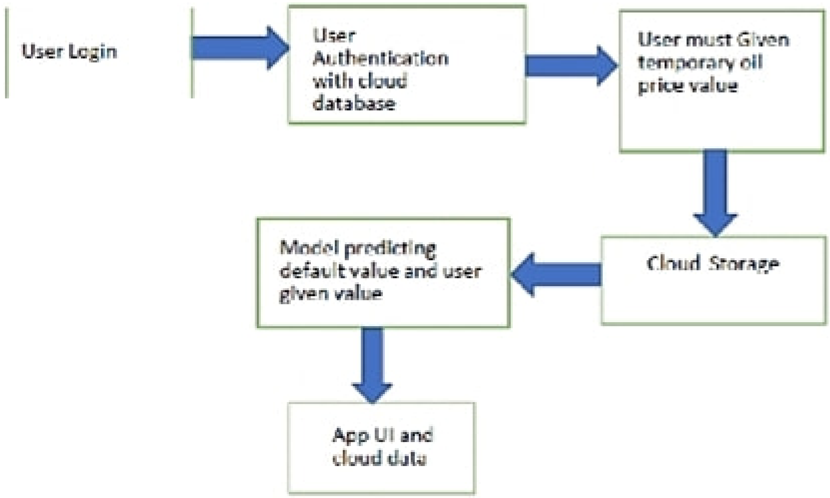
5.2 Solution & Technical Architecture

Oil demand is inelastic, therefore the rise in price is good news for producers because they will see an increase in their revenue.

Oil importers, however, will experience increased costs of purchasing oil.

Because oil is the largest traded commodity, the effects are quite significant.

A rising oil price can even shift economic/political power from oil importers to oil exporters.



TECHNICAL ARCHITECTURE

5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	I can access my account / Displays the line graph / Bar graph	High	Sprint-1
		USN-2	As a user, I will receive confirmation email once I have registered for the application	I can receive confirmation email & click confirm	High	Sprint-1
		USN-3	As a user, I can register for the application through Facebook	I can register & access the account	Low	Sprint-2
		USN-4	As a user, I can register for the application through Gmail	I can register using Gmail account.	Medium	Sprint-1
	Login	USN-5	As a user, I can log into the application by entering email & password	After registration, I can log in by using email & password.	High	Sprint-1
	Line/Bar Graph	USN-6	After entering the inputs, the model will display predictions in Line or Bar Graph Format.	I can get the prediction in various format.	High	Sprint-3
Customer (Web user)	Login	USN-7	As the web user, I can login simply by using Gmail.	Gmail can be used for login	Medium	Sprint-2
Customer Care Executive	Support		The Customer care service will provide solutions for any FAQ and also provide Chatbot.	Solve the problem by support.	Low	Sprint-3
Administrator	News		Admin will Provide the recent news of Oil Prices.	Provide the recent oil prices.	High	Sprint-4
	Notification		Admin will notify when the oil prices changes.	Notification by Gmail or message.	High	Sprint-4
	Access Control		Admin can control the access of users.	Access permission for Users.	High	Sprint-4
	Database		Admin can store the details of users in database.	Stores User details.	High	Sprint-4

6.PROJECT PLANNING & SCHEDULING

6.1 Sprint planning & Estimation

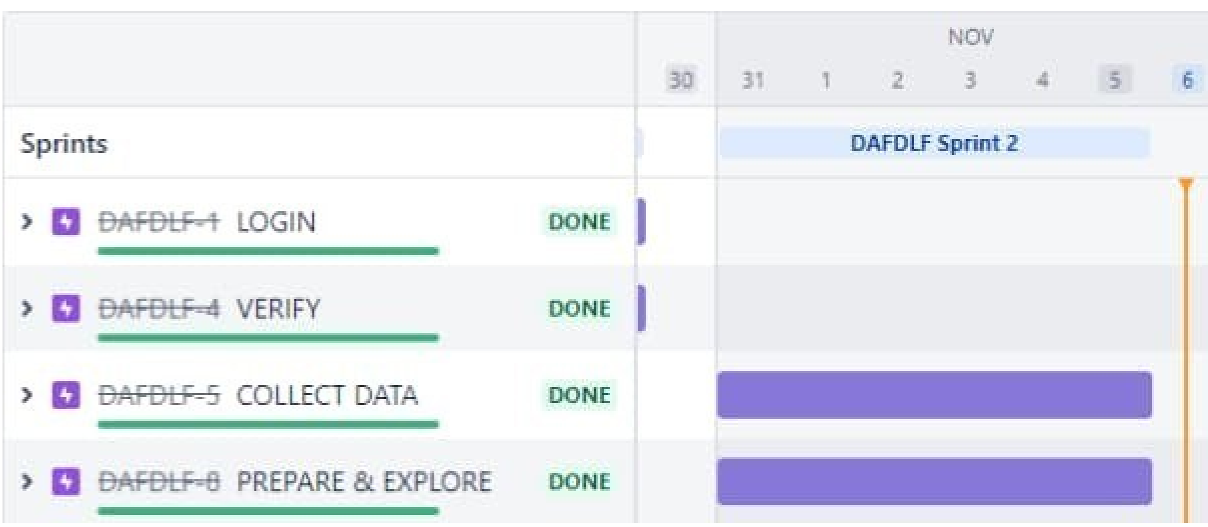
TITLE	DESCRIPTION	DATE
Problem statement for Crude oil prediction	Literature survey on the selected project & gathering information by referring the, technical papers, research publications etc.	19 OCTOBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	19 OCTOBER 2022
Ideation	List the by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	20 OCTOBER 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	23 OCTOBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	25 OCTOBER 2022
Solution Architecture	Prepare solution architecture document.	27 OCTOBER 2022

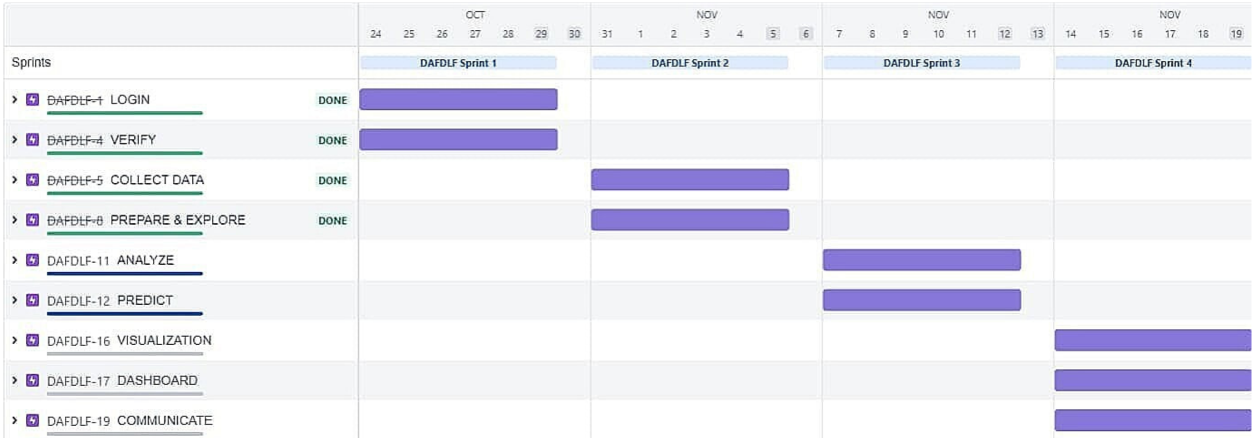
Customer Journey	Prepare the customer journey maps to understand the user interactions & experiences with the application (entry to exit).	28 OCTOBER 2022
Solution-Requirement	Prepare the solution requirement document.	29 OCTOBER 2022
Data Flow Diagrams	Draw the data flow diagrams and submit for review.	29 OCTOBER 2022
Technology Architecture	Prepare the technology architecture diagram.	01 NOVEMBER 2022
Prepare Milestone & Activity List	Prepare the milestones & activity list of the project.	04 NOVEMBER 2022
Project Development - Delivery of Sprint-1, 2, 3 & 4	Develop & submit the developed code by testing it.	IN PROGRESS...

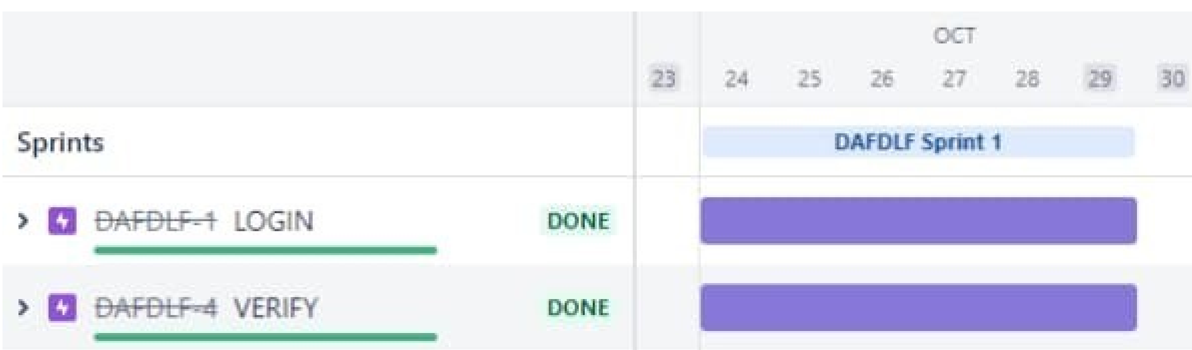
6.2 Sprint Delivery Schedule

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	10	High	Soodappan K
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	10	High	Sathish M
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password.	15	High	Sathish Kumar R
Sprint-2	Input Necessary Details	USN-4	As a user, I can give Input Details to Predict Likeliness of crude oil	15	High	Vijaysankar G
Sprint-2	Data Pre-processing	USN-5	Transform raw data into suitable format for prediction.	15	High	Soodappan K
Sprint-3	Prediction of Crude Oil Price	USN-6	As a user, I can predict Crude oil using machine learning model.	20	High	Sathish M
Sprint-3		USN-7	As a user, I can get accurate prediction of crude oil	5	Medium	Sathish Kumar R
Sprint-4	Review	USN-8	As a user, I can give feedback of the application.	20	High	Vijaysankar G

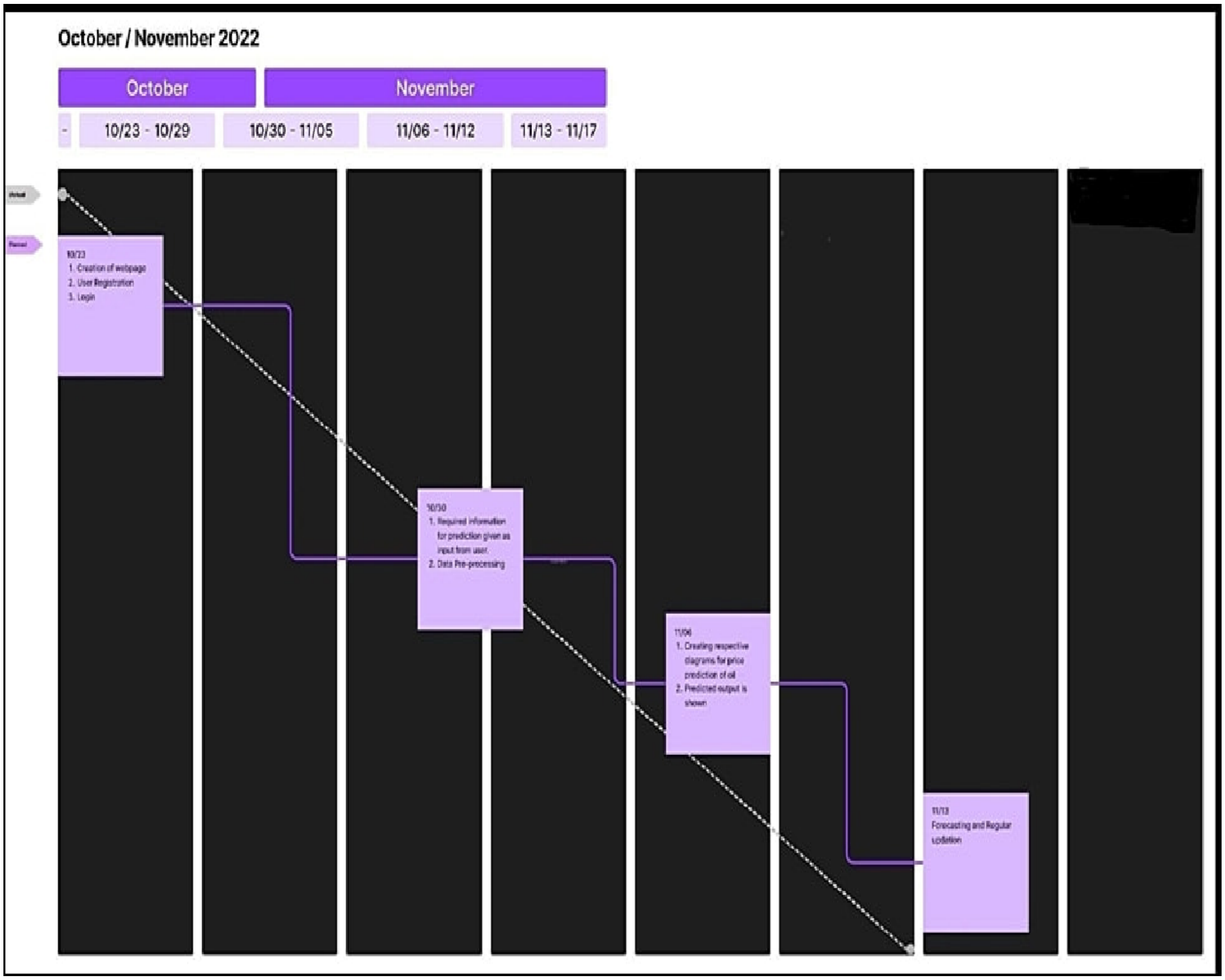
6.3 Report from JIRA




OCT
NOV
NOV
NOV
24 25 26 27 28 29 30 31 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19
Sprints
DAFDLF Sprint 1
DAFDLF Sprint 2
DAFDLF Sprint 3
DAFDLF Sprint 4
~~DAFDLF-1~~ LOGIN DONE
~~DAFDLF-4~~ VERIFY DONE
~~DAFDLF-5~~ COLLECT DATA DONE
~~DAFDLF-8~~ PREPARE & EXPLORE DONE
DAFDLF-11 ANALYZE
DAFDLF-12 PREDICT
DAFDLF-16 VISUALIZATION
DAFDLF-17 DASHBOARD
DAFDLF-19 COMMUNICATE


OCT
23 24 25 26 27 28 29 30
Sprints
DAFDLF Sprint 1
~~DAFDLF-1~~ LOGIN DONE
~~DAFDLF-4~~ VERIFY DONE

BURNDOWN


October / November 2022
October
November
-
10/23 - 10/29
10/30 - 11/05
11/06 - 11/12
11/13 - 11/17
10/23
1. Creation of webpage
2. User Registration
3. Login
10/30
1. Required information for prediction given as input from user.
2. Data Pre-processing
11/06
1. Creating respective diagrams for price prediction of oil
2. Predicted output is shown
11/13
Forecasting and Regular updation

7. CODING & SOLUTIONING

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import datetime
from pylab import rcParams
import matplotlib.pyplot as plt
import warnings
import itertools
import statsmodels.api as sm
from keras.models import Sequential
from keras.layers import Dense
from keras.layers import LSTM
from keras.layers import Dropout
from sklearn.metrics import mean_squared_error
from keras.callbacks import ReduceLROnPlateau, EarlyStopping,
ModelCheckpoint
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
import seaborn as sns
sns.set_context("paper", font_scale=1.3)
sns.set_style('white')
import math
from sklearn.preprocessing import MinMaxScaler
# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter) will list all
files under the input directory
warnings.filterwarnings("ignore")
plt.style.use('fivethirtyeight')
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
```

```
    for filename in filenames:
        print(os.path.join(dirname, filename))
dateparse = lambda x: pd.datetime.strptime(x, '%b %d, %Y')
#Read csv file
from google.colab import files
uploaded = files.upload()
df = pd.read_csv('BrentOilPrices.csv',parse_dates=['Date'], date_parser=dateparse)
#Sort dataset by column Date
df = df.sort_values('Date')
df = df.groupby('Date')['Price'].sum().reset_index()
df.set_index('Date', inplace=True)
df=df.loc[datetime.date(year=2000,month=1,day=1):]
```

Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.

Saving BrentOilPrices.csv to BrentOilPrices (1).csv

```
df.head()
```

	Price
Date	
2000-01-04	23.95
2000-01-05	23.72
2000-01-06	23.55
2000-01-07	23.35
2000-01-10	22.77

```
def DfInfo(df_initial):
    # gives some infos on columns types and numer of null values
    tab_info = pd.DataFrame(df_initial.dtypes).T.rename(index={0: 'column type'})
    tab_info =
tab_info.append(pd.DataFrame(df_initial.isnull().sum()).T.rename(index={0: 'null
values (nb)'}))
    tab_info = tab_info.append(pd.DataFrame(df_initial.isnull().sum() /
df_initial.shape[0] * 100).T.
                              rename(index={0: 'null values (%)'}))
```

```
    return tab_info
DfInfo(df)
Price
column type                         float64
null values (nb)                    0
null values (%)                     0.0
df.index
DatetimeIndex(['2000-01-04', '2000-01-05', '2000-01-06', '2000-01-07',
               '2000-01-10', '2000-01-11', '2000-01-12', '2000-01-13',
               '2000-01-14', '2000-01-17',
               ...
               '2019-09-17', '2019-09-18', '2019-09-19', '2019-09-20',
               '2019-09-23', '2019-09-24', '2019-09-25', '2019-09-26',
               '2019-09-27', '2019-09-30'],
              dtype='datetime64[ns]', name='Date', length=5016, freq=None)
y = df['Price'].resample('MS').mean()
y.plot(figsize=(15, 6))
plt.show()

rcParams['figure.figsize'] = 18, 8
decomposition = sm.tsa.seasonal_decompose(y, model='additive')
fig = decomposition.plot()
plt.show()

sc = MinMaxScaler(feature_range = (0, 1))
df = sc.fit_transform(df)
train_size = int(len(df) * 0.70)
test_size = len(df) - train_size
train, test = df[0:train_size, :], df[train_size:len(df), :]
def create_data_set(_data_set, _look_back=1):
    data_x, data_y = [], []
    for i in range(len(_data_set) - _look_back - 1):
```

```
        a = _data_set[i:(i + _look_back), 0]
        data_x.append(a)
        data_y.append(_data_set[i + _look_back, 0])
    return np.array(data_x), np.array(data_y)
look_back =90
```

8. TESTING

8.1Test Cases

```
X_train,Y_train,X_test,Ytest = [],[],[],[]
X_train,Y_train=create_data_set(train,look_back)
X_train = np.reshape(X_train, (X_train.shape[0], X_train.shape[1], 1))
X_test,Y_test=create_data_set(test,look_back)
X_test = np.reshape(X_test, (X_test.shape[0], X_test.shape[1], 1))
regressor = Sequential()

regressor.add(LSTM(units = 60, return_sequences = True, input_shape =
(X_train.shape[1], 1)))
regressor.add(Dropout(0.1))

regressor.add(LSTM(units = 60, return_sequences = True))
regressor.add(Dropout(0.1))

regressor.add(LSTM(units = 60))
regressor.add(Dropout(0.1))

regressor.add(Dense(units = 1))


regressor.compile(optimizer = 'adam', loss = 'mean_squared_error')
reduce_lr = ReduceLROnPlateau(monitor='val_loss',patience=5)
```

```
history =regressor.fit(X_train, Y_train, epochs = 20, batch_size = 15,validation_data=(X_test, Y_test), callbacks=[reduce_lr],shuffle=False)
```

8.2User Acceptance Testing

Epoch 1/20
228/228 [==============================] - 28s 101ms/step - loss: 0.0053 - val_loss: 0.0737 - lr: 0.0010
Epoch 2/20
228/228 [==============================] - 22s 96ms/step - loss: 0.0142 - val_loss: 0.1127 - lr: 0.0010
Epoch 3/20
228/228 [==============================] - 21s 93ms/step - loss: 0.0241 - val_loss: 0.1022 - lr: 0.0010
Epoch 4/20
228/228 [==============================] - 22s 95ms/step - loss: 0.0191 - val_loss: 0.0532 - lr: 0.0010
Epoch 5/20
228/228 [==============================] - 22s 96ms/step - loss: 0.0044 - val_loss: 0.0023 - lr: 0.0010
Epoch 6/20
228/228 [==============================] - 21s 93ms/step - loss: 0.0018 - val_loss: 0.0028 - lr: 0.0010
Epoch 7/20
228/228 [==============================] - 23s 100ms/step - loss: 0.0016 - val_loss: 0.0040 - lr: 0.0010
Epoch 8/20
228/228 [==============================] - 21s 91ms/step - loss: 0.0015 - val_loss: 0.0040 - lr: 0.0010
Epoch 9/20
228/228 [==============================] - 21s 92ms/step - loss: 0.0015 - val_loss: 0.0074 - lr: 0.0010

Epoch 10/20
228/228 [==============================] - 21s 92ms/step - loss: 0.0017
- val_loss: 0.0043 - lr: 0.0010
Epoch 11/20
228/228 [==============================] - 21s 93ms/step - loss: 0.0014
- val_loss: 4.9764e-04 - lr: 1.0000e-04
Epoch 12/20
228/228 [==============================] - 22s 97ms/step - loss: 0.0011
- val_loss: 4.0066e-04 - lr: 1.0000e-04
Epoch 13/20
228/228 [==============================] - 21s 93ms/step - loss:
9.6524e-04 - val_loss: 3.3321e-04 - lr: 1.0000e-04
Epoch 14/20
228/228 [==============================] - 21s 92ms/step - loss:
9.6356e-04 - val_loss: 2.8505e-04 - lr: 1.0000e-04
Epoch 15/20
228/228 [==============================] - 21s 92ms/step - loss:
9.2024e-04 - val_loss: 2.7974e-04 - lr: 1.0000e-04
Epoch 16/20
228/228 [==============================] - 21s 93ms/step - loss:
9.1895e-04 - val_loss: 2.7104e-04 - lr: 1.0000e-04
Epoch 17/20
228/228 [==============================] - 22s 96ms/step - loss:
8.2996e-04 - val_loss: 2.7737e-04 - lr: 1.0000e-04
Epoch 18/20
228/228 [==============================] - 21s 93ms/step - loss:
8.1935e-04 - val_loss: 2.6434e-04 - lr: 1.0000e-04
Epoch 19/20
228/228 [==============================] - 21s 94ms/step - loss:
8.7719e-04 - val_loss: 2.7174e-04 - lr: 1.0000e-05
Epoch 20/20
228/228 [==============================] - 21s 92ms/step - 8.3641e-04

9. RESULTS

9.1 PERFORMANCE MATRICES

We use two standard performance metrics in the oil price prediction literature for comparing different oil price prediction models. The first metric is Mean Squared Prediction Error (MSPE). MSPE of a prediction model measures the average of the squares of the prediction errors. The prediction error is the difference between the true value and the predicted value. Let $y_1, y_2, \ldots, y_n$ be the true oil prices and , , …, be the predicted oil prices under an oil price prediction model, then the MSPE of that model is:

For comparison purposes, we use the no-change model as the baseline model and express the MSPE of another model as a ratio relative to the MSPE of the no-change model. If the MSPE ratio of a model is less than 1, then the model is more accurate than the no-change model in terms of MSPE.

The second metric is Directional Accuracy Ratio (DAR), which measures the accuracy of predicting the direction of oil price change (i.e., whether oil price increases or decreases in the next time slot). It can be computed as follows:where if and otherwise. Note that if we do a random guess of the oil price direction by tossing a fair coin, the DAR would be 0.5. Thus, if the DAR of a model is greater than 0.5, then the model is better than a random guess.

3.3. Evaluation

Table 1, Table 2 summarize the results for predicting the U.S. refiner acquisition cost for crude oil imports. Our stream learning model achieves the lowest MSPE and the highest DAR among all prediction models for the 1-month, 6-month, 9-month and 12-month forecast time horizons. For example, for the 1-month time horizon, the MSPE of our stream learning model is 11.87, with an error reduction of 31.1% compared with the no-change model (17.23), an error reduction of 13.7% compared with the ANN model (13.75), and an error reduction of 25.3% compared with the forecast combination model in Baumeister and Kilian (2015) (15.89). The DAR of our stream learning model is 0.594, which is higher than both ANN model (0.582) and forecast combination model (0.570). For the 3-month horizon, although our model has a lower accuracy than the forecast combination model, it is more accurate than the no-change model, and more accurate than the ANN model in terms of MSPE.

Table 1. MSPE of different oil price prediction models: U.S. Refiner Acquisition Cost for Crude Oil Imports.

Time horizon	Empty Cell	**No-change**	**ANN**	**Forecast combination**	**Stream learning**
1-month	MSPE	17.23	13.75	15.89	**11.87**
	Ratio	1	0.798	0.922	**0.689**
3-month	MSPE	97.13	96.92	**88.00**	95.37
	Ratio	1	0.998	**0.906**	0.981
6-month	MSPE	221.75	205.21	212.21	**156.07**
	Ratio	1	0.925	0.957	**0.704**
9-month	MSPE	277.09	276.06	262.68	**183.79**
	Ratio	1	0.996	0.948	**0.663**
12-month	MSPE	307.72	303.23	280.64	**176.02**

	Ratio	1	0.985	0.912	**0.572**

Notes: MSPE stands for Mean Squared Prediction Error, lower is more accurate. Boldface indicates the best performance score among all four prediction models for a given forecast time horizon.

Table 2. DAR of different oil price prediction models: U.S. Refiner Acquisition Cost for Crude Oil Imports.

Time horizon	Empty Cell	**Random guess**	**ANN**	**Forecast combination**	**Stream learning**
1-month	DAR	0.5	0.582	0.570	**0.594**
3-month	DAR	0.5	0.586	**0.592**	0.562
6-month	DAR	0.5	0.510	0.556	**0.598**
9-month	DAR	0.5	0.594	0.575	**0.602**
12-month	DAR	0.5	0.643	0.627	**0.667**

Notes: DAR stands for Directional Accuracy Ratio, higher is more accurate. Boldface indicates the best performance score among all four prediction models for a given forecast time horizon.

Table 3, Table 4 summarize the results for predicting the WTI crude oil spot price. Again, our stream learning model achieves the highest accuracy among all prediction models for all forecast time horizons except for the 3-month time horizon. For example, for the 1-month time horizon, the MSPE of our stream learning model is 17.16, with an error reduction of 16.2% compared with the no-change model (20.48), an error reduction of 8.7% compared with the ANN model (18.79), and an error reduction of 8.0% compared with the forecast combination model (18.66). The DAR of our stream learning model is 0.570, which is also more accurate than both the ANN model (0.542) and the forecast combination model (0.521).

Table 3. MSPE of different oil price prediction models: West Texas Intermediate (WTI) Crude Oil Price.

Time	Emp	**No-**	**ANN**	**Forecast**	**Stream**

horizon	ty Cell	change		combination	learning
1-month	MSPE	20.48	18.79	18.66	**17.16**
	Ratio	1	0.917	0.911	**0.838**
3-month	MSPE	101.81	101.77	**92.24**	99.58
	Ratio	1	0.999	**0.906**	0.978
6-month	MSPE	233.03	226.61	224.17	**181.47**
	Ratio	1	0.972	0.962	**0.778**
9-month	MSPE	293.54	290.25	279.45	**214.34**
	Ratio	1	0.989	0.952	**0.730**
12-month	MSPE	328.31	307.62	300.73	**234.98**
	Ratio	1	0.937	0.916	**0.716**

Notes: MSPE stands for Mean Squared Prediction Error, lower is more accurate. Boldface indicates the best performance score among all four prediction models for a given forecast time horizon.

Table 4. DAR of different oil price prediction models: West Texas Intermediate (WTI) Crude Oil Price.

Time horizon	Empty Cell	Random guess	ANN	Forecast combination	Stream learning
1-month	DAR	0.5	0.542	0.521	**0.570**
3-month	DAR	0.5	**0.586**	0.576	0.546
6-month	DAR	0.5	0.469	0.543	**0.602**

9-month	DAR	0.5	0.502	0.571	**0.606**
12-month	DAR	0.5	0.526	0.605	**0.610**

Notes: DAR stands for Directional Accuracy Ratio, higher is more accurate. Boldface indicates the best performance score among all four prediction models for a given forecast tim

10.ADVANTAGES AND DISADVANTAGES

Crude oil is mainly used as a fuel and combustible, but is also indispensable as a chemical raw material. It is the foundation of modern life and in virtually every product around us – from smartphones and vehicle parts to wind turbines.
crude oil is essential for the development of a country as it provides the base of industrial advancement. 97% of the energy required for transportation is provided by crude oil.A windfall from cheaper crude oil will result in drop in production costs for various industries. But there are other drawbacks from lower crude oil prices. The global market is already feeling the pinch of the plunge as prices of most commodities have sunk.

Ethanol, heating oil, natural gas, coal, copper, steel, corn, cotton, rice, rubber, soyabean, sugar, rice and wheat have all dropped by over 20 per cent in the last one year. This means returns to farmers will be lower, thus curbing their purchasing power.

Will that have an effect on consumer goods, electronics and other such products? We will have to wait and see, though it is most like the issue of lower rubber prices figured in the Rajya Sabha. The reality is that with crude oil prices ruling low, natural rubber will be on leash since the user industry has the option to switch over to cheaper synthetic rubber. Corn and vegetable oils will also be under pressure in view of lower prices. Ethanol production at current crude oil level is unviable, while demand for vegetable oils for bio-diesel will be subdued.

DISVANTAGES

Oilisanon-renewablesourceofenergy.

Burning oil produces carbondioxide gas.

Burning oil can pollute the air.

Much of our oil has to be import edanditis becoming more and more expensive as reserves reduce and import sincrease.

oil releases gases such as carbondioxide and other green house emissions.

These gases are very harmful to the ozonelayeras their accumulation encourages its deterioration.

Continued destroying of the ozone layer ultimately leads to globalwarming.

Lower oil prices mean less drilling and exploration activity becausemost of the new oil driving the economic activity is unconventional and hasahigher cost per barrel than a conventional source of oil.

Less activity can lead to layoffs which can hurt the local businesses that catered to these workers.

11.CONCLUSION

Crude oil prices are highly fluctuated time series. It is affected bymanyeconomicandpoliticalfactors.Specially,thereareseveralsuddenincreaseanddecreasethroughoutthetime.

In order to eliminate the irregular trend. We try several methods, hpfilter,loessfilter,log transformation and difference.

It shows that log transformation and difference have the best performance. After this manipulation, the data is generally stationary.

Based on the whole modeling and prediction, we can conclude that our model is relatively easy to handle and capture the mainfeature.

But the crudeoil price is not that stationary for analysis, and exhibits intensive flucatuation even after transformation. The crude oil price can suffer sudden increase and decrease.

Thus, only analysis of the crude oil price itself can hardly predict thesudden change. Maybe, we can find some latent variable to improve modeling and prediction.

Introduce the variable selection before forecasting. In this process,we compare three different methods and analyze core influencing factors based on the literature review from supply and demand, global economic development, financial market, and technology aspects.

12.FUTURESCOPE

It shows that the prediction accuracy of the variable selection-machinelearning integrated model is significantly improved compared with that oftheunivariatemodel.

The number of core variables selected by BMA is neither the most nor theleastamongthethreevariableselectionmodels,indicatingthatthenumberofcorevariableswillalsoaffectthepredictionresults.

Thestatisticaltestresultsshowthatthepredictionof1stepinadvancein-sampleand1stepinadvanceinoutofsample.

wemayintroducemoreindependentvariableswiththehelpofinternetsearchdata,testourframeworkperformance.

Moreover, investor sentiment can be quantified in this process. In addition,differentvariableselectionmethodscanbeintroducedmore.

Thisindicatesthatthevariableselection-basedmachinelearningintegrated research framework proposed in thissignificantly improves theforecastingperformanceofoilprices

13.APPENDIX

Source Code

APP.PY

```
from google.colab import drivedrive.mount('/content/drive')

importnumpyasnp#used for numerical analysis
fromflaskimportFlask,render_template,request#Flaskisaapplicationusedtorun/serve
ouraimporttensorflowastf
```

```
fromtensorflow.keras.modelsimportload_model#we areloading ourmodel from
keras


tf.get_logger().setLevel('ERROR')app=Flask(name)#ourflaskapp

model=load_model('/content/drive/MyDrive/crude_oil.h5',)#loadingthe model
inthe flask app


@app.route('/') #rendering html templatedefhome() :

    return render_template("/content/drive/MyDrive/index.html") #rendering html

template@app.route('/about')


defhome1() :
    return render_template("/content/drive/MyDrive/index.html") #rendering html

template@app.route('/predict')

defhome2() :
returnrender_template("/content/drive/MyDrive/web.html")#rendering
htmltemplate


@app.route('/login',methods = ['POST']) #route for our predictiondeflogin() :

x_input=str(request.form['year']) #requesting the filex_input=x_input.split(',')


print(x_input)
for i in range(0, len(x_input)):x_input[i]=float(x_input[i])

print(x_input)x_input=np.array(x_input).reshape(1,-

1)temp_input=list(x_input)temp_input=temp_input[0].tolist()lst_output=[]

n_steps=10i=0while(i<1):
```

```
if(len(temp_input)>10):#print("temp
    input",temp_input)x_input=np.array(temp_input[1:])
    print("{}dayinput{}".format(i,x_input))x_input=x_input.reshape(1,-1)
    x_input=x_input.reshape((1,n_steps,1))#print(x_input)

    yhat=model.predict(x_input,verbose=0)print("{}dayoutput{}".format(i,yhat))temp
    _input.extend(yhat[0].tolist())temp_input=temp_input[1:]#print(temp_input)lst_out
    put.extend(yhat.tolist())
        i=i+1else:

    x_input=x_input.reshape((1,n_steps,1))yhat = model.predict(x_input,
    verbose=0)print(yhat[0])temp_input.extend(yhat[0].tolist())print(len(temp_input))l
    st_output.extend(yhat.tolist())
            i=i+1print(lst_output):



    returnrender_template("/content/drive/MyDrive/web.html",showcase='Thepredicte
    dvalueis#return str(x)


    if__name__=='__main__';
```

Web.html:

```
<htm l>
```

```
<metacharset="utf-8">
<metaname="viewport"content="width=device-width,initial-scale=1">

<st
yle
>di
v.he
ade
r{

top:0;posi
tion:fixed;
   padding-left:
400px;}div.hea
der1{
top:20;pos
ition:fixe
d;

padding-left:490px;
}

*{
margin:0;
padding:
```

```
0;border:
0;outline
:0;
text-
decoration:none;font-
family:montserrat;
}
.navbar
{
margin-
left:10px;
padding:
10px;
background-color:hsl(180,
96%, 52%);font-
family:'Roboto',sans-
serif;font-style:italic;
border-
radius:30p
x;font-
size:30px;
box-sizing:
border-
box;max-
width:18%;
text-align:center;
```

```
}
a:hover{
background-
color:black;col
or:white;
border-
radius:16px;
0font-
size:30px;p
adding:10p
x;
}


body
{
background-image:url("https://media.gettyimages.com/photos/pouring-oil-to-
car-engine-pictubackground-position:center;
font-
family:sans-
serif;backgrou
nd-
size:cover;mar
gin-top:40px;
}
```

```
.maininput[type="text"],.maininput[type="text"],.maininput[type="text"],.maininput[t
    ypborder:0;
    background:non
    e;display:block;
    margin:20px
    auto;text-
    align:center;
    border:2px solid
    #800080;padding:15px
    3px;width:400px;outlin
    e:none;color:white;
    border-
    radius:100px;trans
    ition:0.25s;font-
    size:20;


    }
    .b
    or
    {
    b
    or
    d
    er
    :
```

```
0;
background:non
e;display:block;
margin:20px
auto;text-
align:center;
border:2px solid
#800080;padding:10px
3px;width:500px;outlin
e:none;color:white;tran
sition:0.25s;
}
.maininput[type="text"]:focus,.maininput[type="text"]:focus,.maininput[type="
text"]:focinput[type="text"]:focus{

width:280px;
border-color:#8e44ad;
}
.logbtn{
display:block;wid
th:35%;height:50
px;border:none;b
order-
radius:24px;
background:linear-
gradient(120deg,#3498db,#8e44ad,#3498db,#8e44ad);backgrou
nd-size:200%;
```

```
color:#fff;outl
ine:none;curs
or:pointer;tra
nsition:.5s;fo
nt-size:25;
}
.logbtn:hover{
background-center;
}

input::placeholder{
color:#F5FFFA;
}
.bottom-text{
margin-
top:60px;text-
align:center;fon
t-size:13px;

}

</style>
<body>
<divclass="navbar">
<ahref="index.html">Home</a>
<br>
```

```
</div>
<center><div><fontcolor="Powderblue"font-family="sans-
serif"size=8><b>CrudeOil



<br><br><br><br>


<formclass="main"action="/login"method="post">
<br>
<fontsize=20><inputtype="text"name="year1"placeholder="Enterprevious10t
<fontsize=20><inputtype="text"name="year2"placeholder="Enterprevious9th
<fontsize=20><inputtype="text"name="year3"placeholder="Enterprevious8th
<fontsize=20><inputtype="text"name="year4"placeholder="Enterprevious7th
<fontsize=20><inputtype="text"name="year5"placeholder="Enterprevious6th
<fontsize=20><inputtype="text"name="year6"placeholder="Enterprevious5th
<fontsize=20><inputtype="text"name="year7"placeholder="Enterprevious4th
<fontsize=20><inputtype="text"name="year8"placeholder="Enterprevious3th
<fontsize=20><inputtype="text"name="year9"placeholder="Enterprevious2nd
<fontsize=20><inputtype="text"name="year10"placeholder="Enterprevious1
st
<center><inputtype="submit"class="logbtn"value="Predict"></center>
<divclass="bor"><b><fontcolor="white"size=5>showcase</font></b></div>
</form>
</div>
</body>


</html>
```

Index.html:

```
<!DOCTYPE
html>


<html>
<head>
<title>Home</title>
<metacharset="utf-8">
<metaname="viewport"content="width=device-width,initial-scale=1">
<style>
body
{
background-image:url("https://media.gettyimages.com/photos/pouring-oil-to-
car-engibackground-position:center;
font-family:Times-new
roman;background-
size:cover;margin-top:40px;
}
.pd{
padding-bottom:100%;}
.navbar
{
margin-
left:10px;paddin
g:10px;
```

```
background-color:hsl(180, 96%,
52%);font-family:'Roboto',sans-
serif;font-style:italic;
border-
radius:30px;font-
size:30px;
box-sizing: border-
box;max-width:18%;
text-align:right;

}
a
{
color:grey;fl
oat:right;
text-
decoration:none;fon
t-style:normal;
padding-right:20px;
}
a:hover{
background-
color:black;color:whit
e;
border-
radius:15px;0font-
```

```
size:30px;padding-
left:10px;
}
p
{
color:turqouise;fon
t-style:italic;font-
size:30px;text-align:
left;padding-left:
500px;text-
align:justify;
}
</style>
</head>
<body>
<divclass="navbar">
<ahref="web.html">Predict</a>
<ahref="index.html">Home</a>
<br>
</div>
<br>
<center><bclass="pd"><fontcolor="white"size="15"font-
family="ComicSansMS">C
<div>
<br>
<center>
<p><fontcolor="white">Demandforoilisinelastic,thereforetheriseinpriceiscosts
of purchasing oil.
```

Because oil is the largest traded commodity, the effects are quite significant.
A rising oil price can even shift economic/political power from oil importers

toThecrude oil price movements are subject to diverse influencing factors.</p>

</center>

</div>

</body>

</html>

GITHUBLINK:

https://github.com/IBM-EPBL/IBM-Project-45700-1660731727.git

PROJECT DEMO LINK:

https://drive.google.com/file/d/1kPSD5owkAswvi8JVJ7Z_9pfu_HQlzXh8/view?usp=drivesdk