Team ID	**PNT2022TMID46996**
Project Name	**Crude Oil Price Prediction**

INTRODUCTION:

1.1 PROJECT OVERVIEW:

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.

1.2 PURPOSE:

Crude oil price prediction is a challenging task in oil producing countries. It has fluctuations such as demand, supply, cost, etc., its prices is among the most complex and tough to model because fluctuation of price in crude oil are highly irregular, nonlinear and varies dynamically with high uncertainty.

In order to overcome all the above mentioned obstacles and problems we are here to provide a user friendly application with high benefits.

In modern predictions the predictors only used historical data but by using algorithms and user interface this crude oil price predictions are made user friendly.

Due to fluctuations it is very difficult for the users to predict the prices and thereby they don’t have correct resources to gather information hence by using this interface the process has been made easier to the users.

LITRATURE SURVEY:

2.1 EXISTING PROBLEM

- Using the proposed model the major crude oil price moment is analysed and model which interest in economic and finance field.
- If it forecast the price with quantitative and qualitative method.
- It focuses only on oil price volatility analysis and oil price determination.
- The algorithm used in this crude oil price prediction are SVM, ARIMA, Back propagation, Random walk model. for classification or regression of data groups. But it does not execute very well when the data set has more sound ie., target classes are over
- **Support vector machine:** It is a type of deep learning algorithm that perform supervised learning lapping.
- **Random walk model:** It is described as a "stochastic" process because it works through the application of random variables. Chances in stock prices have the same distribution and are independent of each other.
- **Back propagation:** It is an algorithm that is designed to test for errors working back from output nodes to input nodes. Back propagation could be rather sensitive to noisy data and irregularity.
- **ARIMA:** Auto Regressive Integrated Moving Average is used to understand past data or to predict future data in a series. It cannot be used for seasonal time series. It is difficult to predict turning points.

2.2 REFERENCES

1. Yang, C.; Lv, F.; Fang, L.; Shang, X. The pricing efficiency of crude oil futures in the Shanghai International Exchange. Financ Res.Lett. 2020, 36, 101329. [CrossRef]

2. Murat, A.; Tokat, E. Forecasting oil price movements with crack spread futures. Energy Econ. 2009, 31, 85–90. [CrossRef]

3. Yu, L.; Zhao, Y.; Tang, L. Ensemble forecasting for complex time series using sparse representation and neural networks. J.Forecast. 2017, 36, 122–138. [CrossRef]

4. Hou, A.; Suardi, S. A nonparametric GARCH model of crude oil price return volatility. Energy Econ. 2012, 34, 618–626. [CrossRef]

5. Lanza, A.; Manera, M.; Giovannini, M. Modeling and forecasting cointegrated relationships among heavy oil and product prices.Energy Econ. 2005, 27, 831–848. [CrossRef]

6. Jammazi, R.; Aloui, C. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. Energy Econ. 2012, 34, 828–841. [CrossRef]

7. Xie, W.; Yu, L.; Xu, S.; Wang, S. A new method for crude oil price forecasting based on support vector machines. In International Conference on Computational Science; Springer: Berlin/Heidelberg, Germany, 2006; pp. 444–451.

8. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. Science 2006, 313, 504–507. [CrossRef] [PubMed]

9. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef] [PubMed]

10. Sutskever, I.; Hinton, G.E.; Taylor, G.W. The recurrent temporal restricted boltzmann machine. Adv. Neural Inf. Process. Syst. 2008, 21, 1601–1608.

2.3 PROBLEM STATEMENT DEFINITION:

PROBLEM:

- Crude oil price fluctuations have a far reaching impact on global economies.
- These fluctuations make a major impact on stake holders and the investors.
- Cause of fluctuation of prices, demand and supply.

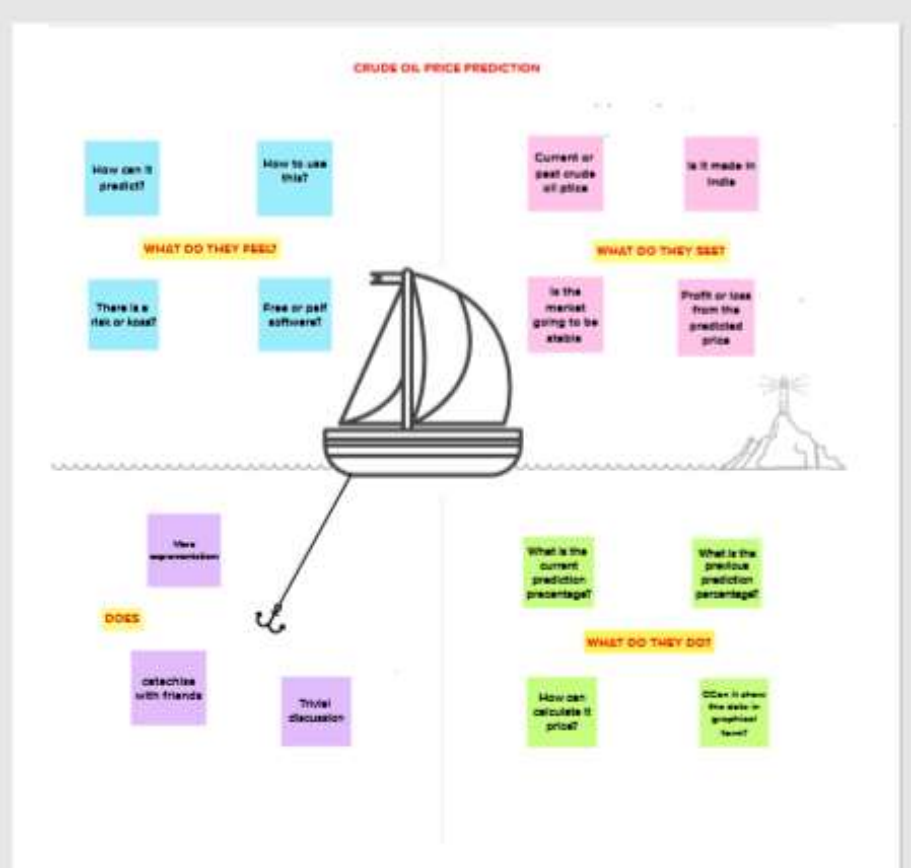
SOULTION:

- In this project we mainly focus on neural networks to predict the crude oil prices.

- We are using the previous history of crude oil prices as the input data to the training model.
- Here data preprocessing, training and testing is done using ANN.
- We use RNN and LSTM algorithm, were the output of one hidden layer act as input of another hidden layer. To evaluate the complex problems into a simpler way.
- The LSTM algorithm is well adapted to categorize, analyze and predict time series of uncertain.

IDEATION & PROPOSED SOLUTION:

3.1 EMPATHY MAP CANVAS



3.2 IDEATION AND BRAINSTROMING

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

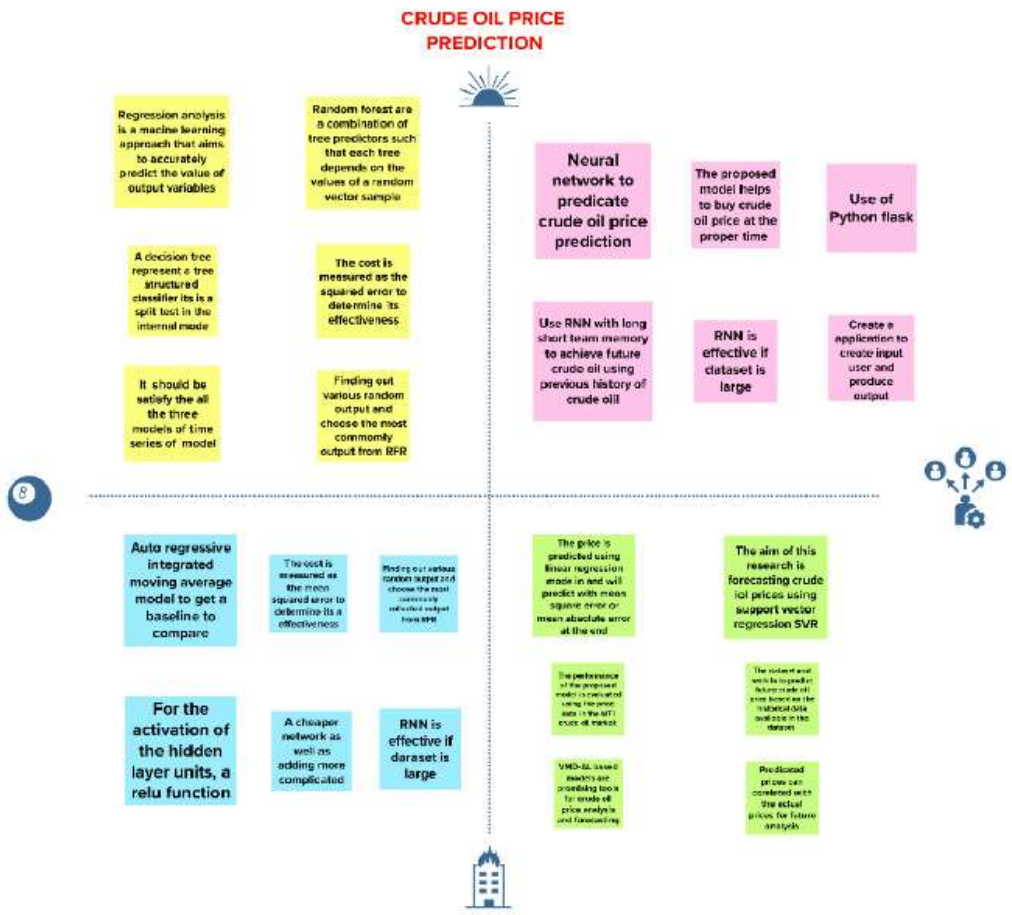
PROBLEM

What is the best solution to analyse and predict the prices of crude oil ?

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.
- If possible, be visual.


CRUDE OIL PRICE PREDICTION
Regression analysis is a machine learning approach that aims to accurately predict the value of output variables
Random forest are a combination of tree predictors such that each tree depends on the values of a random vector sample
A decision tree represent a tree structured classifier its is a split test in the internal node
The cost is measured as the squared error to determine its effectiveness
It should be satisfy the all the three models of time series of model
Finding out various random output and choose the most commomly output from RFR
Neural network to predicate crude oil price prediction
The proposed model helps to buy crude oil price at the proper time
Use of Python flask
Use RNN with long short team memory to achieve future crude oil using previous history of crude oil
RNN is effective if dataset is large
Create a application to create input user and produce output
8
Auto regressive integrated moving average model to get a baseline to compare
The cost is measured as the mean squared error to determine its effectiveness
Finding out various random output and choose the most commomly selected output from RFR
The price is predicted using linear regression mode in and will predict with mean square error or mean absolute error at the end
The aim of this research is forecasting crude oil prices using support vector regression SVR
For the activation of the hidden layer units, a relu function
A cheaper network as well as adding more complicated
RNN is effective if daraset is large

2

Discuss the results

After everyone makes their placement, discuss the results.

① 5 minutes

A)Grouping based on the dataset

RNN is effective if dataset is large

It should satisfy all the three model of time series model

Finding out which a series output and choose the most accurately calculated output from RNN

B)Grouping based on model

Neural networks to predicts crude oil price

Use RNN with long short term memory to achieve future crude oil price prediction, history of crude oil

Auto regressive integrated moving average(ARIMA) model to get a baseline

C)Deploy model

Deploy the model using python flask

Create a application to create input from user and produce output

D)Model evaluation

Draw graphs and plots for analysing the results

Model

After you collaborate

Sometimes a detailed discussion the result is enough, but here are some ways you could build on this discussion.

Quick add-ons

Keep reflecting

Fred Polak claims that a culture's collective image of the future represents the direction that society is heading. Review the collaborate grid, saving only the collection of images, and reflect on these questions:

- What does this collage reveal about the world?
- What are we doing to create or avoid these realities?

Keep moving forward

The Wicked Opportunities® Creator Economy Digital Pack

This edition of the Wicked Opportunities Card Decks explores the Creator Economy.

Open the link →

Resource Center

Looking for resources to support your foresight practice? Browse our Resource Center for free digital readers, trend cards, worksheets, and templates! Printed materials can also be purchased and delivered to your door.

Open the link →

Frameworks for Quality Futures Intelligence

Environmental scanning is the lifeblood of Strategic Foresight. But getting started can feel like information overload.

Open the link →

The Image of the Future by Fred Polak

Download the book →

The Probability/Impact Matrix

Open the template →

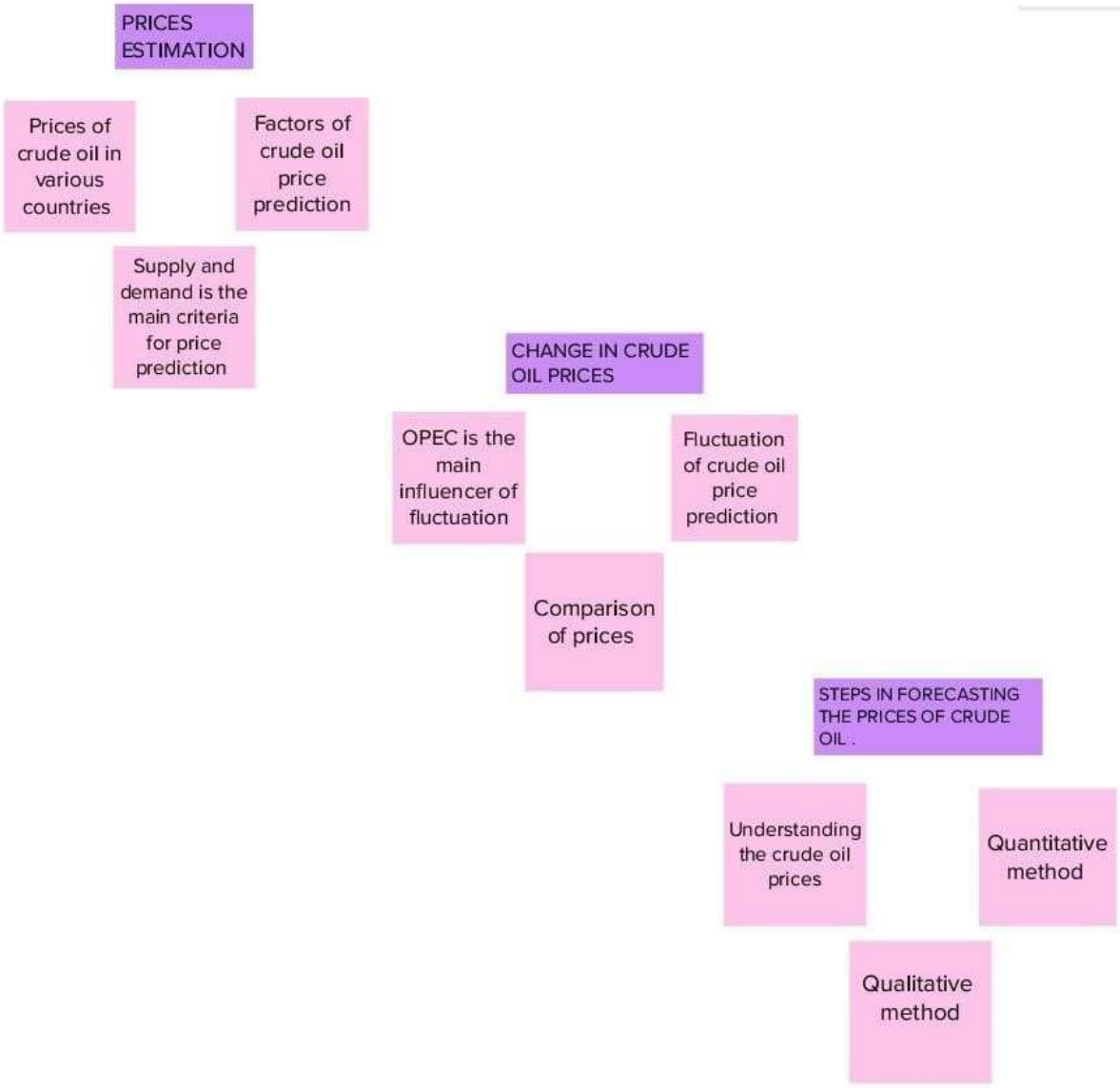
Trend Card Deck Template

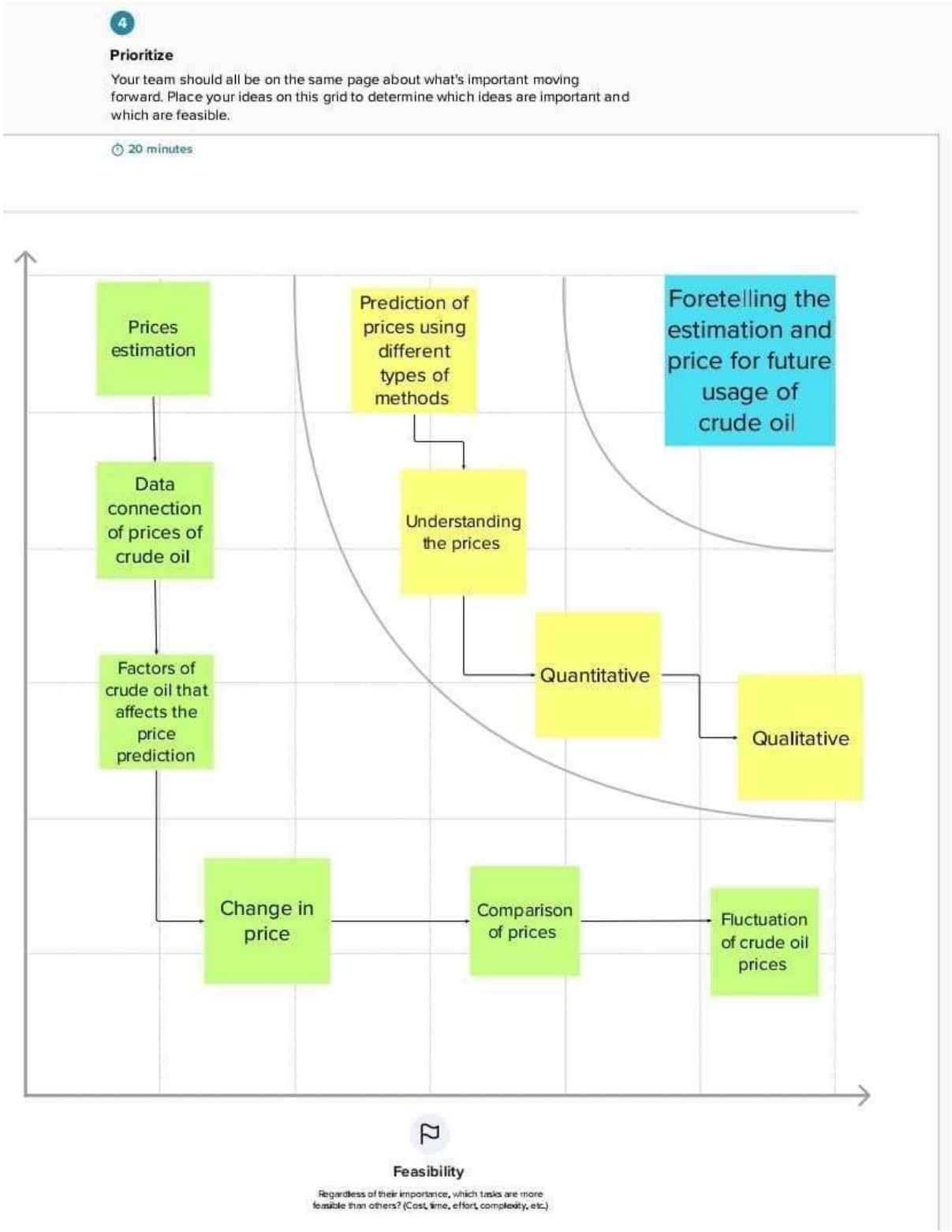
Open the template →

3

Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. In the last 10 minutes, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.





3.3 PROPOSED SOLUTION:

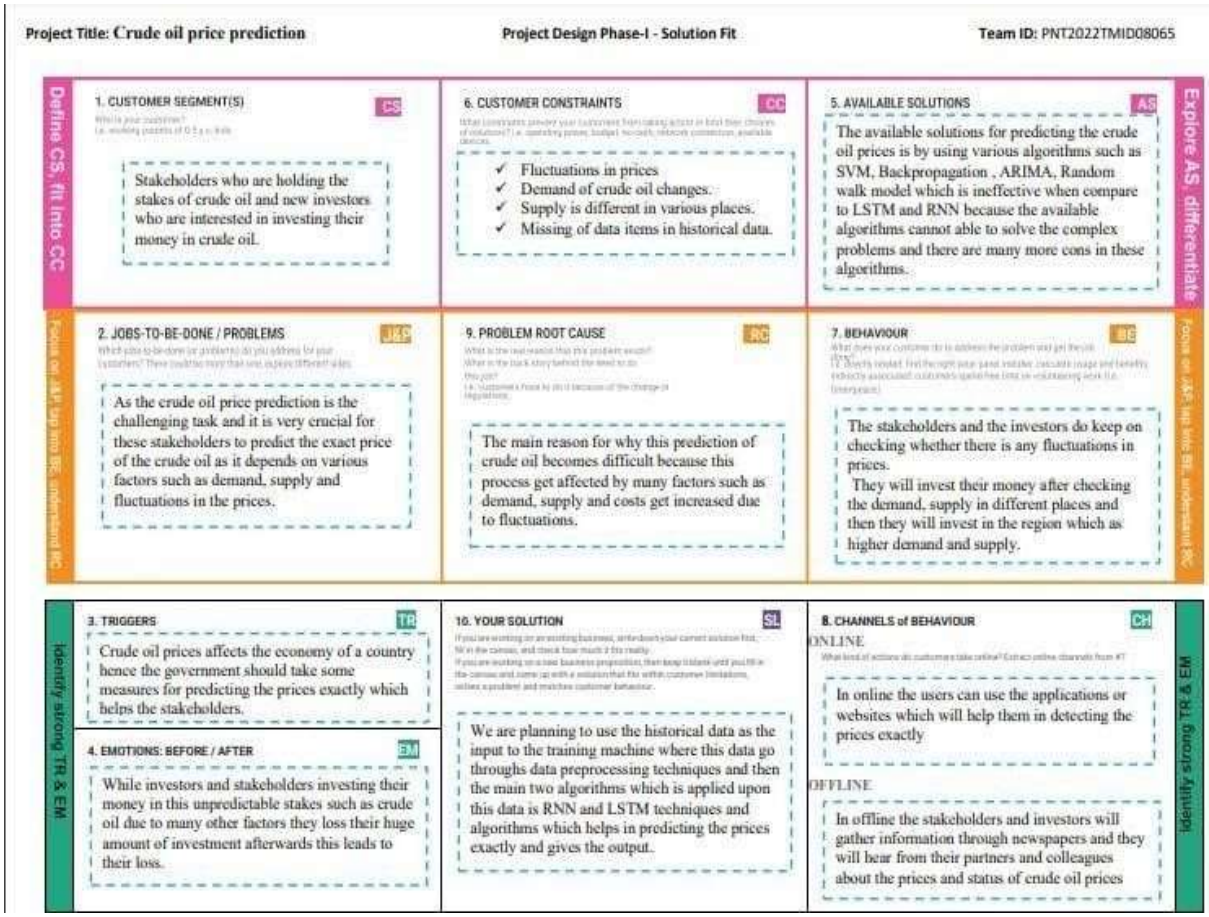
- In this proposed system initially the previous historical data of crude oil prices is given as an input in the csv file format.
- With this input the data pre-processing is done which includes the process of

    1. Quality assessment.
    2. Data cleaning.
    3. Data reduction.

the evaluation is done by deep learning algorithms like RNN and LSTM which makes the complex process easier and involves in time series prediction.

- At the end the trained model predicts the accurate prices of crude oil.
- **LSTM:** Long Short Term Memory is an artificial neural network used in the fields of artificial intelligence and deep learning. Unlike standard feed-forward neural networks, LSTM has feedback connections. Such a recurrent neural network (RNN) can process not only single data points (such as images), but also entire sequences of data (such as speech or video).
- **RNN:** A recurrent neural network (RNN) is a class of artificial neural networks where connections between nodes can create a cycle, allowing output from some nodes to affect subsequent input to the same nodes. This allows it to exhibit temporal dynamic behaviour. Derived from feed-forward neural networks, RNNs can use their internal state (memory) to process variable length sequences of inputs.

SOLUTION FIT



REQUIREMENT ANALYSIS:

4.1 FUNCTIONAL REQUIREMENTS

Following are the functional requirements of the proposed solution.

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	Showing graph by obtaining the data from the dataset
FR-4	Support	Providing answers for the queries asked by users.
FR-5	News	Information of the oil prices will be updated by admin
FR-6	Notification	Notification will be sent for the users price alert
Fr-7	Database	Information of the User will be stored

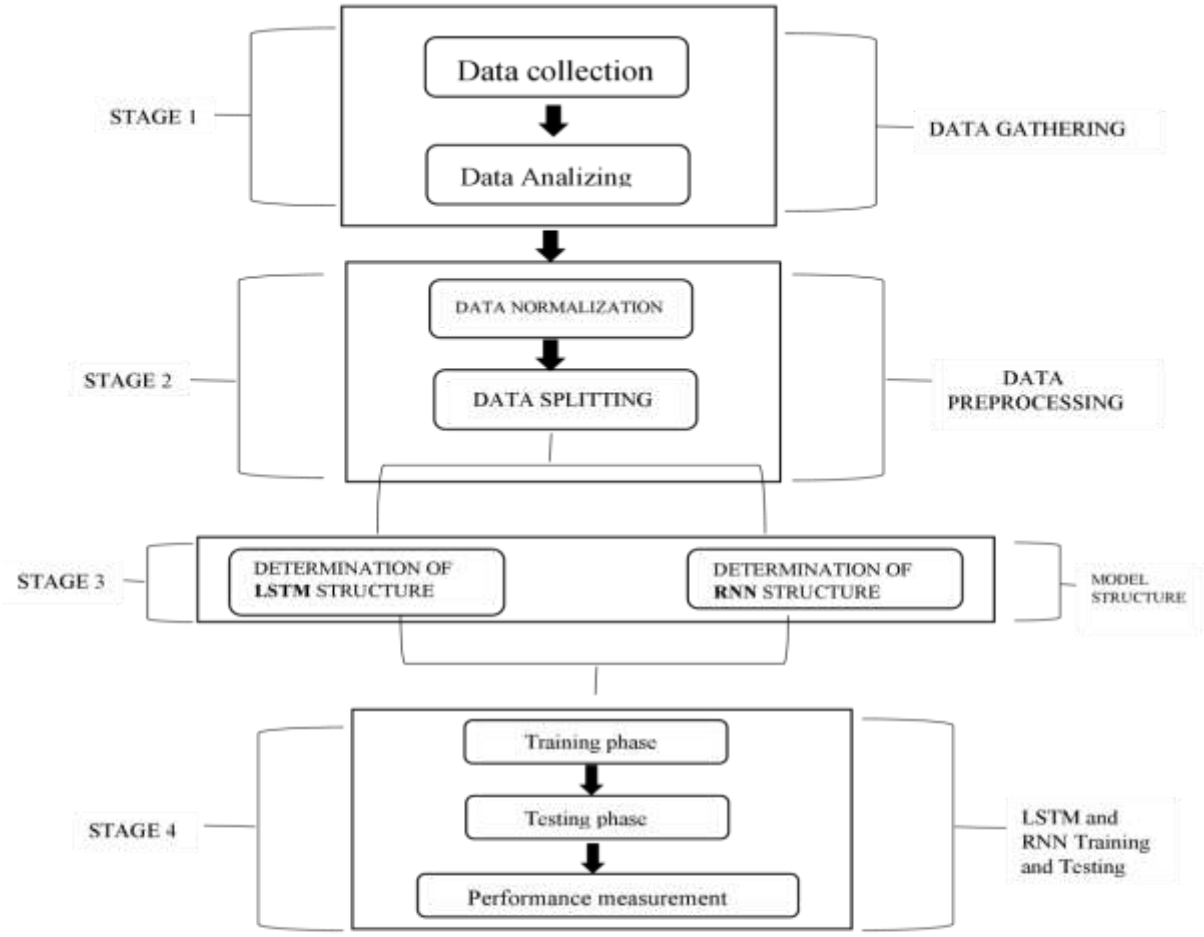
4.2 NON FUNCTIONAL REQUIREMENTS

Following are the non-functional requirements of the proposed solution.

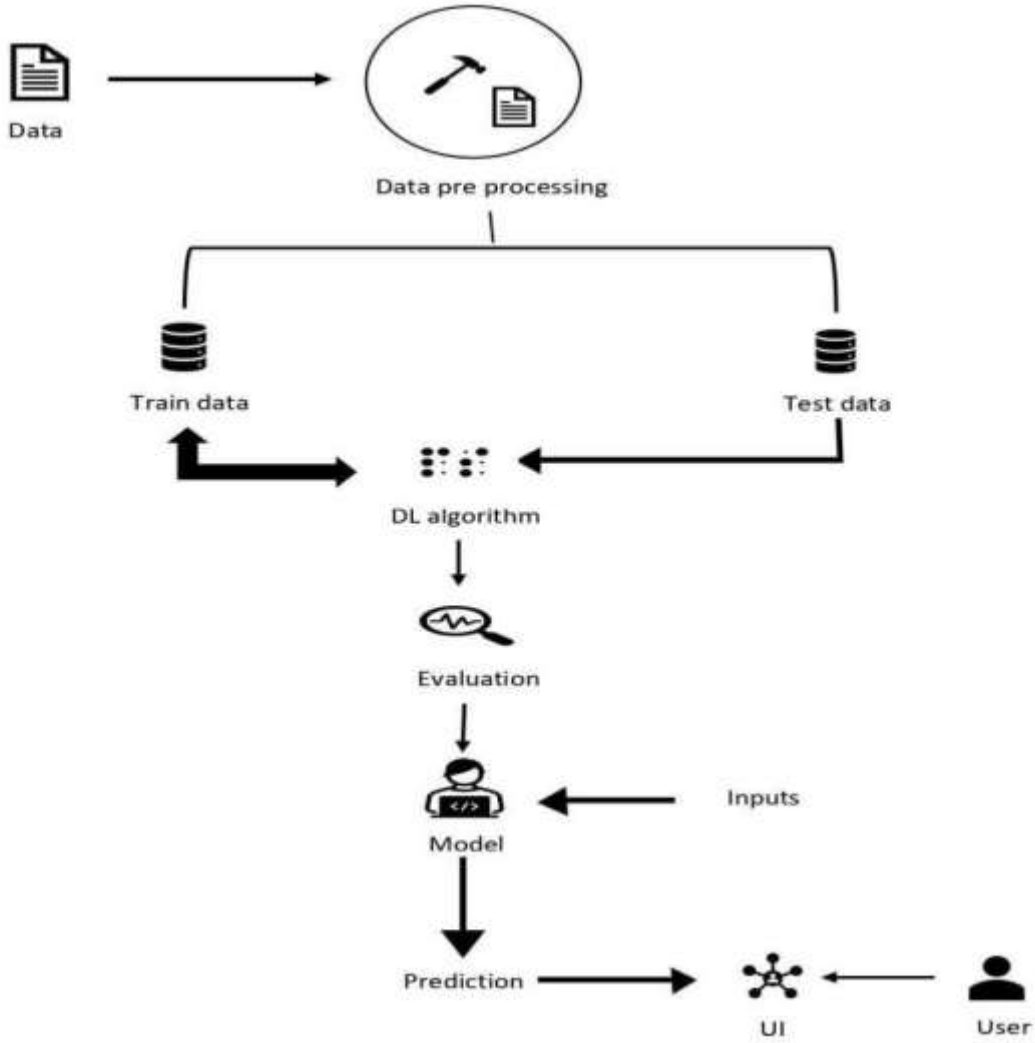
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	It can use by wide variety of client as it is very simple to learn and not complex to proceed.
NFR-2	**Security**	We are using login for the user and the information will be hashed so that it will be very secure to use.
NFR-3	**Reliability**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

PROJECT DESIGN:

5.1 DATA FLOW DIAGRAMS



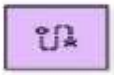
5.2 SOLUTION & TECHNICAL ARCHITECTURE



5.3 USER STORIES

SCENARIO

Browsing, booking, attending, and rating a local city tour



Steps

What does the person (or group) typically experience?

Interactions

What interactions do they have at each step along the way?

- **People:** Who do they see or talk to?
- **Places:** Where are they?
- **Things:** What digital touchpoints or physical objects would they use?

Goals & motivations

At each step, what is a person's primary goal or motivation? ("Help me..." or "Help me avoid...")

Positive moments

What steps does a typical person find enjoyable, productive, fun, motivating, delightful, or exciting?

Negative moments

What steps does a typical person find frustrating, confusing, angering, costly, or time-consuming?

Areas of opportunity

How might we make each step better? What ideas do we have? What have others suggested?

Entice

How does someone initially become aware of this process?

Browsing app

Firstly we should develop a app for the price prediction

Comfortable websites

That app should be very much easy to use for the users

Safing the user details

We have to safe guard the data of the users from the unauthorised persons

The development of the application will be helpful for predicting the crude oil price

Will be very helpful for the daily vehicle users.

is that possible to develop the application based on the AI domain

We must attain the price state of the oil at earliest.

We can achieve this by developing an application under deep learning

It is very easy to develop a application on this problem

On predicting the crude oil price it is very helpful globally.

The price may be unstable due to several factors like climatic change,political crisis.

The crude oil price is different price in different countries.

It is required all over the world

It will provide huge employment opportunities amoung the technical freshers.

Enter

What do people experience as they begin the process?

is it a easy process?	Getting into it	usage for several days
Yes , some experts consider this as a easy process.	In the begining it will be somewhat difficult	After usage of the application it will be very familier for the users.

Can setup the individual setup for the client

New users could interact with the users about the usage

Helper Bots also developed for the user interaction about the application

very helpful for the dailyvehicle users

Users will have strong grid to understand the usage of the application

Gain some knowledge on the new domain

Useful for predicting the price more precisely.

Users in the rural areas are not applicable on this

Un educated people could not get an knowledge on the usage of the application.

Could predict the price of the oil massively

More users are connected with eachother virtually.

Engage

In the core moments in the process, what happens?

Actively Participate

Gain the data

Make use of it

Users can be actively participated in this application

On this they could attain the desired data that they really want

On attaining the data they could make the use of the crude oil more effectively.

User can interact with the app server to attain the desired data

The server helps the user to get the data that he really wants.

Everyone can make use of it.

Make this to reach out everyone.

Crude oil prices will be predicted at the most accurate format.

Very helpful

Easy for everyone to make use of it and attain the desired data.

Some what difficult for all the users to adapt

More compled and may lead to some errors.

Opportunities of all across the world

Helpful for the daily vehicle users.

Engage

In the core moments in the process, what happens?

Actively Participate

Gain the data

Make use of it

Users can be actively participated in this application

On this they could attain the desired data that they really want

On attaining the data they could make the use of the crude oil more effectively.

User can interact with the app server to attain the desired data

The server helps the user to get the data that he really wants.

Everyone can make use of it.

Make this to reach out everyone.

Crude oil prices will be predicted at the most accurate format.

Very helpful

Easy for everyone to make use of it and attain the desired data.

Some what difficult for all the users to adapt

More compled and may lead to some errors.

Opportunities of all across the world

Helpful for the daily vehicle users.

PROJECT PLANNING & SCHEDULING:

6.1 SPRINT PLANNING AND ESTIMATION

Milestone name	Milestone number	Description	Mandatory
Project Objectives	M-01	We will be able to learn to prepare dataset, data pre-processing, working with RNN layers, read data using various libraries and algorithms such as RNN and LSTM for prediction of crude oil prices.	Yes
Project Flow	M-02	A project management process flowchart is a graphical aid, designed to visualise the sequence of steps to be followed throughout the project management process	Yes
Prerequisite	M-03	To complete this project, we should have known the following projects such as Keras, TensorFlow, RNN and LSTM etc.	Yes
Prior Knowledge	M-04	One should have knowledge on the Supervised Learning, RNN, Regression, Classification and Clustering techniques, NLP etc.	Yes
Data collection	M-05	We can collect dataset from different open sources like kaggle.com Deep learning etc.	Yes
Project Structure	M-06	Importing the data pre-processing libraries, Define the Variables for reading the data and classes for storing it . Importing the libraries to train and test the data.	Yes
Model building	M-07	Importing the model building libraries, Initialising the model, Adding RNN layers, Adding Dense layers, Configuring the learning Process, Train the model, Save the model, Predictions.	Yes
Application building	M-08	Create an HTML file ,build the python code,run the application in browser and showcase prediction on UI	Yes
Train RNN model in IBM	M-09	Register for IBM Cloud and train data Classification Model	Yes
Ideation phase	M-10	Prepare Literature Survey on the selected Project and Information Gathering, empathy map and ideation	Yes

Project design phase-I	M-11	Prepare Proposed solution, problem-solution fit and Solution Architecture	Yes
Project Design Phase-II	M-12	Prepare Customer journey functional requirements, Dataflow diagram and Technology Architecture	Yes
Project Planning Phase	M-13	Prepare Milestone list, Activity list and Sprint Delivery Plan	Yes
Project Development Phase	M-14	Project Development delivery of Sprint 1, Sprint 2, Sprint 3, Sprint 4	Yes

Activity List

Activity number	Activity	Sub activity	Assigned member	Status
1.	Project Objectives		Subhiksha	Completed
2.	Project Flow		Sneha	Completed
3.	Prerequisite		Sowjanya	Completed
4.	Prior Knowledge		Suruthi	Completed
5.	Project Structure		Supriya	
5.	Data collection	5.1 Download the Dataset	Subhiksha	Completed
6.	Data pre- processing	6.1. Import the Libraries	Sneha	In-progress
		6.2. Importing the dataset	Sowjanya	In-progress
		6.3. Handling missing data	Suruthi	In-progress
		6.4. Feature scaling	Supriya	
		6.5.Data Visualization	Subhiksha	
		6.6. Splitting Data into Train and Test	Sneha	
		6.7. Creating a dataset with sliding windows.	Sowjanya	
7.	Model Building	7.1. Importing the model building libraries.	Suruthi	In-progress
		7.2. Initialising the model.	Supriya	In-progress

		7.3. Adding LSTM layers.	Subhiksha	In-progress
		7.4. Adding output layers.	Sneha	In-progress
		7.5. Configuring the learning process.	Sowjanya	In-progress
		7.6. Train the model.	Suruthi	In-progress
		7.7. Model Evaluation	Supriya	In-progress
		7.8. Save the model	Sneha	In-progress
		7.9.Test the model	Subhiksha	
8.	Application building	8.1. Create an HTML file	Sowjanya	In-progress
		8.2. Build python code	Supriya	In-progress
		8.3. Run the app in local browser	Suruthi	In-progress
		8.4. Showcasing prediction on UI	Sneha	
9.	Train the Model on IBM	9.1. Register for IBM cloud	Subhiksha	In-progress
		9.2. Train the ML model on IBM	Sneha,Sowjanya	Completed
		9.3. Integrate flask with scoring end point.	Sneha	
10.	Ideation phase	10.1. Literature Review.	Suruthi	Completed
		10.2. Empathy map.	Supriya	Completed
		10.3. Ideation.	Subhiksha	Completed
11.	Project design phase-I	11.1 Proposed Solution	Subhiksha,Sneha, Suruthi	Completed
		11.2 Problem solution fit.	Subhiksha, Sowjanya	Completed
		11.3 Solution Architecture.	Suruthi, Supriya	Completed
12.	Project Design Phase-II	12.1 Customer journey.	Subhiksha,Sneha	Completed
		12.2. Functional requirement.	Sowjanya,Suruthi	Completed

		12.3. Data flow Diagram.	Supriya	Completed
		12.4. Technology Architecture.	Subhiksha, Sowjanya,Supriya	Completed
13.	Project Planning Phase	13.1. Prepare milestone and activity list.	Subhiksha,Suruthi	Completed
		13.2. Sprint delivery plan	Subhiksha,Suruthi	Completed
14.	Project Development Phase	14.1. Project development-Delivery of Sprint-1.	All	In-progress
		14.2. Project development-Delivery of Sprint-2.	All	In-progress
		14.3. Project development-Delivery of Sprint 3.	All	In-progress
		14.4. Project development-Delivery of Sprint-4.	All	In-progress

6.2 SPRINT DELIVERY SCHEDULE

Milestone name	Milestone number	Description	Mandatory
Project Objectives	M-01	We will be able to learn to prepare dataset, data pre-processing, working with RNN layers, read data using various libraries and algorithms such as RNN and LSTM for prediction of crude oil prices.	Yes
Project Flow	M-02	A project management process flowchart is a graphical aid, designed to visualise the sequence of steps to be followed throughout the project management process	Yes
Prerequisite	M-03	To complete this project, we should have known the following projects such as Keras, TensorFlow, RNN and LSTM etc.	Yes
Prior Knowledge	M-04	One should have knowledge on the Supervised Learning, RNN, Regression, Classification and Clustering techniques, NLP etc.	Yes
Data collection	M-05	We can collect dataset from different open sources like kaggle.com Deep learning etc.	Yes
Project Structure	M-06	Importing the data pre-processing libraries, Define the Variables for reading the data and classes for storing it . Importing the libraries to train and test the data.	Yes
Model building	M-07	Importing the model building libraries, Initialising the model, Adding RNN layers, Adding Dense layers, Configuring the learning Process, Train the model, Save the model, Predictions.	Yes
Application building	M-08	Create an HTML file ,build the python code,run the application in browser and showcase prediction on UI	Yes
Train RNN model in IBM	M-09	Register for IBM Cloud and train data Classification Model	Yes
Ideation phase	M-10	Prepare Literature Survey on the selected Project and Information Gathering, empathy map and ideation	Yes

Project design phase-I	M-11	Prepare Proposed solution, problem-solution fit and Solution Architecture	Yes
Project Design Phase-II	M-12	Prepare Customer journey functional requirements, Dataflow diagram and Technology Architecture	Yes
Project Planning Phase	M-13	Prepare Milestone list, Activity list and Sprint Delivery Plan	Yes
Project Development Phase	M-14	Project Development delivery of Sprint 1, Sprint 2, Sprint 3, Sprint 4	Yes

Activity List

Activity number	Activity	Sub activity	Assigned member	Status
1.	Project Objectives		Subhiksha	Completed
2.	Project Flow		Sneha	Completed
3.	Prerequisite		Sowjanya	Completed
4.	Prior Knowledge		Suruthi	Completed
5.	Project Structure		Supriya	
5.	Data collection	5.1 Download the Dataset	Subhiksha	Completed
6.	Data pre- processing	6.1. Import the Libraries	Sneha	In-progress
		6.2. Importing the dataset	Sowjanya	In-progress
		6.3. Handling missing data	Suruthi	In-progress
		6.4. Feature scaling	Supriya	
		6.5.Data Visualization	Subhiksha	
		6.6. Splitting Data into Train and Test	Sneha	
		6.7. Creating a dataset with sliding windows.	Sowjanya	
7.	Model Building	7.1. Importing the model building libraries.	Suruthi	In-progress
		7.2. Initialising the model.	Supriya	In-progress

FEATURES:

7.1 FEATURE

MinMax scaler subtracts the minimum value in the feature and then divides it by the range. The range is the difference between the original maximum and the original minimum. MinMax scaler preserves the shape of the original distribution. It does it meaningfully change the information embedded in the original data this scaling comprises all the inliers in the narrow range.

```
    the file content from Bluemix Object Storage."""

    url1 = ''.join(['https://identity.open.softlayer.com', '/v3/auth/tokens'])
    data = {'auth': {'identity': {'methods': ['password'],
            'password': {'user': {'name': 'member_adcb54bd899a7e39e31583bccad1577f68f1992f','domain': {'id': '4619da3fa8524bed
            'password': 'P^/m0,!#7s0H9psc'}}}}}
    headers1 = {'Content-Type': 'application/json'}
    resp1 = requests.post(url=url1, data=json.dumps(data), headers=headers1)
    resp1_body = resp1.json()
    for e1 in resp1_body['token']['catalog']:
        if(e1['type']=='object-store'):
            for e2 in e1['endpoints']:
                        if(e2['interface']=='public'and e2['region']=='dallas'):
                            url2 = ''.join([e2['url'],'/', container, '/', filename])
    s_subject_token = resp1.headers['x-subject-token']
    headers2 = {'X-Auth-Token': s_subject_token, 'accept': 'application/json'}
    resp2 = requests.get(url=url2, headers=headers2)
    return StringIO(resp2.text)


df_data_1 = pd.read_csv("crude_oil_prices_daily.csv")
df_data_1.head()
```

Out[7]:

	Date	Closing Value
0	02-01-1986	25.56
1	03-01-1986	26.00
2	06-01-1986	26.53
3	07-01-1986	25.85
4	08-01-1986	25.87

7.2 FEATURE 2

Removing null values from the data set is one of the important steps in data wrangling. These null values adversely affect the performance and accuracy of any machine learning algorithm. So it is very important to remove null values from the data set before applying any machine learning algorithm to the dataset.

In [3]:
```
data= pd.read_csv("Crude_Oil_Prices_Daily.csv")
data
```

Out[3]:

	Date	Closing Value
0	02-01-1986	25.56
1	03-01-1986	26.00
2	06-01-1986	26.53
3	07-01-1986	25.85
4	08-01-1986	25.87
...	...	...
8218	03-07-2018	74.19
8219	04-07-2018	NaN
8220	05-07-2018	73.05
8221	06-07-2018	73.78
8222	09-07-2018	73.93

8223 rows × 2 columns

In [4]:
```
data.isnull().any()
```

Out[4]:
```
Date             False
Closing Value     True
dtype: bool
```

In [5]:
```
data.isnull().sum()
```

Out[5]:
```
Date             0
Closing Value    7
```

In [6]:
```
data.dropna(axis=0,inplace=True)
data.isnull().sum()
```

Out[6]:
```
Date             0
Closing Value    0
dtype: int64
```

In [7]:
```
data_oil=data.reset_index()['Closing Value']
```

In [8]:
```
data_oil
```

Out[8]:
```
0       25.56
1       26.00
2       26.53
3       25.85
4       25.87
        ...
8211    73.89
8212    74.19
8213    73.05
8214    73.78
8215    73.93
Name: Closing Value, Length: 8216, dtype: float64
```

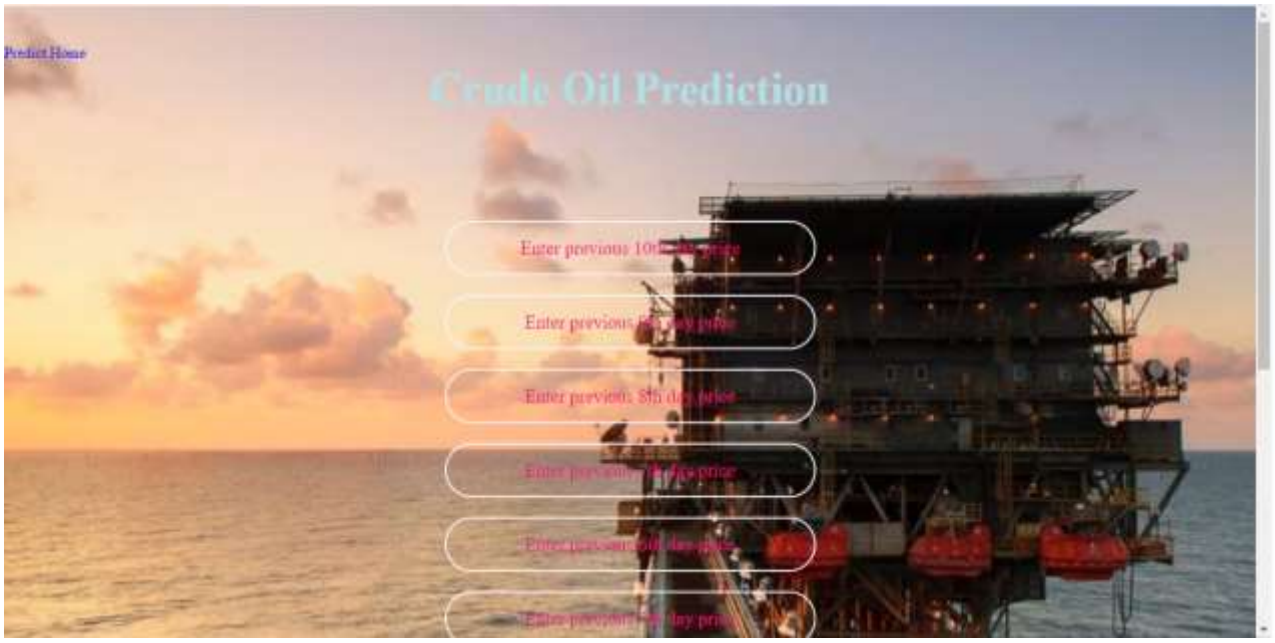
TESTING:

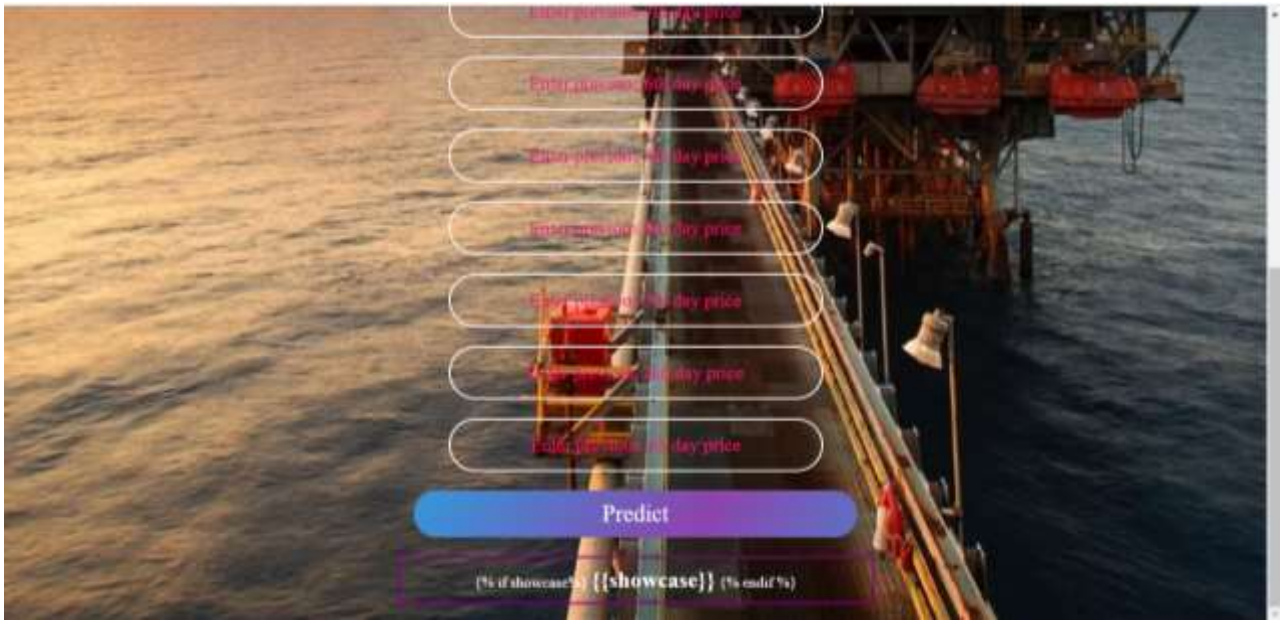
8.1 TEST CASES

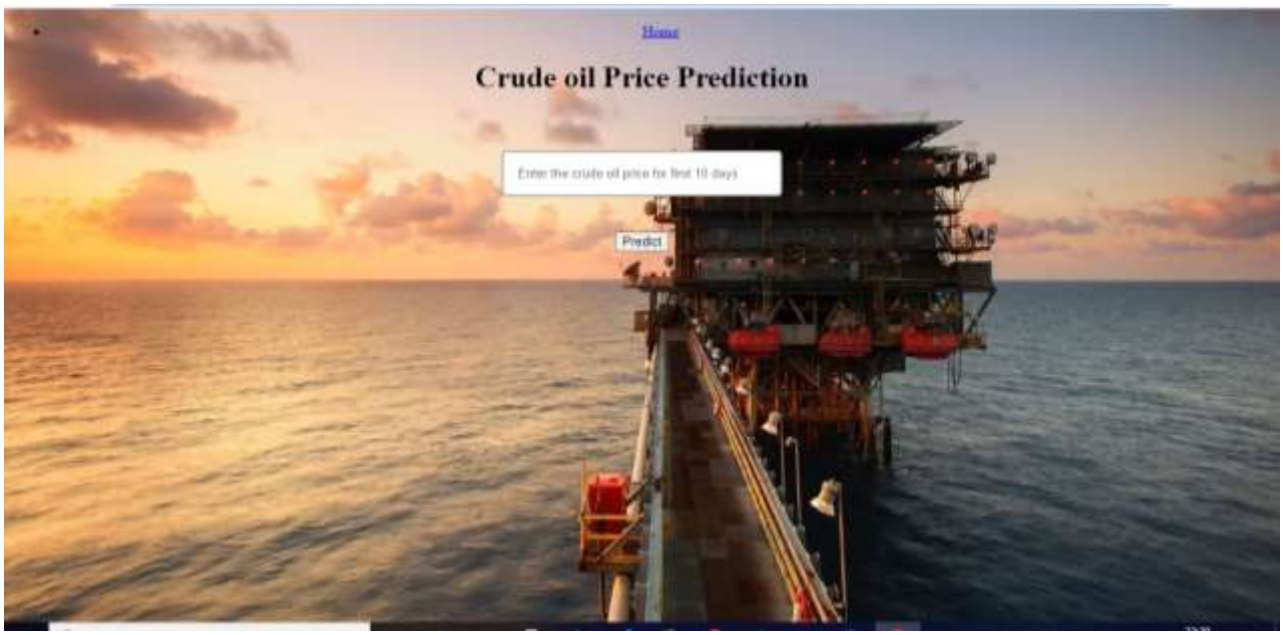
In [39]:
```
lst_output=[]
n_steps=10
i=0
while(i<10):
    if(len(temp_input)>10):
        X_input=np.array(temp_input[1:])
        print("{} day input {}".format(i,X_input))
        X_input=X_input.reshape(1,-1)
        X_input=X_input.reshape((1,n_steps,1))
        yhat=model.predict(X_input,verbose=0)
        print("{} day output {}.format(i,yhat)")
        temp_input.extend(yhat[0].tolist())
        temp_input=temp_input[1:]
        lst_output.extend(yhat.tolist())
        i=i+1
    else:
        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))
        lst_output.extend(yhat.tolist())
        i=i+1
```

```
[0.1642093]
11
1 day input [0.4811195  0.49726048 0.46794017 0.47297497 0.47119799 0.47341922
 0.46497853 0.47038353 0.47149415 0.16420931]
{} day output {}.format(i,yhat)
2 day input [0.49726048 0.46794017 0.47297497 0.47119799 0.47341922 0.46497853
 0.47038353 0.47149415 0.16420931 0.16367218]
{} day output {}.format(i,yhat)
3 day input [0.46794017 0.47297497 0.47119799 0.47341922 0.46497853 0.47038353
 0.47149415 0.16420931 0.16367218 0.16097896]
{} day output {}.format(i,yhat)
4 day input [0.47297497 0.47119799 0.47341922 0.46497853 0.47038353 0.47149415
 0.16420931 0.16367218 0.16097896 0.15659374]
```

8.2 USER ACCEPTANCE TESTING


Predict Home
Crude Oil Prediction
Enter previous 10th day price
Enter previous 9th day price
Enter previous 8th day price
Enter previous 7th day price
Enter previous 6th day price
Enter previous 5th day price


Enter previous 4th day price
Enter previous 3rd day price
Enter previous 2nd day price
Enter previous 1st day price
Predict
{% if showcase %} {{showcase}} {% endif %}



RESULTS:

9.1 PERFORMACE METRICS

We have used Root Mean Square Error (RMSE) as the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how to spread out these residuals. In other words, it tells you how concentrated the data is around the line of best fit. Root mean square error is commonly used in climatology, forecasting, andregression analysis to verify experimental results.

The formula is

$$RMSE = \sqrt{(f - o)^2}$$

Where:
f = forecasts (expected values or unknown results), o
= observed values (known results).

The bar above the squared differences is the mean (similar to $\bar{x}$). The same formula can be written with the following, slightly different, notation (Barnston, 1992):

$$\mathbf{RMSE}_{fo} = [\sum_{i=1}^{N} (z_{f_i} - z_{o_i})^2 / N]^{1/2}$$

Where:
Σ = summation ("add up") (zfi – Zoi)2 = differences, squared N = sample size.

```
In [27]: train_predict=scaler.inverse_transform(train_data)
         test_predict=scaler.inverse_transform(test_data)
         ### Calculate RMSE performance metrics
         import math
         from sklearn.metrics import mean_squared_error
         math.sqrt(mean_squared_error(train_data,train_predict))

Out[27]: 29.347830443269938
```

ADVANTAGES AND DISADVANTAGES:

10.1 ADVANTAGES

- An RNN remembers each and every information through time where it is useful for time series prediction.
- The main advantage of this approach of **RNN** is that it continuously captures the unstable pattern of the crude oil prices.
- LSTM is capable of learning long term dependencies, especially in sequence prediction problems.
- Minimizing the risk associated with long time volatility in oil prices.
- Price forecasts are very important to various stake holders; government, public and private enterprises, policy makers and investors.
- LSTM units include a 'memory cell' that can maintain information in memory for long period of time.
- RNN can model a collection of record (ie., time collection)so that each pattern can be assumed to be dependent on previous ones.
- This model is highly suggested because investors can use it not only to initiate trades but also has an effective tool.

10.2 DISADVANTAGES

- The previous algorithms does not execute very well solution when the dataset sounds high.
- It will take more time while training the data.
- The changes in the stack prices have the same distribution and on independent of each other.
- ARIMA models are simply checked for their adequacy. However, the disadvantages of this method include the need for a large number of initial data and the absence of a simple method of adjusting the parameters of the model.
- The performance of back propagation relies very heavily on the training data. Back propagation needs a very large amount of time for training.
- The past movement or trend of a stock price or market cannot be used to predict its future movement.
- As the support vector classifier works by placing data points, above and below the classifying hyper plane there is no probabilistic clarification for the classification.

11. CONCLUSION:

In this project, an LSTM (Long Short Term Memory) model is presented with the task of determining the most favourable lag in the crude oil price data. The prediction is accurate till there is a massive and sudden change in the actual data, where it becomes challenging to predict the exact new price with the change, however, the proposed model has efficiently taken into consideration these patterns. Else ways, this also proves the theory that financial markets are unpredictable and change anytime because of known and unknown factors. This work indicates that the LSTM model is an effective tool for crude oil price prediction and can be efficiently used for long-term price forecasting by determining the optimal lags. The proposed model is powerful and highly suggested because investors can use it not only to initiate trades but also as a effective tool to judge various strategies relating to investments.

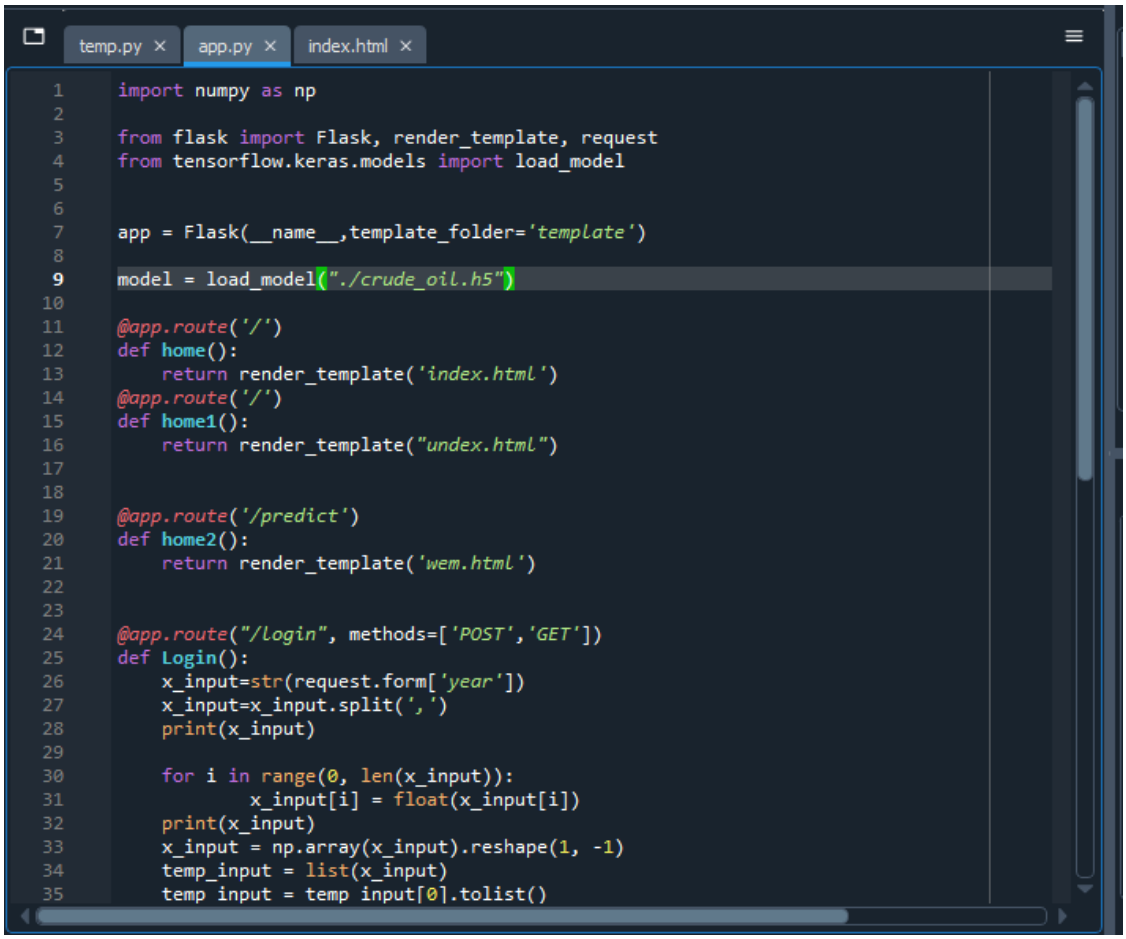
12. FUTURE SCOPE:

In the coming future, fundamental indicators and market trends have been plan to be In-corporated into a model which will help the proposed model perform more efficient.

13.APPENDIX:

SOURCE CODE:

App.py:



C:\Users\ELCOT\Desktop\crude oil\app.py

temp.py × app.py × index.html ×

```
        x_input = np.array(x_input).reshape(1, -1)
        temp_input = list(x_input)
        temp_input = temp_input[0].tolist()
        lst_output = []
        n_steps = 10
        i = 0
        while (i < 1):

                if (len(temp_input) > 10):
                    x_input = np.array(temp_input[1:])
                    print("{} day input {}".format(i, x_input))
                    x_input = x_input.reshape(1, -1)
                    x_input = x_input.reshape((1, n_steps, 1))
                    yhat = model.predict(x_input, verbose=0)
                    print("{} day output {}".format(i, yhat))
                    temp_input.extend(yhat[0].tolist())
                    temp_input = temp_input[1:]
                    lst_output.extend(yhat.tolist())
                    i = i + 1
                else:
                    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))
                    lst_output.extend(yhat.tolist())
                    i = i + 1

        print(lst_output)

        return render_template("predict.html", showcase='The next day predicted value is:
```

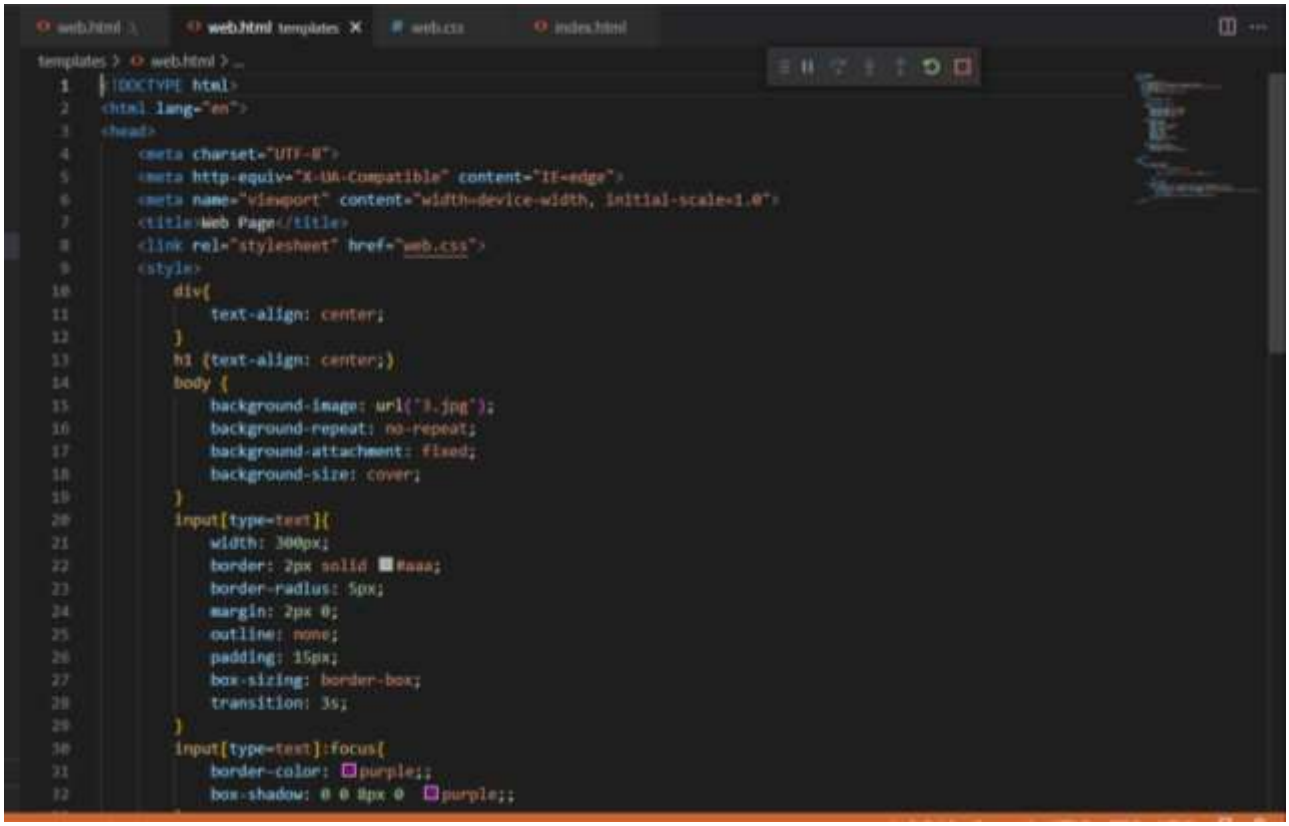
INDEX.HTML:

web.html web.html index.html ×

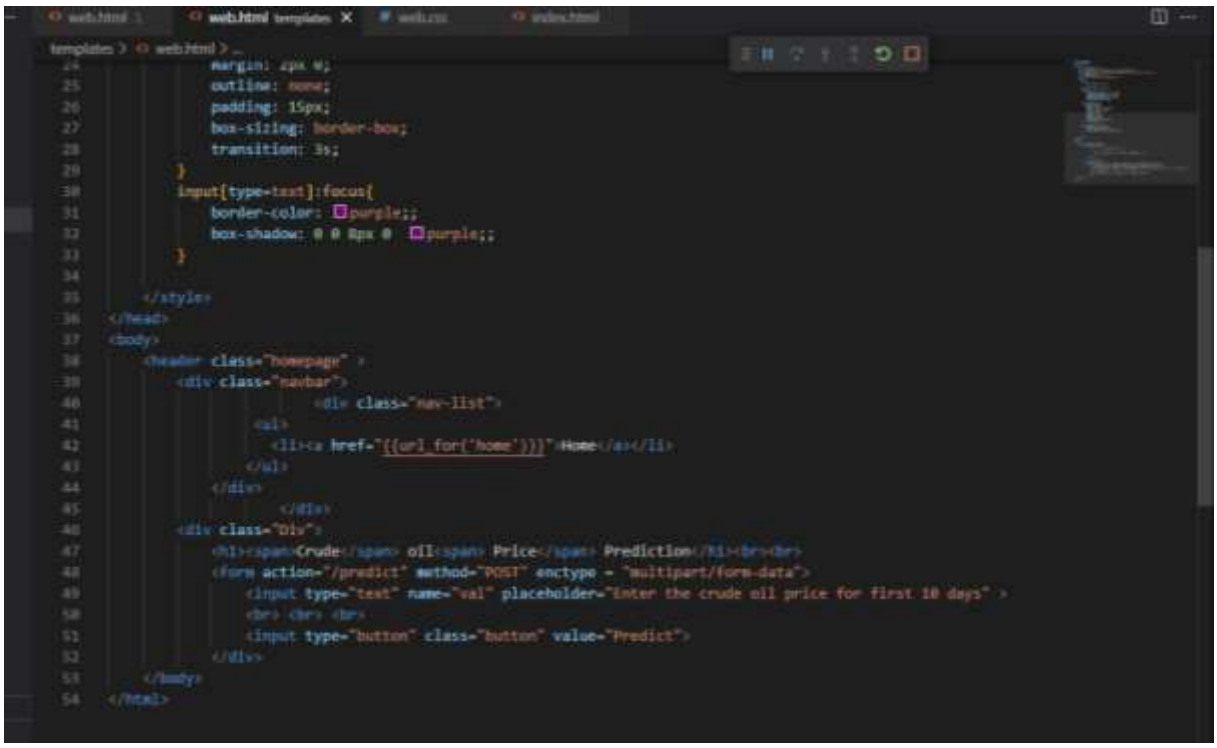
templates > index.html > ...

```
<!DOCTYPE html>
<html>
<head>
<style>
    h1 {text-align: center;}
    p {text-align: center;}
body {
  background-image: url('3.jpg');
  background-repeat: no-repeat;
  background-attachment: fixed;
  background-size: cover;
}
</style>
</head>
<body>
<h1 style="color:black;">Crude oil price prediction</h1>
<p style="color:blue;"> Demand for oil is inelastic, therefore the rise in price is good news
    for producers because they will see an increase in their revenue.</p>
    <p style="color:blue;">Oil importers, however, will experience increased costs of purchasing oil,</p>
    <p style="color:blue;"> Because oil is the largest traded commodity, the effects are quite significant.</p>
        <p style="color:blue;"> A rising oil price can even shift economic/political power from oil importers
            to oil exporters.</p>
            <p style="color: blue;">The crude oil price movements are subject to diverse
                influencing factors</p>
</body>
</html>
```

WEB.HTML:



```
templates > web.html > ...
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta http-equiv="X-UA-Compatible" content="IE=edge">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
    <title>Web Page</title>
    <link rel="stylesheet" href="web.css">
    <style>
        div{
            text-align: center;
        }
        h1 {text-align: center;}
        body {
            background-image: url('3.jpg');
            background-repeat: no-repeat;
            background-attachment: fixed;
            background-size: cover;
        }
        input[type=text]{
            width: 300px;
            border: 2px solid #aaa;
            border-radius: 5px;
            margin: 2px 0;
            outline: none;
            padding: 15px;
            box-sizing: border-box;
            transition: 3s;
        }
        input[type=text]:focus{
            border-color: purple;;
            box-shadow: 0 0 8px 0  purple;;
```



```
            margin: 2px 0;
            outline: none;
            padding: 15px;
            box-sizing: border-box;
            transition: 3s;
        }
        input[type=text]:focus{
            border-color: purple;;
            box-shadow: 0 0 8px 0  purple;;
        }

    </style>
</head>
<body>
    <header class="homepage" >
        <div class="navbar">
                    <div class="nav-list">
                <ul>
                  <li><a href="{{url_for('home')}}">Home</a></li>
                </ul>
            </div>
                    </div>
        <div class="Div">
            <h1><span>Crude</span> oil<span> Price</span> Prediction</h1><br><br>
            <form action="/predict" method="POST" enctype = "multipart/form-data">
                <input type="text" name="val" placeholder="Enter the crude oil price for first 10 days" >
                <br> <br> <br>
                <input type="button" class="button" value="Predict">
            </div>
    </body>
</html>
```

GIT HUB LINK:
https://github.com/IBM-EPBL/IBM-Project-5917-1658819801

DEMO VEDIO LINK:
https://www.youtube.com/watch?v=0Aj4lcSvV0M&feature=youtu.be

GIT HUB LINK:
https://github.com/IBM-EPBL/IBM-Project-13250-1659515191

DEMO VEDIO LINK:
https://www.youtube.com/watch?v=0Aj4lcSvV0M&feature=youtu.be