IBM-Project-PNT2022TMID38581

CRUDE OIL PRICE PREDICTION

Bonafide record of work done by

DIVYA P	**(420419104011)**
JEFFRI MEGDALIN J	**(420419104023)**
KIRUTHIKA S	**(420419104030)**
PRATHISHA K	**(420419104040)**

Professional Readiness for Innovation,

Employability, and Entrepreneurship

BACHELOR OF ENGINEERING

In

COMPUTER SCIENCE AND ENGINEERING

NOVEMBER 2022

ADHIPARASAKTHI ENGINEERING COLLEGE

MELMARUVATHUR-603319

CHAPTER-1
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.

Crude oil price fluctuations have a far reaching impact on global economies and thus price forecasting can assist in minimizing the risks associated with volatility in oil prices. Price forecasts are very important to various stakeholders: governments, public and private enterprises, policymakers, and investors

1.2 Purpose

Crude oil is amongst the most important resources in today'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 prices has become the need of the hour, it is a boon to many large and small industries, individuals, the government.

The evaporative nature of crude oil, its price prediction becomes extremely difficult and it is hard to be precise with the same. Several different factors that affect crude oil prices. We propose a contemporary and innovative method 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 crude oil prices which have been incorporated by finding out the optimal lag and number of the delay effect that controls the prices of crude oil.

Variation of lag in a period of time has been done for the most optimum and close results, we then have validated our results by evaluating the root mean square error and the results obtained using the proposed model have significantly outperformed.

CHAPTER-2
LITERATURE SURVEY

1	**Paper title**	Crude Oil Price Prediction Using Deep Learning
	Problem definition	Predict the crude oil prices and evaluate the model
	Methodology/ Algorithm	Using LSTM(Long Short based on deep learning
	Advantages	• The model is assessed by utilizing the valuable information in the WTI unrefined petroleum markets • The model achieves increments in the expected precision of results
	Disadvantages	The crude oil price depends on several external factors and high volatility
2	**Paper title**	Multi-step-ahead Crude Oil Price Forecasting Based on Autoregressive Integrated Moving Average and Improved Optimization enhanced Gated Recurrent Unit
	Problem definition	Crude oil price volatility has a strong influence on the stability of the global energy market. Therefore, both traders and policy makers have been interested in the accurate forecast of crude oil price so as to prevent large losses and to stabilize the market
	Methodology/ Algorithm	Using ARMA, IPSO, GRU

	Advantages	● To prevent large losses and to stabilize the market ● Crude oil spot prices covering a period of 714 days ● For twenty-step forecasting, the overall reduction of RMSE is as much as 53%, which significantly raises the prediction accuracy.
	Disadvantages	In the overall reduction of RMSE is not an

		100%, only 53% prediction accuracy was raises for 20th step forecasting
3	**Paper title**	A Novel Hybrid Approach with A Decomposition Method and The RVFL Model for Crude Oil Price Prediction
	Problem definition	Volatility of international crude oil prices is influenced by various external factors on different time scales. User search data (USD) which reflects investor attentions has been widely researched and proved to be associated with crude oil price change at different frequency bands.
	Methodology/ Algorithm	Random vector functional link (RVFL) ,Bivariate empirical mode decomposition (BEMD)
	Advantages	● Third, Brent crude oil spot price is used to test the proposed approach empirically ● Forecasting results are analyzed with various evaluation criteria and verified robustness.
	Disadvantages	The proposed approach statistically outperforms traditional forecasting machine learning techniques and similar counterparts (with USD or EMD-based method) in terms of prediction accuracy.

4	**Paper title**	Forecasting Crude Oil Price Using Event Extraction
	Problem definition	Econometric Variable Prediction Problem
	Methodology/ Algorithm	Event Extraction Algorithm
	Advantages	● It uses textual contents and relation between entities ● High quality features
	Disadvantages	Crude oil prices are largely influenced by various factors, such as economic

		development, financial markets, conflicts, wars, and political events.
5	**Paper title**	Predictive Analytics for Crude Oil Price Using RNN-LSTM Neural Network
	Problem definition	This study aims to certify the capability of a prediction model built based on the RNN LSTM network to predict the future price of crude oil.
	Methodology/ Algorithm	Using Recurrent Neural Network and Long Short Term Network
	Advantages	● The capability of the network to provide an improvement of the accuracy of crude oil price prediction ● Millions of traders investing the crude oil prediction
	Disadvantages	● It dynamic nature ● It complex to predict the price of crude oil

CHAPTER-3
IDEATION & PROPOSED SOLUTION

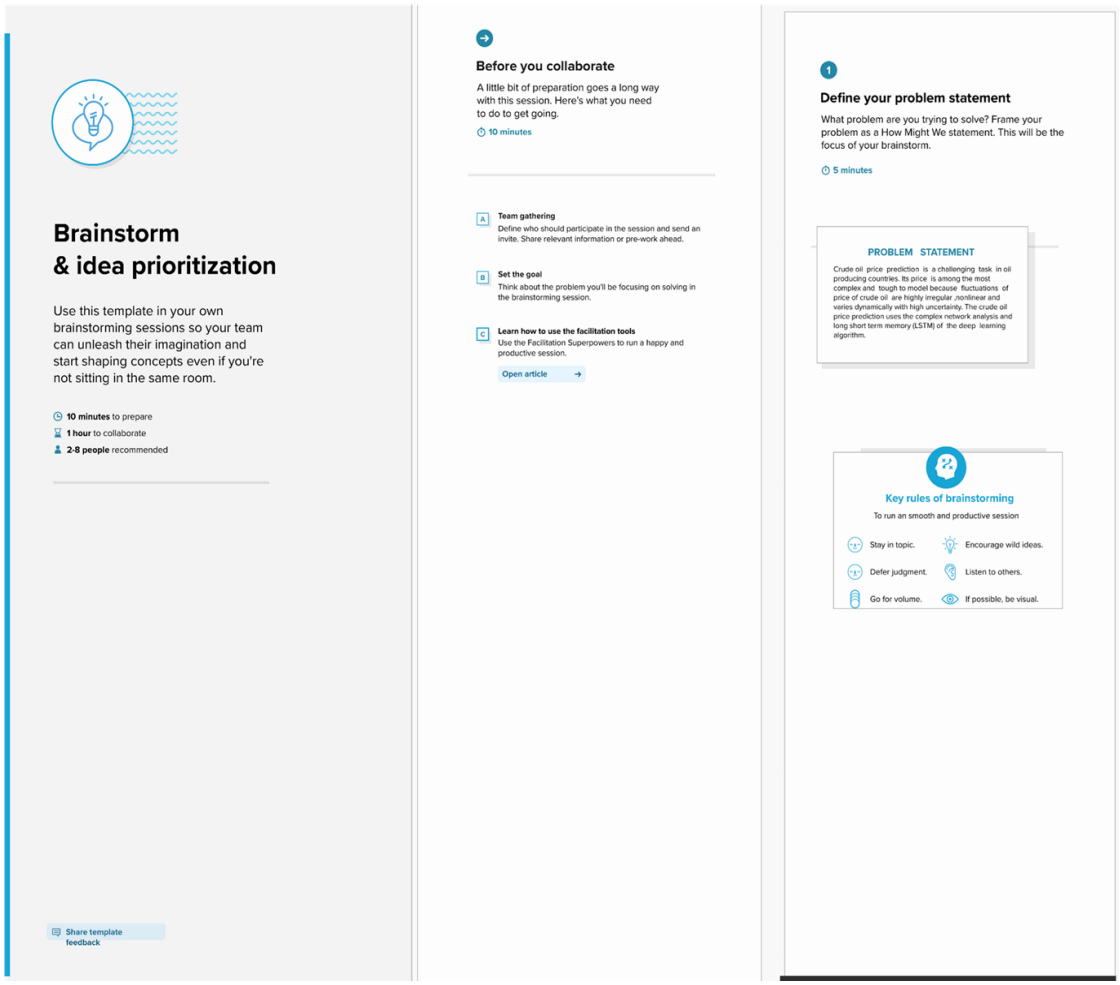
3.1 Empathy Map Canvas

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user's behaviors and attitudes. It is a useful tool to help teams better understand their users. Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user's perspective along with his or her goals and challenges**.**

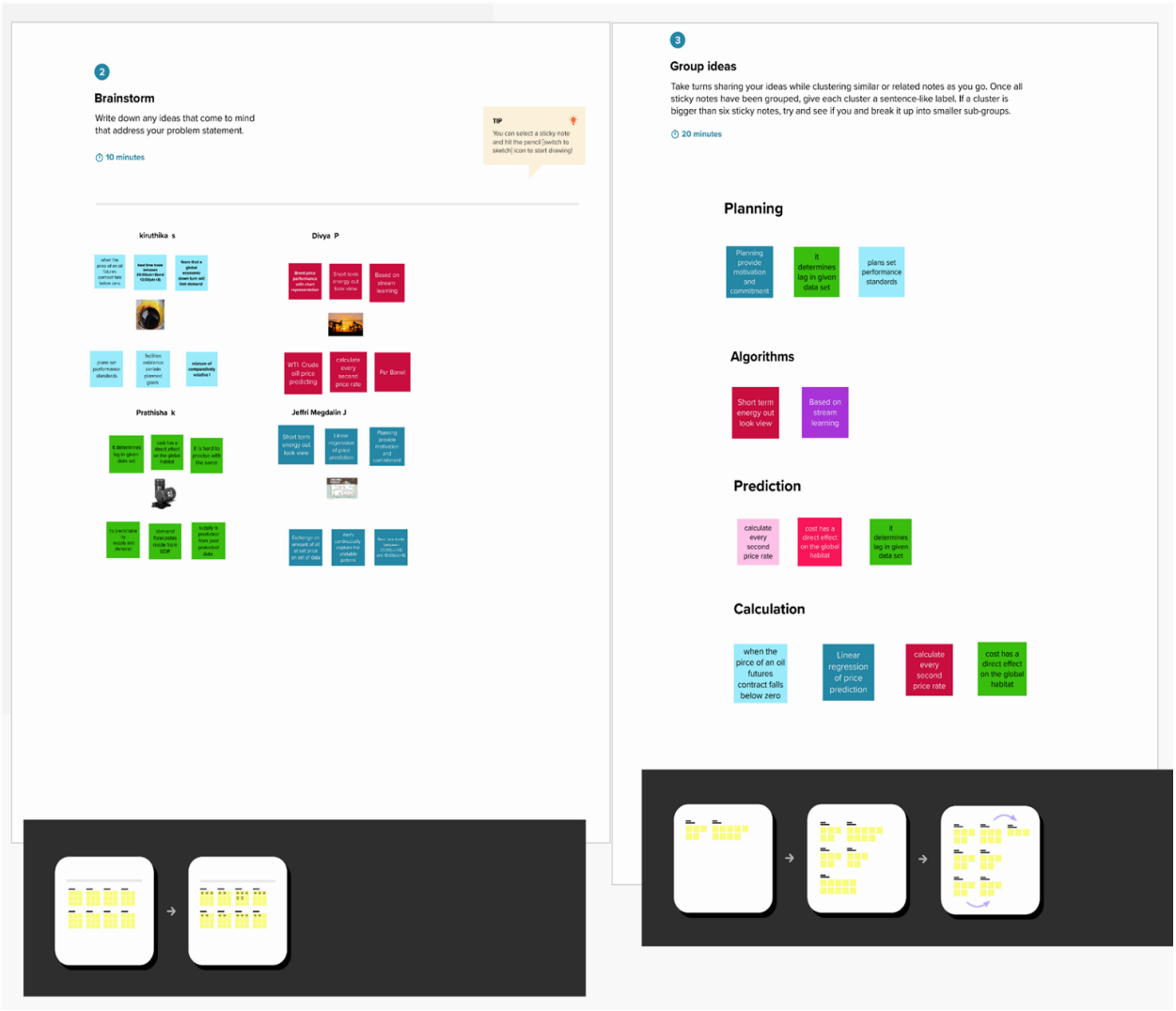
3.2 Ideation & Brainstorming

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

Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization

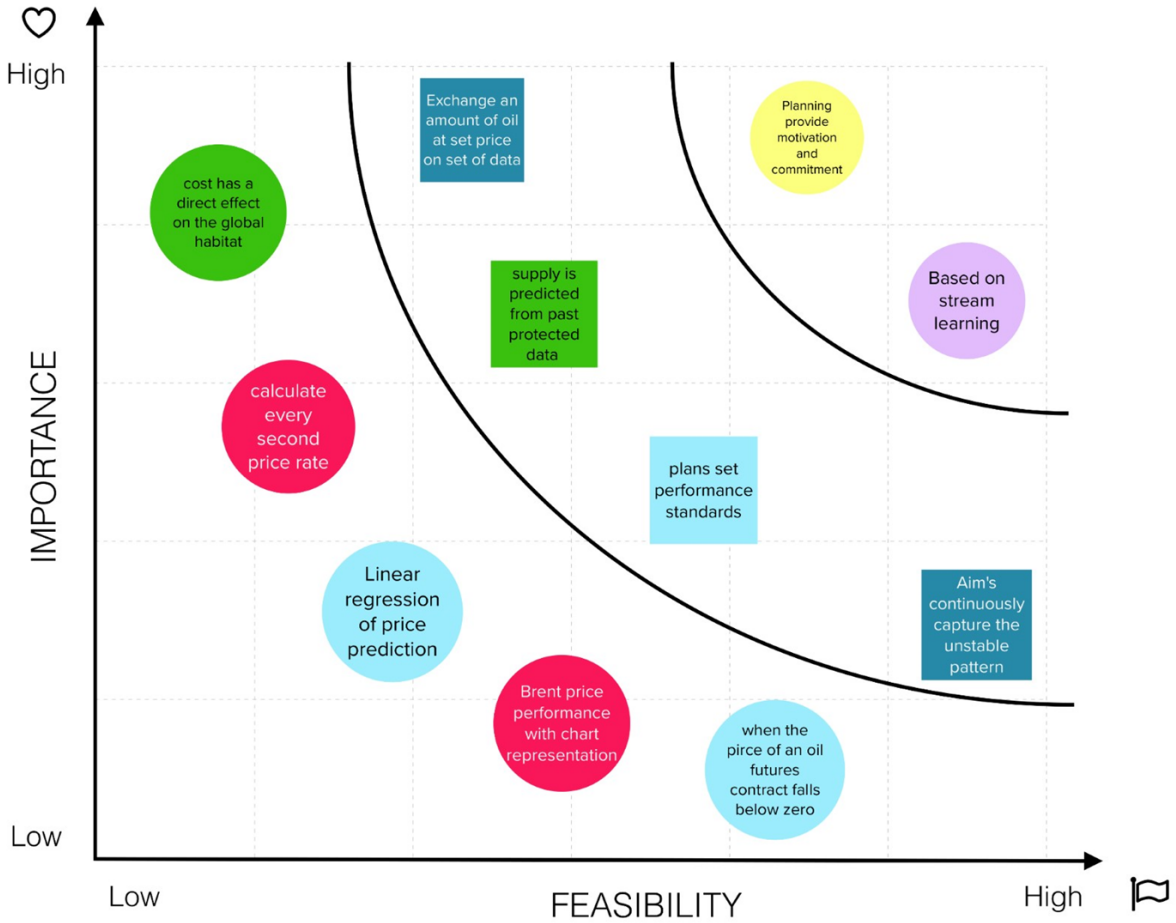
4

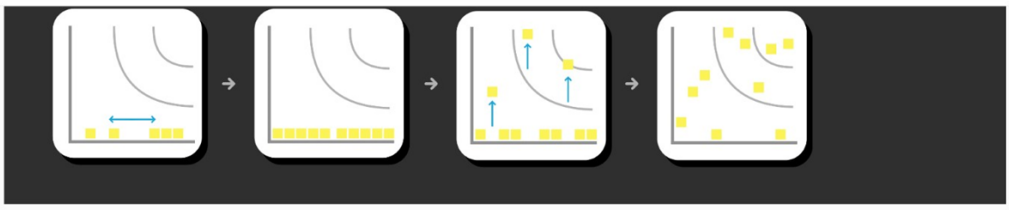
Prioritize

Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.

20 minutes

Idea Prioritization



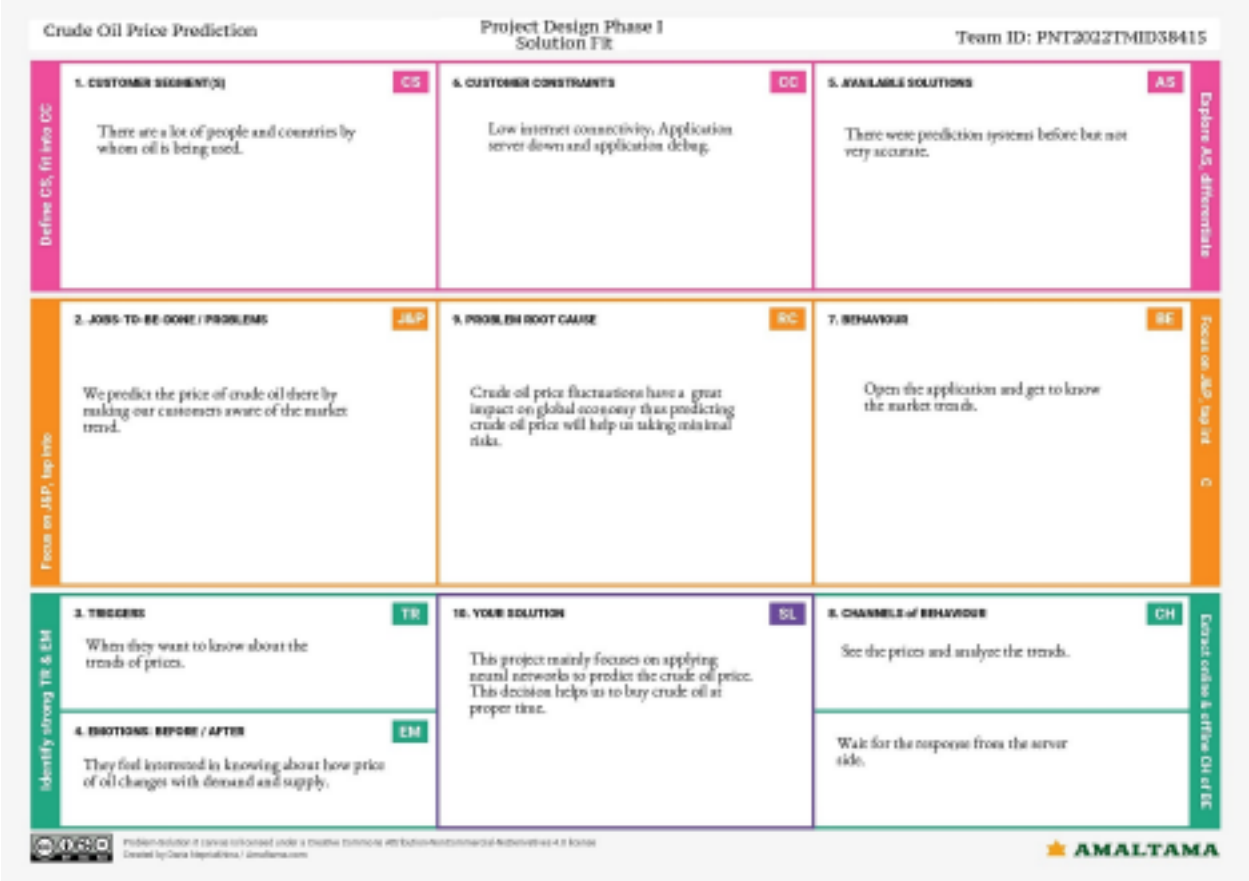


3.3 Proposed Solution

S.No	Parameter	Description
1	**Problem Statement (Problem to be solved)**	Oil price increases are generally thought to increase inflation and reduce economic growth. Oil prices directly affect the prices of goods made with petroleum products. As mentioned above oil prices indirectly affect costs such as transportation, manufacturing, and heating.
2	**Idea / Solution description**	Crude oil is a raw natural resource that is extracted from the earth and refined into products such as gasoline and petroleum products. Crude oil is a global trade in markets around the world.
3	**Novelty / Uniqueness**	Supply, demand, and sentiment towards oil futures contracts, which are traded heavily by speculators play a dominant role in price determination.
4	**Social Impact / Customer Satisfaction**	Oil spills can damage the environment and the wildlife and marine life that depends on it they can also cause physical, mental, and financial stress to people as individuals.
5	**Business Model (Revenue Model)**	Models general includes information like products or services the business plans to sell, target markets, and any anticipated expenses. There are dozens of types of business models including retailers, manufacturers, fee-for services, or freemium providers.
6	**Scalability of the Solution**	Crude oil prices are determined by global supply and demand. Economic growth is one of the biggest factors affecting

		retailers. The impact on crude oil prices can name some of them as the US economy, US dollar exchange, supply and demand statistics, and crude oil and petroleum distillates inventory.

3.4 Problem Solution Fit



CHAPTER-4
REQUIREMENT ANALYSIS

4.1 Functional Requirement

Following are the functional requirements of the proposed solution.

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Homepage	Registration through Form Registration through Gmail Registration through LinkedIN
FR-2	Prediction	The User can enter the required data and the predicted price of the crude oil will be displayed.
FR-3	Display	The predicted graph will be displayed in the user's screen.

4.2 Non-functional Requirements:

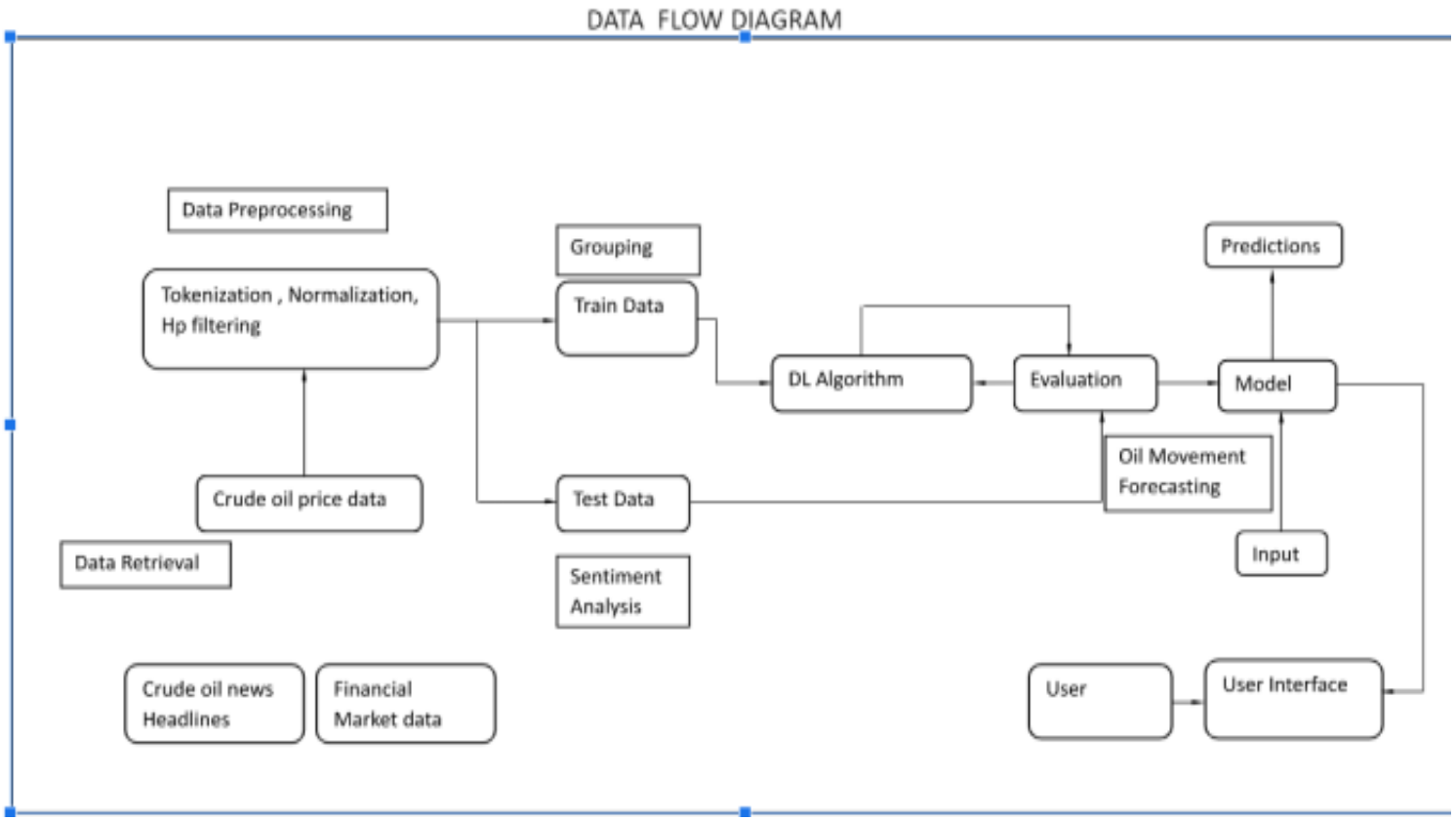
Following are the non-functional requirements of the proposed solution. FR

FR No.	Non-Functional Requirement	Description
NFR-1	Usability	The predictor is user friendly and easy to use it.
NFR-2	Security	The model is tested for it security before the deployment and it is highly secure.
NFR-3	Reliability	Highly reliable.
NFR-4	Performance	It is efficient and optimized method to predict the crude oil price.
NFR-5	Availability	Accessible at any time.
NFR-6	Scalability	It will perform well for many number of users with the same speed.

CHAPTER-5
PROJECT DESIGN

5.1 Data Flow Diagram

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



User Stories

Use the below template to list all the user stories for the product.

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 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 my Account	Low	Sprint-2
		USN-4	As a user,I can register for the application through Gmail	I can register through I already logged into my 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 only email & password.	High	Sprint-1
	Line\Bar graph		After entering the inputs,the model will display predictions inLine\BarGraph Format.	I can get the expected prediction in various formats.	High	Sprint-3

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 Line graph / Bar graph.	High	Sprint-1
		USN-2	As a user,I will receive confirmation email once I have registered for the	I can receive confirmation email & click confirm	High	Sprint-1

		application			
		USN-3 As a user,I can register for the application through Facebook	I can register & access the my Account	Low	Sprint-2
		USN-4 As a user,I can register for the application through Gmail	I can register through I already logged into my 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 only email & password.	High	Sprint-1
	Line\Bar graph	After entering the inputs,the model will display predictions inLine\BarGraph Format.	I can get the expected prediction in various formats.	High	Sprint-3
Customer (Web user)	Login	USN-1 As a web user,Can login simply by using a Gmail or Facebook account.	Already created 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.	I can solve the problems arised by Support.	Low	Sprint-3

Administrator	News	Admin will give the recent news of Oil Prices.	Provide the recent oil prices.	High	Sprint-4
	Notification	Admin will notify when the oil prices change.	Notification by Gmail.	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.	Stores User details.	High	Sprint-4

5.3 Solution & Technical Architecture

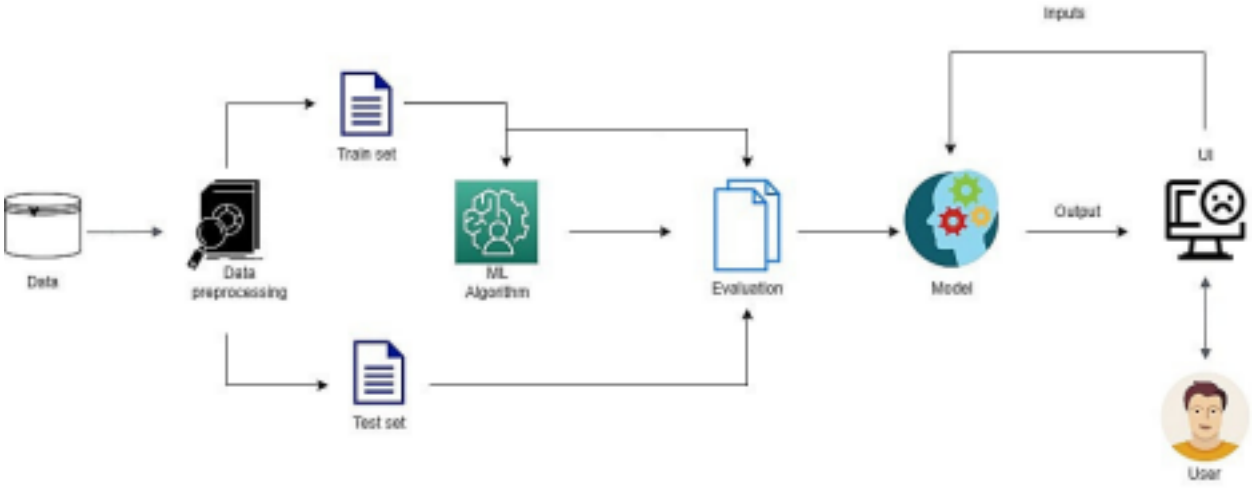
A Solution architecture (SA) is an architectural description of a specific solution. SAs combine guidance from different enterprise architecture viewpoints (business, information and technical), as well as from the enterprise solution architecture (ESA). Ultimately, solution architecture is aimed at the following overarching goals:

(i) Streamlining of day-to-day activities

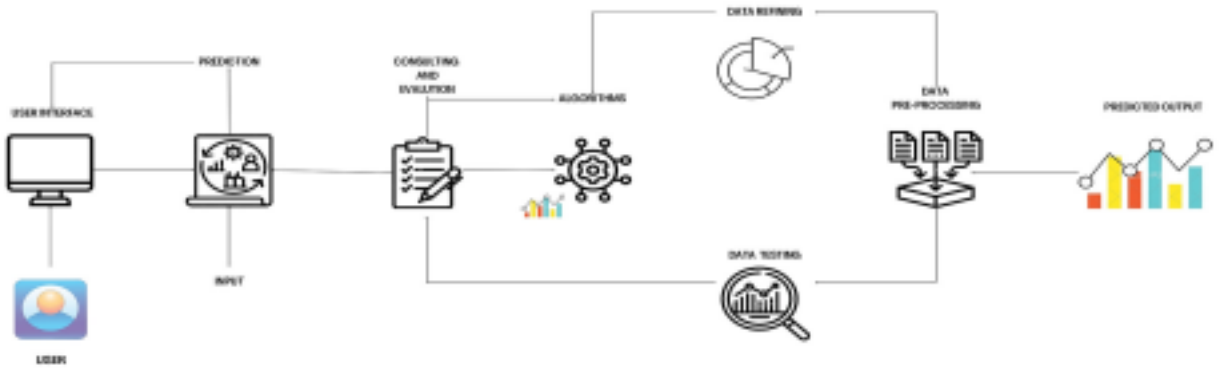
(ii) Providing a more efficient production environment

(iii) Lowering costs and gaining cost-effectiveness

(iv) Providing a secure, stable,and supportable environment



SOLUTION ARCHITECTURE



TECHNICAL ARCHITECTURE

CHAPTER-6
PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring the, technical papers, research publications etc.	28 SEPTEMBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	24 SEPTEMBER 2022
Ideation	List the by organizing the brainstorming session and prioritize the top 3 ideas based on the feasibility & importance.	25 SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	23 SEPTEMBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	30 SEPTEMBER 2022
Solution Architecture	Prepare solution architecture document.	28 SEPTEMBER 2022

6.2 Sprint Delivery Schedule

Use the below template to create product backlog and sprint schedule

Product Backlog, Sprint Schedule, and Estimation (4 Marks) Use the below template to

create product backlog and sprint 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	DIVYA P
Sprint-1	Login	USN-2	As a user, I will receive confirmation email once I have registered for the application	10	High	JEFFRI MEGDALIN J
Sprint-2	Input Necessary Details	USN-4	As a user, I can give Input Details to Predict Likeliness of crude oil	9	High	PRATHISHA K
Sprint-2	Data Pre-processing	USN-5	Transform raw data into suitable format for prediction.	8	High	JEFFRI MEGDALIN J
Sprint-3	Predict	USN-6	As a user, I can predict Crude oil using a machine learning model.	9	High	DIVYA P
Sprint-3		USN-7	As a user, I can get accurate prediction of crude oil	8	Medium	PRATHISHA K

Project Tracker, Velocity & Burndown Chart: (4 Marks)

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	20	6 Days	24 Oct 2022	29 Oct 2022	20	29 Oct 2022
Sprint-2	18	6 Days	31 Oct 2022	05 Nov 2022	18	31 Oct 2022
Sprint-3	17	6 Days	07 Nov 2022	12 Nov 2022	17	07 Nov 2022
Sprint-4	17	6 Days	14 Nov 2022	19 Nov 2022	17	14 Nov 2022

Velocity:

Imagine we have a 10-day sprint duration, and the velocity of the team is 20 (points per sprint). Let's

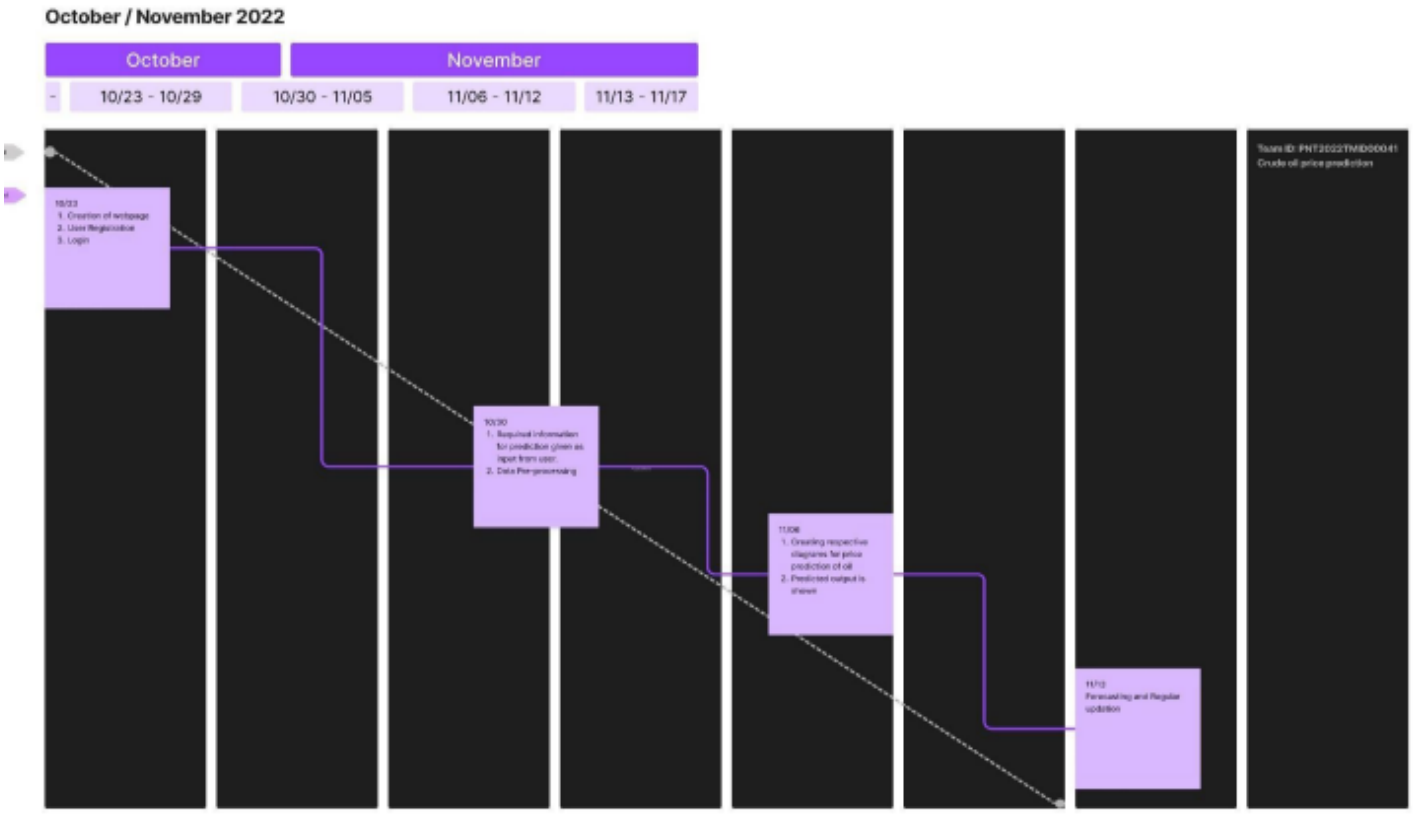
calculate the team's average velocity (AV) per iteration unit (story points per day)

$$AV = \frac{sprint\ duration}{velocity} = \frac{20}{10} = 2$$

Sprint 1 AV = sprint duration / velocity = 20/6 = 3.33 **Sprint 3 AV = sprint duration / velocity = 17/6 = 2.63**
Sprint 2 AV = sprint duration / velocity = 18/6 = 3 **Sprint 4AV = sprint duration / velocity = 17/6 = 2.6**

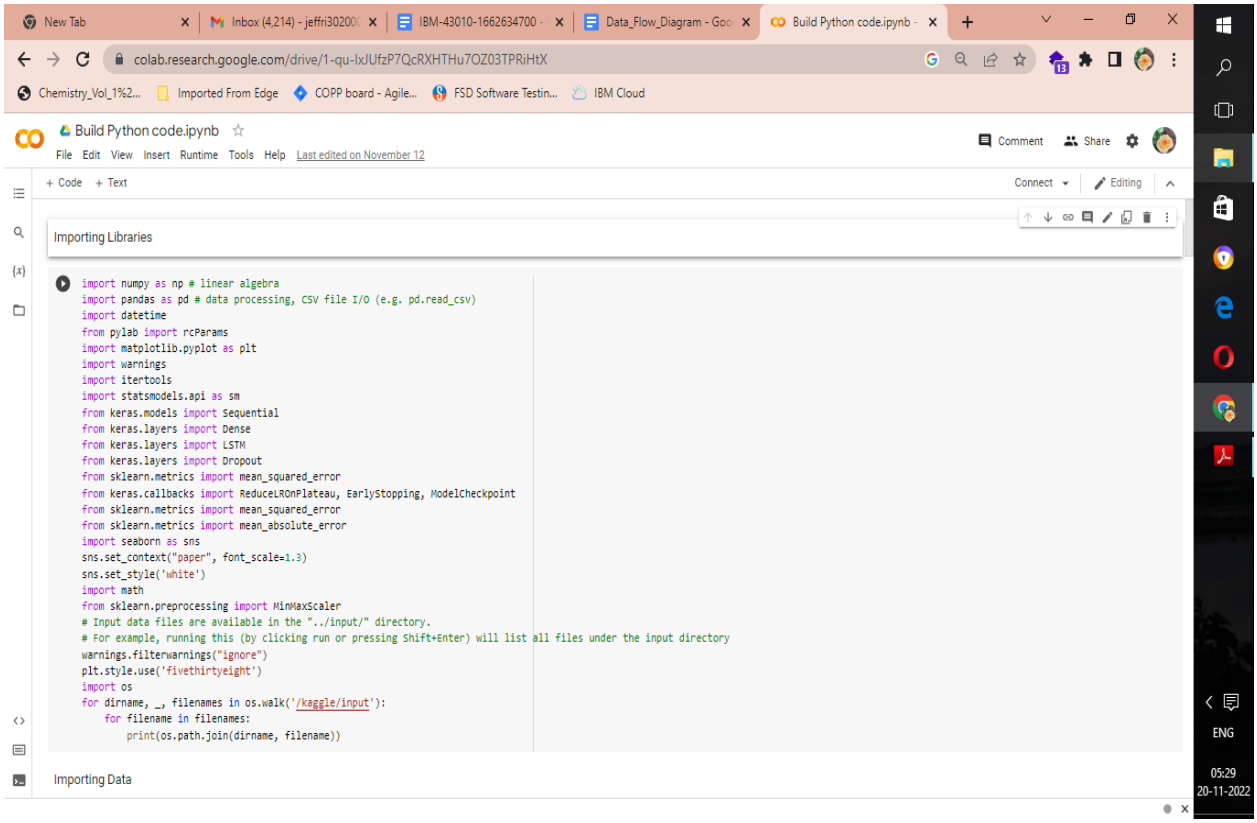
Burndown Chart:

A burndown chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such scrum.However,burndown chart can be applied to an project containing measurable progress overtime.



CHAPTER-7
CODING & SOLUTIONING

7.1 Feature 1



CHAPTER-9

RESULTS

9.1 Performance Metrics

Input

Input measures monitor the amount of resources being used to develop, maintain, or deliver a product, activity or service. Examples include:

- Money spent on equipment
- Number of employee hours worked

- Number of vehicles
- Facility costs
- Total operating expenditures
- Rental fees
- Number of full-time employees

Output

Output measures monitor “how much” was produced or provided. They provide a number indicating how many items, referrals, actions, products, etc. were involved. Examples include:

- Number of permits issued
- Number of pavement miles resurfaced
- Number of people trained
- Number of water leaks fixed
- Number of cases managed
- Number of arrests made
- Number of documents processed
- Number of clients served

Efficiency

Efficiency measures are used to monitor the relationship between the amount produced and the resources used. This means that efficiency measures are created by comparing input and output, see . There are two general types of efficiency measures: unit cost and productivity.
Unit cost is a comparison of an input to an output (i.e. resources used/number produced).
Productivity is a comparison of an output to an input (i.e. number produced/resources used). Examples include:

Unit Cost

- Cost per license issued
- Cost per employee taught
- Cost per lane-mile paved
- Cost per client served
- Cost per document

Productivity

- Licenses processed per employee-hour
- Units produced per week
- Students taught per instructor

● Cases resolved per agent
● Calls handled per hour

Quality

Quality measures are used to determine whether customer expectations are being met. These expectations can take many forms, including: timeliness, accuracy, meeting regulatory requirements, courtesy, and meeting customer needs.

The expectations can be identified as a result of internal or external feedback.

The comparison of outputs is often used to create measures of quality. It may be important to identify certain aspects (aspects / total outputs) about the services, products or activities produced by an organization that are important to its customers.

This comparison of specific outputs to total outputs is used to create measures of accuracy, timeliness and to determine the extent regulatory requirements are met.

Quality measures can also be derived from the evaluation of customer feedback data.

Timeliness

● Busy signal rate
● Percent of drivers licenses issued within one hour.

Accuracy

● Percent of applications requiring rework due to internal errors.
● Taxpayer error rate on tax returns.

Requirements

● Percent of wells meeting minimum water quality requirements.
● Percentage of clients that rated themselves as successfully rehabilitated.

CHAPTER-10

ADVANTAGES & DISADVANTAGES

ADVANTAGES:

Crude oil generates heat. Heating this material and other petroleum products can warm homes in colder weather, making modern living possible even in colder climates. This oil also produces energy.
This product makes machinery move by providing the lubrication oil that modern industrial equipment depends on to run smoothly.
Crude is also used to create the asphalt that cars and trucks move on. Asphalt is easier to lay than concrete and is generally cheaper as a result.
Crude is also an integral part of modern textile production, with 40 percent of textiles now containing some petroleum by product.

DISADVANTAGES:

- Oil is a non-renewable source of energy.
- Burning oil produces carbon dioxide gas.
- Burning oil can pollute the air.
- Much of our oil has to be imported and it is becoming more and more expensive as reserves reduce and imports increase

CHAPTER-11

CONCLUSION

- Crude oil prices have fluctuated over time series. It is affected by many economic and political factors.Specially, there are several sudden increases and decrease throughout the time.
- In order to eliminate the irregular trend. We try several methods, hp filter, loess filter, log transformation and difference.

- Thus, only analysis of the crude oil price itself can hardly predict the sudden 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.

CHAPTER-12

FUTURE SCOPE

It shows that the prediction accuracy of the variable selection-machine learning integrated model is significantly improved compared with that of the univariate model.

we may introduce more independent variables with the help of internet search data, test our framework performance.
Moreover, investor sentiment can be quantified in this process. In addition, different variable selection methods can be introduced. This indicates that the variable selection-based machine learning integrated research framework proposed in this significantly improves the forecasting performance of oil prices

Github Project:
https://github.com/IBM-EPBL/IBM-Project-17840-1659676677

Youtube link:https://youtu.be/XBt73U1lpVA

CHAPTER-13

APPENDIX

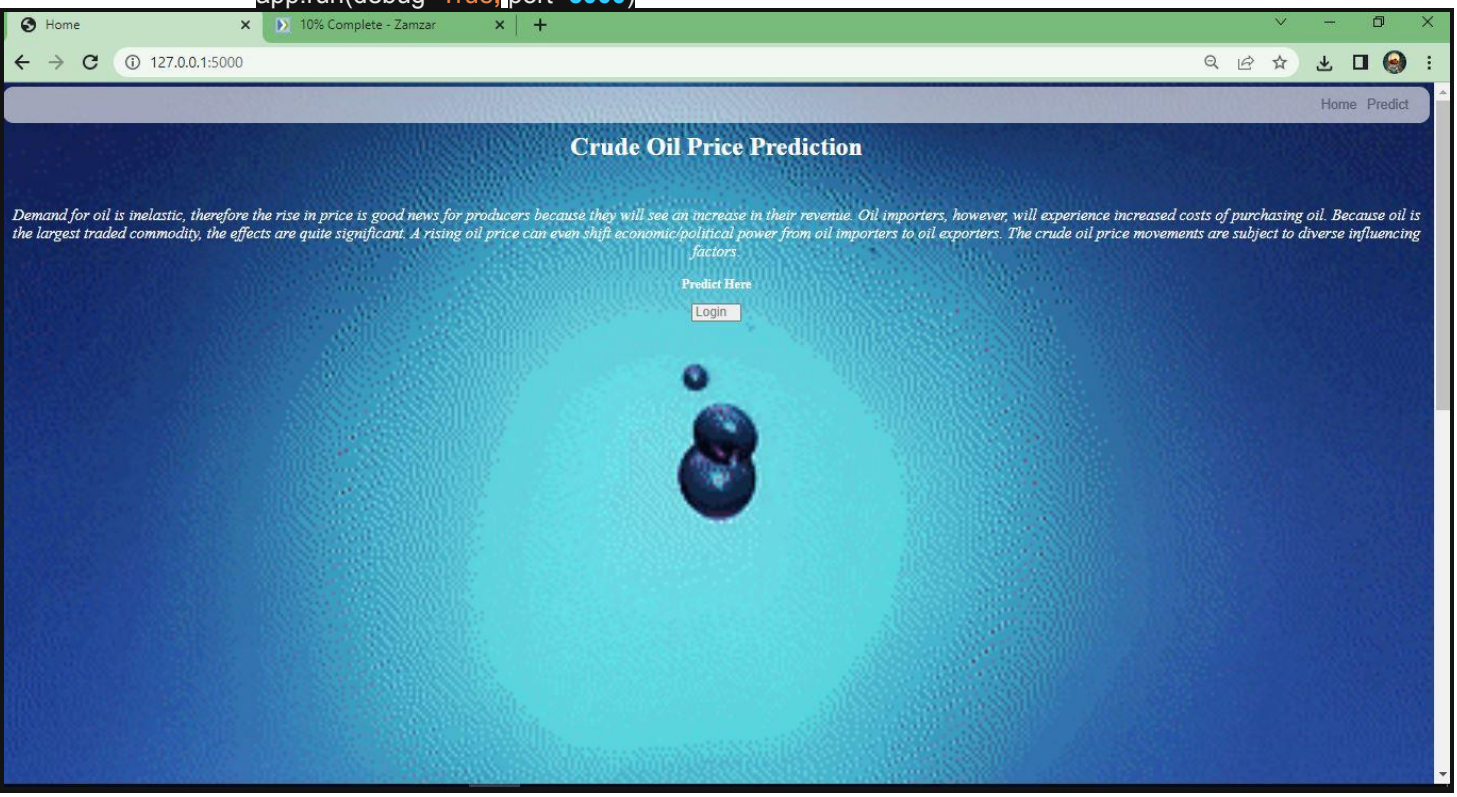
SOURCE CODE:

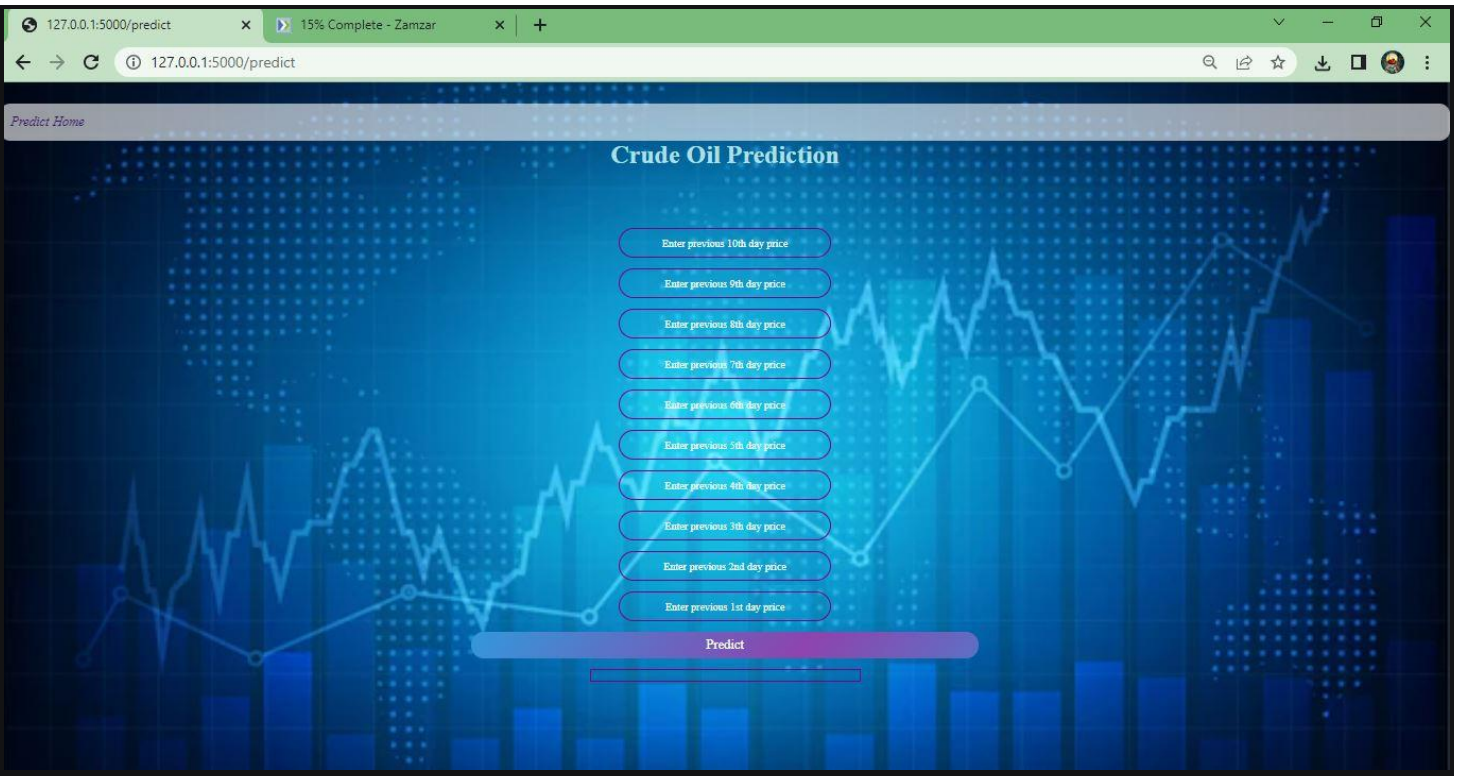
APP.PY

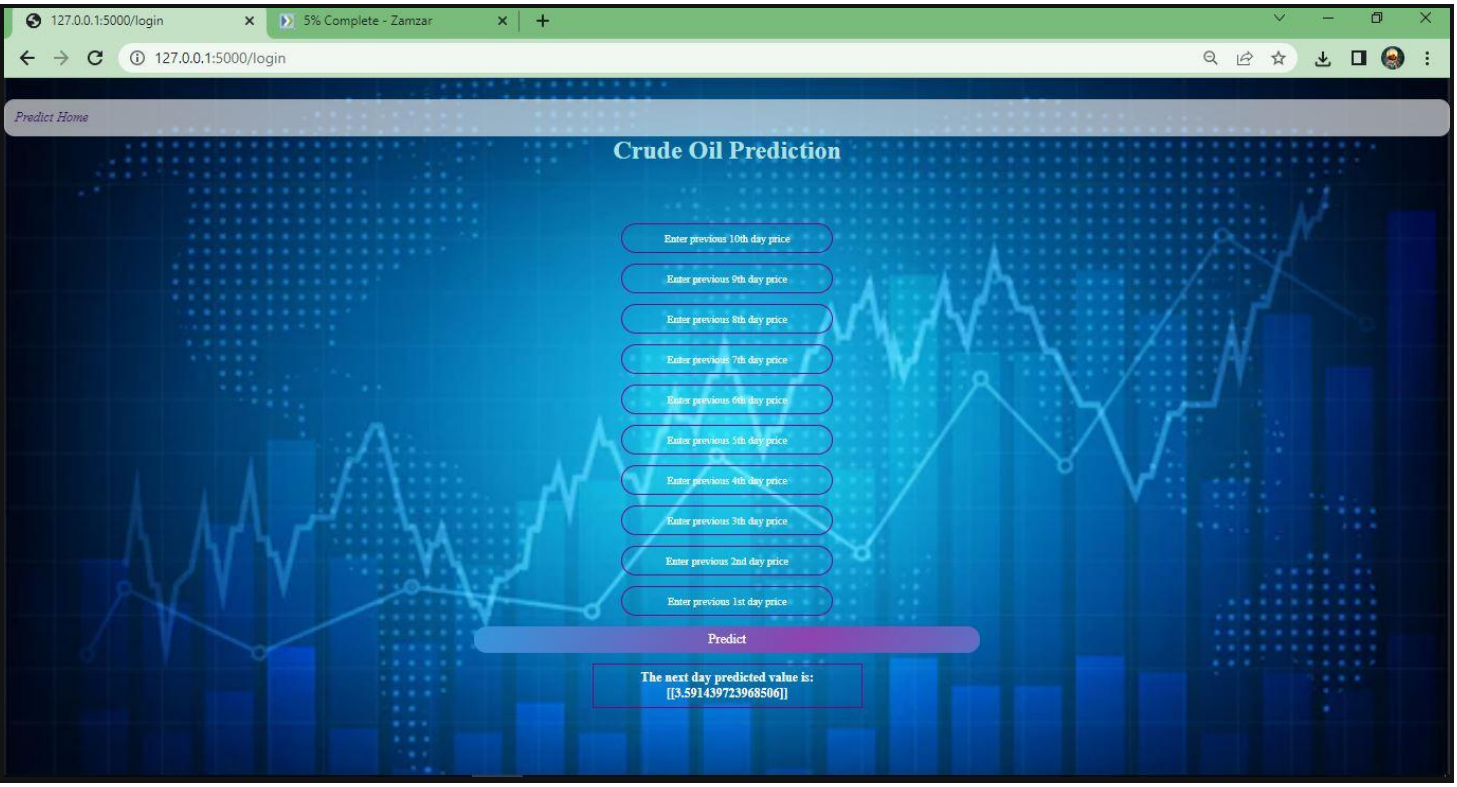
```
import numpy as np

from flask import Flask, render_template, request
from tensorflow.keras.models import load_model


app = Flask(__name__,template_folder='template')

model = load_model("./model/crude_oil.h5")

@app.route('/')
def home():
    return render_template('index.html')
```

```
@app.route('/predict')
def home2():
        return render_template('predict.html')


@app.route("/login", methods=['POST','GET'])
def Login():
        if request.method=='POST':
                a = request.form['year1']
                b = request.form['year2']
                c = request.form['year3']
                d = request.form['year4']
                e = request.form['year5']
                f = request.form['year6']
                g = request.form['year7']
                h = request.form['year8']
                i = request.form['year9']
                 j = request.form['year10']
                x_input = [a, b, c, d, e, f, g, h, i, j]
                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 = 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:' +
str(lst_output))



if __name__ == '__main__':


        app.run(debug=True, port=5000)
```




127.0.0.1:5000/predict
15% Complete - Zamzar
127.0.0.1:5000/predict
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 3th day price
Enter previous 2nd day price
Enter previous 1st day price
Predict


127.0.0.1:5000/login
5% Complete - Zamzar
127.0.0.1:5000/login
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 3th day price
Enter previous 2nd day price
Enter previous 1st day price
Predict
The next day predicted value is:
[[3.591439723968506]]