PROJECT REPORT

Crude oil price prediction

Team ID:	PNT2022TMID53264	**Project ID:**	17241-1659632038
Team Members:	Karthik varunn.S Manasi.R Sayeeshwar.K Vedh Vjay.V	**Roll Numbers:**	195002059 195002070 195002104 195002125
Industry Mentor(s):	Shanthi	**Faculty Mentor**:	Dr. P.Vasuki

1. **INTRODUCTION**

1.1 Project Overview

There are millions of people who invest/trade in the stock market. Many people make a living out of trading daily, this requires them to make very careful trades with minimal risks to avoid incurring huge losses. Price actions of stocks heavily depend on many factors. One major factor is the crude oil prices, recently we also witnessed how the stock market suffered because of a dip in the crude oil prices as a consequence of the Russia-Ukraine war. Traders/Investors in the stock market or the commodity market need an effective and accurate tool to predict this volatility in the crude oil prices.

1.2 Purpose

Crude oil is one of the most essential commodities for daily living, and its price fluctuations have a significant effect on the environment globally. Governments, businesses, and individuals can all benefit from price predictions. Continuous use of statistical and econometric methods, including AI, for predicting crude oil prices may show declines in forecast accuracy. In order to purchase crude oil at the best time, the price may be used as a time analysis prediction. The industries can enhance their profit rate as a result.

2. LITERATURE SURVEY

2.1 Existing problem

The most important strategic resource on the planet, crude oil is the "key" commodity for the global economy. As a result, it has been challenging to anticipate since a variety of factors influence its price, making it difficult to predict. Crude oil's pricing is extremely unstable and variable. Numerous studies have lately been carried out to analyze the challenges of oil price forecasting and identify the best solutions. Anticipating its demands will be advantageous for our government, businesses, and investors. Long short-term memory, RNN(LSTMs) will be constructed as part of this research to estimate crude oil prices. In this paper, we propose an innovative analytical approach for forecasting the price of crude oil. Keywords: Economy, energy, gasoline, pricing, and crude oil.

2.2 References

1. Yu Runfang, Du Jiang and Liu Xiaotao, “Improved Forecast Ability of Oil Market Volatility Based on combined Markov Switching and GARCH-class Model,Procedia Computer Science, vol. 122, pp. 415-422, 2017.
2. K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber,"LSTM: A Search Space Odyssey," IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 10, pp. 2222-2232, Oct. 2017.
3. Mohammad Reza Mahdiani and Ehsan Khamehchi, “A modified neural network model for predicting the crude oil price”, Intellectual Economics, vol. 10, no. 2, pp. 71-77, Aug. 2016.
4. Manel Hamdi and Chaker Aloui, "Forecasting Crude Oil Price Using Artificial Neural Networks: A Literature Survey," Economics Bulletin, AccessEcon, vol. 35,no. 2, pp. 1339-1359, 2015.
5. Aloui, Chaker & Hamdi, Manel. (2015). Forecasting Crude Oil Price Using Artificial Neural Networks: A Literature Survey. Economics Bulletin. 35. 1339-1359.

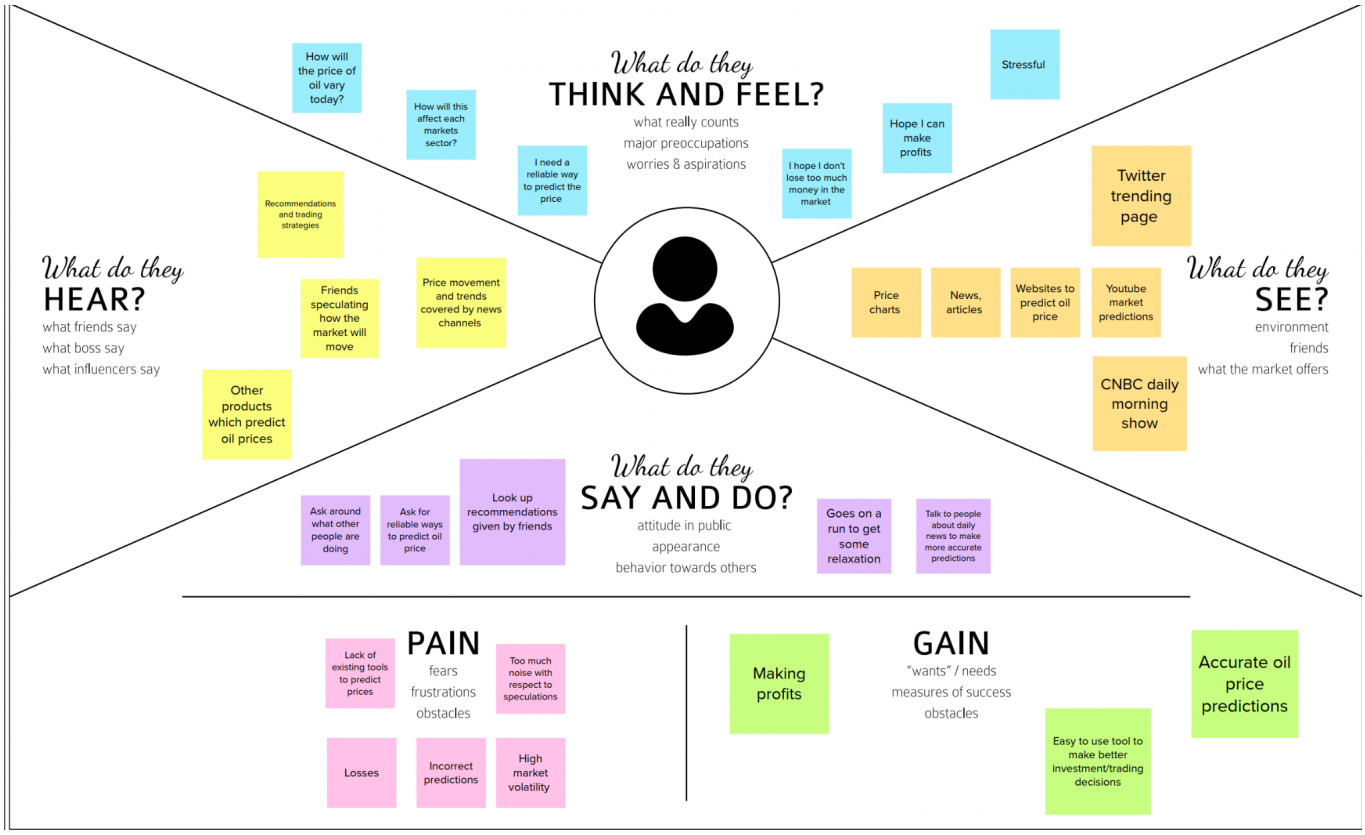
2.3 Problem Statement Definition

Academic academics are starting to use the usual statistical and econometric approaches among the various forecasting models created to anticipate the price of "black gold." Amano offers the first study on oil market forecasts(1987). To forecast the oil market, the author employed a small-scale econometric model. Huntington (1994) used an advanced econometric model to forecast the price of oil in the 1980s. Gulen (1998) used cointegration analysis in a different study to forecast the price of WTI crude oil. Barone-Caddesi et al. (1998) proposed a semi-parametric approach based on the filtered historical simulation technique to predict the price of oil. Morana (2001) used a semi-parametric technique based on the GARCH features of the volatility of oil prices, which were studied by Barone-Caddesi et al (1998) suggested a semi-parametric

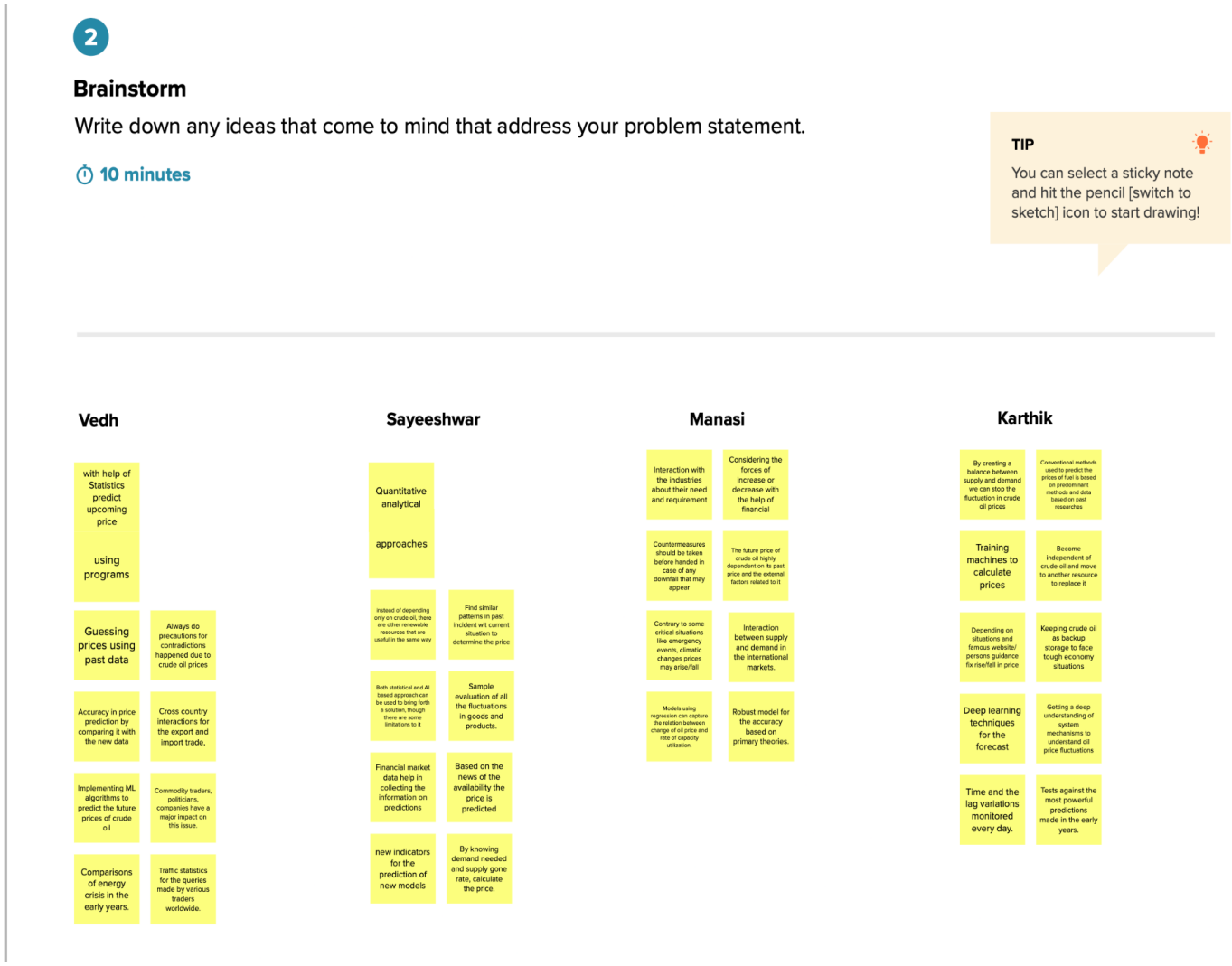
approach based on the filtered historical simulation technique to forecast oil prices. Based on the GARCH properties of the oil price volatility, Morana (2001) employed a semi-parametric approach investigated by Date 10 September 2022 Project Name Crude Oil Prediction Maximum Marks 4 Marks Barone-Caddesi et al. (1998) to short-term Brent crude oil price forecast. In another work, Tang and Hammoudeh (2002) utilized a nonlinear regression to predict OPEC basket prices. Using OECD petroleum inventory levels and relative stock inventories, Ye et al. (2002, 2005) adopted a simple linear regression model for short-term monthly prediction of WTI crude oil spot price. In a related study, Ye et al. (2006) included nonlinear variables such as low- and high-inventory variables to the linear forecasting model suggested by Ye et al. (2002, 2005) to predict short-run WTI crude oil prices. Zamani (2004) used an econometrics forecasting model to anticipate the short-term quarterly WTI crude oil spot price using OECD stocks, non-OECD demand, and OPECsupply. Using error correction models, Lanza et al. (2005) looked at the pricing of products and crude oil. Sandusky (2006) used GARCH, TGARCH, AR, and BIGARCHstatistical models, among others, to forecast daily volatility in petroleum futures price returns. To predict oil demand, supply, and prices, Dees et al. (2007) created a linear model of the global oil market with a primary focus on OPEC behavior. Murat and Tokat (2009) looked into the connection between futures and spot crude oil prices and used the random walk model to test if futures prices might predict changes in spot prices. However, more recent research has used GARCH and several models from the GARCH family to forecast oil prices. For instance, the GARCH model was employed by Narayan and Narayan (2007) and Agnolucci (2009) to forecast spot and futures crude oil prices. In a related study, Mohammadi and Su (2010) investigated the crude oil price-predicting outcomes of various GARCH-type models. CGARCH, FIGARCH, and IGARCH modelswere suggested by Kang et al. (2009) to predict the volatility of crude oil markets. Wei et al. (2010) enhanced the work of Kang et al. (2009) towards the same goal by using linear and nonlinear GARCH-class models. As a result of the application of linear techniques, a sizable difference between the projected and real price of oil has been demonstrated. The most often utilized exogenous variables in these models for predicting oil prices are inventories, supply, and demand. The fact that supply and demand are relatively inelastic to price changes and that inventory adjustments can take time to materialize account for a considerable share of the difference between actual and predicted prices, especially in the near run (Hamilton, 2008). However, traditional statistical and economic techniques frequently only detect linear processes in data. data time series. (Weigend and Gershenfeld, 1994). However, the oil price behavior is characterized by high nonlinearity and irregularity. Therefore, the mentioned models are not the appropriate choice to forecast the oil price.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming



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.

20 minutes

TIP

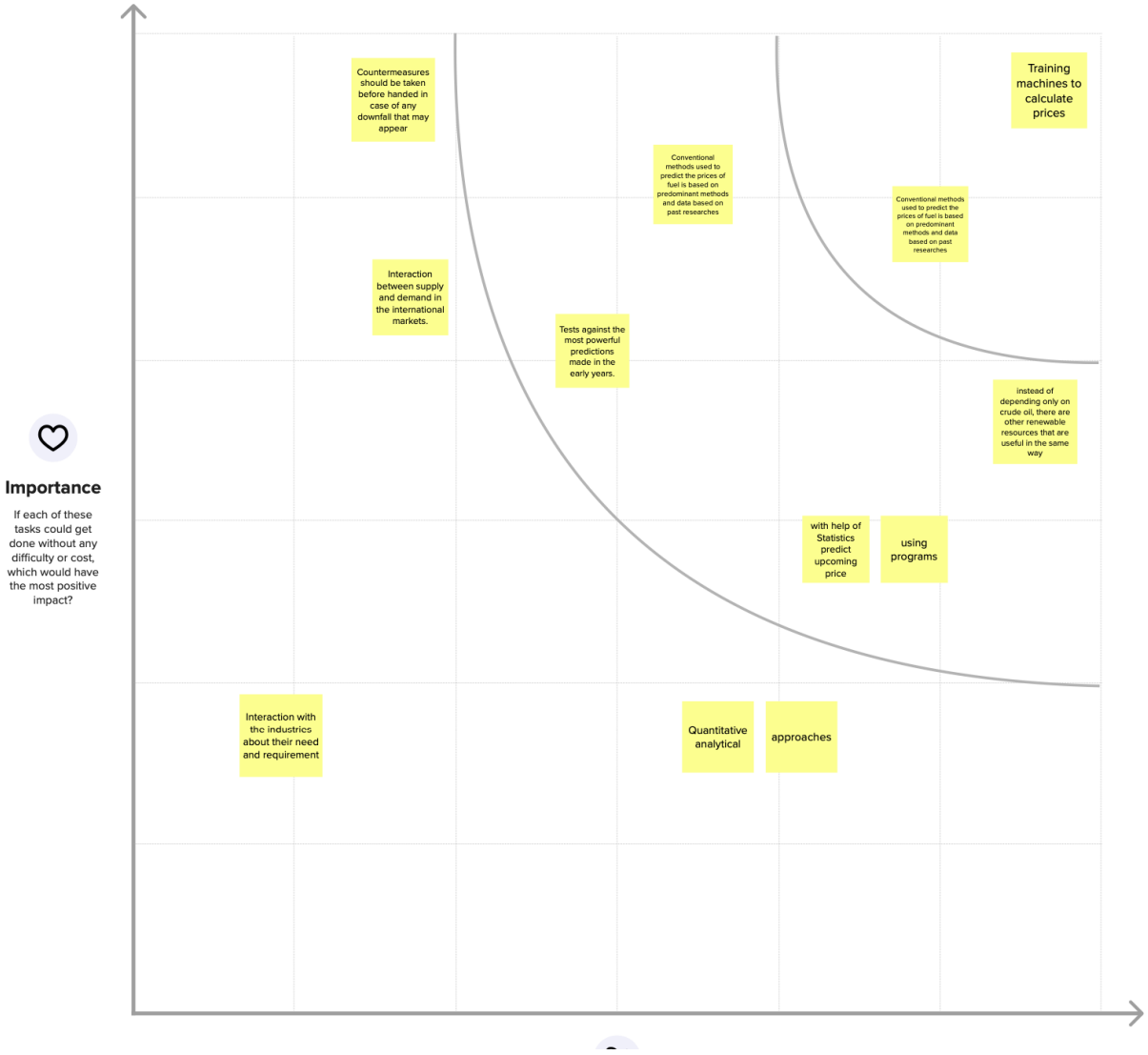
Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.

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

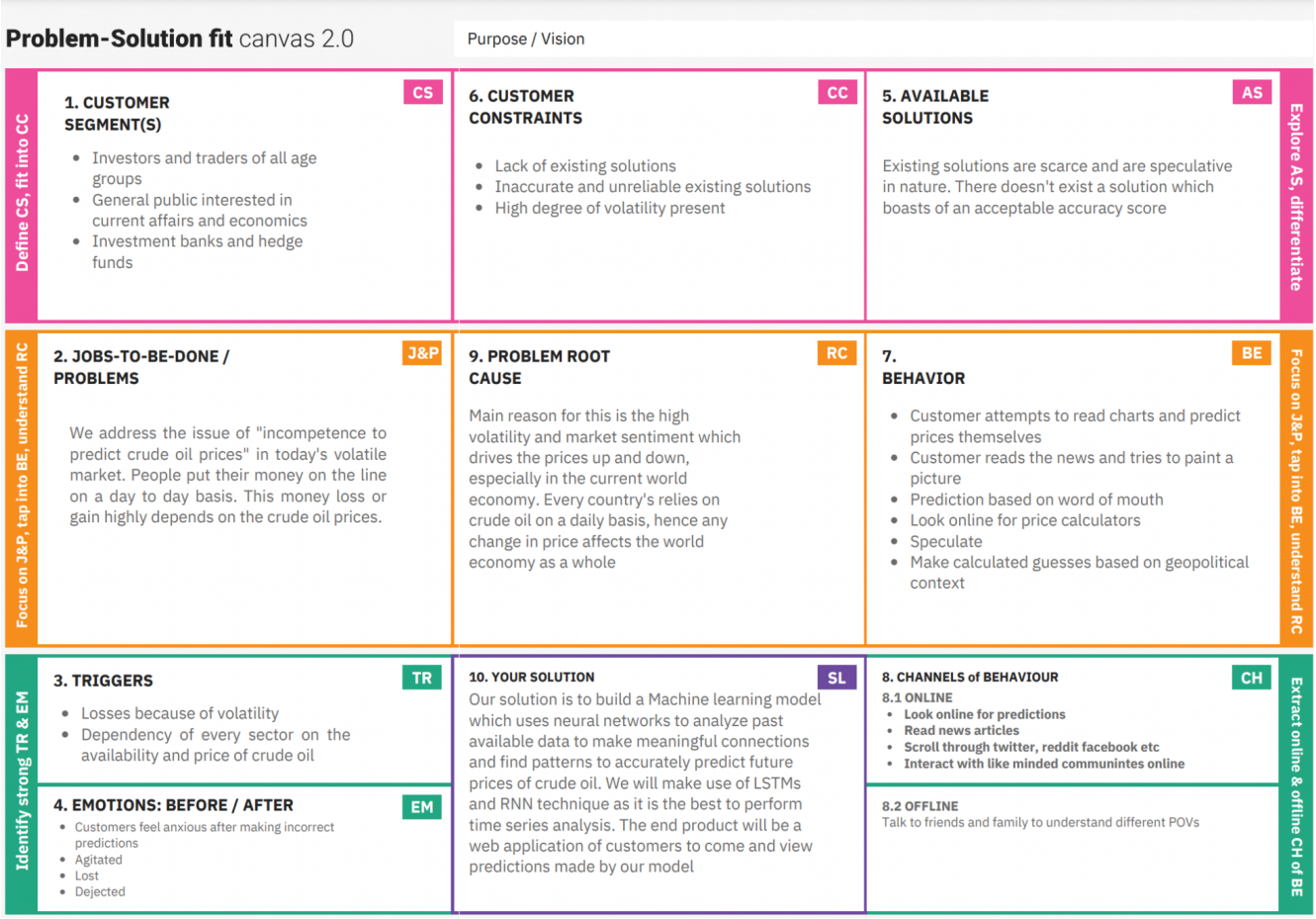


3.3 Proposed Solution

S.No	Parameter	Description
1.	Problem Statement (Problem to be solved)	Crude oil is one of the most essential commodities for daily living, and its price fluctuations have a significant effect on the environment globally. Governments, businesses, and individuals can all benefit from price predictions. Continuous use of statistical and econometric methods, including AI, for predicting crude oil prices may show declines in forecast accuracy. In order to purchase crude oil at the best time, the price may be used as a time analysis prediction. The industries can enhance their profit rate as a result.
2.	Idea / Solution description	A data driven approach is used to predict the prices. RNN is used to achieve future crude oil prices using the previous history of crude oil. The cost is measured to determine its effectiveness. The performance of the proposed model is evaluated using the price data and other materials.
3.	Novelty / Uniqueness	Creating a user-friendly user interface that will allow any user to use the application. The forecast will take into account additional variables influencing price variation in addition to prices. Applications will provide numerous advantages to the government and enterprises.

4.	Social Impact / Customer Satisfaction	The users' performance will be enhanced by the ability to estimate crude oil prices, which will enable them to be ready for unforeseen circumstances. Additionally, it will aid in economic stabilization. Numerous goods and products are directly impacted by price, and changes in price have an impact on the stock markets. Other factors also have an impact on the costs.
5.	Business Model (Revenue Model)	These days the businesses using the supply chain rely on crude oil for transportation. So the prediction of crude oil prices will help the firms increase the profit rate and also result ingrowth of the firm. It can help decision makers such as firms, private investors, or individuals choose to buy or sell crude oil . Crude oil is one of the most profitable trading commodities for traders. RNN and LSTM models are used as the key model to predict crude oil prices.
6.	Scalability of the Solution	User friendly application to enable users to access from anywhere. Algorithms and methods are used to improve the accuracy and prediction.

3.4 Problem Solution fit



4. REQUIREMENT ANALYSIS

4.1 Functional requirement

FR No.	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	User Application	The application will be a web based application. Users can access the system on the internet.
FR-2	User Products available	The prices of the products that are available (crude oil) is constantly updated and the predicted price will be displayed.

FR-3	User Additional Features	The user will be able to see information about the increase or drop in prices of crude oil.
FR-4	User Exceptions	The user has access to/can view the predicted price of crude oil.

4.2 Non-Functional requirements

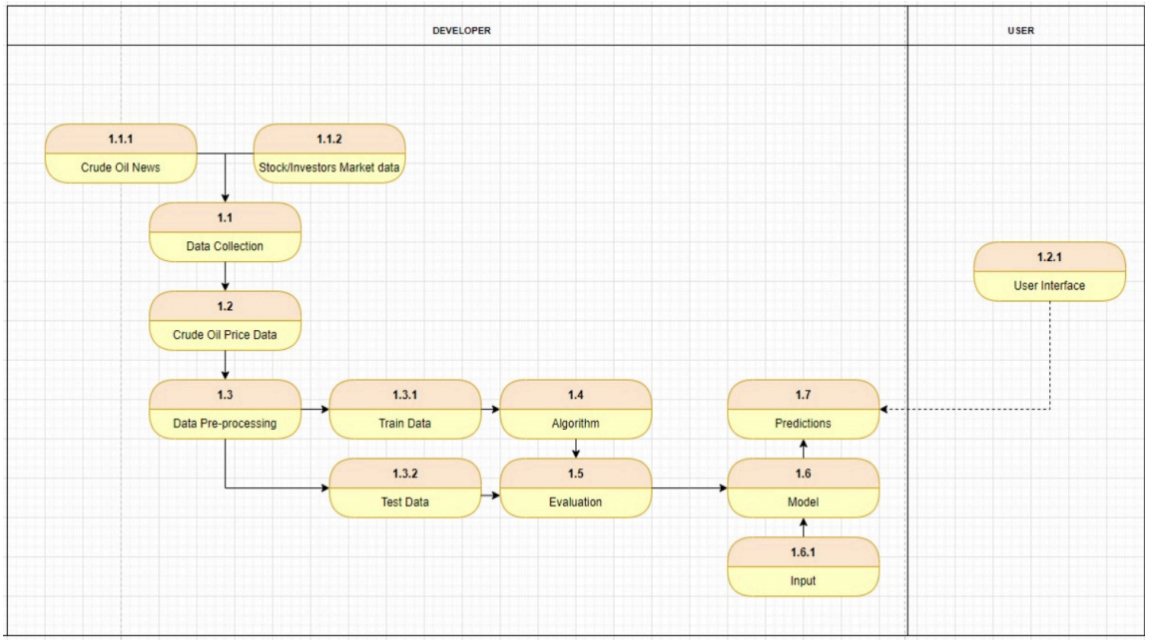
NFR No.	**Non-Functional Requirement**	**Description**
NFR-1	Usability	We can select whether to invest in or buy the product based on the expected statistics which will be predicted a logical user interface will be provided to utilize the system
NFR-2	Security	The integrity of the historical data sets will be maintained, security will be provided for the users as specific cryptographic methods can be applied.
NFR-3	Reliability	The necessary analysis of the data will be done, which enhances the accuracy, thus this product is trustworthy and the predicted value we receive is reliable. An internet connection is also required to get the best output, data corruption will also be avoided.
NFR-4	Performance	The system will permit several users to utilize the resources at once while maintaining the highest level of price prediction accuracy.
NFR-5	Availability	The system will always be available to the users as long as they have a suitable and stable internet connection.
NFR-6	Scalability	The LSTM network model works well for a large number of users.

		Hence the system has room for a bigger audience and can be scaled up easily

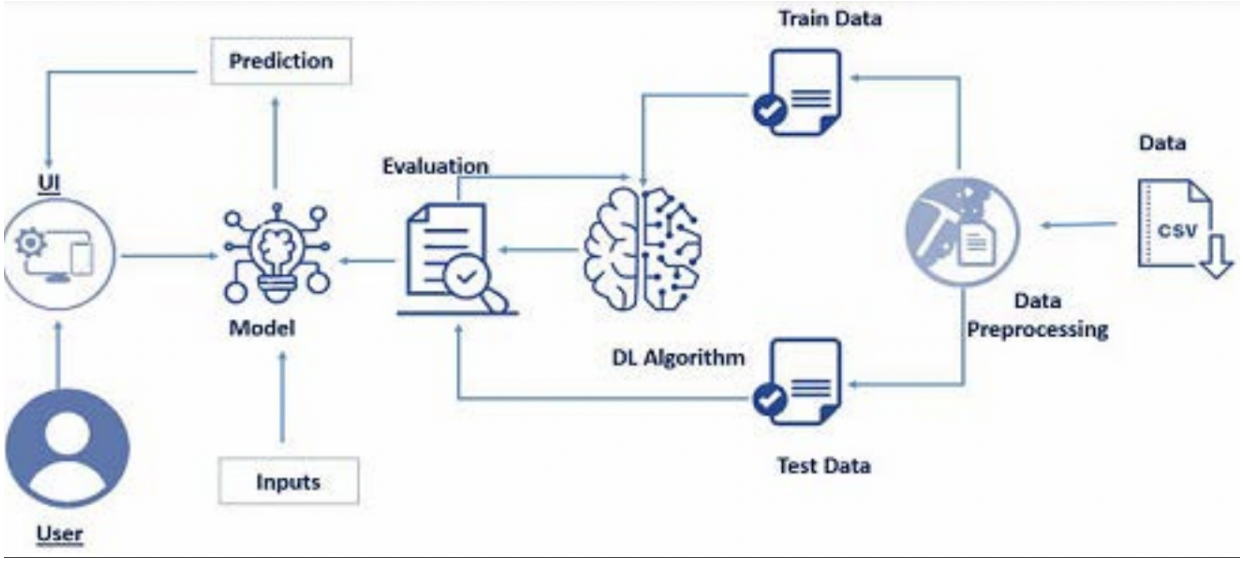
5. PROJECT DESIGN

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 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.

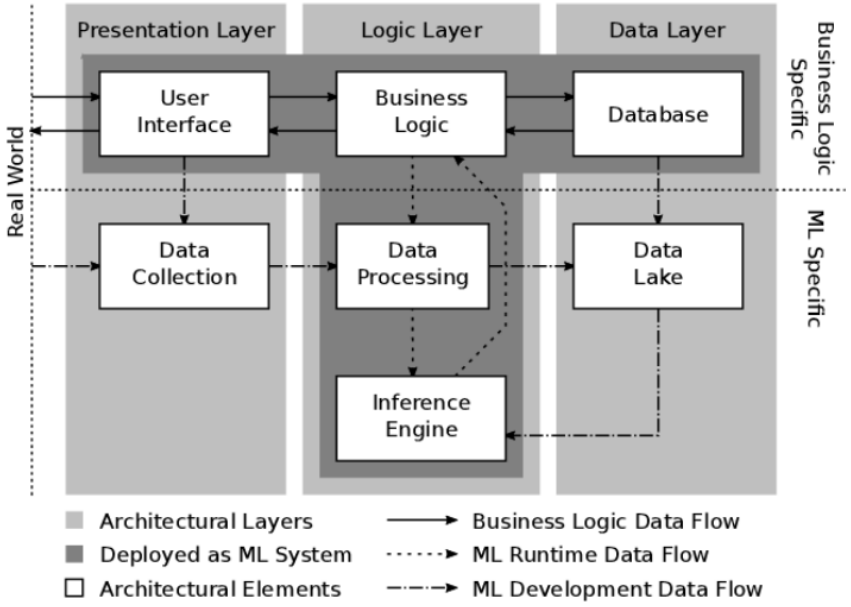


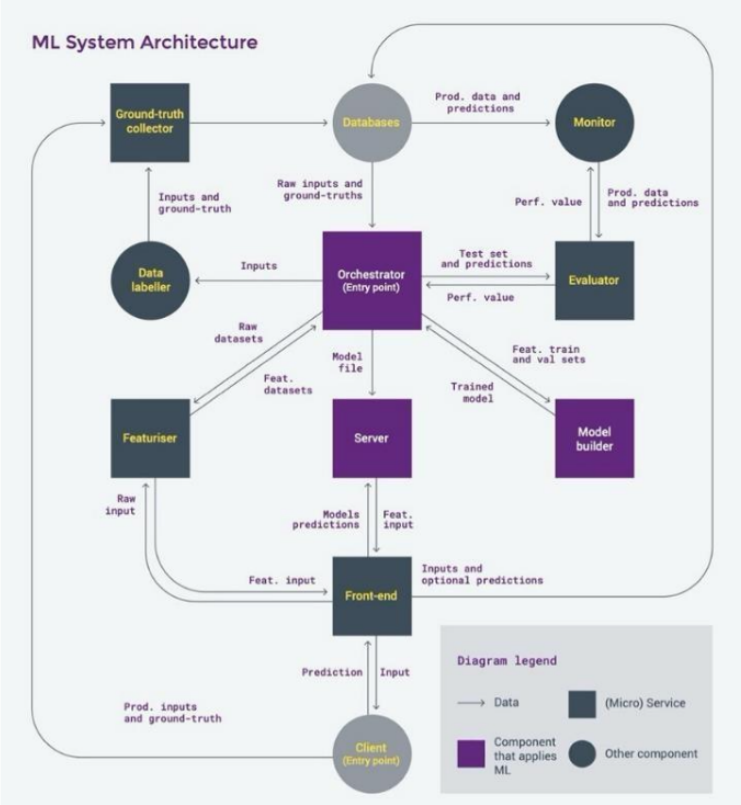
5.2 Solution



Technical Architecture

TECHNICAL ARCHITECTURE:





SNO	COMPONENT	DESCRIPTION	TECHNOLOGY
1.	User interface	UI of the web app	HTML,CSS, Javascript
2.	Cloud database	Database service on cloud	IBM cloud
3.	File storage	File storage requirements	IBM block storage
4.	ML model	Predict crude oil price	RNN and LSTM
5.	Infrastructure	App deployment on local server	Local host
6.	Open-Source-Framework-1	Python	Pandas, flask, numpy, tensorflow
7.	Open-Source-Framework-2	Angular JS	App module, component module
8.	Open-Source-Framework-3	HTML & CSS	and flex model
9.	Performance	Handle 100 to 1000 users to use server at a time	Flask

5.3 User Stories

User Type	Functional Requirements	User Story Number	User Story/Task	Acceptance Criteria	Priority	Release
Customer (Mobile User or Web User)	Registration	USN-1	As a user, I can register for the application by entering my email, password and reconfirming my password.	I can access my dashboard from my mobile device or web browser.	High	Sprint-1
		USN-2	As a user, I will receive a confirmation email once I have registered for the application.	I receive a confirmation for my registration.	High	Sprint-1
		USN-3	As a user, I can register for the application through social media using a pre-existing Google account.	I receive an authentication mail asking if I am the user registering for the application.	Medium	Sprint-2
	Login	USN-4	As a user, I can log into the application using my credentials.	I can now access the application if my credentials are correctly entered.	High	Sprint-1
	Dashboard	USN-5	As a user, I can navigate through the dashboard and view the crude oil predictions.	Access to the dashboard is granted.	High	Sprint-1
	Entering parameters	USN-6	As a user, I can enter the values of the parameters required for the crude oil prediction,	I can view the predictions based on my choice.	High	Sprint-3
	Download Report	USN-7	As a user, I can download a report on to my local system.	I can now view the report on my local system giving me insights about the prediction	Low	Sprint-4
Admin	Registration	USN-6	As the admin, I can register for the application by entering my email, password and reconfirming my password.	I can access my dashboard from my mobile device or web browser.	High	Sprint-1
	Login	USN-7	As the admin, I can log into the application using my credentials.	I can now access the application if my credentials are correctly entered.	High	Sprint-1
	Upload Dataset	USN--8	As the admin, I can upload the dataset on which the predictions have to be performed.	Dataset is uploaded successfully and the model is trained to predict.	High	Sprint-1

6. PROJECT PLANNING & SCHEDULING

6.1 Sprint Planning & Estimation

Product Backlog, Sprint Schedule, and Estimation:

SPRINT	FUNCTIONAL REQUIREMENT	USE STORY NUMBER	USER STORY/ TASK	STORY POINTS	PRIORITY	TEAM MEMBERS
Sprint-1	Data collection	USN-1	Download Crude Oil Price Dataset	2	Medium	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-1	Data Preprocessing	USN-2	Importing The Dataset into Workspace	1	Low	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-1		USN-3	Handling Missing Data	3	Medium	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-1		USN-4	Feature Scaling	3	Low	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-1		USN-5	Data Visualization	3	Medium	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-1		USN-6	Splitting Data into Train and Test	4	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-1		USN-7	Creating A Dataset with Sliding Windows	4	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-2	Model Building	USN-8	Importing The Model Building Libraries	1	Medium	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-2		USN-9	Initializing The Model	1	Medium	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-2		USN-10	Adding LSTM Layers	2	High	Sayeeshwar

						Karthik Varunn Manasi Vedh Vijay
Sprint-2		USN-11	Adding Output Layers	3	Medium	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-2		USN-12	Configure The Learning Process	4	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-2		USN-13	Train The Model	2	Medium	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-2		USN-14	Test The Model	3	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-3	Application Building	USN-15	Build Python Code and Run The App in Local Browser	8	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-3		USN-16	Showcasing Prediction On UI	4	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-3		USN-17	Register for application using Email and Password	3	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-3		USN-18	Log in to application using registered account	1	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-3		USN-19	Download the predictions as PDF file	2	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-4	Train the Model On IBM	USN-20	Register For IBM Cloud	4	Medium	Sayeeshwar Karthik Varunn

						Manasi Vedh Vijay
Sprint-4		USN-21	Train The ML Model On IBM	8	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay
Sprint-4		USN-22	Integrate Flask with Scoring End Point	8	High	Sayeeshwar Karthik Varunn Manasi Vedh Vijay

6.2 Sprint Delivery Schedule

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 Releas (Actual)
Sprint-1	20	7 Days 24 Oct 2022 30 Oct 2022	20	30 Oct 2022
Sprint-2	20	7 Days 31 Oct 2022 06 Nov 2022	16	06 Nov 2022
Sprint-3	20	7 Days 07 Nov 2022 13 Nov 2022	18	13 Nov 2022
Sprint-4	20	7 Days 14 Nov 2022 20 Nov 2022	20	20 Nov 2022

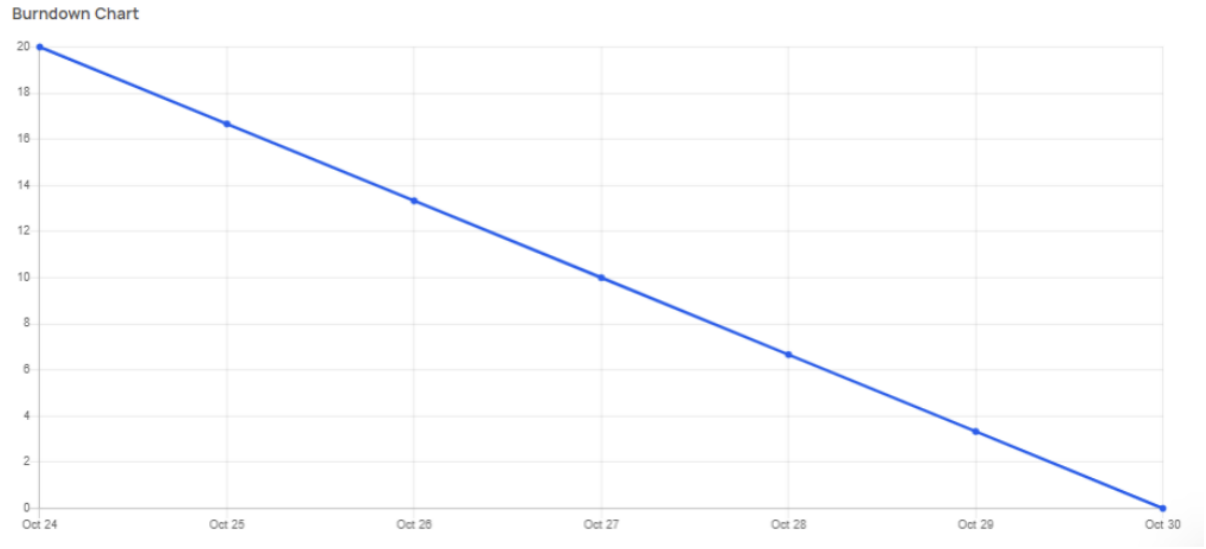
Velocity:

Sprint	Average Velocity
Sprint 1	2.8
Sprint 2	2.2
Sprint 3	2.5
Sprint 4	2.8

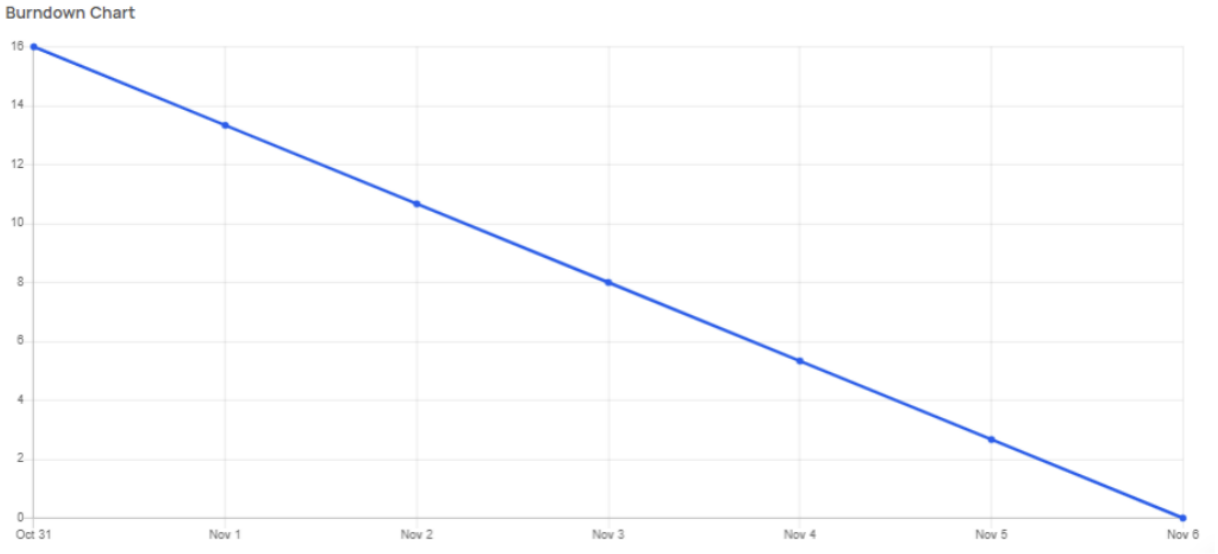
Burndown Chart:

A burndown chart is a graphical representation of work left to do versus time. It is often used in agile software development methodologies such as Scrum. However, burn down charts can be applied to any project containing measurable progress over time.

Burndown Chart - Sprint 1

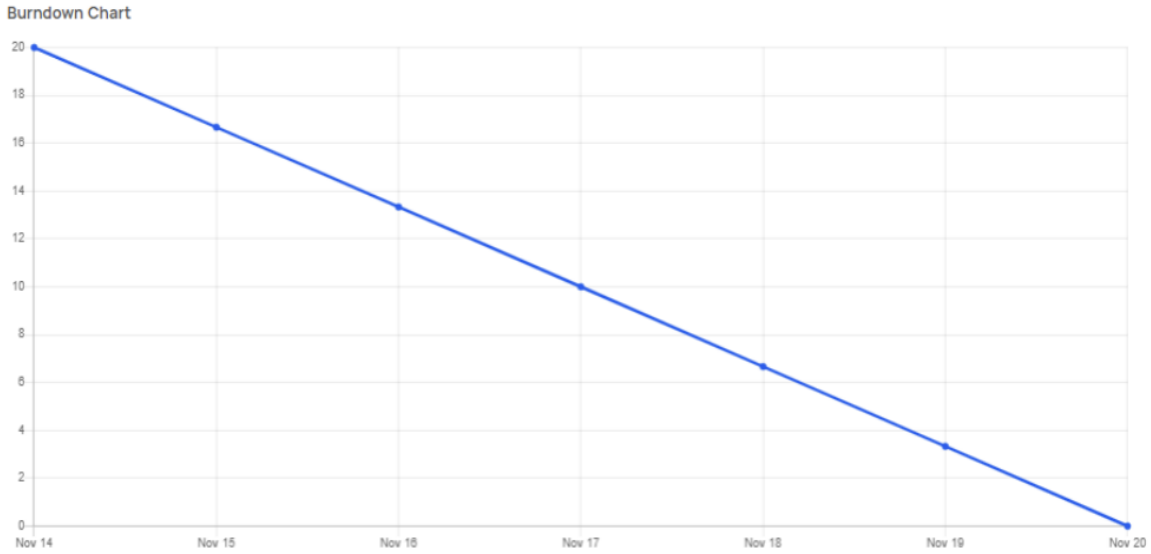


Burndown Chart - Sprint 2



Burndown Chart - Sprint 3

Burndown Chart - Sprint 4

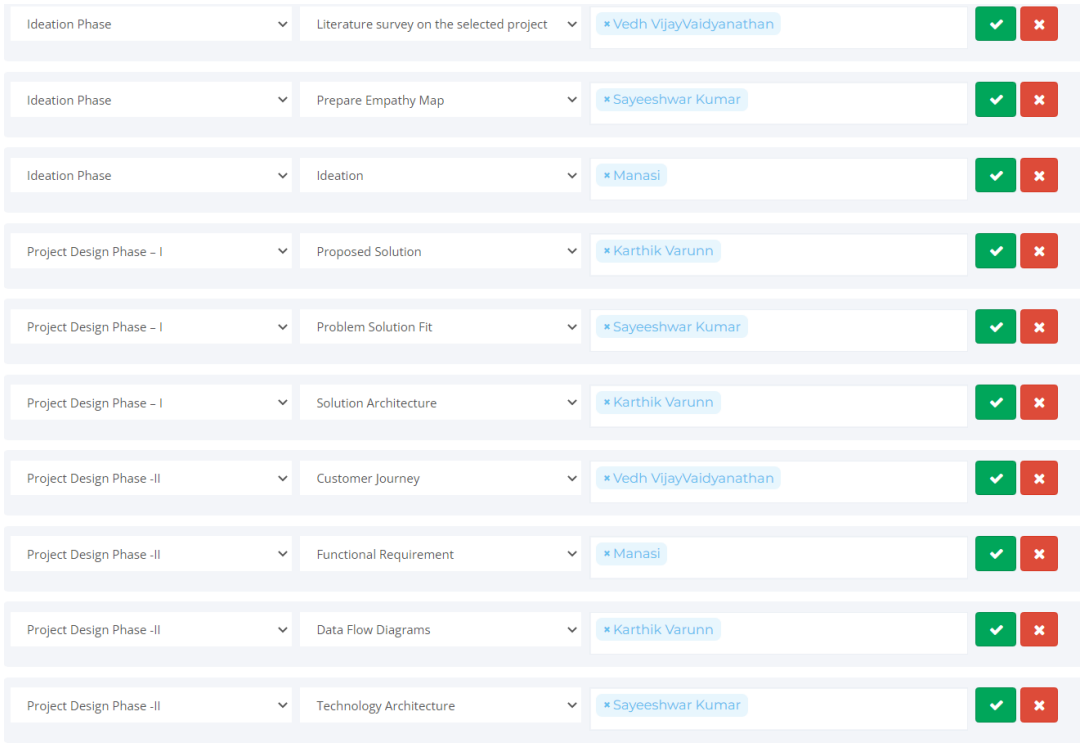


Milestone and activity list

TITLE	DESCRIPTION	DATE
Literature Survey & Information Gathering	Literature survey on the selected project & gathering information by referring to the technical papers, research publications etc.	28th SEPTEMBER 2022
Prepare Empathy Map	Prepare Empathy Map Canvas to capture the user Pains & Gains, Prepare list of problem statements	17th SEPTEMBER 2022
Ideation	List them by organizing the brainstorming session and prioritize the top 3 ideas based on feasibility & importance.	17th SEPTEMBER 2022
Proposed Solution	Prepare the proposed solution document, which includes the novelty, feasibility of idea, business model, social impact, scalability of solution, etc.	5th OCTOBER 2022
Problem Solution Fit	Prepare problem - solution fit document.	5th OCTOBER 2022
Solution Architecture	Prepare the solution architecture document.	12th OCTOBER 2022

Customer Journey	Prepare the customer journey maps to understand the user interactions & experiences with the application (entry to exit).	14th OCTOBER 2022
Functional Requirement	Prepare the functional requirement document.	14th OCTOBER 2022
Data Flow Diagrams	Draw the data flow diagrams and submit for review.	14th OCTOBER 2022
Technology Architecture	Prepare the technology architecture diagram.	14th OCTOBER 2022
Prepare Milestone & Activity List	Prepare the milestones & activity list of the project.	22nd OCTOBER 2022
Project Development - Delivery of Sprint-1, 2, 3 & 4	Develop & submit the developed code by testing it.	IN PROGRESS...

6.3 Reports from JIRA



Application Building	Build python code	× Karthik Varunn
Application Building	Run The app in local browser	× Karthik Varunn
Application Building	Showcasing prediction on UI	× Vedh VijayVaidyanathan
Train The Model on IBM	Register for IBM Cloud	× Manasi
Train The Model on IBM	Train the ML Model on IBM	× Manasi
Train The Model on IBM	Integrate Flask with Scoring End Point	× Sayeeshwar Kumar
Project Planning Phase	Prepare Milestone & Activity List	× Manasi
Project Planning Phase	Sprint Delivery Plan	× Manasi
Project Development Phase	Project Development - Delivery of Sprint-	× Sayeeshwar Kumar
Project Development Phase	Project Development - Delivery of Sprint-	× Vedh VijayVaidyanathan
Project Development Phase	Project Development - Delivery of Sprint-	× Karthik Varunn
Project Development Phase	Project Development - Delivery of Sprint-	× Manasi
Project Objectives	Select Activities	× Vedh VijayVaidyanathan
Data Collection	Download dataset /create dataset	× Sayeeshwar Kumar
Data Preprocessing	Import the Libraries	× Manasi
Data Preprocessing	Importing the dataset	× Karthik Varunn
Data Preprocessing	Handling Missing Data	× Vedh VijayVaidyanathan
Data Preprocessing	Feature Scaling	× Manasi
Data Preprocessing	Data Visualization	× Sayeeshwar Kumar
Data Preprocessing	Splitting Data into Train and Test	× Vedh VijayVaidyanathan
Data Preprocessing	Creating a dataset with sliding windows	× Karthik Varunn

7. CODING & SOLUTIONING

7.1 Feature 1

Home Page

When our application loads, the project title and a brief description of the project are displayed on the home page. To move between the application's pages, use the navigation bar at the top of the screen.

The "Predict Price" button follows the description on the bottom left and brings the application to the prediction page.

Feature 2

Prediction page

Users must enter 10 past prices for crude oil into a form that is presented on the page. The application will display the future anticipated value as soon as the user clicks "predict." The system will display an error message if the user enters either insufficient or excessive data, indicating that the proper information wasn't provided.

Feature 3

DL model

RNN(LSTMs) is the foundation of our prediction model, which is a very good representation of this system.

In time series modelling, the RNN is a deep neural network type that is frequently utilised. The RNN model includes a hidden state that is produced by the sequential information of a time series, and the output is dependent on the hidden state. The vanishing gradient problem, which is resolved by LSTM, makes it challenging for RNNs to learn long-term dependence even though they replicate time series data well.

Since LSTM models have good long- and short-term memory capabilities, thus historical state information on crude oil prices won't be lost. The historical price data for crude oil can be fully extracted, and the features of the most recent data are taken into account. As a result, the LSTM model we chose is quite effective and beneficial.

GithubCodeLinks-

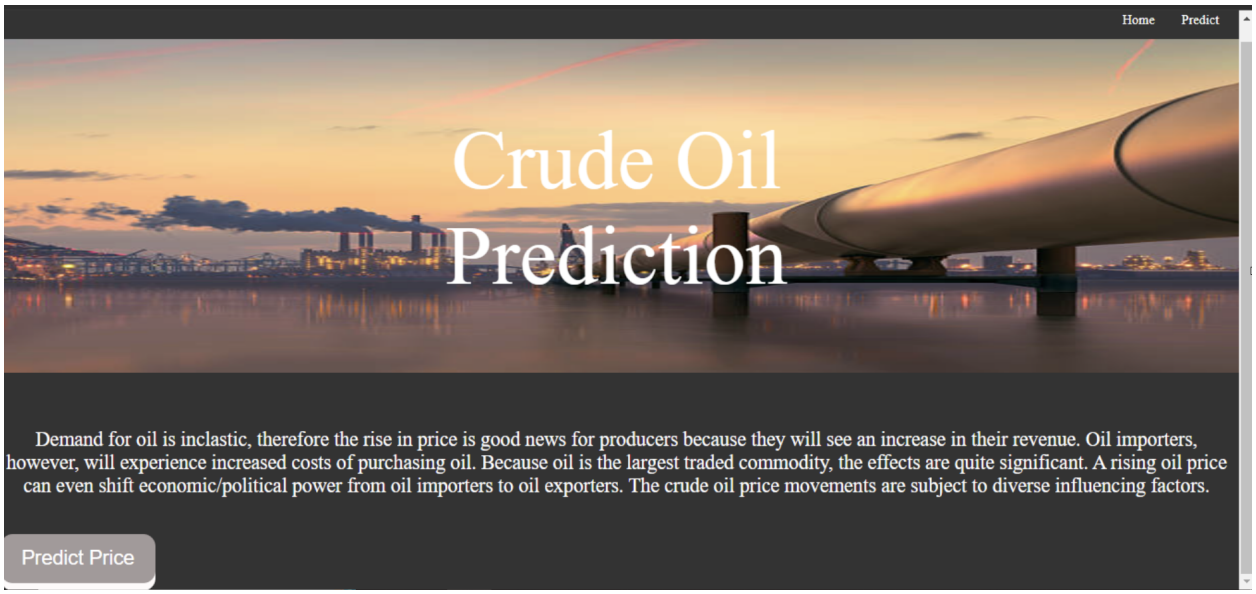
https://github.com/IBM-EPBL/IBM-Project-17241-1659632038/tree/main/code/template

https://github.com/IBM-EPBL/IBM-Project-17241-1659632038/tree/main/code/static

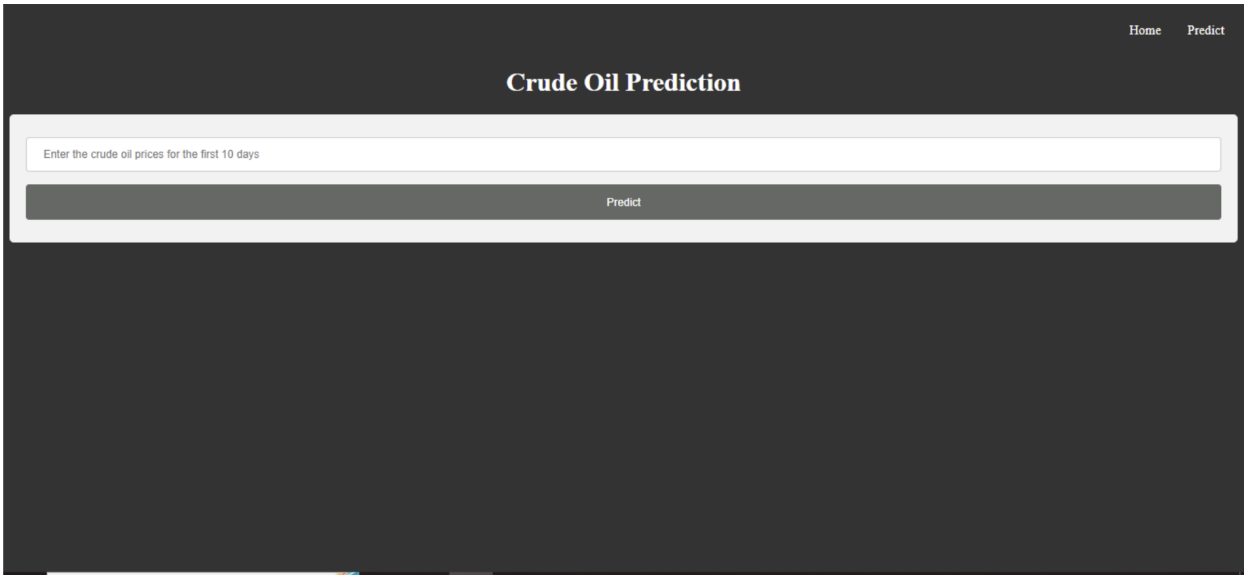
8. TESTING

8.1 Test Cases

Home Page

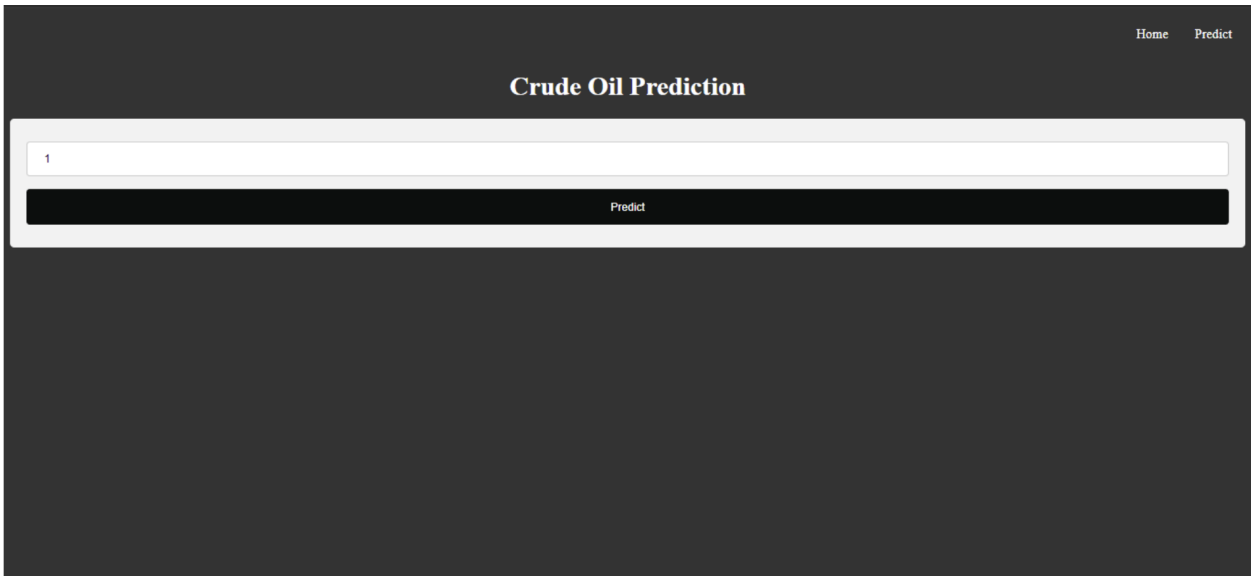


Prediction Page



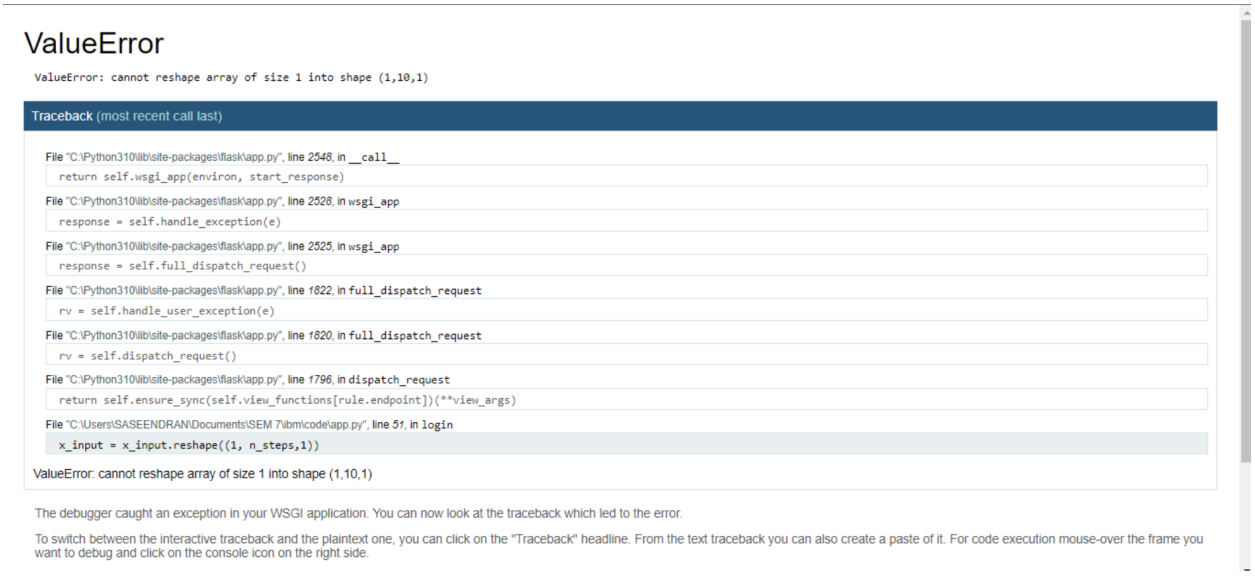
Test Case-1

When the user inputs insufficient data-



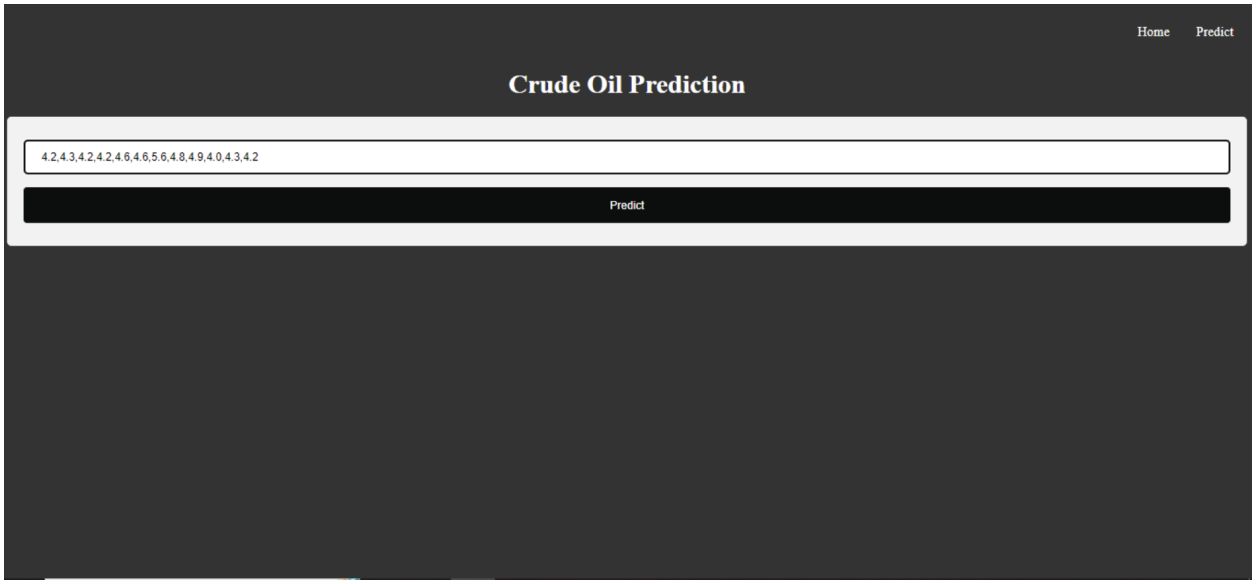
Output

The application displays an error message-



Test Case-2

When the user inputs excess data-

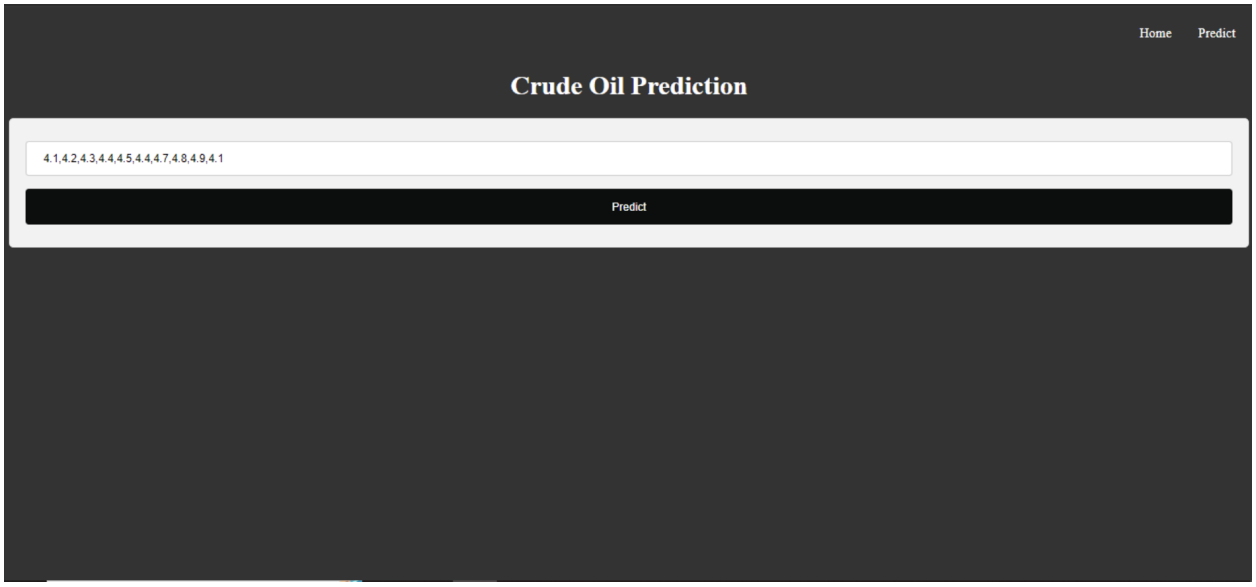


Output

The application displays an error message-

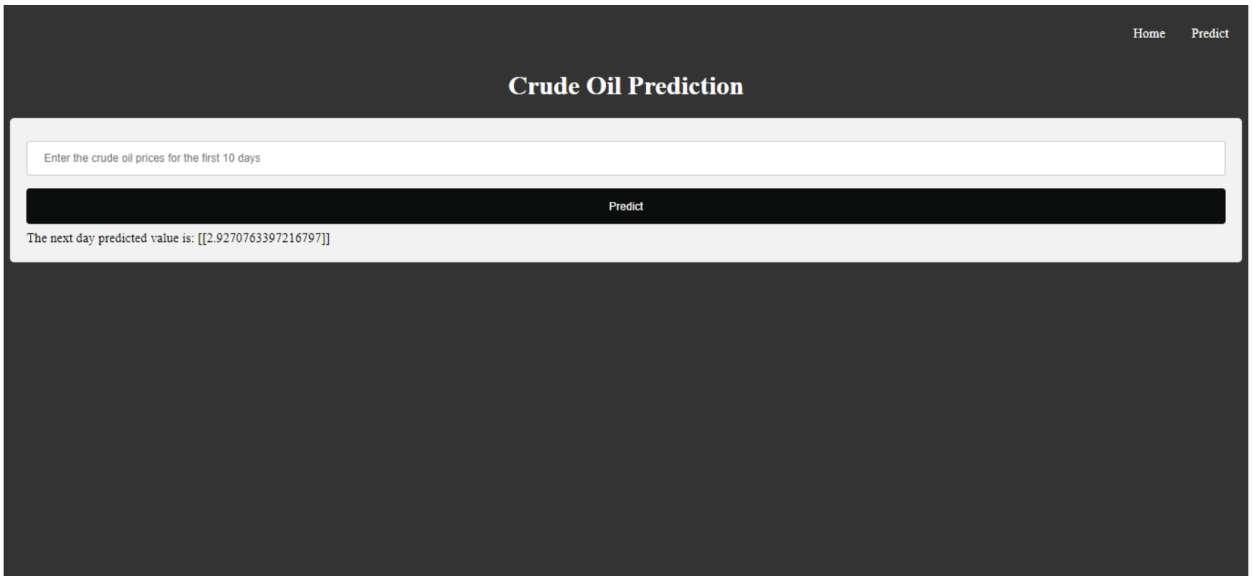
Test Case-3

When the user inputs the correct number of values



Output

The application displays the next day predicted values-



8.2 User Acceptance Testing

Multiple users enter the data into the application to test the model based on the input given, the application calculates the projected values of crude oil.

The following were the results after the application user tested various applications- The end user who uses our application received insightful knowledge on crude oil pricing, which is highly helpful. We draw a conclusion on the acceptance of our system based on the feedback provided.

9. RESULTS

9.1 Performance Metrics

```
math.sqrt(mean_squared_error(Y_train,train_predict))
```

```
37.98840576260146
```

```
mean_absolute_error(Y_train, train_predict)
```

```
30.73885923443683
```

10. ADVANTAGES & DISADVANTAGES

10.1 ADVANTAGES

A. **Scalable application**
   Our application is based on Python and trains and tests the model using data from a CSV file. This strategy is easily scalable and usable for larger institutions with access to numerous databases as a CSV file has a large number of rows and can thus accommodate this.

B. **Ease of use**
   The system is based on a web application with a very user-friendly design that makes it simple for anyone to use, including those without a technical experience. It is very easy to navigate and the predicted price is immediately reflected to the end user in a clear and practical way.

C. **Minimizing risk associated with fluctuating oil prices**
   Crude oil is one of the most important commodities in the world and is used as a raw material for a variety of products. Thus, crude oil price fluctuations have a far reaching impact on global economies. Hence our prediction model which effectively performs price forecasting can assist in minimizing the risks associated with volatility in oil prices and proves to be advantageous to various stakeholders such as - governments, public and private enterprises, policymakers, and investors.

D. ML technique used for prediction

Our prediction model is based on RNN(LSTMs) which is a highly effective model for this system.

The RNN is a type of deep neural network that has been widely used in time series modeling. The RNN model adds a hidden state generated by the sequential information of a time series, with the output dependent on the hidden state. Although RNNs simulate time series data well, these are difficult to learn long-term dependence because of the vanishing gradient issue - which is overcome by LSTM.

LSTM models have excellent long-term and short-term memory ability, which will not lead to the loss of more historical state information on crude oil prices. It can fully extract historical information on the crude oil price and consider the characteristics of the current data for price information. Thus our chosen LSTM model is highly efficient and advantageous.

E. Time analysis prediction

As our system uses time series forecasting and price is used as a time analysis prediction. Thus this is advantageous in order to purchase crude oil at the best time, from which all stakeholders can enhance their profit rate.

F. Dataset

The dataset used by the model in our system can be constantly updated to make the predictions more accurate for a stronger hold over price forecasting and thus our system's scope is endless in this respect.

G. Automated choosing of weights of factors for enhanced accuracy.

The end user must only input past 10 prices of crude oil and the weights for the individual factors are chosen by the model and not by the person inputting data. This reduces human error and also increases the accuracy of the model overall.

10.2 DISADVANTAGES

A. Feature weights

A disadvantage is that, this system predicts prices based on past features that have had an impact on the price and only accounts for a feature as a weight in the model once it has resulted in the fluctuation of crude oil price. That is, it cannot foresee all possible reasons for price fluctuation until certain events have occurred.

B. Complexity

Another drawback is that, despite being simple to use, the system's code is difficult for the average user to understand in the event of mistakes.

C. Maintenance

Due to the complexity of the code, changing it to make use new technologies and updating it when there are faults can be seen as a negative.

11. CONCLUSION

A system that is used to predict crude oil prices was designed, to make easy - the forecasting of the oil prices which is a very challenging problem due to its high volatility. Multiple machine learning methods and processes were used to build the system. The web application was built using HTML, CSS, and JavaScript and the primary coding and integration was done using Python. This application can be used by Traders/Investors in the stock market, governments and businesses to accurately predict crude oil prices with minimal risk, while accounting for day to day price fluctuations. Our model will be updated whenever new oil price dataset is available, so the model continuously evolves over time, and can capture the changing pattern of oil prices. This system can be used by anyone owing to its simplicity and ease of access.

The application was created after extensive research and understanding the current model and how there was room for significant improvement. The final system is deployed on IBM Watson and is available for use.

12. FUTURE SCOPE

The system that has been built is a very scalable and deployable model. This broadens the scope of the project considerably. As the model is an effective tool for crude oil price prediction and can be efficiently used for short term price forecasting, investors can use this model to initiate trades and to judge various strategies relating to investments. As crude oil prices are so volatile, this prediction model will help individuals and businesses purchase crude oil at the right time based on the dataset and this dataset used can be constantly updated to make the predictions more accurate for a stronger hold over the banking loan industry and the systems scope is endless in this respect.

13. APPENDIX

Source Code -

Sprint3 -

https://github.com/IBM-EPBL/IBM-Project-17241-1659632038/tree/main/Project%20Development%20Phase/Sprint-3

Sprint4 -

https://github.com/IBM-EPBL/IBM-Project-17241-1659632038/tree/main/Project%20Development%20Phase/Sprint-4

GitHub Link - https://github.com/IBM-EPBL/IBM-Project-17241-1659632038

Project DemoLink -

https://drive.google.com/file/d/1BEPle0QJiy700JoT8h18VA5YVzbCTDcn/view?usp=share_link