CRUDE OIL PRICE PREDICTION

Academic Year 2019-2023

IBM-Project : 42892-1660710919

TEAM ID : PNT2022TMID44242

PROJECT MEMBERS:

S.NO	NAME OF THE STUDENT	REGISTER NUMBER
1.	INDHUMATHI T	730319106017
2.	LACKSHMIPRABHA E	730319106020
3.	SABITHA C	730319106039
4.	SIVAGNANAM A V	730319106044
5.	VENKATRAMANAN K	730319106047

SUBMITTED IN THE PARTIAL FULFILLMENT FOR THE REQUIREMENTS

FOR THE AWARD OF

BACHELOR OF ENGINEERING

IN

ELECTRONICS AND COMMUNICATION ENGINEERING

BUILDERS ENGINEERING COLLEGE

NATHAKADAIYUR, KANGAYAM-638 108.

TABLE OF CONTENTS

1.ABSTRACT

Over millions of years ago, the remains of these animals and plants were covered by layers of sand, slit and rock. Heat and Pressure from these layers turned the remains into what we now call crude oil or petroleum. Crude oil is a naturally occurring petroleum product composed of hydrocarbon deposits and other organic materials. A type of fossil fuel, crude oil is refined to produce usable products including gasoline, diesel and various other forms of petrochemicals and it is a limited resource.

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, and 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.

2.INTRODUCTION

2.1 PROJECT OVERVIEW

Globally, crude oil is one of the important fuel sources and historically, has contributed to over a third of the world's energy. Many economists view crude oil as the single most important commodity in the world as it is currently the primary source of energy production.Oil is especially important to businesses that rely on fuel, such as airlines, plastic producers, and agricultural business.

In 2019, global oil consumption reached 10075 million barrels per day considering data from the International Energy Agency(IEA). The crude oil demand has increased due to expeditious economic growth.Since crude oil price series are generally considered to be non linear and non stationary time series, they can be accurately influenced by several factors; therefore, accurately predicting the price of oil can be quite challenging.

There are innumerable ways and approaches which are being used and have been used for predicting the prices of crude oil,one of the most common techniques which are used is by accepting the current statistics and prices and assuming it to be the same for future without any change at all in the prices, however, such a prediction never works now as the nature of oil is extremely volatile. This study aims to design a method of estimating the price level for crude oil. The proposed method is the prediction of crude oil using Artificial Intelligence.

2.2 PURPOSE

Price differences can stem from various reasons, such as where the oil is produced, transportation costs, political and economic conditions in the regions where the oil is sold, and refining costs.

- Because of Political issues and climatic conditions, the price will fluctuate according to the events and it becomes hard to predict its price.
- Prior knowledge of crude oil price must be considered among the key parameters needed to make a proper decision towards development, production processes and government for short term and long-term planning, hence any rise or fall in crude oil price has a measurable effect on the economy.

3.LITERATURE SURVEY

3.1 EXISTING PROBLEM

3.1.1 TITLE: Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization, 2021

AUTHOR NAME: Aytac Altan

Estimating the price of crude oil, which is seen as an important resource for economic development and stability in the world, is a topic of great interest by policy makers and market participants. However, the chaotic and nonlinear characteristics of crude oil time series (COTS) make it difficult to estimate crude oil prices with high accuracy. To overcome these challenges, a new crude oil price prediction model is proposed in this study, which includes the long short-term memory (LSTM), technical indicators such as trend, volatility and momentum, and the chaotic Henry gas solubility optimization (CHGSO) technique. In the proposed model, features based on trend, momentum and volatility technical indicators are utilized. The features are obtained by using the trend indicators such as exponential moving average (EMA), simple moving average (SMA) and Kaufman's adaptive moving average (KAMA), the momentum indicators such as commodity channel index (CCI), rate of change (ROC) and relative strength index (RSI), and the volatility indicators such as average true range (ATR), volatility ratio (VR) and highest high-lowest low (HHLL). These indicators are obtained separately for the West Texas Intermediate (WTI) and Brent COTS. Especially, including the volatility indicator in the model is important in terms of the robustness of the proposed model. The results show that the proposed prediction model copes with the chaoticity and nonlinear dynamics of both WTI and Brent COTS.

2.1.2 TITLE: Prediction Model for the Viscosity of Heavy Oil Diluted with Light Oil Using Machine Learning Techniques, 2022

AUTHOR NAME: Xiaodong Gao

Due to the presence of asphaltene, the flow assurance of high viscosity crude oil
becomes more challenging and costly to produce in wellbores and pipelines. One of the most effective ways to reduce viscosity is to blend heavy oil with light oil. However, the viscosity measurement of diluted heavy crude is either time-consuming or inaccurate. This work aims to develop a more accurate viscosity model of diluted heavy crude based on machine learning techniques. A multilayer neural network is used to predict the viscosity of heavy oil diluted with lighter oil. The input data used in the training include temperature, light oil viscosity, heavy oil viscosity, and dilution ratio. In this modeling process, 156 datasets were retrieved from the available literature of various heavy-oil fields in China. Part of the data (80%) is used to train the developed models using Adam optimizer algorithms, while the other part of The data (20%) is used to predict the viscosity of heavy oil diluted with lighter. The performance and accuracy of the machine learning models were tested and compared with the existing viscosity models. It was found that the new model can predict the viscosity of diluted heavy oil with higher accuracy, and it performs better than other models. The absolute average relative error is 10.44%, the standard deviation of the relative error is 8.45%, and The coefficient of determination is R2 = 0.95.

The viscosity predicted by the neural network outperformed existing correlations by the statistical analysis used for the datasets available in the literature. Therefore, the method proposed in this paper can better estimate the viscosity of diluted heavy crude oil and has important promotion value.

2.1.3 TITLE: Application of machine learning and artificial intelligence in oil and gas industry, 2021

AUTHOR NAME: Anirbid Sircar.

Oil and gas industries are facing several challenges and issues in data processing and handling. Large amount of data bank is generated with various techniques and processes. The proper technical analysis of this database is to be carried out to improve performance of oil and gas industries. This paper provides a comprehensive state-of-art review in the field of machine learning and artificial intelligence to solve oil and gas industry problems. It also narrates the various types of machine learning and artificial intelligence techniques which can be used for data processing and interpretation in different sectors of upstream oil and gas

industries. The achievements and developments promise the benefits of machine learning and artificial intelligence techniques towards large data storage capabilities and high efficiency of numerical calculations. In this paper a summary of various researchers work on machine learning and artificial intelligence applications and limitations is showcased for upstream and sectors of the oil and gas industry. The existence of this extensive intelligent system could really eliminate the risk factor and cost of maintenance. The development and progress using these emerging technologies have become smart and makes the judgment procedure easy and straightforward. The study is useful to access intelligence of different machine learning methods to declare its application for distinct tasks in the oil and gas sector.

3.2 REFERENCES

[1]. Aytac Altan, Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization, 2021

[2]. Xiaodong Gao, Prediction Model for the Viscosity of Heavy Oil Diluted with Light Oil Using Machine Learning Techniques, 2022

[3]. Xinran Gao, An Explainable Machine Learning Framework for Forecasting Crude Oil Price during the COVID-19 Pandemic, 2022

[4]. Quanying Lu1, Analysis and forecasting of crude oil price based on the variable selection-LSTM integrated model, 2021

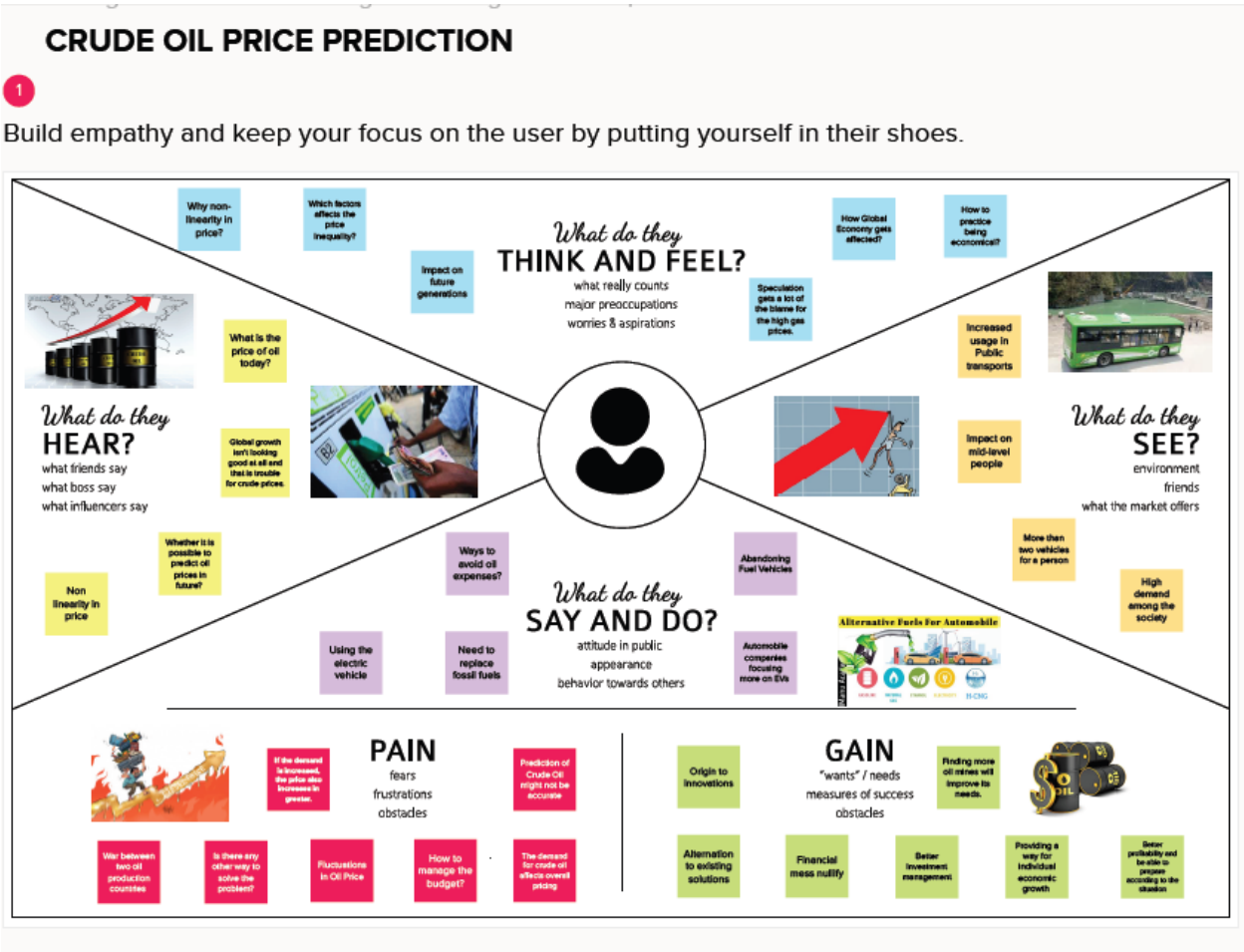
[5]. Anirbid Sircar, Application of machine learning and artificial intelligence in oil and gas industry, 2022

3.3 PROBLEM STATEMENT DEFINITION

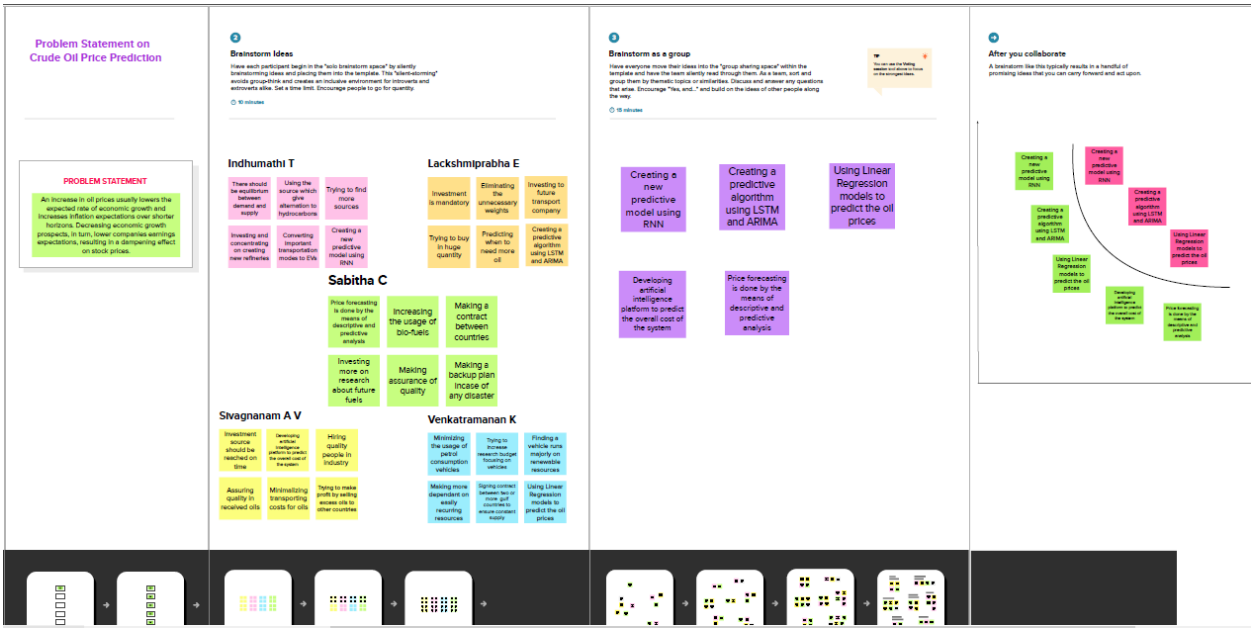
In the existing system, the oil market's internal and external environments are evolving, and there are now a wide variety of complicated impacting elements. It is getting harder to identify practical elements and forecast oil prices as the variables driving global oil prices become increasingly complicated. Numerous previous types of research on the forecasting of crude oil prices demonstrate that the forecasting outcomes depend on the frequency and interval choice of the modeling sample data. Since crude oil's financial characteristics are steadily improving, the volatility of crude oil prices will inevitably have an impact on the profits of oil firms and the actions of investors. Therefore, it is crucial to accurately capture the current trend in global oil prices as well as conduct a systematic examination of the features of complicated international oil markets. It is challenging to identify the elements that most strongly influence the price of oil.

4.IDEATION PROPOSED SOLUTION

4.1 EMPATHY MAP CANVAS



4.2 IDEATION & BRAINSTORMING



4.3 PROPOSED SOLUTION

PROBLEM STATEMENT :

- Oil price prediction is not capable enough to deliver the accurate predicted prices as expected.
- Now the problem arising with the current ANN and CNN models that are used as prediction

models that they can’t provide accurate results when the data is too big.

- Get the desired output and it can be compared with real value to fetch the errors occurring in the model.

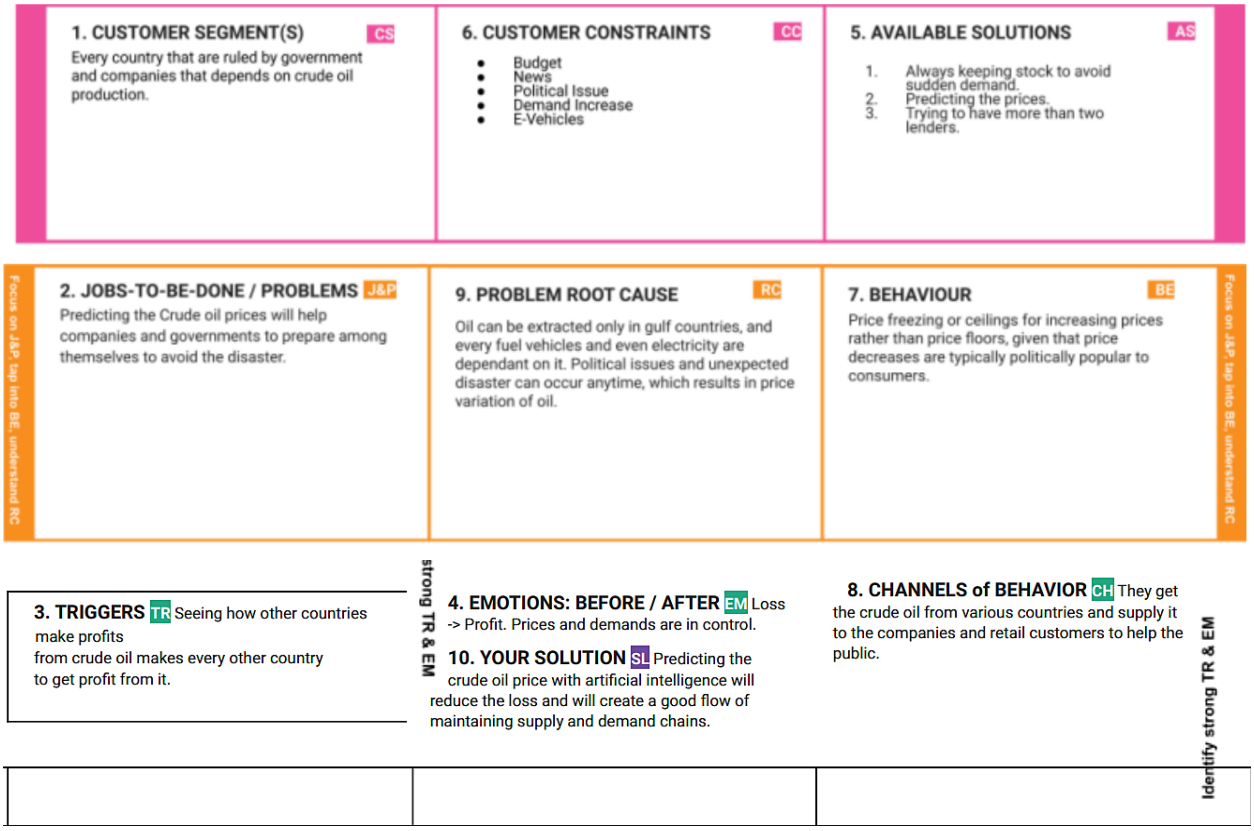
IDEA/SOLUTION DESCRIPTION :

- Factors lead to the successful prediction of the oil prices and provide more accurate results from the model.
- WTI and used the LST model to predict the oil prices with the modal function as the input.
- A LSTM algorithm was proposed which uses backward and feed-forward propagation which helps to get more accurate results.

UNIQUENESS :

- Recurrent Neural Network refers to attacking the sequential problem or temporal aspects of data as time series which is a powerful tool in stock price prediction.
- To extract the feature automatically, and incorporate exogenous variables very easily.
- RNN can connect the previous information to the current task.
- 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 price.

4.4 PROBLEM SOLUTION FIT



5.REQUIREMENT ANALYSIS

5.1 FUNCTIONAL REQUIREMENT

Following are the functional requirements of the proposed solution.

SI NO	FUNCTIONAL REQUIREMENT(EPIC)	SUB REQUIREMENT(STORY/SUB-TASK)
01	User application	User can download this application in the play store or he/she can view it in the browser itself.
02	User products available	The factors that determine the crude oil prices are demand, supply, quality and speculation and the product prices are updated constantly.
03	User additional features	Updating the model whenever new oil price data are available to capture the changing pattern of oil prices.
04	User expectations	Achieving the highest accuracy.

5.2 NON-FUNCTIONAL REQUIREMENT

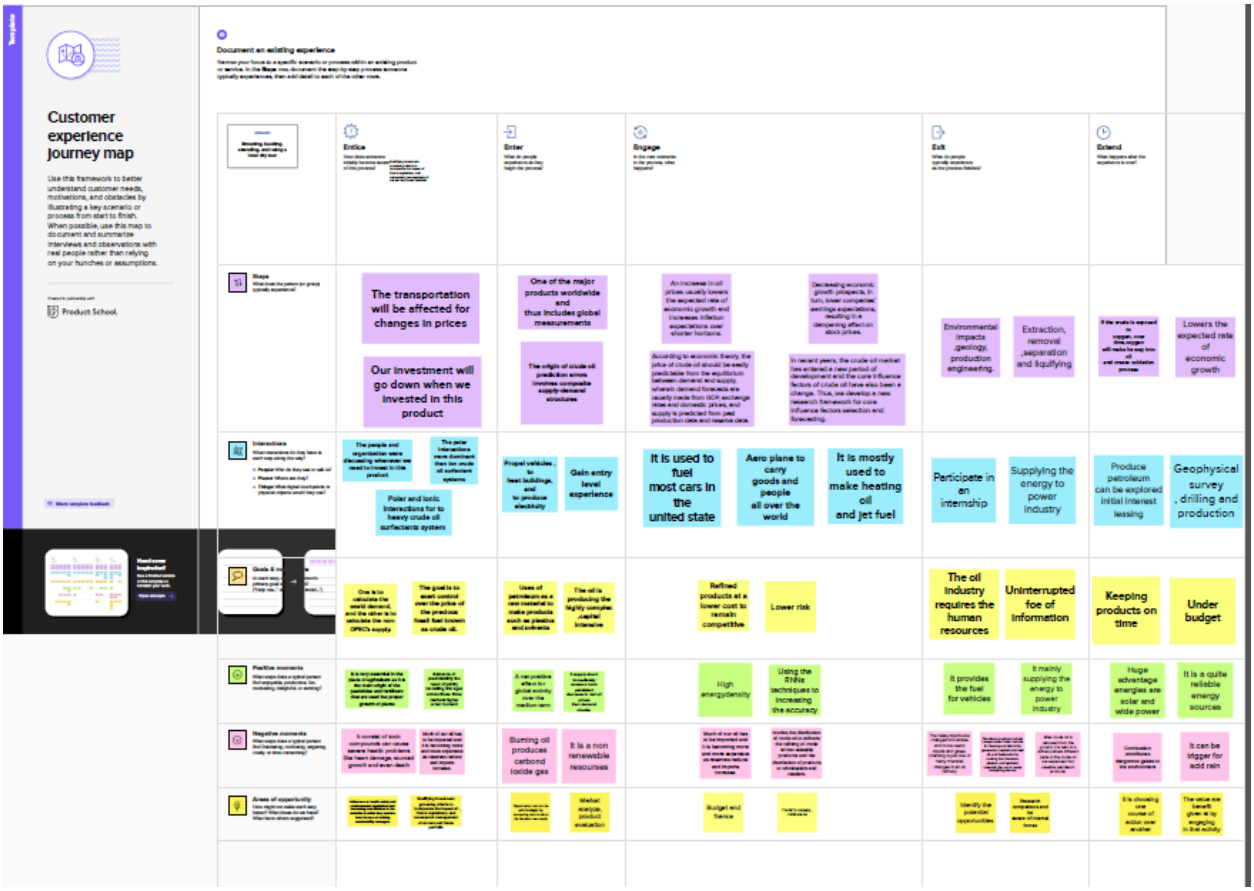
Following are the non functional requirements of the proposed solution.

SI NO	NON FUNCTIONAL REQUIREMENT	DESCRIPTION
01	Usability	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**
02	Security	Price forecasts are very important to various stakeholders: governments, public and private enterprises, policymakers,

		and investors.
03	Reliability	The predicted price of crude oil can not be trusted because the accuracy of crude oil price is not stable.
04	Performance	The performance of crude oil price price prediction is highly complicated since a greater number of test cases is done.
05	Availability	Availability of solutions is more beneficial for importers and exporters and as well as people around the world.
06	Scalability	Scalability is much greater(90% - 95%).

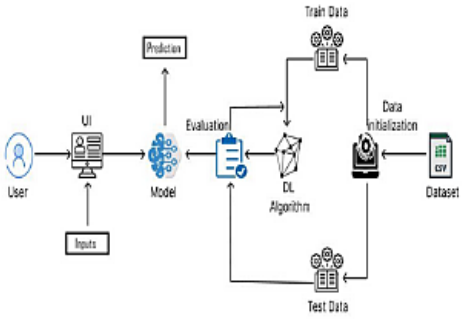
6.PROJECT DESIGN

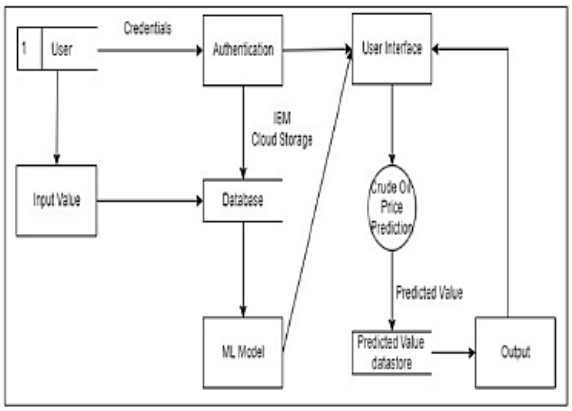
6.1 CUSTOMER JOURNEY MAP



6.2 DATA FLOW DIAGRAMS

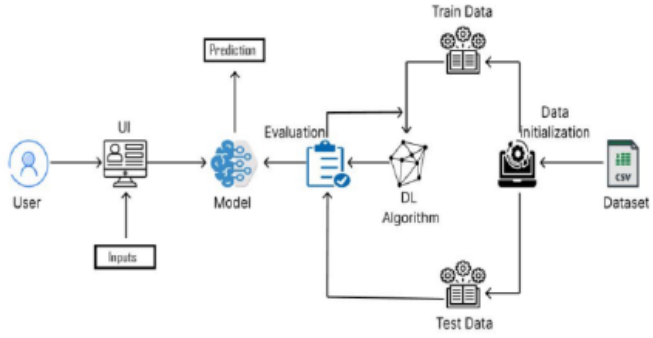
Data Flow Diagrams:





6.3 SOLUTION & TECHNICAL ARCHITECTURE

Technical Architecture:



- **Dataset is split into train and test datasets.**
- **Then, the training data is inputed LSTM model, and the model parameters are adjusted to achieve the optimal training model.**
- **The aim of this dataset and work is to predict future Crude Oil Prices based on the historical data available in the dataset.**
- **A comprehensive crude oil evaluation based on a detailed analysis is necessary to determine the value of these crude oils to the refinery.**
- **Once an ML algorithm is trained on a particular dataset and if you test it on the same dataset, it's more likely to have high accuracy because the model knows what to expect.**
- **User can know predicted price through the user interface.**

Table-1 : Components & Technologies:

S.No	Component Description	Technology
1.	User Interface User interacts with the application using website UI, which is used to get the various user needed various user information details from the website UI	HTML, CSS, JavaScript.
2.	Application Logic-1 This logic depends on the extracting the needed contents into the dataset.	Python
3.	Application Logic-2 This logic depends on training the dataset to get the accuracy by predicting the value.	Anaconda Jupyter or Google colab.
4.	Database Dataset is download , loaded and split into training and testing	Anaconda Jupyter
5.	Cloud Database Database Service on Cloud	IBM cloud,IBM Watson studio
6.	File Storage Massive amounts of data that the cloud environment must process in real time and store for later use.	IBM Block Storage or Other Storage Service or Local Filesystem
7.	Machine Learning Model It allows the user to feed a computer algorithm an immense amount of data and have the computer analyze and make data-driven recommendations and decisions based on only the input data.	Long Short Term Memory(LSTM)
8.	Infrastructure (Server / Cloud) Application Deployment on Local System / Cloud Local Server Configuration: Google server (Collab)	Local, Cloud Foundry.

Table-2: Application Characteristics:

S.No	Characteristics Description	Technology
1.	Open-Source Frameworks A software for which the original source code is made freely available and may be redistributed and modified according to the requirement of the user.	Python, Google colab, Anaconda Jupyter.
2.	Security Implementations IBM Watson studio Application Firewall provides security features that are in addition to many of the components identified in the recommended security framework. Firewall architecture is based on a shared library that can be easily updated when new security threats are identified.	Encryptions, Data isolation,Data protection,Transport layer security (TLS)protocol.
3.	Scalable Architecture Python is one of the pioneers of programming languages that developers can use to do all the scaling work. To improve scalability, you can enable or disable services run by the dispatcher on individual servers to balance the load for a given computer by request type.	Technology used in the architecture is that with Python and the IBM Watson studio.

6.4 USER STORIES

User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (common People, Business man)	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 / dashboard	High	Sprint-1
		USN-2	As a user, I can register for the application by entering my Facebook id.	I can register and access the dashboard with Gmail	High	Sprint-1
	Login	USN-3	As a user, I can log into the application by entering email & password	I can login and access using the email and password	Low	Sprint-2
	Dataset Management	USN-4	As a user, I split the dataset for training and testing.	Dataset split for training and testing.	High	Sprint-1
	Model Building	USN-5	As a user I build a model and train the model, test the model to predict the future price.	The model is build and model is trained and tested.	High	Sprint-1
Administrator	Login	USN-1	As an Administrator, I can login into the analysis page.	I can login using email and password.	High	Sprint-1
	Dashboard	USN-2	As an Administrator, I can access the Dashboard.	I can update the details of the features.	Medium	Sprint-2
	Authentication	USN-3	As an Administrator, I can verify the identity of the user.	I can check if the email and password is correct or not.	High	Sprint-1
	Authorization levels	USN-4	As an Administrator, I can determine the extent of system rights that the user has access to.	I can verify the user has been properly identified and authenticated.	Medium	Sprint-2

7. PROJECT PLANNING AND SCHEDULING

7.1 Sprint Planning & Estimation

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	7 Nov 2022	13 Nov 2022	20	29 Oct 2022
Sprint-2	20	6 Days	9 Nov 2022	15 Nov 2022		
Sprint-3	20	6 Days	17 Nov 2022	22 Nov 2022		
Sprint-4	20	6 Days	22 Nov 2022	19 Nov 2022		

7.2 Sprint Delivery Schedule

Product Backlog, Sprint Schedule, and Estimation (4 Marks)

Use the below template to create product backlog and sprint schedule

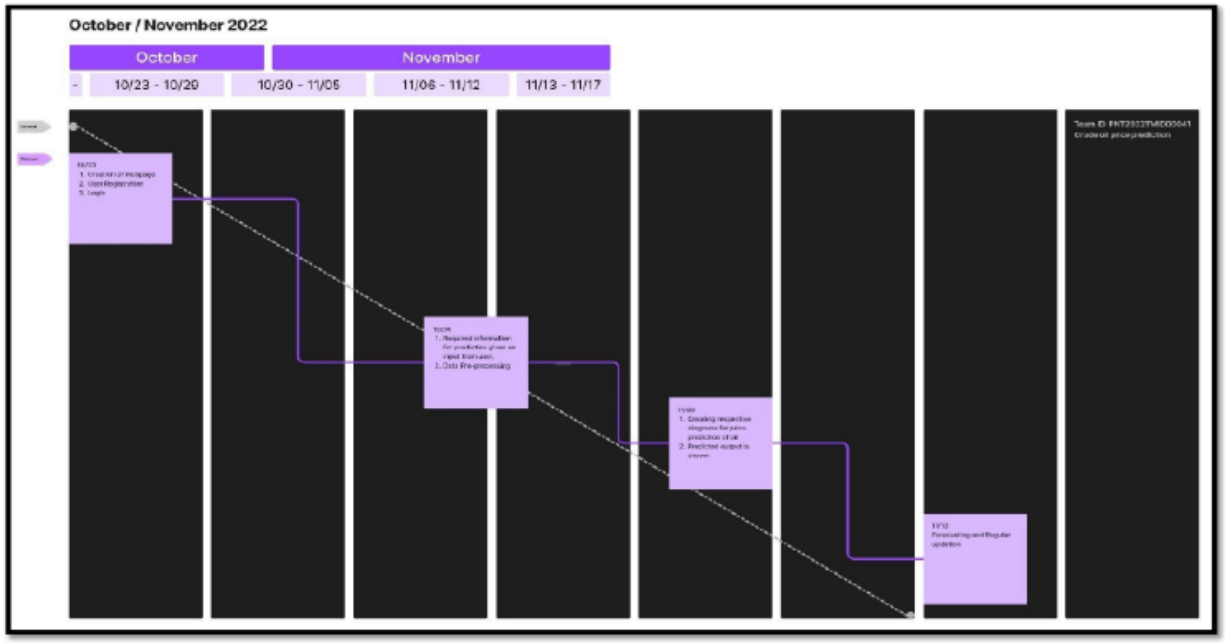
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Registration	USN-1	As a user, I can register for the application by entering my email, password, and confirming my password.	10	High	Sabitha C
Sprint-1		USN-2	As a user, I will receive confirmation email once I have registered for the application	10	High	Sivagnanam A V
Sprint-1	Login	USN-3	As a user, I can log into the application by entering email & password.	15	High	Indhumathi T
Sprint-2	Input Necessary Details	USN-4	As a user, I can give Input Details to Predict Likeliness of crude oil	15	High	Venkatramanan K
Sprint-2	Data Pre-processing	USN-5	Transform raw data into suitable format for prediction.	15	High	Lackshmi Prabha E

Sprint-3	Prediction of Crude Oil Price	USN-6	As a user, I can predict Crude oil using a machine learning model.	20	High	Venkatramanan K
Sprint-3		USN-7	As a user, I can get accurate prediction of crude oil	5	Medium	Sivagnanam A V
Sprint-4	Review	USN-8	As a user, I can give feedback on the application.	20	High	Indhumathi T

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	7 Nov 2022	13 Nov 2022	20	29 Oct 2022
Sprint-2	20	6 Days	9 Nov 2022	15 Nov 2022		
Sprint-3	20	6 Days	17 Nov 2022	22 Nov 2022		
Sprint-4	20	6 Days	22 Nov 2022	19 Nov 2022		

7.3 Reports from JIRA



8. CODING AND SOLUTIONING

templates > index.html > html > head > title

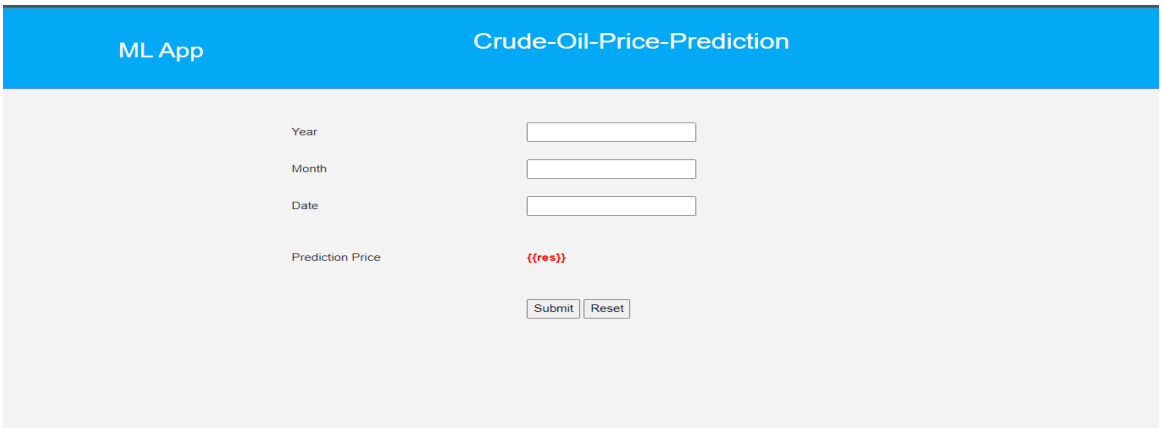
```
<!DOCTYPE html>
<html lang="en" dir="ltr">
  <head>
    <meta charset="utf-8">
    <title>Login & Signup for Crude Oil Prediction</title>
    <link rel="stylesheet" href="static/style.css">
    <meta name="viewport" content="width=device-width, initial-scale=1.0">
  </head>
  <body>
    <div class="wrapper">
      <div class="title-text">
        <div class="title login">Login Form</div>
        <div class="title signup">Signup Form</div>
      </div>
      <div class="form-container">
        <div class="slide-controls">
          <input type="radio" name="slide" id="login" checked>
          <input type="radio" name="slide" id="signup">
          <label for="login" class="slide login">Login</label>
          <label for="signup" class="slide signup">Signup</label>
          <div class="slider-tab"></div>
        </div>
        <div class="form-inner">
          <form action="predict.html" class="login">
            <div class="field">
              <input type="text" placeholder="Email Address" required>
            </div>
            <div class="field">
              <input type="password" placeholder="Password" required>
            </div>
            <div class="pass-link"><a href="#">Forgot password?</a></div>
            <div class="field btn">
```

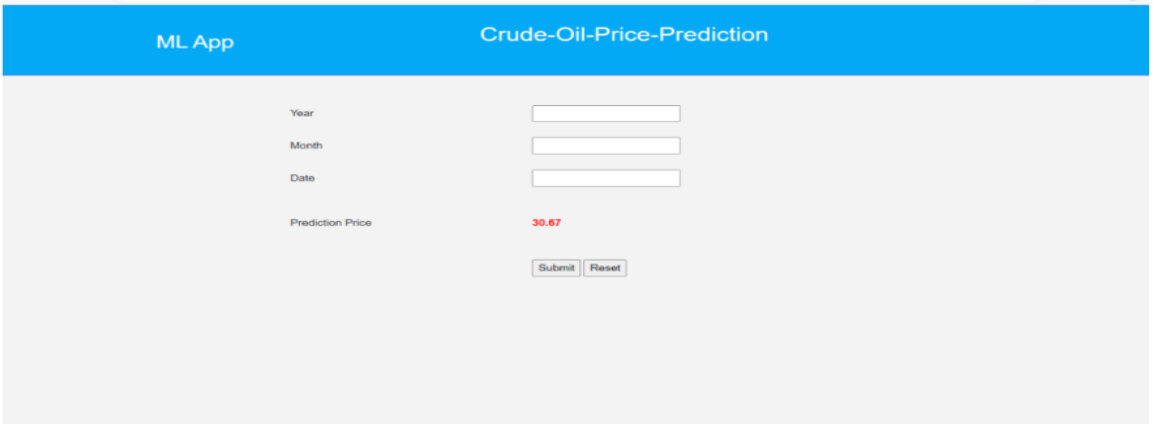
templates > index.html > html > head > title

```
            <form action="predict.html" class="login">
              <div class="field">
                <input type="text" placeholder="Email Address" required>
              </div>
              <div class="field">
                <input type="password" placeholder="Password" required>
              </div>
              <div class="pass-link"><a href="#">Forgot password?</a></div>
              <div class="field btn">
                <div class="btn-layer"></div>
                <input type="submit" value="Login">
              </div>
              <div class="signup-link">Not a member? <a href="">Signup now</a></div>
            </form>
            <form action="#" class="signup">
              <div class="field">
                <input type="text" placeholder="Email Address" required>
              </div>
              <div class="field">
                <input type="password" placeholder="Password" required>
              </div>
              <div class="field">
                <input type="password" placeholder="Confirm password" required>
              </div>
              <div class="field btn">
                <div class="btn-layer"></div>
                <input type="submit" value="Signup">
              </div>
            </form>
          </div>
        </div>
      </div>
```

```
    <script>
      const loginText = document.querySelector(".title-text .login");
      const loginForm = document.querySelector("form.login");
      const loginBtn = document.querySelector("label.login");
      const signupBtn = document.querySelector("label.signup");
      const signupLink = document.querySelector("form .signup-link a");
      signupBtn.onclick = (()=>{
        loginForm.style.marginLeft = "-50%";
        loginText.style.marginLeft = "-50%";
      });
      loginBtn.onclick = (()=>{
        loginForm.style.marginLeft = "0%";
        loginText.style.marginLeft = "0%";
      });
      signupLink.onclick = (()=>{
        signupBtn.click();
        return false;
      });
    </script>

  </body>
</html>
```


ML App
Crude-Oil-Price-Prediction
Year
Month
Date
Prediction Price
{{res}}
Submit
Reset


ML App
Crude-Oil-Price-Prediction
Year
Month
Date
Prediction Price
30.67
Submit
Reset

9. TESTING

9.1 TEST CASES

A test case has components that describe input, action and an expected response, in order to determine if a feature of an application is working correctly. A test case is a set of instructions on “HOW” to validate a particular test objective/target, which when followed will tell us if the expected behavior of the system is satisfied or not.

Characteristics of a good test case:

- Accurate: Exacts the purpose.
- Economical: No unnecessary steps or words.
- Traceable: Capable of being traced to requirements.
- Repeatable: Can be used to perform the test over and over.
- Reusable: Can be reused if necessary.

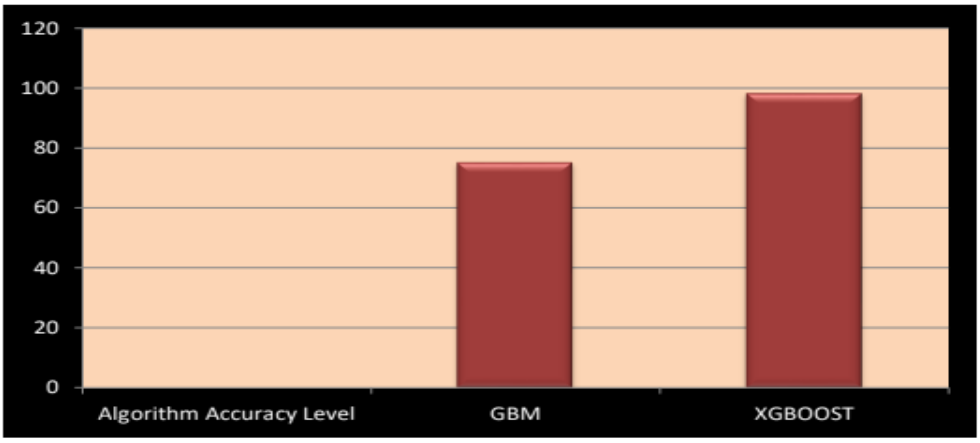
S.NO	Scenario	Input	Output	Status
1	User login	User name and password	Login	Login success.
2	Price Prediction	Give input in the user defined parameters	Predicting the crude oil prices by getting previous prices.	Details are stored in a database.

9.2 USER ACCEPTANCE TESTING

This sort of testing is carried out by users, clients, or other authorized bodies to identify the requirements and operational procedures of an application or piece of software. The most crucial stage of testing is acceptance testing since it determines whether or not the customer will accept the application or programme. It could entail the application's U.I., performance, usability, and usefulness. It is also referred to as end-user testing, operational acceptance testing, and user acceptance testing (UAT).

10. RESULTS

10.1 PERFORMANCE METRICS



11. ADVANTAGES & DISADVANTAGES

11.1 ADVANTAGE

- Give accurate result
- Easy to access and get the price
- Effective with large datasets

11.2 DISADVANTAGE

- Hard to find oil price
- Inefficient in accuracy
- Poor Customer support

12. CONCLUSION

Predicting Crude Oil prices is a very challenging problem due to the high volatility of oil prices. In this paper, we developed a new oil price prediction approach using ideas and tools from stream learning, a machine learning paradigm for analysis and inference of continuous flow of non-stationary data. Our stream learning model will be updated whenever new oil price data are available, and provided to model, so the model continuously evolves over time, and can capture the changing pattern of oil prices. In addition, updating the model requires only a small constant time per new data example, the experiment results show that our stream learning model outperformed four other popular oil price prediction models over a variety of

forecast time horizons. This process is used to Predict the oil Prices. The prediction model predicts continuous valued functions.

13. FUTURE SCOPE

Future research may extend our work by considering a richer set of market variables, such as political or commercial factors and phases of economic instability, which are often determinants of crude oil price. Moreover, another direction for future research is the application of the proposed model to forecast the price of other commodities. Moreover, it is a worthwhile direction to explore the consideration of one or more computational cost factors when comparing different forecasting models. Therefore, calculations based on operational research methods might be a good direction.

14. APPENDIX

GITHUB LINK

https://github.com/IBM-EPBL/IBM-Project-42892-1660710919

DEMONSTRATION