

**Crude Oil Price Prediction**

**SUBMITTED BY**

Team ID: PNT2022TMID31699

Team Size: 4

Team Lead: NAWIN RAJ KUMAR S (711719106029)

Team Member: DEVANSH SHUKLA R (711719106011)

Team Member: BHUVANESHWARI K(711719106008)

Team Member: RITHICKA S (711719106037)

|  | **TABLE OF CONTENTS** |  |
| --- | --- | --- |
| **CHAPTER** | **TITLE** | **PAGE.** |
| **NO** |  | **NO** |
| 1. | **INTRODUCTION** | 1 |
|  | 1.1 PROJECT OVERVIEW | 2 |
|  | 1.2 PURPOSE | 3 |
| 2. | **LITERATURE SURVEY** | 4 |
|  | 2.1 EXISTING PROBLEM | 8 |
|  | 2.2 REFERENCES | 8 |
|  | 2.3 PROBLEM STATEMENT DEFINITION | 9 |
| 3. | **IDEATION & PROPOSED SOLUTION** | 10 |
|  | 3.1 EMPATHY MAP CANVAS | 10 |
|  | 3.2 IDEATION & BRAINSTORMING | 11 |
|  | 3.3 PROPOSED SOLUTION | 12 |
|  | 3.4 PROPOSED SOLUTION FIT | 13 |
| 4. | **REQUIREMENT ANALYSIS** | 14 |
|  | 4.1 FUNCTIONAL REQUIREMENT | 14 |
|  | 4.2 NON-FUNCTIONAL REQUIREMENT | 15 |
| 5. | **PROJECT DESIGN** | 16 |
|  | 5.1 DATA FLOW DIAGRAMS | 16 |
|  | 5.2 SOLUTION & TECHNICAL ARCHITECTURE | 17 |
|  | 5.3 USER STORIES | 20 |
| 6. | **PROJECT PLANNING & SCHEDULING** | 22 |
|  | 6.1 SPRINT PLANNING & ESTIMATION | 22 |
|  | 6.2 SPRINT DELIVERY SCHEDULE | 23 |

i

| 7. | **RESULTS** | 24 |
| --- | --- | --- |
|  | 7.1 PERFORMANCE METRICS | 25 |
| 8. | **ADVANTAGES & DISADVANTAGES** | 26 |
| 9. | **CONCLUSION** | 28 |
| 10. | **FUTURE SCOPE** | 29 |
|  | **APPENDIX** | 30 |

ii

**CHAPTER – 1**

**INTRODUCTION**

Crude oil is a yellow black naturally occurring liquid found in geological formations beneath the Earth's surface, it can be separated into various kinds of consumer fuels through the process of fractional distillation. Crude oil is the most important energy resource on the Earth right now. So far, it remains the world’s leading fuel, with nearly one-third of global energy consumption. So, forecasting the price of crude oil is of great significance for energy policymakers, market participants, portfolio diversification, and energy risk management. There are many factors influencing the crude oil price, and the influence period of each factor on the crude oil prices is not consistent, so the crude oil prices have nonlinear characteristics. However, identifying the formation process of crude oil prices is of significance for accurate prediction, but this process is complicated. Due to strong chain effects owned by this crude oil market, any changes in the factors involved will have an exclusive impact on the price. Furthermore, the crude oil price contributes over 50% of the average price of petroleum and it is one of the most used commodities around the globe. Therefore, every increment and decrement that occurs to the crude oil price will then also give an impact to the price of petroleum and later correspond to the global economy. A good prediction tool is crucial to be developed for this matter. Therefore, we try to use the machine learning methods to deal with the vague influence among various factors. The formation process of crude oil prices can lead the traditional econometrics model to have a large error in crude oil price prediction, but the RNN and LSTM models can fit well. We have considered financialization of crude oil markets. The commodity attributes form the long-term trend of crude oil prices, and its financial attributes cause short-term fluctuation. In this paper, we try to forecast the price of crude oil from both spatial perspective and historical perspective.

1

**1.1 OVERVIEW**

Crude oil is the world's most leading fuel. The main advantages of crude oil are it has high density, it is easily available. Oil is used in almost all the industries. Oil is a Constant Power Source. Oil energy is very reliable when compared to other sources such as solar and wind energy. Some machine learning models fit the dataset efficiently depending upon the type of data points provided. The main aim of this project is to find the different models that efficiently fit the data points and predict the price of fuel with the help of machine learning models. This project works on comparing the different supervised learning models and brings a conclusion based on the efficiency. We have used the LSTM network to know which gives the best in terms of accuracy and performance. These algorithms give a numeric value as output. So we can compare the output of these models with the actual models. Now-a-days the oil price has been increasing in leaps and bounds due to certain reasons like inflation throughout the world. Hence these are derived or extracted from petroleum. The sources of crude oil for India come from neighboring countries such as Dubai and Saudi-Arabia. To predict the values of petroleum like petroleum and Diesel within the future, we've decided to use the Machine Learning algorithms and apply ensemble learning. Ensemble learning is a technique where we use different algorithms or single algorithms many times. In this way we can compare different algorithms and find the best one for our problem statement.

2

**1.2 PURPOSE**

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.

With high oil prices (and high gasoline prices), people will drive less - staying closer to home for shopping, combining various errands to be more efficient, and so on. Likewise, they will spend less on oil-derived products whose prices rise with higher oil prices.

If increased exploration and production is a normal byproduct of higher oil prices, so too is substitution. When Nazi Germany faced oil shortages in World War II, methods of producing oil, diesel and gasoline substitutes from vegetable oils, animal fats and coal were thoroughly explored. Likewise, the oil crisis of the 1970s gave the development of ethanol in Brazil a major boost.

Over time, though, more and more options become viable and greater changes in behavior are possible. Given time, people will drive less, take better care of their cars (to increase mileage), switch to more fuel-efficient car models and/or use more public transportation. Likewise, companies will find limits on just how much they can pass on higher input costs and will seek to reduce their usage of oil and oil by-products as well.

3

**CHAPTER – 2**

**LITERATURE SURVEY**

# Literature Survey

**Application of Traditional and Statistical Econometric Models:**

Academic scholars were the first to utilize conventional statistical and econometric approaches among the many forecasting models created to anticipate the price of "black gold." Amano suggests the initial investigation into oil market forecasting (1987). The author used a small-scale econometric model to forecast oil markets. Huntington (1994) utilized an advanced econometric model to predict oil prices in the 1980s. In a different study, Gulen (1998) forecasted the price of WTI crude oil using cointegration analysis. Barone-adesi et al. (1998) suggested a semi-parametric method based on the filtered historical simulation technique to forecast oil prices.

In a different study, Gulen (1998) forecasted the price of WTI crude oil using cointegration analysis. Barone-adesi et al. (1998) suggested a semi-parametric method based on the filtered historical simulation technique to forecast oil prices. In order to estimate the short-term price of Brent crude oil, Morana (2001) employed a semi-parametric technique based on the GARCH features of oil price volatility that were studied by Barone-adesi et al (1998). To predict OPEC basket prices, Tang and Hammoudeh (2002) employed a nonlinear model. Ye et al. (2002, 2005) created a straightforward linear regression model for the short-term.

monthly forecast of WTI crude oil spot price using OECD petroleum inventory levels and relative stock inventories. A related investigation by Ye et.

(2002, 2005) created a straightforward linear regression model for the short-term monthly forecast of the spot price of WTI crude oil using the OECD petroleum inventory levels and relative stock inventories. In a related study, Ye et al. (2006) modified the linear forecasting model presented by Ye et al. (2002, 2005) to predict short-term WTI crude oil prices by including nonlinear variables such low- and high-inventory variables. Using OECD stocks, non-OECD demand, and OPEC supply, Zamani (2004) forecasted the quarterly WTI crude oil spot price using an econometric forecasting model. Lanza and associates.

They is usage of error correction methods to analyse the price of crude oil and other products. Sadorsky (2006) used a variety of univariate and multivariate statistical models, including GARCH, TGARCH, AR, and BIGARCH, to predict daily volatility in petroleum futures price returns. With an emphasis on OPEC behaviour, Dees et al. (2007) created a linear model of the global oil market to forecast oil demand, supply, and prices. Murat and Tokat (2009) examined the connection between futures and spot crude oil prices, putting a random walk model to the test to see how well futures prices could predict spot price changes. Cheong (2009) used ARCH models to forecast the crude oil markets.

On the other hand, GARCH and different models from the GARCH family have been utilised in recent studies to forecast oil prices. For instance, using the GARCH model, Narayan and Narayan (2007) and Agnolucci (2009) predicted spot and futures crude oil prices. In a related study, Mohammadi and Su (2010) investigated the forecasting outcomes of various GARCH-type models in order to estimate the price of crude oil. CGARCH, FIGARCH, and IGARCH models were proposed by Kang et al. (2009) to predict the volatility.

The study by Kang et al. (2009) was expanded upon by Wei et al. (2010) utilising both linear and nonlinear GARCH-class models with the same objective. The use of linear approaches indicated a large discrepancy between oil price predictions and real prices. However, inventory, supply, and demand are the most often employed exogenous variables in these models to forecast oil prices. Inventory adjustments can be slow due to supply and demand being very inelastic to price changes, which accounts for a major share of the difference between actual and

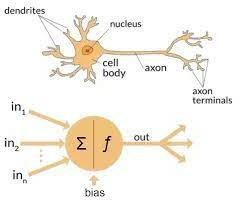
predicted prices, especially in the near run (Hamilton, 2008). On the other hand, conventional statistical and economic methods often can only capture linear processes in data time series (Weigend and Gershenfeld, 1994).

**Artificial Neural Network (ANN):**

**Definition and Neuron Model Evolution**

## Definition

ANN is an input-output mathematical model that mimics how the human brain functions by adopting the same strategy for learning new things. An equivalence between a biological and an artificial neuron is shown in Fig.1

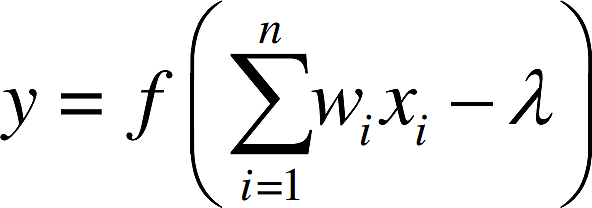


Comparison of the biological neuron (a) with the synthetic neuron (b)

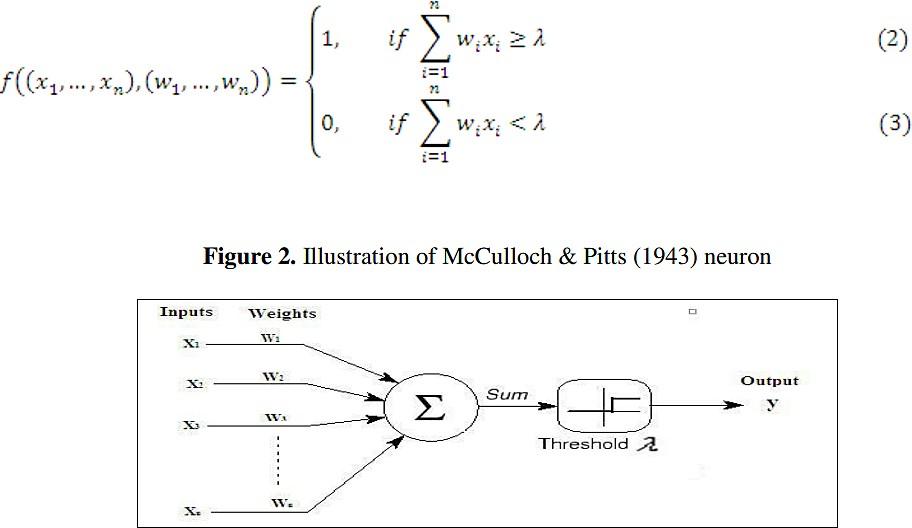
**Evolution of the Neuron Model**

1. **McCulloch & Pitts' Neuron Model (1943)**

McCulloch and Pitts' model of a neuron (1943). McCulloch and Pitts proposed the first synthetic neuron, sometimes known as a formal neuron (1943). The mathematical formulation of the McCulloch-Pitts neuron model is as follows:



The McCulloch-Pitts neuron's inputs, which are just binary numbers (zeroes or ones), are represented by 1 2 x, x,..., n x, while the weights it receives from connections are 1 2,,..., w w wn. The McCulloch-Pitts neuron's output is denoted by y, the threshold is denoted by, and the sign function is denoted by f.

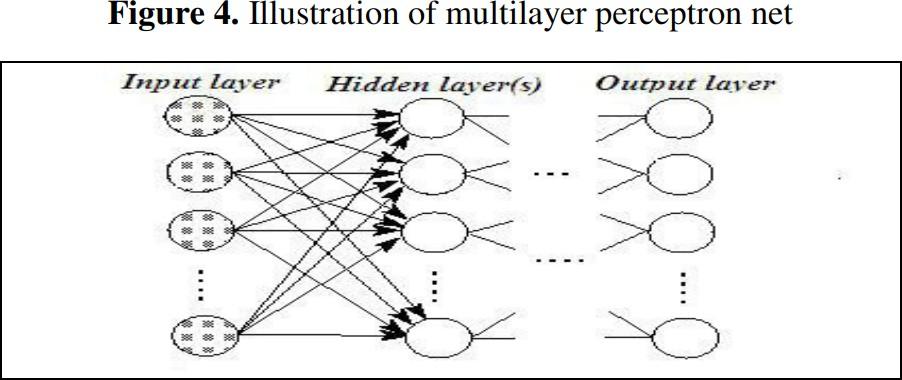


## Multilayer perceptron model

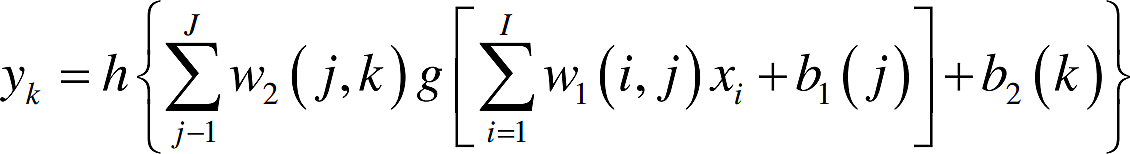
The model can only handle linearly separable functions because perceptron neural nets with no hidden layers assume only binary input-output values and only two layers. The delta rule was first put forth by Windrow and Hoff (1960). It

entails changing the connection weights in order to narrow the gap between the desired and actual output value. The output value can therefore accept any value in place of 0 and 1.

Minsky and Papert (1969) emphasized the value of incorporating one or more hidden layers to identify the complex properties contained in the inputs in their book. The learning algorithm developed by Rumelhart et al. backpropagation has traditionally been used to train multilayer perceptron nets (described in the following section) (1986).



In this network structure, information spreads in a single direction—forward—from the input units to the neurons in the first hidden layer, and then from the first hidden layer's outputs to the next layer and so on. The network output as a result is as follows (for instance, with one hidden layer):



Where I x represents the input variables for the network, I represents the total number of input variables, J represents the total number of nodes in the hidden layer, and K represents the total number of neurons in the output layer. The first and second layers' respective transfer/activation functions are denoted by the letters g and h; The weights matrix for the hidden layer is w1, while the weights matrix for the output layer is w2. It should be noted that at least one transfer

function of the hidden layer (which is further defined in the following section) must be nonlinear. The bias vectors of the hidden layer are 1 b and 2 b. (Hornik et al., 1989).

7

**2.1 EXISTING PROBLEM**

Several machine learning techniques were proposed for oil price prediction, such as artificial neural networks and support vector machine These are nonlinear models which may produce more accurate predictions if the oil price data are strongly nonlinear. However, these machine learning techniques, like other traditional machine learning techniques, rely on a fixed set of training data to train a machine learning model and then apply the model to a test set. Such an approach works well if the training data and the test data are generated from a stationary process but may not be effective for non -stationary time series data such as oil price data.

**REFERENCE**

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

8

**2.3 PROBLEM STATEMENT DEFINITION**

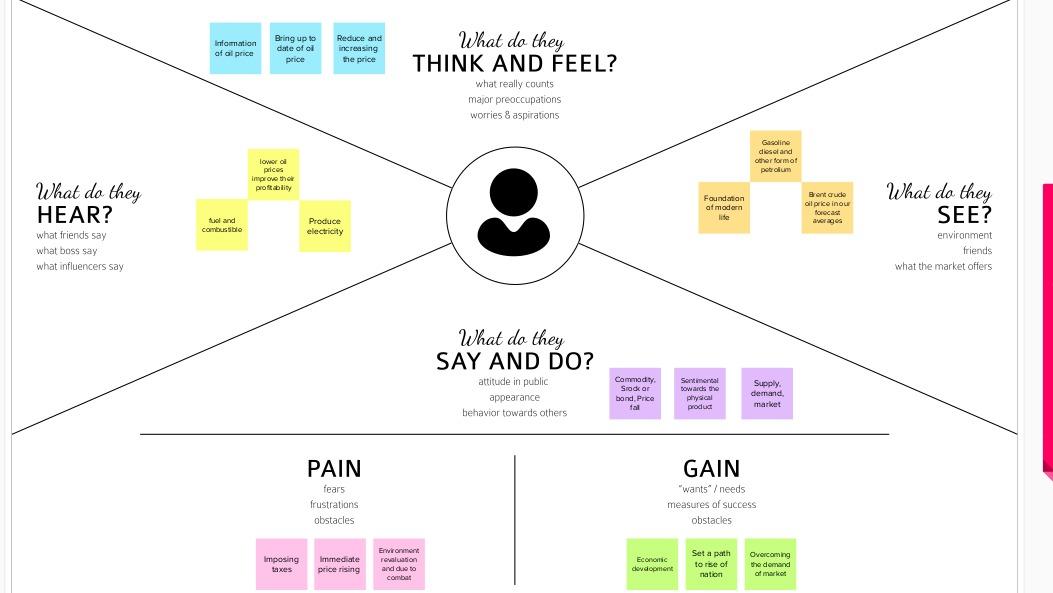
The crude oil price prediction task is interesting as well as divides researchers and academics into two groups those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction.

9

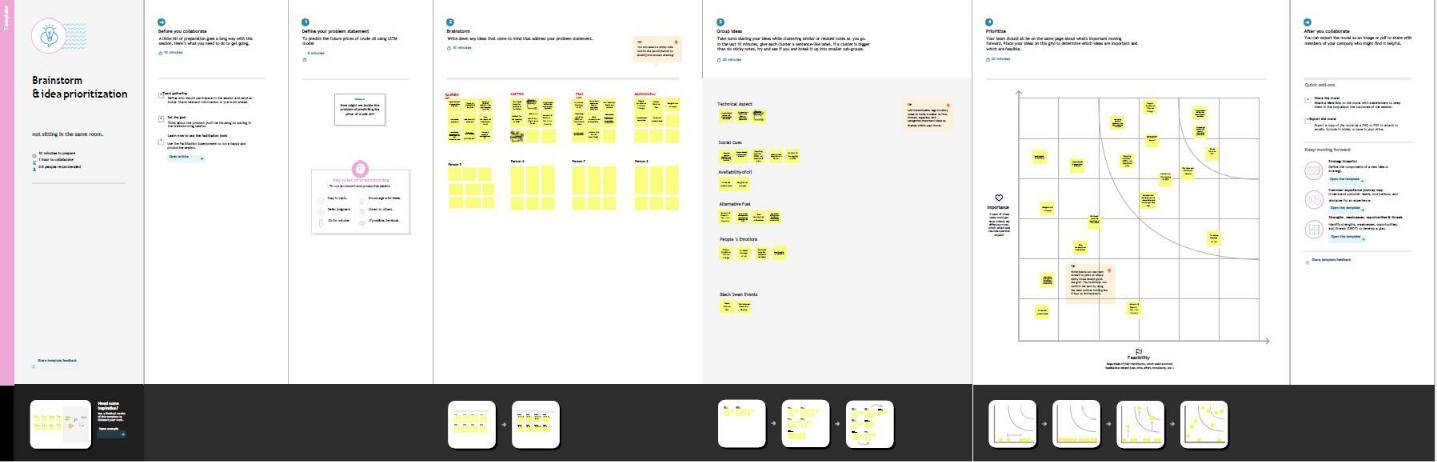
**CHAPTER-3**

**IDEATION & PROPOSED SOLUTION**

**3.1 EMPATHY MAP CANVAS**



**3.2 IDEATION & BRAIN STROMING**



11

**3.2 PROPOSED SOLUTION**

This paper describes the system that overcomes the problem faced while predicting the price of crude oil. Here we’ve compared machine learning algorithms using crude oil daily price dataset. We performed experiments with various algorithms on crude oil daily price dataset and observed the mean square error to predict accuracy using two algorithms namely Linear regression, Long Term Short memory. First, importing all the necessary libraries needed. Then importing the dataset of the respective company using csv read function. After that, building a Linear Regression model and predicting the future stock price. If we did not get the most accuracy, then we must build an LSTM model. LSTM model is considered as one of the effective ways to predict the future Stock prices. To build an LSTM model, we must split the dataset into Train and Test dataset. Then we must normalize the dataset. After the Future Scaling process, using plotting libraries, the model will predict the Future crude oil price with the most possible accuracy.

12

**3.3 PROPOSED SOLUTION FIT**



13

**CHAPTER – 4**

**REQUIREMENT ANALYSIS**

**4.1 FUNCTIONAL REQUIREMENT**

Following are the functional requirements of the proposed solution.

| **FR No.** | **Functional** | **Sub Requirement (Story / Sub-** |
| --- | --- | --- |
|  | **Requirement (Epic)** | **Task)** |
| FR-1 | User Application | User Direct Open with Google |
|  |  | Play Store App User Can |
|  |  | Download the Crude Oil Price |
| FR-2 | User Products Available | User Using the Application |
|  |  | There Are So Many Products in |
|  |  | Crude Oil Price App |
|  |  | User Update the Energy and Oil |
|  |  | Price Instant the Application |
| FR-3 | User Additional | User Can Read Latest News and |
|  | Features | View Oil Price Charts User View |
|  |  | Major Energy Quotes |
|  |  | User Can Using a Multiple |
|  |  | Colour Themes |
| FR-4 | User Exceptions | User Can Exchange Rates and |
|  |  | Currency Converter |

**4.2 NON-FUNCTIONAL REQUIREMENTS**

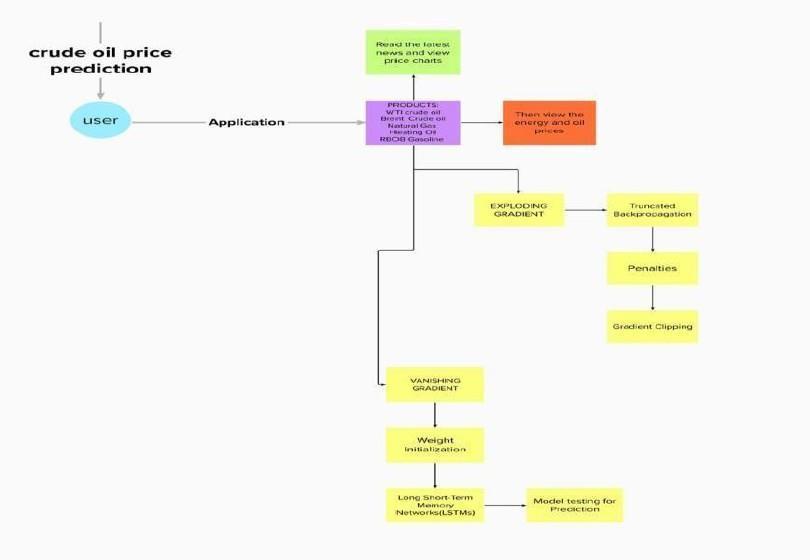
| **FR No.** | **Non-Functional** | **Description** |
| --- | --- | --- |
|  | **Requirement** |  |
| NFR-1 | **Usability** | Used to improve to the Accuracy |
|  |  | of crude oil price prediction |
| NFR-2 | **Security** | In the rising oil price can even |
|  |  | shift economical/political power |
|  |  | from oil importers to oil |
|  |  | exporters communications will be |
|  |  | secured |
| NFR-3 | **Reliability** | Reliability of the pointing towards |
|  |  | high –risk components |
|  |  |  |
| NFR-4 | **Performance** | Performance of this project is to |
|  |  | improve to the accuracy of crude |
|  |  | oil price prediction |
| NFR-5 | **Availability** | The Availability Solution is More |
|  |  | Benefit for and the Importers and |
|  |  | exporters in the crude oil price |
|  |  | prediction. |
| NFR-6 | **Scalability** | The scalability is 90%-95% |

**CHAPTER-5**

**PROJECT DESIGN**

**5.1 DATA FLOW DIAGRAMS**

The classic visual representation of how information moves through a system is a data flow diagram (DFD). A tidy and understandable DFD can graphically represent the appropriate quantity of the system demand. It demonstrates how information enters and exits the system, what modifies the data, and where information is kept.



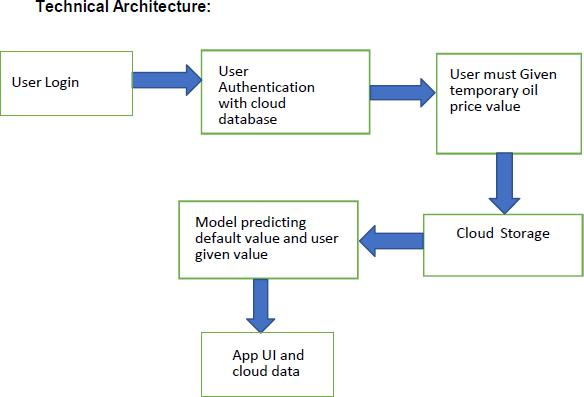
**5.2 SOLUTION & TECHNICAL ARCHITECTURE**

**Application Characteristics:**



| **S.No** | **Characteristics** | **Description** |  | **Technology** | | |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1. | Open-Source | Python, |  | Pandas, | | flask, | numpy, |
|  | Frameworks-1 |  |  | tensorflow | | |  |
|  |  |  | |  | | | |
| 2. | Open-Source | JavaScript, Angular Js. | | App module, component | | | |
|  | Frameworks-2 |  |  | module | |  |  |
| 3. | Security | User data will be stored | | End | to | end encryption | |
|  | Implementations | according to CIA | | (SHA- 256) | | |  |
|  |  | model. |  |  |  |  |  |
| 4. | Scalable Architecture | IBM cloud and firebase | | IBM | watson, | | Firebase, |
|  |  | both used for better | | Mysql | |  |  |
|  |  | performance in storage | |  |  |  |  |
|  |  | and authentication. | |  |  |  |  |
| 5. | Availability | Handle | huge | Effective | | coding and | |
|  |  | requests,avoid | DDOS | restrictive user | | |  |
|  |  | and XSS |  | access based on need | | | |
|  |  | attack. |  |  |  |  |  |
| 6. | Performance | Handle more than 1000 | | Flask |  |  |  |
|  |  | users to use server at a | |  |  |  |  |
|  |  | time. |  |  |  |  |  |

| **Components & Technologies:** | |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| **S.No** | **Component** | **Description** | **Technology** |
| 1. | User Interface | Web application | HTML, CSS, |
|  |  |  | JavaScript |
|  |  |  | ,Angular Js |
| 2. | Application Logic-1 | Logic for a process in the | Python |
|  |  | application |  |
| 3. | Application Logic-2 | Logic for a process in the | IBM Watson |
|  |  | application | Assistant |
| 4. | Database | Data Type, Configurations | MySQL |
| 5. | Cloud Database | Database Service on Cloud | IBM cloud |
| 6. | File Storage | File storage requirements | IBM Block |
|  |  |  | Storage,Local |
|  |  |  | Filesystem |
| 7. | External API-1 | Purpose of External API | Firebase |
|  |  | used in the |  |
|  |  | application |  |
| 8. | Machine Learning | Purpose of Machine | Recurrent neural |
|  | Model | Learning Model | network & LSTM |
|  |  |  |  |
| 9. | Infrastructure (Server | Application Deployment on | Local, Firebase. |
|  | / Cloud) | Local System / Cloud |  |
|  |  | Local Server Configuration: |  |
|  |  | Cloud Server Configuration: |  |



**5.3 USER STORIES**

| **User Type** | **Functional** | **User** | **User Story /** | **Acceptance** | **Priority** | **Release** |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Requirement** | **Story** | **Task** | **criteria** |  |  |
|  | **(Epic)** | **Number** |  |  |  |  |
|  |  |  |  |  |  |  |
| Customer | Application | USN-1 | You can | I can access | High | Sprint- |
| (Mobile user) |  |  | download the | own |  | 1 |
|  |  |  | crude oil price | decisions. |  |  |
|  |  |  | by opening the |  |  |  |
|  |  |  | Google Play |  |  |  |
|  |  |  | Store app |  |  |  |
|  |  |  | directly as a |  |  |  |
|  |  |  | user. |  |  |  |
|  |  |  |  |  |  |  |
|  | Available | USN-2 | Users of the | I can | High | Sprint- |
|  | Products |  | application may | receive the |  | 1 |
|  |  |  | instantly update | data once |  |  |
|  |  |  | the energy and | click then |  |  |
|  |  |  | oil prices while | confirm |  |  |
|  |  |  | using it because |  |  |  |
|  |  |  | there are so |  |  |  |
|  |  |  | many different |  |  |  |
|  |  |  | products |  |  |  |
|  |  |  | in the crude oil |  |  |  |
|  |  |  | price app. |  |  |  |
|  | Additional | USN-3 | Users can read | I can view | High | Sprint- |
|  | Features |  | the most recent | then read |  | 2 |
|  |  |  | news and see oil | the price |  |  |
|  |  |  | price charts. | prediction. |  |  |
|  |  |  | Major Energy |  |  |  |
|  |  |  | Quotes User |  |  |  |
|  |  |  | View |  |  |  |
|  |  |  | The user may |  |  |  |
|  |  |  | use many colour |  |  |  |
|  |  |  | schemes. |  |  |  |
|  | Expectations | USN-4 | User Can | I can expect | Medium | Sprint- |
|  |  |  | Convert |  |  | 2 |
|  |  |  | Currency And |  |  |  |
|  |  |  | Exchange Rates |  |  |  |
|  |  |  |  |  |  |  |

|  | Login | USN-5 | Log in as a user |  | High | Sprint 2 |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  | without using |  |  |  |
|  |  |  | your email |  |  |  |
|  |  |  | address, |  |  |  |
|  |  |  | username, |  |  |  |
|  |  |  | or password. |  |  |  |
| Customer | I can see the | USN-6 |  | I can view | High | Sprint 3 |
| (Web user) | price of crude |  |  | the price |  |  |
|  | oil as a |  |  | directly |  |  |
|  | consumer. |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Customer | I am the user | USN-7 |  | I can accept | medium | Sprint 4 |
| Care | and I |  |  | the terms |  |  |
| Executive | executive the |  |  |  |  |  |
|  | pricing |  |  |  |  |  |
|  | history. |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Administrator | As a | USN-8 |  | Show the | High | Sprint 4 |
|  | manager, it |  |  | result |  |  |
|  | anticipates |  |  |  |  |  |
|  | the results. |  |  |  |  |  |
|  |  |  |  |  |  |  |

**CHAPTER-6**

**PROJECT PLANNING & SCHEDULING**

**6.1 SPRINT PLANNING & ESTIMATION**

| **Sprint** | **Functional** | **User** | **User Story / Task** | **Story** | **Priority** | **Team Members** |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Requirement** | **Story** |  | **Points** |  |  |
|  | **(Epic)** | **Number** |  |  |  |  |
| Sp rint- | Registration | USN-1 | As a user, I can | 2 | High | NAWIN RAJ KUMAR |
| 1 |  |  | register for the |  |  |  |
|  |  |  | application by |  |  |  |
|  |  |  | entering my email, |  |  |  |
|  |  |  | password, and |  |  |  |
|  |  |  | confirming my |  |  |  |
|  |  |  | password. |  |  |  |
| Sprint- |  | USN-2 | As a user, I will | 2 | High | RITHICKA |
| 1 |  |  | receive |  |  |  |
|  |  |  | confirmation email |  |  |  |
|  |  |  | once I have |  |  |  |
|  |  |  | registered for the |  |  |  |
|  |  |  | application |  |  |  |
| Sprint- | Login | USN-3 | As a user, I can log | 2 | High | DEVANSH SHUKLA |
| 1 |  |  | into the application |  |  |  |
|  |  |  | by entering email & |  |  |  |
|  |  |  | password |  |  |  |
| Sprint- | Input | USN-4 | As a user, I can give | 14 | High | BHUVANESHWARI |
| 2 | Necessary |  | Input Details to |  |  |  |
|  | Details |  | Predict Likeliness |  |  |  |
|  |  |  | of crude Oil |  |  |  |
| Sprint- | Data Pre- | USN-5 | Transform raw data | 14 | High | NAWIN RAJ KUMAR |
| 2 | Processing |  | into suitable format |  |  |  |
|  |  |  | for prediction |  |  |  |
| Sprint- | Prediction of | USN-6 | As a user, I can | 18 | High | RITHICKA |
| 3 | Crude Oil |  | predict Crude Oil |  |  |  |
|  | Price |  | using machine |  |  |  |
|  |  |  | learning model. |  |  |  |
| Sprint- |  | USN-7 | As a user, I can get |  | Medium | DEVANSH SHUKLA |
| 3 |  |  | accurate prediction |  |  |  |
|  |  |  | of crude oil. |  |  |  |
| Sprint- | Review | USN-6 | As a user, I can give | 20 | High | BHUVANESHWARI |
| 4 |  |  | feedback of the |  |  |  |
|  |  |  | application |  |  |  |
|  |  |  | 22 |  |  |  |

**6.2 SPRINT DELIVERY SCHEDULE**

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

**CHAPTER-7**

**RESULTS**

**7.1 PERFORMANCE METRICS**

**Sensitivity:** Sensitivity is defined as the true-positive recognition rate, number of true positives / (number of true positives + number of false negatives) which is shown in equation 7.1.

| Sensitivity (TPR) = | (7.1) |
| --- | --- |

TP



TP+FN

**Specificity:** Specificity is defined as the proportion of actual negatives, which got predicted as the negative (or true negative) which is shown in equation in 7.2.

| Specificity (TNR) = | (7.2) |
| --- | --- |

TN



TN+FP

**Accuracy**: Accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training, data. Accuracy is shown in equation in 7.3.

| Accuracy = | (7.3) |
| --- | --- |

TP+TN



TP+TN+FP+FN

**Precision**: It is the quality of a positive prediction made by the model. Precision refers to the number of true positives divided by the total number of positive predictions. Precision is shown in equation 7.4.

| Precision = | (7.4) |
| --- | --- |

TP



TP+FP

**Recall:** Recall literally is how many of the true positives were recalled (found), such that how many of the correct hits were also found. Recall is shown in equation 7.5.

| Recall = | (7.5) |
| --- | --- |

TP



TP+FN

24

**F1-Score:** F1 Score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. F1 -Score is shown in equation 5.6.

| F1-Score = | (7.6) |
| --- | --- |

2



2 + +

| **Performance measure** | **LSTM** | **Proposed Network** |
| --- | --- | --- |
|  |  |  |
| **Accuracy** | 97.57% | 92.62% |
|  |  |  |
| **F-score** | 94.06% | 89.78% |
|  |  |  |
| **Recall** | 93.87% | 93.26% |
|  |  |  |
| **Sensitivity** | 86.79% | 82.61% |
|  |  |  |
| **Specificity** | 98.52% | 89.78% |
|  |  |  |
| **Precision** | 94.27% | 87.87% |
|  |  |  |

**CHAPTER 8**

**ADVANTAGES & DISADVANTAGES**

**ADVANTAGES:**

LSTM models have great advantages in terms of mining the long-term dependence of crude oil price sequence data. Furthermore, LSTM models can automatically search for nonlinear features and complex patterns of crude oil prices, which shows excellent forecasting performance in crude oil price prediction. As a very powerful prediction tool, LSTM has been widely used in prediction-related fields. Therefore, to forecast crude oil price more accurately, we have selected the LSTM model for this study.

The different gates inside LSTM boost its capability for capturing nonlinear relationships for forecasting. Causal factors generally have non -linear impact on demand. When these factors are used as part of the input variable, the LSTM could learn the nonlinear relationship for forecasting.

It is natural that events would impact demand on the day when it is happening as well as the days before and after the event is happening. For example, people would book more days of accommodation to attend a sports event. The LSTM could triage the impact patterns from different categories of events.

26

**DISADVANTAGES:**

They became popular since they solved the issue of gradients disappearing. However, they are unable to eliminate the problem. The issue lies in that data needs to be moved between cells for its analysis. Furthermore, the cell is becoming extremely complex with the addition of functions (such as the forget gate) that are now part of the picture.

LSTMs are affected by various random weights and behave similarly to neural networks that feed forward. They favor small initialization over large weights.

With the growing technology of data, mining scientists are searching for a system that can store past data for more extended periods of time than LSTMs. The motivation behind the development of such a model is the habit of humans of dividing a particular chunk of information into smaller parts to facilitate recollection.

**CHAPTER 9**

**CONCLUSION**

Determining an effective and efficient approach in predicting highly complex and volatile prices like crude oil is a critical and challenging task in an economy of a nation. Most of the prediction techniques are designed focusing on statistical and econometrics point of view which has been helpful in numerous scenarios, however prediction using powerful [AI tool](https://www.sciencedirect.com/topics/engineering/artificial-intelligence-tool) like the LSTM is very rare. In this paper, we proposed a new crude oil price prediction technique based on complex network analysis and LSTM. To evaluate the effectiveness and robustness of the technique, we conducted the experiment on ten different prices of crude oil across the world used by other researchers. From the experiment conducted we can conclude that, during the training process, the selection of batch size and number of LSTM layers has a great influence on the objective function value, fitting effect, and running time. The appropriate batch size and number of LSTM layers can effectively improve the model. Compared with the traditional and classic econometric prediction method, the model selects more datasets over a longer period as training samples. The LSTM prediction model has higher precision and wider application scenarios. The LSTM model can clearly predict the trend of crude oil price next time.

28

**CHAPTER 10**

**FUTURE SCOPE**

This paper only considers crude oil price in India, without necessarily considering other factors such as, financial market, economic growth, dollar exchange rate, demand and supply etc. The model proposed in this thesis is built based on monthly data, which restricts the prediction horizons to months. The proposed technique can be extended by considering other factors that affect crude oil price volatilities such as, financial market, economic growth, exchange rate, demand and supply and the weather. And the horizon of the prediction can be widened by considering daily data. The proposed technique can be implemented with different datasets such as the stock market data in the future to further check the validity of the proposed technique.

**APPENDIX**

**SOURCE CODE**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

import chart\_studio.plotly as py

import plotly.graph\_objs as go

from plotly.offline import plot

#for offline plotting

from plotly.offline import download\_plotlyjs, init\_notebook\_mode, plot, iplot

init\_notebook\_mode(connected=True)

oil = pd.read\_csv('COPP.csv')

oil.head()

oil.info()

oil['Date'] = pd.to\_datetime(oil['Date'])

print(f'Dataframe contains crude oil prices between {oil.Date.min()} {oil.Date.max()}')

print(f'Total days = {(oil.Date.max() - oil.Date.min()).days} days')

oil.describe()

oil[['Open','High','Low','Close','Adj Close']].plot(kind='box')

layout = go.Layout(

title=' Prices of crude oil ',

xaxis=dict(

title='Date',

titlefont=dict(

family='Courier New, monospace',

size=18,

color='#7f7f7f'

)

),

yaxis=dict(

title='Price',

titlefont=dict(

family='Courier New, monospace',

size=18,

color='#7f7f7f'

)

)

)

oil\_data = [{'x':oil['Date'], 'y':oil['Close']}]

plot = go.Figure(data = oil\_data, layout=layout)

iplot(plot)

LSTM

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

from sklearn.preprocessing import MinMaxScaler from keras.models import Sequential

from keras.layers import Dense,LSTM,Dropout data = pd.read\_csv('COPP TRAIN.csv') data.head()

data.info()

data["Close"]=pd.to\_numeric(data.Close,errors='coerce')

data = data.dropna()

trainData = data.iloc[:,4:5].values

data.info()

sc = MinMaxScaler(feature\_range=(0,1))

trainData = sc.fit\_transform(trainData)

trainData.shape

X\_train = []

y\_train = []

for i in range (60,165):

X\_train.append(trainData[i-60:i,0])

y\_train.append(trainData[i,0])

X\_train,y\_train = np.array(X\_train),np.array(y\_train)

X\_train = np.reshape(X\_train,(X\_train.shape[0],X\_train.shape[1],1)) #adding the batch\_size axis

X\_train.shape

model = Sequential()

model.add(LSTM(units=100, return\_sequences = True, input\_shape =(X\_train.shape[1],1)))

model.add(Dropout(0.2))

model.add(LSTM(units=100, return\_sequences = True))

model.add(Dropout(0.2))

model.add(LSTM(units=100, return\_sequences = True))

model.add(Dropout(0.2))

model.add(LSTM(units=100, return\_sequences = False))

model.add(Dropout(0.2))

model.add(Dense(units =1))

model.compile(optimizer='adam',loss="mean\_squared\_error")

hist = model.fit(X\_train, y\_train, epochs = 70, batch\_size = 32, verbose=2)

plt.plot(hist.history['loss'])

plt.title('Training model loss')

plt.ylabel('loss')

plt.xlabel('epoch')

plt.legend(['train'], loc='upper left')

plt.show()

testData = pd.read\_csv('COPP TEST.csv')

testData["Close"]=pd.to\_numeric(testData.Close,errors='coerce')

testData = testData.dropna()

testData = testData.iloc[:,4:5]

y\_test = testData.iloc[60:,0:].values

#input array for the model inputClosing = testData.iloc[:,0:].values inputClosing\_scaled = sc.transform(inputClosing) inputClosing\_scaled.shape

X\_test = []

length = len(testData)

timestep = 60

for i in range(timestep,length):

X\_test.append(inputClosing\_scaled[i-timestep:i,0])

X\_test = np.array(X\_test)

X\_test = np.reshape(X\_test,(X\_test.shape[0],X\_test.shape[1],1)) X\_test.shape

y\_pred = model.predict(X\_test)

y\_pred

predicted\_price = sc.inverse\_transform(y\_pred)

plt.plot(predicted\_price, color = 'green', label = 'Predicted crude oil Price')

plt.title('Crude Oil price prediction')

plt.xlabel('Time')

plt.ylabel('Price')

plt.legend()

plt.show()

print(predicted\_price)

**GITHUB LINK:** https://github.com/IBM-EPBL/IBM-Project-24723-1659947879

34