

Crude Oil Price Prediction

# SUBMITTED BY

Team ID: PNT2022TMID19030

Team Size: 4

Team Lead: RAJENDRAN S (1919103090)

Team Member: PRASANTH M (1919103087)

Team Member: KARTHIK PRASANNA V (1919103050)

Team Member: MADHAVARAJ A (1919103065)

# TABLE OF CONTENTS

|  |  |  |
| --- | --- | --- |
| **CHAPTER**  **NO** | **TITLE** | **PAGE.**  **NO** |
| 1. | **INTRODUCTION**   * 1. PROJECT OVERVIEW   2. PURPOSE | 1  2  3 |
| 2. | **LITERATURE SURVEY**   * 1. EXISTING PROBLEM   2. REFERENCES   3. PROBLEM STATEMENT DEFINITION | 4  8  8  9 |
| 3. | **IDEATION & PROPOSED SOLUTION**   * 1. EMPATHY MAP CANVAS   2. IDEATION & BRAINSTROMING   3. PROPOSED SOLUTION   4. PROPOSED SOLUTION FIT | 10  10  11  12  13 |
| 4. | **REQUIREMENT ANALYSIS**   * 1. FUNCTIONAL REQUIREMENT   2. NON-FUNCTIONAL REQUIREMENT | 14  14  15 |
| 5. | **PROJECT DESIGN**   * 1. DATA FLOW DIAGRAMS   2. SOLUTION & TECHNICAL ARCHITECTURE   3. USER STORIES | 16  16  17  20 |
| 6. | **PROJECT PLANNING & SCHEDULING**   * 1. SPRINT PLANNING & ESTIMATION   2. SPRINT DELIVERY SCHEDULE | 22  22  23 |

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

**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 resources 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 exclusive impact to the price. Furthermore, the crude oil price contributes over 50% on 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 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.

# 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 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 neighbouring 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.

# PURPOSE

Crude oil price fluctuations have a far-reaching impact on global economies and thus price forecasting can assist in minimising 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 by product 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 behaviour 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.

# CHAPTER – 2 LIERATURE SURVEY

**Crude Oil Price Prediction using Artificial Neural Network**

Crude oil is amongst the most important resources in today’s world, it is the chief fuel, and its cost has a direct effect on the global habitat, our economy and oil exploration, exploitation, and other activities. Prediction of oil prices has become the need of the hour, it is a boon to many large and small industries, individuals, the government. The evaporative nature of crude oil, its price prediction becomes extremely difficult, and it is hard to be precise with the same. Several different factors that affect crude oil prices. We propose a contemporary and innovative method of predicting crude oil prices using the artificial neural network (ANN). The main advantage of this approach of ANN is that it continuously captures the unstable pattern of the crude oil prices which have been incorporated by finding out the optimal lag and number of the delay effect that controls the prices of crude oil. Variation of lag in a period 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.

# Crude oil price prediction model with long short-term memory deep learning based on prior knowledge data transfer

Energy resources have acquired a strategic significance for economic growth and social welfare of any country throughout the history. Therefore, the prediction of crude oil price fluctuation is a significant issue. In recent years, with the development of artificial intelligence, deep learning has attracted wide attention in various industrial fields. Some scientific research about using the deep learning model to fit and predict time series has been developed. In an attempt to increase the accuracy of oil market price prediction, [Long Short Term](https://www.sciencedirect.com/topics/engineering/long-short-term-memory)

[Memory](https://www.sciencedirect.com/topics/engineering/long-short-term-memory), a representative model of deep learning, is applied to fit crude oil prices in this paper. In the traditional application field of long short term memory, such as [natural language processing](https://www.sciencedirect.com/topics/engineering/natural-language-processing), large amount of data is a consensus to improve training accuracy of long short term memory. To improve the prediction accuracy by extending the size of training set, transfer learning provides a heuristic data extension approach. Moreover, considering the equivalent of each historical data to train the long, short-term memory is difficult to reflect the changeable behaviours of crude oil markets, a very creative algorithm named data transfer with prior knowledge which provides a more availability data extension approach (three data types) is proposed. For comparing the predicting performance of initial data and data transfer deeply, the ensemble empirical mode decomposition is applied to decompose time series into several [intrinsic mode functions](https://www.sciencedirect.com/topics/engineering/intrinsic-mode-function), and these intrinsic mode functions are utilized to train the models. Further, the empirical research is performed in testing the prediction effect of West Texas Intermediate and Brent crude oil by evaluating the predicting ability of the proposed model, and the corresponding superiority is also demonstrated.

# Global crude oil price prediction and synchronization-based accuracy evaluation using random wavelet neural network

In the present paper, a new neural network is developed to improve the prediction accuracy of crude oil price fluctuations. The proposed model combines wavelet neural network (WNN) with random time effective function. WNN is a predictive system with the ability to implement strong nonlinear approximation. The random time effective function is applied to formulate the varied impact of historical data on current market, which endows historical data with time-variant weights to make them affect differently on the training process of WNN. Besides, the multiscale composite complexity synchronization (MCCS) is used as the new method to evaluate the predictive performance. The empirical experiments are implemented in predicting crude oil prices and moving average absolute return

series of WTI and BRE. Through comparing with the traditional [back](https://www.sciencedirect.com/topics/engineering/backpropagation) [propagation](https://www.sciencedirect.com/topics/engineering/backpropagation) neural network (BPNN), [support vector machine](https://www.sciencedirect.com/topics/engineering/support-vector-machine) (SVM) and WNN models, the empirical results demonstrate that the proposed model has a higher accuracy in crude oil price fluctuations predicting and is advantageous in improving the precision of prediction.

# Crude oil price prediction using complex network and deep learning algorithms

Crude oil price prediction is a challenging task in [oil producing countries](https://www.sciencedirect.com/topics/engineering/oil-producing-country). Its price is among the most complex and tough to model because fluctuations of price of crude oil are highly irregular, nonlinear and varies dynamically with high uncertainty. This paper proposed a hybrid model for crude oil price prediction that uses the complex [network analysis](https://www.sciencedirect.com/topics/engineering/electric-network-analysis) and long short-term memory (LSTM) of the deep learning algorithms. The complex network analysis tool called the visibility graph is used to map the dataset on a network and K-core centrality was employed to extract the non-linearity features of crude oil and reconstruct the dataset. The complex network analysis is carried out to pre-process the original data to extract the non-linearity features and to reconstruct the data. Thereafter, LSTM was employed to model the reconstructed data. To verify the result, we compared the empirical results with other research in the literature. The experiments show that the proposed model has higher accuracy and is more robust and reliable.

# [Crude Oil Price Prediction with Decision Tree Based Regression Approach](https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1391&context=jitim)

Crude oil is an essential commodity for industry and the prediction of its price is crucial for many business entities and government organizations. While there have been quite a few conventional statistical models to forecast oil prices, we find that there is not much research using decision tree models to predict crude oil prices. In this research, we develop decision tree models to forecast crude oil

prices. In addition to historical crude oil price time series data, we also use some predictor variables that would potentially affect crude oil prices, including crude oil demand and supply, and monthly GDP and CPI during the period 1992 through 2017 with a total of 312 observations. In this research, we use decision tree models to predict crude oil price. We find that the decision tree models developed in this research are expected to have higher forecasting accuracy than that of such benchmark models as multiple linear regression and time series autoregressive integrated moving average.

# 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.

# REFERENCES

1. Zhao, Y., Li, J. and Yu, L., 2017. A deep learning ensemble approach for crude oil price forecasting. Energy Economics, 66, pp.9-16.
2. Xie, W., Yu, L., Xu, S. and Wang, S., 2006, May. A new method for crude oil price forecasting based on support vector machines. In International conference on computational science (pp. 444-451). Springer, Berlin, Heidelberg.
3. Zhang, J.L., Zhang, Y.J. and Zhang, L., 2015. A novel hybrid method for crude oil price forecasting. Energy Economics, 49, pp.649-659.
4. Bashiri Behmiri, N. and Pires Manso, J.R., 2013. Crude oil price forecasting techniques: a comprehensive review of literature. Available at SSRN 2275428.
5. Li, X., Shang, W. and Wang, S., 2019. Text-based crude oil price forecasting: A deep learning approach. International Journal of Forecasting, 35(4), pp.1548- 1560.

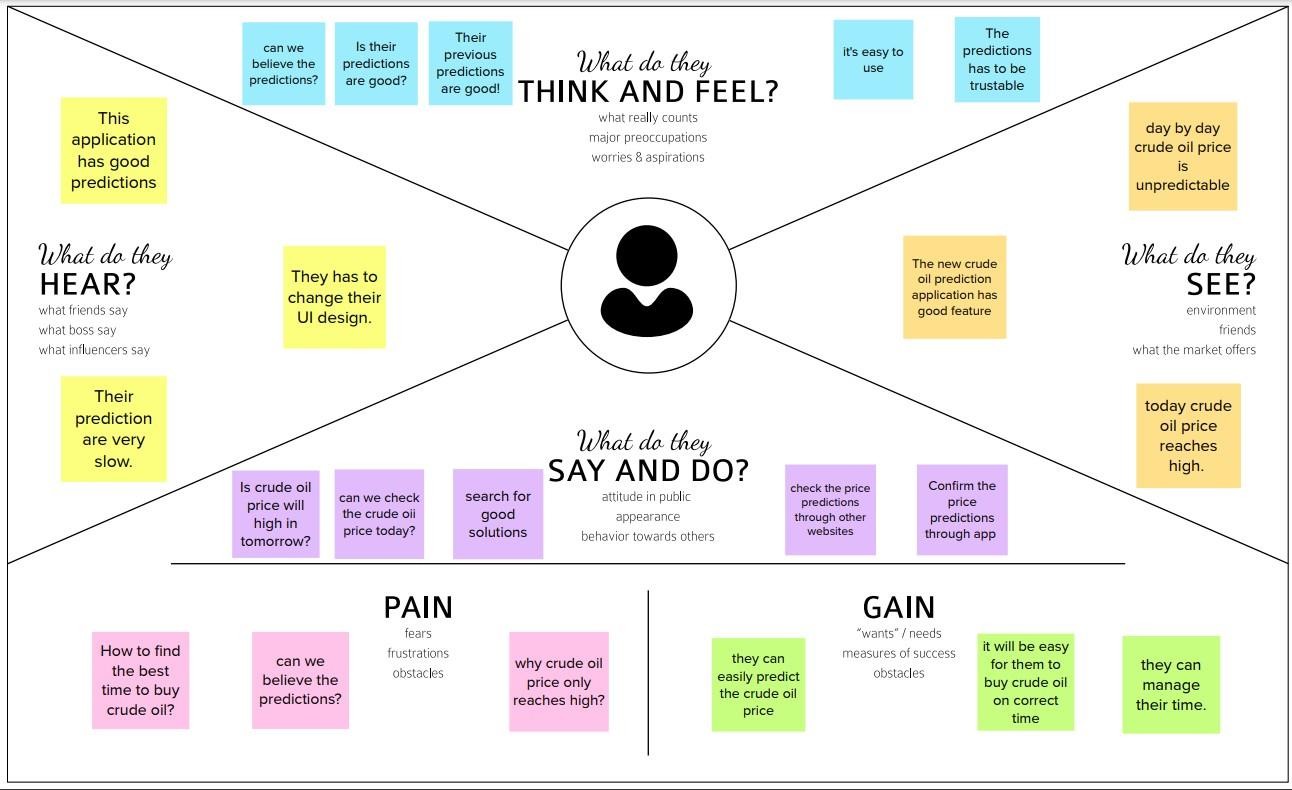
# PROBLEM STATEMENT DEFINITION

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.

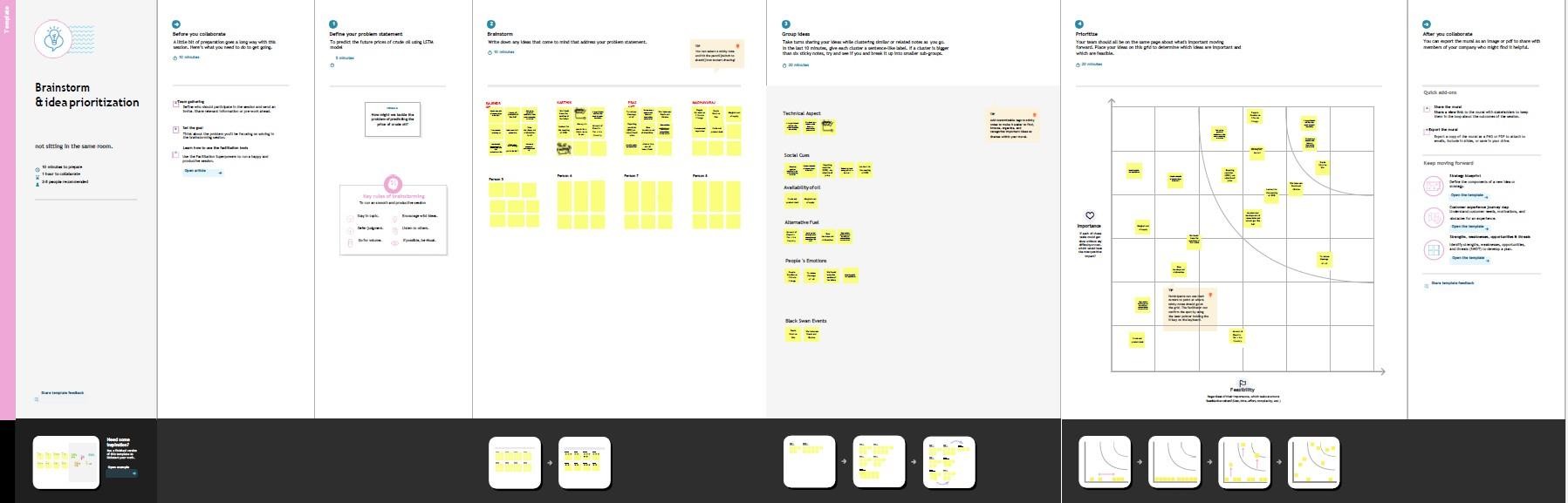
# CHAPTER-3

**IDEATION & PROPOSED SOLUTION**

# EMPATHY MAP CANVAS



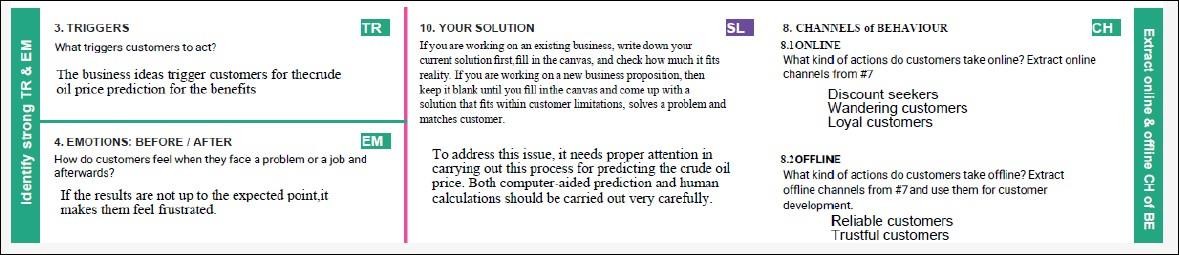
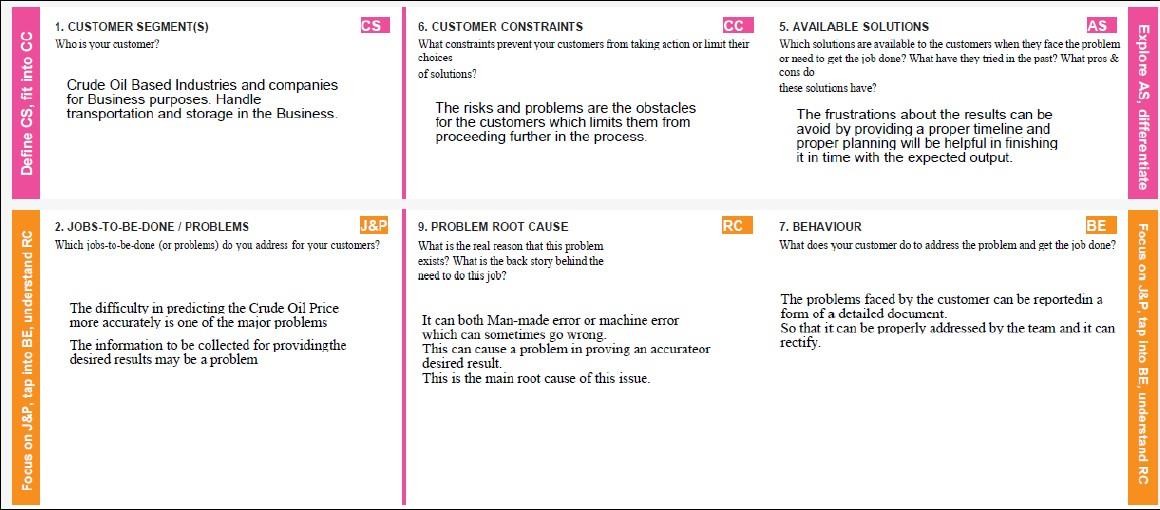
* 1. **IDEATION & BRAIN STROMING**



# 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 Future Scaling process, using plotting libraries, the model will predict the Future crude oil price with the most possible accuracy.

# PROPOSED SOLUTION FIT



**CHAPTER – 4 REQUIREMENT ANALYSIS**

# FUNCTIONAL REQUIREMENT

Following are the functional requirements of the proposed solution.

|  |  |  |
| --- | --- | --- |
| **FR No.** | **Functional**  **Requirement (Epic)** | **Sub Requirement (Story / Sub-**  **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 Features | User Can Read Latest News and 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 |

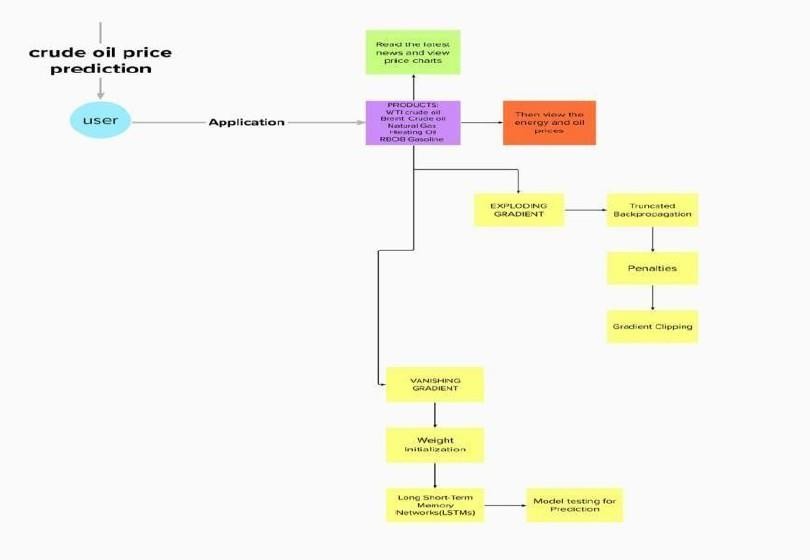
# NON-FUNCTIONAL REQUIREMENTS

|  |  |  |
| --- | --- | --- |
| **FR No.** | **Non-Functional Requirement** | **Description** |
| 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**

# 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.

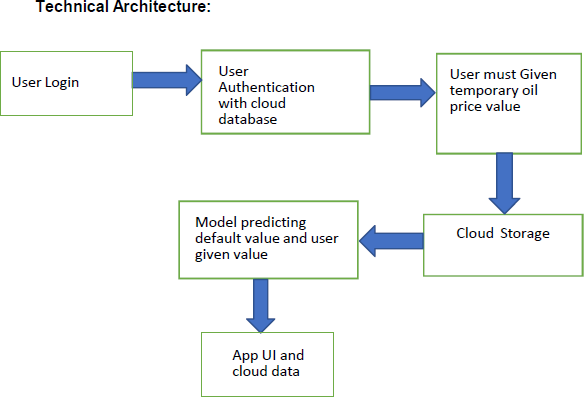


# SOLUTION & TECHNICAL ARCHITECTURE Application Characteristics:

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

**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 application | Python |
| 3. | Application Logic-2 | Logic for a process in the application | IBM Watson 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 used in the  application | Firebase |
| 8. | Machine Learning Model | Purpose of Machine Learning Model | Recurrent neural network & LSTM |
| 9. | Infrastructure (Server  / Cloud) | Application Deployment on Local System / Cloud  Local Server Configuration: Cloud Server Configuration: | Local, Firebase. |



# USER STORIES

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **User Type** | **Functional Requirement (Epic)** | **User Story Number** | **User Story / Task** | **Acceptance criteria** | **Priority** | **Release** |
| Customer (Mobile user) | Application | USN-1 | You can download the crude oil price by opening the Google Play Store app directly as a user. | I can access own decisions. | High | Sprint- 1 |
|  | 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 without using your email address, username,  or password. |  | High | Sprint 2 |
| Customer (Web user) | I can see the price of crude oil as a consumer. | USN-6 |  | I can view the price directly | High | Sprint 3 |
| Customer Care Executive | I am the user and I executive the pricing history. | USN-7 |  | I can accept the terms | medium | Sprint 4 |
| Administrator | As a manager, it anticipates the results. | USN-8 |  | Show the result | High | Sprint 4 |

**CHAPTER-6**

# PROJECT PLANNING & SCHEDULING

* 1. **SPRINT PLANNING & ESTIMATION**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **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. | 2 | High | RAJENDRAN |
| Sprint- 1 |  | USN-2 | As a user, I will receive confirmation email once I have registered for the  application | 2 | High | PRASANTH |
| Sprint- 1 | Login | USN-3 | As a user, I can log into the application  by entering email & password | 2 | High | KARTHIK PRASANNA |
| Sprint- 2 | Input Necessary Details | USN-4 | As a user, I can give Input Details to  Predict Likeliness of crude Oil | 14 | High | MADHAVARAJ |
| Sprint- 2 | Data Pre- Processing | USN-5 | Transform raw data  into suitable format for prediction | 14 | High | RAJENDRAN |
| Sprint- 3 | Prediction of Crude Oil Price | USN-6 | As a user, I can predict Crude Oil  using machine learning model. | 18 | High | PRASANTH |
| Sprint- 3 |  | USN-7 | As a user, I can get accurate prediction  of crude oil. |  | Medium | KARTHIK PRASANNA |
| Sprint- 4 | Review | USN-6 | As a user, I can give  feedback of the application | 20 | High | MADHAVARAJ |

# SPRINT DELIVERY SCHEDULE

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

TP+FN

(7.1)

**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) = TN

TN+FP

(7.2)

**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 = TP+TN

TP+TN+FP+FN

(7.3)

**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 = TP

TP+FP

(7.4)

**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 = TP

TP+FN

(7.5)

**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 = 2𝑇𝑃

2𝑇𝑃+𝐹𝑃+𝐹𝑁

(7.6)

|  |  |  |
| --- | --- | --- |
| **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.

# DISADVANTAGES:

They became popular since they solved the issue of gradients disappearing. However, that 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 favour 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 effective and efficient approach in predicting highly complex and volatile price 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 in the next time.

# CHAPTER 10 FUTURE SCOPE

This paper only considers crude oil price in India, without necessary considering other factors such as, financial market, economic growth, dollar exchange rate, demand and supply etc. The model proposed in this thesis in build based on monthly data, which restrict 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 dataset 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-24663-](https://github.com/IBM-EPBL/IBM-Project-24663-1659947207) [1659947207](https://github.com/IBM-EPBL/IBM-Project-24663-1659947207)