Republic of Iraq
Ministry of Higher Education and Scientific Research
University of Babylon
College of Science for Women
Department of Computer Science

Oil Price Prediction Using Deep Neural Network Technique of Gated Recurrent Unit (GRU) and Multivariate Analysis

A Project

Submitted to the Council of the College of Science for Women at University of Babylon in Partial Fulfillment of the Requirements for the Degree of High Diploma in Science \ Computer Science

By

Hanan Ghanim Abdulwahed

Supervised By

Prof. Dr. Samaher Al-Janabi

2022 A.D **1444 A.H**

بســـم الله الرحمــن الرحيم

﴿ يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنْكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ وَاللَّهُ بِمَا تَعْمَلُونَ خَبِيرٌ ﴾

صدق الله العلي العظيم

سورة المجادلة : أية [١١]

Supervisor Certification

I certify that this thesis entitled "**Oil Price Prediction Using Deep Neural Network Technique of Gated Recurrent Unit (GRU) and Multivariate Analysis**" was prepared at the Department of computer Sciences/ College of Science for Women/ University of Babylon, by (**Hanan Ghanim Abdulwahed**) as partial fulfillment of the requirements for the degree of Higher Diploma of Science in Computer Science.

Signature:

Name: Prof. Dr Samaher Al-Janabi

Date: / / 2022

Address : University of Babylon/College of Science for Women

The Head of the Department Certification

In view of the available recommendations, I forward the dissertation entitled "**Oil Price Prediction Using Deep Neural Network Technique of Gated Recurrent Unit (GRU) and Multivariate Analysis**" for debate by the examination committee.

Signature:

Name: **Asst. prof. Dr Saif Mahmood Alalak**

Date: / / 2022

Address : University of Babylon/College of Science forWomen

Certification of the Examining Committee

We, are the chairman and members of the examination committee, certify that we have read this project entitled (**Oil Price Prediction Using Deep Neural Network Technique of Gated Recurrent Unit (GRU) and Multivariate Analysis)** and after examining the higher diploma student **(Hanan Ghanim Abdulwahed)** in its contents in **16/10 /2022** , and that in our opinion and it is accepted as a project for the degree of high diploma in science/computer science with degree (**Excellent**) .

Committee Chairman :

Signature:

Name : Mohammed Abdullah Naser

Scientific order : Prof. Dr.

Date: / / 2022

Committee Member :

Signature:

Name: Mahdi Abed Salman

Scientific order : Asst. Prof. Dr.

Date: / / 2022

Committee Member (Supervisor) :

Signature:

Name: Samaher Hussein Ali Al-Janabi

Scientific order : Prof. Dr.

Date: / / 2022

Date of examination: 16\10\2022

Deanship authentication of college of science for women.

Approved for the college committee of grade studies.

Signature:

Name: Abeer Fauzi Murad Al- Rubaye

Scientific order : Prof. Dr.

Address: Dean of college science for women .

Date: / / 2022

Dedication

To "ALLAH" my Lord, my Creator, my Dependents and my Hope.

To the messenger who reached the valley of safety and advised the nation.

To the Prophet of mercy and the light of the worlds "Prophet Muhammad peace be upon him".

To my dear father, who taught me tender without waiting for someone who always prides me … asking "ALLAH" to extend in his life to witness my superiority and success after a long wait.

To my dear mother, her invitations are the secret of my success and her tenderness.

To my dear husband, support me and he was the secret of my success and to my children.

To my brothers and sisters who supported me all the way.

To all members of my family and friends who have spared no effortin my encouragement and support.

Hanan Ghanim Abdulwahed

2022

Acknowledgments

All Praises be to Almighty "ALLAH" who has enabled me to complete this task successfully and our utmost respect to His last Prophet"Mohammad",

I would like to express my gratitude and thanks to my supervisor **"Dr. Samaher Al-Janabi"** who has never stopped encouraging me and always guided me throughout this thesis.

I also would like to express my wholehearted thanks to my family **(father, mother, husband, children, sisters, and brother)** for their support they provided me

Last, I would also like to thank staff of Computer Science Department who have helped me and inspired me during my High Diploma study.

Hanan Ghanim Abdulwahed

2022

Abstract

The prediction of oil prices and determining the percentage of increase or decrease in their prices is one of the most important that one factors influencing the economy of countries and an effective impact on the budget of any oil country, including Iraq, and a key factor in the labor market in general. Prediction techniques represent an effective tool for discovering knowledge from large and complex databases in several areas, including oil prices. In this work, a prediction model will be designed and implemented that determines oil prices based on seven basic characteristics(Date, WTI, GOLD, SP 500, US DOLLAR INDEX, US 10YR BOND, DJU). The proposed model is based on deep neurocomputing technique through analysis is the 7 features to ten years. The proposed model consists of three main stages : The first stage called preprocessing includes checking outliers , missing values , and normalization the data to preparing it of the next stage , that determined the important of each feature through compute the correlation ,entropy and information gain . After that split the dataset into two parts ; training and testing. The first part of dataset is used to build the predictor called Hybrid Model to Oil Price Based on NeuroComputing Technique (HMOP–NCT) , While the second part is used to evaluate HMOP-NCT through three measures called (R2 , MSE and MAE). The HMOP-NCT prove its ability to providing good accept result of oil price with low error rates. The multivariate analysis prove with WTI,GOLD,USA DOLAR more affect in oil price and information gain for it are (WTI=11.272 , GOLD= 11.227 and DJU = 11.614).While Gate recurrent unit neurocomputing technique shown their ability to handle dataset have features different in behaviors for multi years. And give accept result with small prediction error in short time (R2=0.945,MSE=0.0505,MAE=0.1948).

Table of Contents

List of Tables

List of Figures

List of Algorithms

List of Abbreviations

DL	Deep Learning
GRU	Gated Recurrent Unit
IDA	Intelligent Data Analysis
LSTM	Long Short Term Memory
MAE	Mean Absolute Error
MSE	Mean Squared Error
MVA	Multivariate Analysis

Chapter One:

General Introduction

Chapter One: General Introduction

1.1 Introduction

Data represent one of the most precious items in the world and the foundations of various fields of computer science. Data refer to any object with such a defined set of features, a particular trait that is unique to that thing, or a collection of objects with their features. It needs several kinds and can be generated from observation, studies, or recording. In general, academics work with a concept known as data science, which combines three domains: data, intelligence, and statistics (Al.Janabi et al.,(2020)). There are many different types of data science, including small, normal, and big /huge datasets. Small data are sorted in a consistent way, such as a table or list, and their size is restricted with no and over 30 samples, As a result, it is not subject to the normal distribution and cannot be used to make any decisions. In contrast, normal data are structured data that are subject to the normal distribution and may make a variety of decisions, including (clustering, classification, prediction, optimization, etc.). Big data, on the other hand, can take on a variety of formats, including structured, semi-structured, or unstructured, and its size ranges from 1TB to 1ZB. By combining two key ideas—machine learning techniques and cloud computing (Al.Janabi et al.,(2021)).

Intelligent analysis of data is the study of analysing different forms of data effectively, concerned with extracting useful knowledge from this data and drawing techniques from a variety of fields, including artificial intelligence. IDA, in general, includes three stages: Preparation of data, data mining, data validation, and explanation, Where Preparation of data is a critical step in data analysis since it means converting the data into forms that can be used for very profitable purposes (Zabezhailo.et al.,(2019)).Second stage build model of data (.e., clustering , classification , predicted , etc.). Finally, data validation and explanation data is a final stage used to validate data and reliable tools. The advantage of IDA is, It works on a real problem, that is understood and worked on to obtain high-accuracy results. The drawback of IDA is The size of data sets in the real world is increasing, which leads to complex calculations that require a lot of time to reach the desired results. As a result, the main goal of IDA is to obtain knowledge. Based on machine learning, artificial intelligence, mainly, IDA helps to obtain useful information, necessary data, and interesting models with high accuracy(Peres.et.al.,(2018)).

The project that deals with subset of artificial intelligence is called deep learning. That uses algorithms to make computers able to work in a way similar to a human brain. It's a combination of three types: supervised learning, unsupervised learning, and reinforcement learning (Li .et.al,(2019)).

Prediction is one of the main principle related to design model of real problem that can guess what will happen in the future or find the unknown based on existing data or evidence. Prediction can happen in three different ways prediction methods are based on inputs: that depend on personal opinions and do not need data. Time-dependent prediction methods (various times). Past statements are used assuming that the present will be the same as the past. yet, prediction methods: are based on correlated models. Explanatory variables are used to predict the future (Alkaim A.F. et.al,(2020)).

Multivariate Analysis (MVA) is defined as a statistical procedure for analysing data, or as a process that includes multiple variables that lead to a single result. (MVA) is part of the exploratory data, meaning that there are different ways to conduct an (MVA) depending on the type of data and the problem. It is trying to solve. An (MVA) technique can be divided into two parts: The dependence technique, and the Interdependence technique. Where the dependence technique means that when one or some of the variables depends on other variables. This is seen in machine learning, where these methods are applied to create prediction models. The analyst enters the input data, selects the independent variables, and decides which ones to trust (Vighnesh D,(2021)) .While Interdependence techniques mean that no variables depend on other variables, so this technique is not dependent on cause and effect, but instead looks for correlations to explain several variables or groups them in meaningful ways (Hair, J.F., (2011)).The fundamental benefit of multivariate analysis is that it draws more accurate conclusions since it takes into account multiple independent variables that can affect the variability of dependent variables.

Since the price of oil is one of the key factors that influence the global economy, changes in the price have a significant impact on the nation's economy. Therefore, accurate forecasting of oil prices is a very difficult task. This study aims to define the variables or characteristics that influence oil prices. As result, the researcher strives to build a reliable prediction model to study the factors affecting oil and collect and analyze data, to obtain results of high accuracy(Al-Janabi.et.al,(2020)).

1.2 Statement of the Problem

The problem of this work is separated into parts : the first part is related to programming challenges while; the second part is related to application challenges; In general; the prediction techniques are split into two fields; prediction techniques related to data mining and predictions related to neurocomputing; this work deal with the second type of prediction technique.

- Gate recurrent unit (GRU) is one of the neurocomputing prediction techniques that characterized by many features that make it the best. these features.(i.e. ,GRU gives high accuracy results and work with real data but on other hand; the core of that algorithm is based on (try and error principle) in determined their parameters, also it have high computation. Therefore; the first challenge of this project is how can avoid these limitations (i.e., high computation and befit from their features).

- The problem of prediction the oil price need to high efficiency techniques; Therefore, the second challenge of this project is how can avoid these limitations thought build an efficient technique to predict oil price based on many features different in their behaviors ; it capture through ten years.

1.3 The Objectives of Project

The objective of this project is to design and implement a prediction model of oil prices based on multivariate variables through combination between Multivariate analysis represent by(correlation with information gain) with one of the deep neurocomputing technique called GRU .

1.4 Related Works

The prediction of oil prices and determining the percentage of increase or decline is one of the most significant things impacting in the economy of countries and an effective influence on the budget of any oil country, including Iraq. Therefore ;this section of the project will attempt to review the previous researchers' work in the same field as our problem and compare it based on five points: author(s) name, database/dataset used, Pre-processing and methodology measure, as explained in table 1.1 :

Xu .et.al,(2022)[1], investigated the relationship between China's actual economic activity and the volatility of the world oil price using a vector auto-regression model with stochastic volatility in the mean.. When it was discovered that an increase in the uncertainty of the international oil price by one standard deviation lowers electricity generation by roughly 0.2%. These research work is similar with that work through working on the same types of dataset but using different pre-processing and methodology.

Abbas’s .et.al,(2022)[2], used (QARDL) approach to calculate the monthly average empirically return of the traditional stock market index. Geopolitical oil price risk, the main variable of concern, has considerable results for both assets in a bullish scenario. Additionally, when there is a bearish trend, the Islamic stock market's worldwide interest rate and exchange rate both move in the same way. our work differed from this work in terms of the method of data processing , deep neurocomputing technique with evaluation measures.

Sun.et.al,(2022)[3], compared crude oil price with the current price in China to see if there is any correlation. The model that was used reveals that the price of crude oil has changed over the previous time period. The results of the dummy tests show that it has clear effects on the introduction of crude oil futures contracts. The USD/CNR exchange rate has a beneficial effect on crude oil futures prices. While , our work determine the oil price through analysis seven different features based on multivariate analysis .

He et al. (2022)[4], proposed a new hybrid forecasting model to anticipate the trend in the price of crude oil using VMD and ML algorithms. The hybrid prediction model with a support vector machine classifier displays greater predictive power when compared to the other classifiers .Also, It is give more accurate at forecasting high volatility than low volatility in crude oil prices.

Gupta et al. (2020)[5], proposed a contemporary and novel approach to predict oil prices using an artificial neural network that is characterized as continually consider the shifting pattern of oil price variations by determining the appropriate delay and the number of delay effects that regulate oil prices. For the most accurate and near results, the lag in time1 was varied. Next tested our findings by calculating the RMSE, and the results provided by the proposed model far outperformed those of the control group. our work is similar to this work in terms of data analysis and the use of intelligent methods of processing to obtain the results, but we differ in terms of the prediction model to reach the desired values and results.

Carpio et al. (2019)[6], analyzed the long-term and short-term implications of oil prices on ethanol, gasoline, and sugar price forecasts. The model of vector error correction with exogenous variable (VECX) is employed. The data show that the oil price projection has long-term implications for the other three price estimates, In the near run, Forecasts for ethanol and gasoline costs are more susceptible to changes in upcoming oil prices than forecasts for sugar prices. while sugar's future volatility is lower than the other two prices. These statistics can help sugarcane and energy industry representatives make strategic decisions. Our work is similar to that work by using the same type of dataset ,but , we used different prediction technique called Gate recurrent unit.

Table 1.1 Comparison among Previous Works

Name	Dataset/database	Preprocessing	Methodology
Xu, Q., Fu, B., & Wang, B. (2022)	The consequences of oil price volatility on China's economy https://doi.org/10.1016/j.eneco.2022.105840	autoregression model (AR-model)	New-Keynesian model
Abbass, K., Sharif, A., Song, H., Ali, M. T., Khan, F., & Amin, N. (2022)	Geopolitical oil price risk https://doi.org/10.1016/j.resourpol.2022.102730	autoregressive distributed lag (ARDL) models	QARDL
Sun, C., Zhan, Y., Peng, Y., & Cai, W. (2022)	Crude oil price and exchange rate https://doi.org/10.1016/j.eneco.2021.105707	MS-VAR model	Markov-switching vector autoregressive

Name	Dataset/database	Preprocessing	Methodology
			model
He, H., Sun, M., Li, X., & Mensah, I. A. (2022)	Crude oil price https://doi.org/10.1016/j.energy.2021.122706	Multimodal data	Variational mode decomposition (VMD) , (ML) algorithm .
Gupta, N., & Nigam, S. (2020).	Predictions for Oil Prices https://doi.org/10.1016/j.procs.2020.03.136	Artificial neural network (ANN) model	Oil Price Prediction Model
Carpio, L. G. T. (2019)	The impact of fluctuating oil prices on estimates of the prices of ethanol, gasoline, and sugar https://doi.org/10.1016/j.energy.2019.05.067	long-term and short-term model	(VECX) model

1.5 Project Layout

The current and remaining chapters of this project are organized in the following sentences:

- Chapter one makes a sense of the general concepts of the main tasks of the problem and introduces the aims of the project.
- Chapter two focuses on the theoretical concepts related to this work.
- Chapter three explains the proposed model. Also shown each phases of the model and analysis it .
- Chapter four demonstrates the model's application and the case results study's.
- Chapter five shows the conclusion of this study and some suggestions for further research in this area.

Chapter Two:

Theoretical Background

Chapter Two: Theoretical Background

2.1 Introduction

One of the driving forces behind the development of the world economy is crude oil. Consequently, crude oil price oscillations must occur and greatly affect the economies of countries (Li et al.,2021). Therefore, forecasting oil prices is a key factor in the lab or market. As a result, the researcher will discuss how to design and implement a forecasting model that can determine oil prices based on basic characteristics in this chapter, Given the advantage of forecasting techniques for their ability to discover knowledge from large and complex databases in several areas, including oil prices. As a result, to solve the problems of determining oil prices ,the researcher will employ a multivariate forecast that takes into account a variety of indicators, including crude oil prices, macroeconomic indicators, political indicators, and technical indicators (Good, 2005). It is feasible to build or modify data based on forecast findings since a variety of factors may have an impact on crude oil prices. However, compared to the rapid changes in crude oil, other indicators lag behind them significantly, which results in multivariate forecasting being preferred for daily oil price predictions. the long term. To this end, deep learning forecasting techniques have been used to forecast the pricing of crude oil in recent years (Wang et al.,2018). To profit from it in the following stage, which modifies the new data, a prediction model (GRU) with an internal memory capable of preserving crucial information from the previous step was chosen. Missing data (any external event such as data entry error or data collection problems or respondent-related activity (e.g., refusal to answer) that results in missing data. Extraordinary events can be explained, while extraordinary observations have no explanation and fall within the normal range of values on each of the variables. Where Missing data has negative effects.

2.2 Multivariate Analysis

Multivariate Analysis (MVA) is defined as a statistical procedure for analyzing data, or a process dedicated to analyzing data sets with more than one variable. This means that takes a whole host of variables Which makes it a complicated as well as an essential tool. Where The major advantage of such a model is that it takes into account as many aspects as feasible. As a consequence, bias is reduced and the outcome is as near to reality as possible (Vighnesh D,2021). MVA entails the observation and analysis of many statistical result variables at the same time. Where the approach is employed across several dimensions while accounting for the impacts of all factors on the answers of interest, and it is especially useful when working with correlated data (Grace et al.,2018) , Due to the fact that everything that occurs in the world uses multivariate equations, several issues exist is the result of multiple reasons(Dutta,2021).

Dependence and interdependence techniques make up the two categories of an (MVA) technique. Where the dependency approach may be characterized as one in which one or more variables are labelled as a dependent variable or set of variables to be anticipated, explained by other variables known as independent variables, or may be characterized as one or more variables depending on other variables. For example machine learning, where these techniques are used to build predictive models, where the analyst enters the input data and determines the independent variables and which ones to depend on. An example of a dependence technique is Multiple regression(Gottumukkala ,2018).

Dependence techniques can be classified depending on the number of dependent variables and the measurement scales the variables use (Hair et al., 2011) :

Dependence approaches can be split into three categories, depending on the number of dependent variables: those with a single dependent variable, several dependent variables, or multiple dependent or independent connections (Gottumukkala,2018).

The second category of dependent strategies is those with either metric (quantitative or numeric) or non-metric (qualitative or categorical) dependent variables (Gottumukkala,2018).

Where If there is just one metric-dependent variable in the study, the best strategy to employ is :

1- Multiple regression analysis be according to this equation :

R1 = K1 + K2 + …… + Kn ……………(2.1)

(metric) (metric, nonmetric)

2- Conjoint analysis be according to this equation :

R1 = K1 + K2 + …… + Kn ……………(2.2)

(metric, nonmetric) (nonmetric)

While, If a single dependent variable is nonmetric (categorical) the appropriate methods are:

3- Multiple discriminant analysis be according to this equation:

R1 = K1 + K2 + …… + Kn ……………(2.3)

(nonmetric) (metric)

4- Linear probability models may be according to this equation:

R1 = K1 + K2 + …… + Kn …………(2.4)

(nonmetric) (metric, nonmetric)

5- Logistic regression be according to this equation:

$$F(R1) \quad = \quad K1 + K2 + \ \ldots\ldots + Kn \ \ldots\ldots\ldots\ldots\ldots \ (2.5)$$

(nonmetric) (metric, nonmetric)

If the problem had several dependent variables, there are other techniques used, if several dependent variables are metric and independent variables are nonmetric are :

6- In the case of a single dependent variable Analysis of Variance (ANOVA) is used:

$$R1 \quad = K1 + K2 + \ \ldots\ldots + Kn \ \ldots\ldots\ldots\ldots\ldots \ (2.6)$$

(metric) (nonmetric)

7- Multivariate analysis of variance (MANOVA) is used :

$$R1 + R2 + \ldots. + Rn \quad = K1 + K2 + \ \ldots\ldots + Kn \ \ldots\ldots\ldots (2.7)$$

(metric) (nonmetric)

8- Canonical correlation is suitable when there are numerous metric-dependent variables and metric independent variables. Dummy variable coding can be used to change non-metric dependent variables into metric ones so that canonical correlation can be employed once more. Together it can be represented as below:

$$R1 + R2 + \ldots. + Rn \quad = \quad K1 + K2 + \ \ldots\ldots + Kn \ \ldots\ldots..(2.8)$$

(metric, nonmetric) (metric, nonmetric)

9- Structural Equation Modelling (SEM) is used when a collection of dependent or independent variable connections must be assumed (Gottumukkala,2018) according to:

$$R1 = K11 + K12 + \ldots\ldots + K1n \ldots\ldots\ldots\ldots(2.9)$$
$$R2 = K21 + K22 + \ldots\ldots + K2n \ldots\ldots\ldots\ldots(2.10)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$Rn = Km1 + Km2 + \ldots\ldots + Kmn$$
(metric) (metric, nonmetric)

Metric data, also known as quantitative data, interval data, or ratio data, are measurements that identify or describe people (or things) based not only on the presence of an attribute but also on the extent to which the characteristic may be utilized to define the subject. Metric data examples include age and weight (Chiou,2005).

All the structured data that market researchers utilize that is not metric data is referred to as nonmetric data. For instance, nonmetric data comprises information that is rated (known as ordinal data) and information without a clear linear trend (which is called nominal or categorical) (Ray Poynter, 2017).

While, When using interdependence approaches, variables cannot be categorized as dependent or independent. Without formally assuming any particular distributions for the variables, it seeks to identify correlations between variables (Nimkar , Great Learning Team,2020). Therefore it can be classified :

1- To make it easier for us to interpret the data, factor analysis is an interdependent approach that unites several variables into a small number of components (Erickson,2014).

2- Using an interdependent approach called cluster analysis, items (respondents, goods, businesses, variables, etc.) are grouped so that each object is both similar to and distinct from the others in the cluster (Erickson,2014).

3- Multidimensional Scaling (MDS) is a method that generates a map representing the relative locations of several objects given just a distance table. The map might be one, two, three, or even four dimensions. The software computes the metric or non-metric answer. The proximity matrix refers to the distance table. It can be derived directly from trials or indirectly using a correlation matrix (Nimkar, Great Learning Team,2020).

4- A method known as correspondence analysis uses points on a map to represent the locations of the rows and columns of a table of non-negative data(Nimkar ,Great Learning Team,2020).

The fundamental benefit of multivariate analysis is that the inferences made are more correct since more independent factors are taken into account that affects the variability of the dependent variables. The results are more reasonable and more in line with the actual situation; The primary drawback of MVA is that it necessitates quite intricate calculations to reach a decent result.

2.3 Predicate Techniques (Deep Learning)

It is a new field of research that deals with finding theories and algorithms that allow the machine to learn on its own by simulating neurons in the human body. It is also one of the branches of science that deals with artificial intelligence, one of the most important techniques used as follows:

2.3.1 Zeiler & Fergus Net (ZF NET)

ZFNet is a classic convolutional neural network. It was created by (Zeiler and Fergus ,2013). was introduced for enhancing AlexNet by changing the architecture's settings, particularly by increasing the size of the intermediate convolutional layer. Where, ZFNet has made some detailed changes based on AlexNet, and there is no breakthrough in the network architecture, but some changes have been made to the convolution kernel and stepping. ZFNet is a CNN model,Also, ZFNet is a modified version of AlexNet which gives a better accuracy (Zeiler et al.,2014).

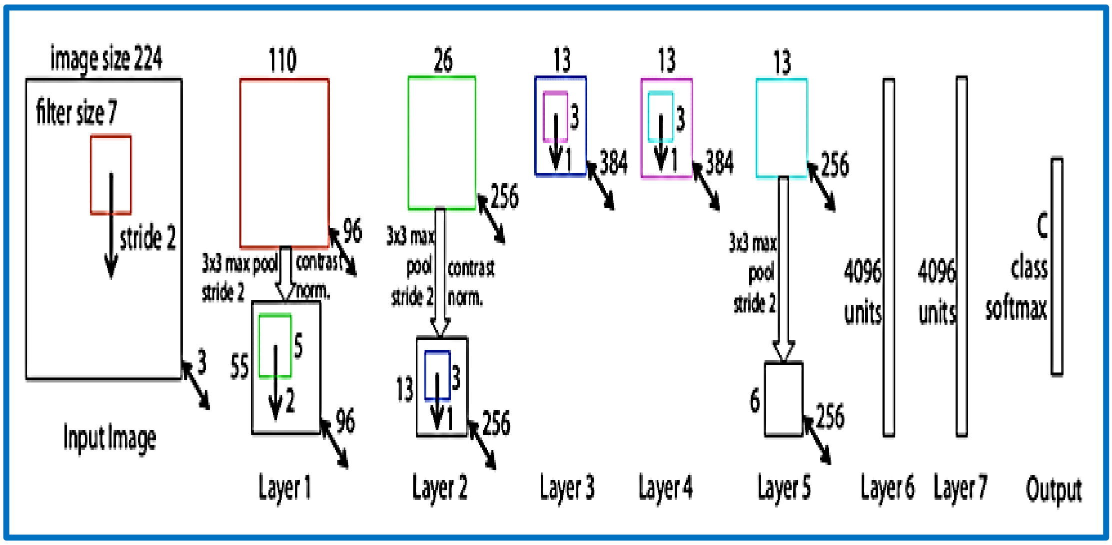


Figure (2.1) ZFNet Architecture

The most important feature of this network is reduced filter size. To enhance the number of features that the network is capable of detecting. In the third, fourth, and fifth convolutional layers, ZFnet increased the number of activation maps from (385,384,256) to (512,1024,512).ZFnet has significantly improved the image classifier error rate compared with AlexNet.

2.3.2 Alex Net

AlexNet is a well-known deep CNN structure proposed by Krizhevsky and Sutskever (2012). On the ImageNet dataset, AlexNet obtained excellent classification accuracy, which was a big achievement in the machine learning area. As a result, the input distributions of the layers of AlexNet differ from one another. This can make parameter training extremely difficult and time-consuming, necessitating proper setup. AlexNet, which used an 8-layer CNN. The convolutional neural network known as AlexNet has significantly influenced machine learning, especially when deep learning is used to improve vision of machines The network's design was comparable to Yann LeCun et al.'a1 LeNet, but more sophisticated, with convolutional layers and more filters per layer (Russakovsky et al.,2015).

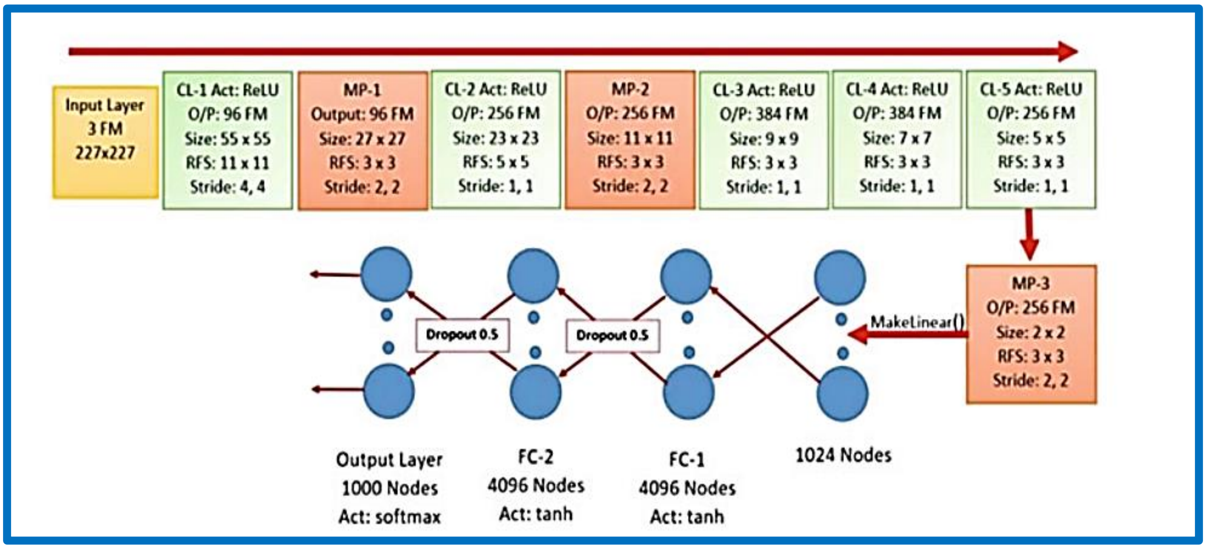


Figure (2.2) Alex Net Architecture

The first notable CNN model that was trained on GPUs was AlexNet. Due to this faster model training. It has a deeper architecture with 8 layers, which means it can extract more features than LeNet. At the time, it also worked well with color photographs. The ReLu activation function of this network has two advantages. It does not limit the output, unlike other activation functions. This suggests that there isn't much feature loss. It only cancels the negative output of gradient summation, not the dataset itself. Because not all perceptions are active, model training speed will be increased even further (Shree et al.,2019).

2.3.3 Google Net

Google Net was proposed by Google research (with the collaboration of various universities) in the research paper "Going Deeper with Convolutions" in 2014. The ILSVRC 2014 image classification challenge was won by this architecture. It has a significantly lower error rate than previous winners AlexNet and ZF-Net.This architecture employs techniques convolutions in the architecture's center and global average pooling (Sam et al.,2019), According to the following figure (Szegedy et al., 2015).

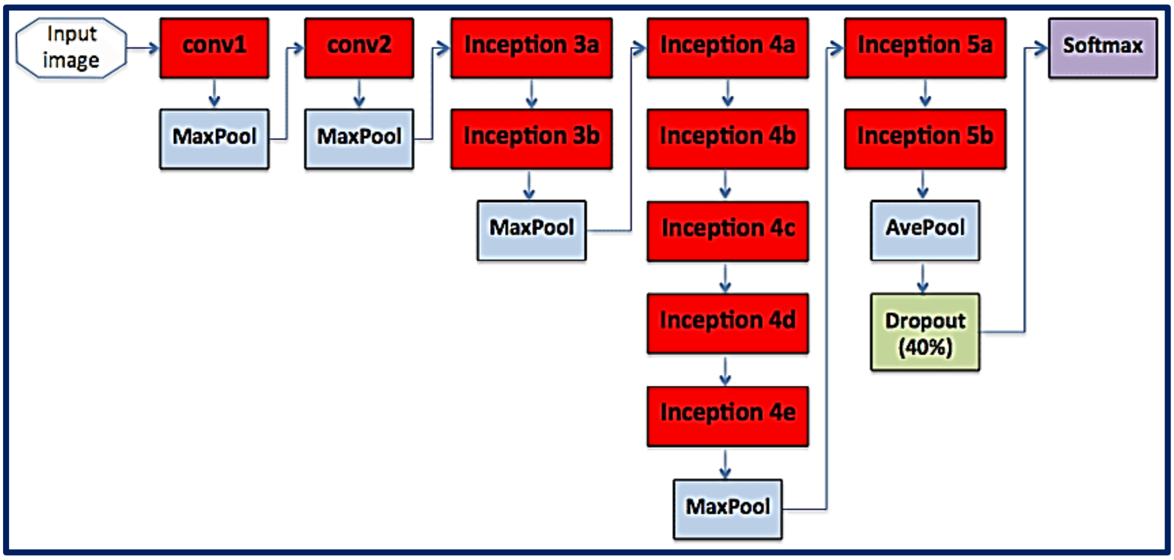


Figure (2.3) GoogleNet Architecture

While maintaining a constant computational budget, the GoogLeNet method expands the network's width and depth (Arora et al.,2013). The layer-by-layer construction can assess the last layer's correlation statistics and group them. This net enables the number of layers at each level to be raised without causing an uncontrollable growth in computational complexity. Compared to other networks with comparable performance, this network is three times faster. Even so, configuring the design of this network is difficult (Lu et al.,2018).

2.3.4 Long Short Term Memory (LSTM)

2.3.4.1 Traditional Long Short Term Memory (Trad-LSTM)

Trad-LSTM is a deep learning artificial recurrent neural network (RNN) architecture. Unlike traditional feedforward neural networks, LSTM has feedback connections. It can analyze whole data sequences as well as single data points (such as photos) (such as speech or video). LSTM can learn long-term dependencies. Its function is similar to that of a simple RNN in that it implements the recurrent structure. As a result, LSTM contains three major gates for input, hidden, and output states, which are referred to as the input gate, forget gate, and output gate, respectively.. The sigmoid function is used to modulate these gates' outputs(Shao et al., 2019), According to the following figure (Rahman & Siddiqui,2019):

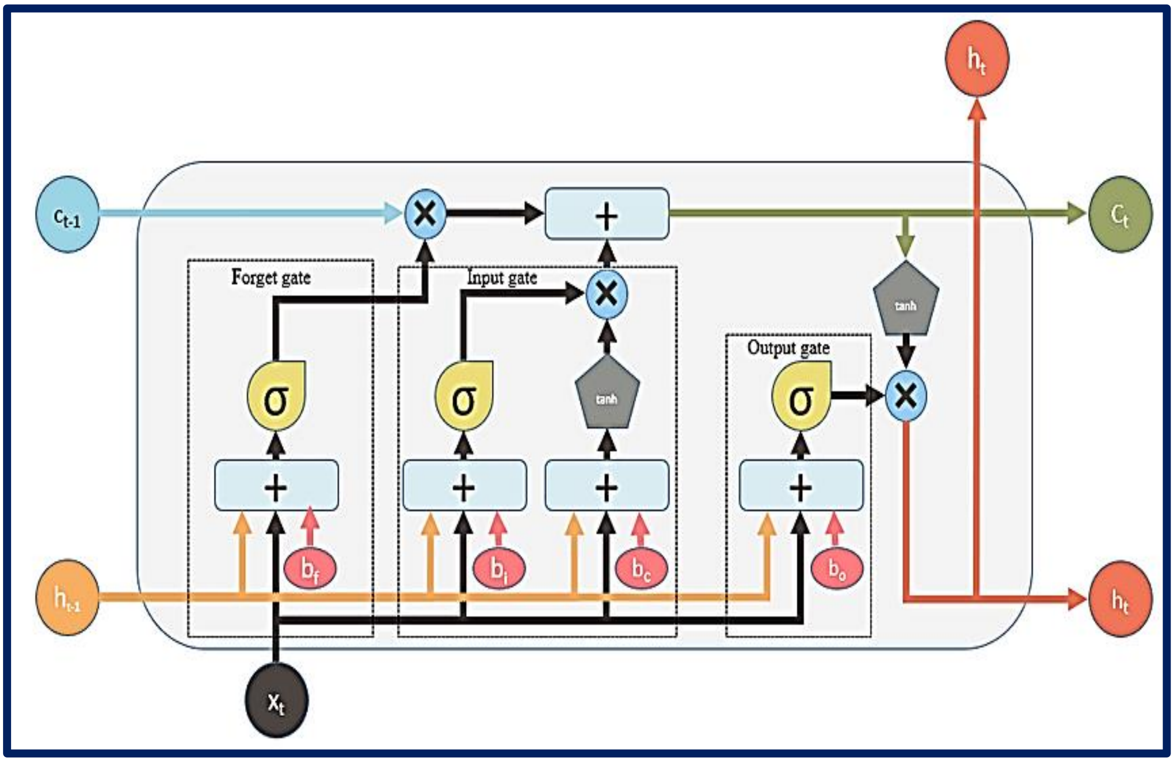


Figure (2.4) Trad-LSTM Architecture

The first gate is a forget that allows a sigmoid layer to decide what data from cell state to discard (Abedin et al., 2021) :

$$F_t = \sigma\,(W_{kf} * k_t + U_{hf} * h_{(t-1)} + b_f\,) \;\ldots\ldots\ldots\ldots\ldots\; (2.11)$$

The second gate is an input that uses a sigmoid layer to determine which values will be updated and a tanh layer to generate a vector of newly updated values as :

$$I_t = \sigma(W_{ki} * k_t + U_{hi} * h_{t-1} + b_i) \;\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.12)$$

$$\tilde{C}_t = \tanh(W_c * k_t + U_c * h_{t-1} + b_c) \;\ldots\ldots\ldots\ldots\ldots\ldots (2.13)$$

Afterward, the cell's state was modified by above equations to

$$C_t = f_t * C_{(t-1)} + i_t * \tilde{C}_t \qquad \ldots\ldots\ldots\ldots\ldots\ldots (2.14)$$

The output of the current state will be determined using the updated cell state and a sigmoid layer that decides which elements of the cell state will be the final output, as detailed in equations 5 and 6.

$O_t = \sigma (W_{ko} * k_t + U_{ho} * h_{(t-1)} + b_o)$ …………..(2.15)

$h_t = o_t * \tanh(c_t)$ ………….. (2.16)

R=h

Primary parameters

F: forget gate at t.

I: input gate at t.

C: current memory function.

secondary parameters

K: input vector.

h,R: output vector.

o: output gate at t.

h(t-1) : previous hidden state .

T: time step.

W, U: weight matrices.

B: bias vector.

Activation functions

σ : sigmoid function is defined as ($\sigma(x) = \frac{1}{1+e^{-x}}$) , {0,1}

tanh : hyperbolic tangent , {-1,1}.

There is no connection between a gate and its preceding memory cell in the traditional LSTM architecture. The output gate is frequently found closed during training because it is impossible to access the previous memory cell state when it is closed. This issue can be solved using the LSTM variant peephole convolutional LSTM (Rahman & Siddiqui,.2019).

The primary purpose for which the LSTM network was created is related to its most significant feature. Due to the vanishing, the RNNs were unable to capture long-term dependencies. This type of model can input and output sequences step by step throughout time, allowing for variable-length inputs and/or outputs. With this trait, they avoid one of the fundamental limitations of conventional feedforward neural networks.

2.3.4.2 Bi-Directional- Long Short Term Memory (Bi-LSTM)

Any neural network can be made to have sequence information in both the forward (future to past) and backward (bi-LSTM) directions (past to future). The performance of models on problems involving sequence classification can be enhanced by the use of bi-LSTM, an extension of the classic LSTM. Because of its bidirectional nature input pattern, it analyzes any time series data more effectively than any conventional model, including (ARIMA, SARIMA, and ARIMAX) (Sezer .et.al.,(2020)). The Bi-LSTM model is more effective than LSTM because it takes into account both past and future data while training. Although using Bi-LSTMs may not be appropriate for all sequence prediction issues, it can have some advantages in terms of better outcomes in those areas where it is (Datta et al.,2021).

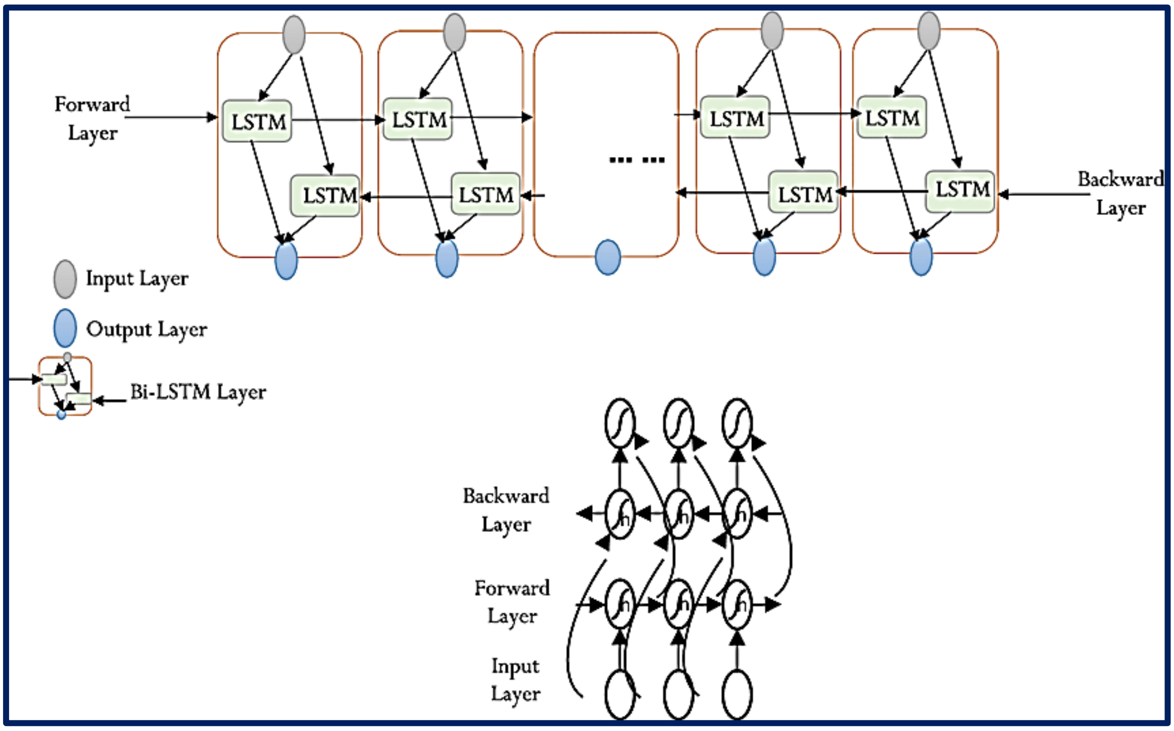


Figure (2.5) Bi-LSTM Architecture

With the help of the following equations, We compute and merge the forward and backward hidden states:

$$hf = \sigma(Wf * K + hf + bf) \quad \text{............ (2.17)}$$

$$hb = \sigma(Wb * K + hb + bb) \quad \text{............ (2.18)}$$

$$R = (hf * Wf + hb * Wb + b) \quad \text{............(2.19)}$$

Where :

h: hidden state.

W: weight vector.

B: bias vector.

σ: sigmoid function.

K: input vector.

R: output.

The most crucial aspect of this network is how it behaves, which enables it to learn about the current state of data from previous and future data via its forward and backward layers and produce more insightful output. Between the forward and backward layers of a bi-LSTM layer, there are no hidden to hidden connections. This makes it easier to comprehend data from each Bi-LSTM unit's forward layer as well as its backward layer. Bi-LSTM, therefore, performs better than LSTM when analyzing time series data.

2.3.4.3 Multi-Time scale Long Short Term Memory(MT-LSTM)

A generalization of LSTMs called MT-LSTM allows for the capture of data over a range of timescales. Both short and long texts are well modelled by MT-LSTM. Using memories of various timescales. MT-LSTM is therefore capable of transmitting important data over a long distance with ease. MT-LSTM offers a faster convergence speed than normal LSTM since the error signal can be back propagated across multiple timescales during the training phase (Liu .et.al.,(2015)).

A diagram showing the unfolded LSTM and MT-LSTM is shown in Fig. (2.6). The unit that is currently activated is shown by the dotted node, while the unit that is activated is shown by the solid node. Units that remained unchanged are represented by dotted lines, and those that will be updated at the following time step are represented by solid lines.

Where we note that:

- LSTM units are divided into three groups: G1, G2, and Gg. At different times Tk is used to activate each group Gk, (1 gk)

- To maintain the corresponding LSTM groups, the gates, and weight matrices are also divided. The standard LSTM and the MT-LSTM have the same number of groups.

- The only group's Gk that are executed at each time step t satisfy (t MOD Tk) = 0.

- The selection of the set of periods Tk, T1, and Tg is arbitrary. The exponential series of periods is used here: The period of group Gk is Tk = 2k-1.

- The fastest group, G1, can be executed at every time step and functions similarly to the conventional LSTM. The Gk group moves the least quickly. (Liu et al . , (2015)).

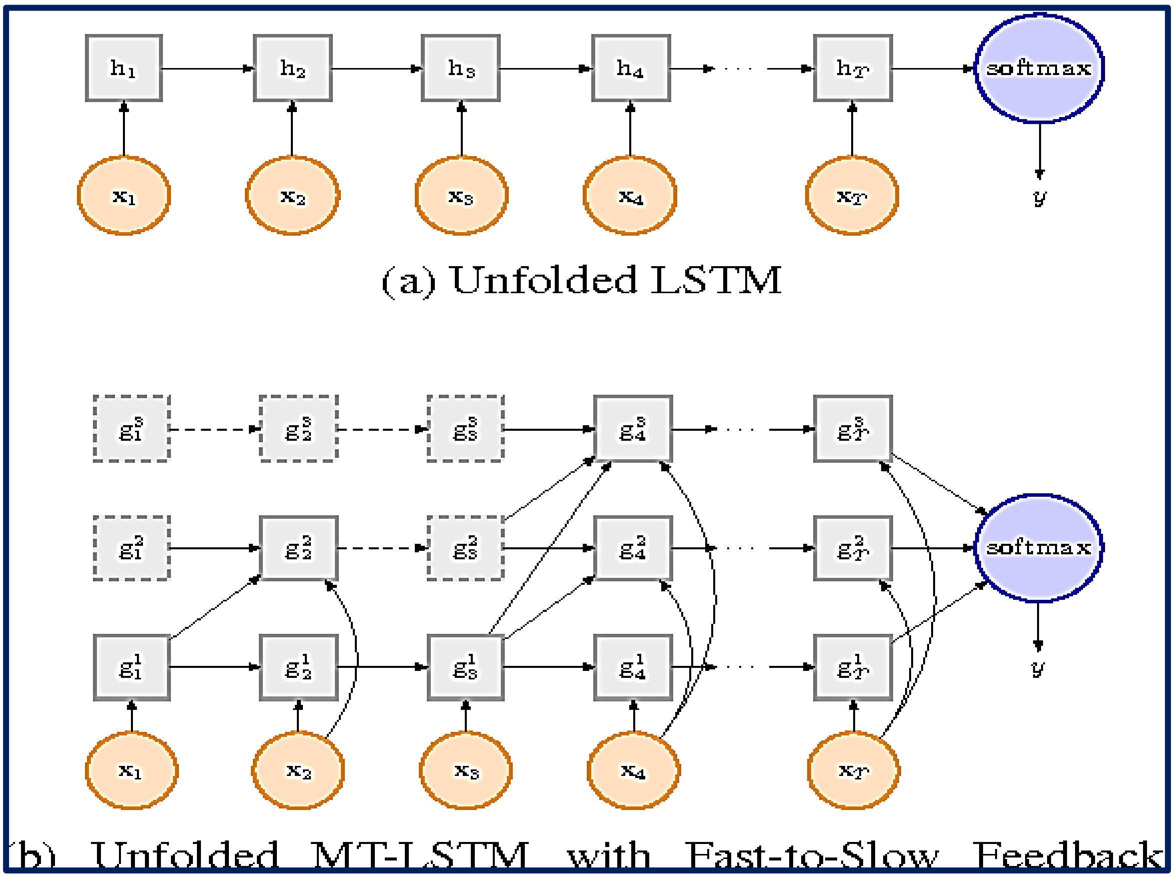


Figure (2.6) MT-LSTM Structure

- The hidden state vector and memory cell of group Gk are calculated in two step at time t ((Liu .et.al.,(2015)) :

1- When group Gk is activated at time step t, the LSMT units of this group are calculated by the following equations:

$I_{kt} = \sigma (W_{ki} x_t + \sum_{j=1}^{g} U_{j\to k}^{i} h_{j}^{t-1} + \sum_{j=1}^{g} V_{j\to k}^{i} c_{j}^{t-1})$...(2.20)

$F_{kt} = \sigma (W_{kf} x_t + \sum_{j=1}^{g} U_{j\to k}^{f} h_{j}^{t-1} + \sum_{j=1}^{g} V_{j\to k}^{f} c_{j}^{t-1})$..(2.21)

$O_{kt} = \sigma (W_{ko} x_t + \sum_{j=1}^{g} U_{j\to k}^{o} h_{j}^{t-1} + \sum_{j=1}^{g} V_{j\to k}^{o} c_{j}^{t})$..(2.22)

$\tilde{C}_{kt} = \tanh (W_{kc} x_t + \sum_{j=1}^{g} U_{j\to k}^{c} h_{j}^{t-1})$ ………………(2.23)

$C_{kt} = f_{kt}\ c_{kt-1} + i_{kt} \odot \tilde{c}_{kt}$ ………………(2.24)

$H_{kt} = o_{kt} * \tanh (c_{kt})$ ………………(2.25)

Where:

I_{kt} = input gate.

F_{kt} = forget gate.

O_{kt} = output gate.

C_{kt} = memory cell.

H_{kt} = hidden state.

(2) When group Gk is deactivated at time t, its LSMT units remain unaffected:

$C_{kt} = c_{kt-1}$ ……………………………(2.26)

$H_{kt} = h_{kt-1}$ ...…………………………(2.27)

2.3.4 Gate Recurrent Unit (GRU)

A form of recurrent neural network called a GRU deals with the problem of long-term dependencies, which can cause larger vanilla RNN networks to experience vanishing gradients (Jiao .et.al.,(2020)). GRUs address this issue by storing "memory" from earlier time points to assist inform the network for future predictions (Sparkle Russell-Puleri and Dorian Puleri,(2019)). However, In the field of machine translation, GRU has received a lot of attention. GRUs can alternatively be seen as a more straightforward variation of LSTMs (Long Short-Term Memory). It was claimed to be motivated by the Long Short-Term Memory unit (Ambika Choudhury,(2020)). Where, It was introduced by (Kyunghyun Cho .et.al.,(2014)). The GRU is comparable to a long short-term memory (LSTM) with a forget gate, but it has fewer parameters because it lacks an output gate (DENNY BRITZ,(2015)). GRUs have been shown to perform better on smaller, less frequent datasets (Chung .et.al.,(2014)).

Gated Recurrent Unit (GRU) has an internal mechanism called gates(update and reset gates)(Xin Wang .et.al.,(2019)). These gates control the flow of information by determining what information should be saved or deleted at each time step. Also, Only one concealed state is communicated between time steps in GRUs. Due to the calculations and gating mechanisms that the input data and hidden state undergo together, This hidden state can accommodate both long-term and short-term dependencies (Wang .et.al.,(2018)). Therefore according to the bellow diagram, which shows the architecture of GRU, We can clarification or interpretation of the mechanism of GRU actions by making a statement of the work of each gate according to the following equations (Gabriel Loye,(2019)),(Zhou et.al.,(2016)) :

In the first step, The update gate is calculated by combining the prior concealed state and the current input data, According to the following equation:

$$Z_t = \sigma (W_z \,.\, K_t + U_z \,.\, h_{t-1} + b_z) \qquad \text{…………………(2.28)}$$

Following that, the current time step's input data and the previous time step's hidden state are both used to derive and compute the reset gate, According to the following equation:

$$R_t = \sigma (W_r \,.\, K_t + U_r \,.\, h_{t-1} + b_r) \qquad \text{……………………(2.29)}$$

In the intermediate memory unit or candidate hidden state, information from the previous hidden state is mixed with input, According to the following equation:

$h_{\sim} = \tanh(W_h \cdot K_t + r \cdot U_h \cdot h_{(t-1)} + b_r)$ ……………(2.30)

lastly, we find the resulting, According to the following equation:

$h = z \cdot h_{t-1} + (1-z) \cdot h_{\sim}$ ……………(2.31)

$y = h$ ……………(2.32)

Primary parameters

Z: update gate at t.

R: reset gate at t.

$h^{\sim}$: current memory.

secondary parameters

K: input vector.

h,y: output vector

T: timestep.

H_{t-1} : previous hidden state .

W, U: weight matrices.

B: bias vector.

Activation functions

σ: sigmoid function as ($\sigma(x) = \frac{1}{1+e^{-x}}$).

tanh: the hyperbolic tangent.

Activation functions are possible, provided that σ (x) € { 0 ,1} , tanh € { -1 ,1}, (Dey et.al.,(2017)) .According to the following figure(Sparkle Russell-Puleri and Dorian Puleri , (2019)):

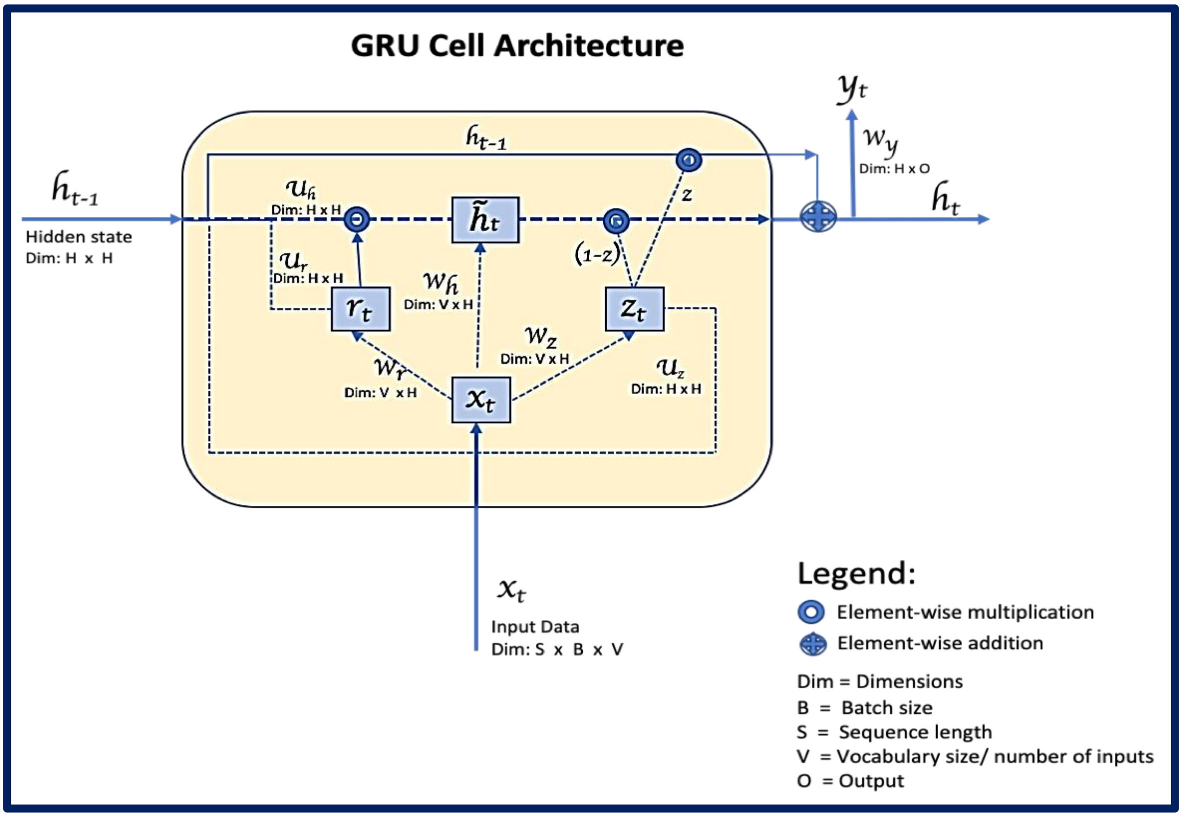


Figure (2.7) GRU Architecture

GRU's objective is to develop a model that can correctly forecast the price of oil through train this model on previous data and evaluating this model based on results. To achieve this, we'll begin by selecting features and preprocessing the data, then define, train, and ultimately evaluate the models. In general, our project will be implemented according to the following mechanism (Gabriel Loye ,(2019)) :

Table 2.1: Comparison among Main Techniques of Predication

Techniques	Advantages	Disadvantages	primary parameters	secondary parameters
GRU	-GRUs are not as complex as LSTM and computing them does not take too much time, because of the lower number of gates. - it is easy to modify and doesn't need memory units. -It gives highly accurate results. -GRU requires a short time for training compared to the other methods, Therefore, It is faster to train -It retains previous important information due to the presence of hidden memory. - GRU has smaller errors. - GRU is less complex than LSTM because it has less number of gates.	- its limited ability to predict the nonlinear relationship data. - It lacks an output gate.	h\` Z R	W U B K Y H
Trad-LSTM	-The purpose for which the LSTM model was created is related to that. Due to the disappearance, the RNNs were unable to capture the long-term dependencies - This model's capacity to input and output sequences time after time step, allowing for variable length inputs and/or outputs, is another crucial feature. They circumvent one of the primary drawbacks of traditional feedforward neural networks because of this characteristic.	- that inputs and outputs must both be fixed in size. Therefore, there are significant limits to this type of model for various applications involving variable length sequences of input and/or output. - To be taught and prepared for use in the real world, they need a lot of resources and time. - Because LSTMs are prone to overfitting, using the dropout strategy to them is difficult.	C\` f I	W U B K R H O C

		Dropout is a regularization strategy in which input and recurrent connections to LSTM units are probabilistically deleted from activation and weight updates when training a network.		
Bi-LSTM	- It helps to learn about the current state of data from previous and future data through its forward and backward layers, therefore it can produce a more meaningful output. - A bi-LSTM layer contains no hidden-to-hidden linkages between the forward and backward layers, therefore This facilitates comprehension of data from both the backward and forward layers in each Bi-LSTM unit. - When analyzing time series data, Bi-LSTM performs more precisely than LSTM. - It solves the fixed sequence to sequence prediction problem.	- Since Bi-LSTM has double LSTM cells so it is costly. Not a Good fit for Speech Recognition. - Bi-LSTM is a much slower model and requires more time for training.	H R	W K B
MT-LSTM	- A generalization of LSTMs, MT-LSTMs gather data over a range of timescales. - Both short and long texts can be effectively modelled using MT-LSTM. Considering the various timeframe memories. - Over a long distance, MT-LSTM can transmit important data easily.	-It requires more memory to train. -It is easy to over fit. -It sensitive to different random weight initialization.	C` f I	W U C K R H O

2.4 Evaluation Measure

Evaluation measures are metrics used to gauge the effectiveness of a specific model. The efficiency of the crude oil price prediction model is assessed using these standards. These metrics include mean squared error (MSE), mean absolute error (MAE), and R2 score. These measures are used to assess the fraction of the dependent variable's volatility that may be attributed to the independent variable (s). This method is frequently used to assess the ability of a linear regression model to predict. It is also a statistic that is used in connection with statistical models whose primary goal is either to forecast future events or to evaluate hypotheses using other relevant data (Chicco et al., 2021). An estimate of the frequency with which the observed results of the model. As shown below:

1- Coefficient of determination(R2 Score)

$$R^2 = 1 - \frac{\sum_{i=1}^{m}(Ki - Yi)^2}{\sum_{i=1}^{m}(\hat{Y} - Yi)^2} \qquad \text{.......................(2.33)}$$

Worst case = - ∞ , Best case = +1

The amount of the dependent variable's volatility that can be predicted from the independent variables is represented by the coefficient of determination.

2- Mean Squared Error (MSE)

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(Ki - Yi)^2 \qquad \text{......................(2.34)}$$

Worst case = + ∞ , Best case = 0

If there are outliers to be found, MSE can be employed. In fact, MSE works well for giving these points more weight. Obviously, the squared element of the function multiplies the mistake if the model ultimately yields a single extremely poor forecast.

3- Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{m}\sum_{i=1}^{m} |Ki - Yi|^{2} \qquad \text{.......................(2.35)}$$

Worst case = + ∞ , Best case = 0

If the outliers show poor portions of the data, MAE can be applied. As a matter of fact, MAE rarely penalizes outliers during training, giving the model a generic and constrained performance measure. On the other hand, the model's performance will be average if the test set also includes a lot of outliers.

Where:

K: The item's expected value for i.

Y:The element's actual value for i.

Chapter Three:

Building HMOP-NCT

Chapter Three: Hybrid Model For Prediction Oil Price

Based On NeuroComputing Technique

3.1 Introduction

This chapter presents how to design and implement a predicting model capable of determining oil prices, because determining the rate of increase or decrease in its prices is one of the most important factors that affect the economies of countries, and it is also considered an effective influence on the budget of any country. The proposed model called hybrid model for prediction oil price based on neurocomputing technique (HMOP-NCT) consist of three stages: The first stage called preprocessing includes checking outlier , missing values and normalization the data to preparing it of the next stage , that determined the important of each features through compute the correlation ,entropy and information gain .After that split the dataset into parts ; training and testing. The first part of dataset used the build the predictor called Hybrid model to oil price based on neuro computing technique (HNOP – NCT) , While the second part used to evaluated HMOP-NCT through three error measures called (R2 , MSE and MAE), as explain in Figure 3.1 and Algorithm 3.1 .

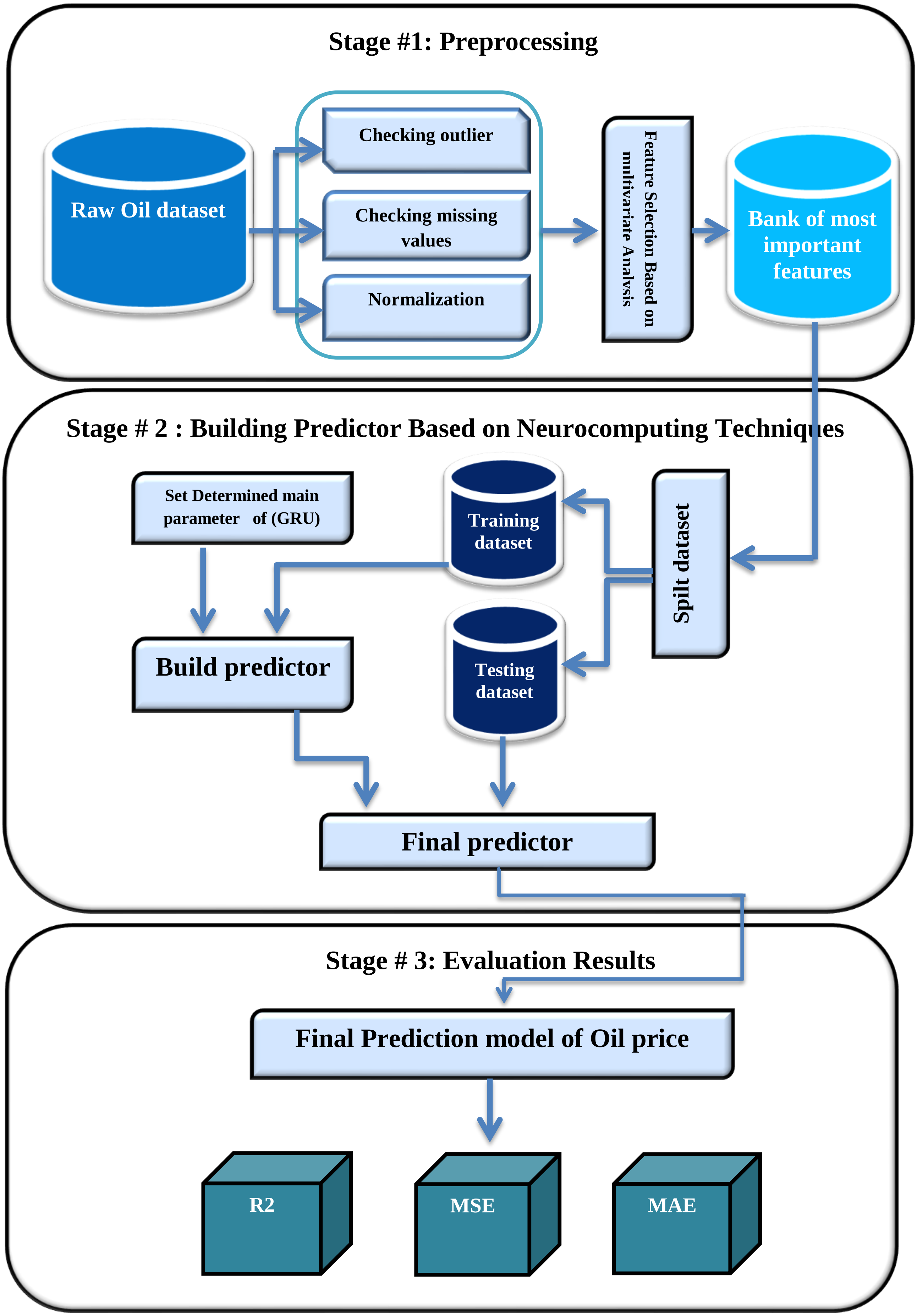


Figure 3:1 Block Diagram of Oil Price Prediction

Algorithm #3.1: HMOP-NCT Model

Input: *Dataset of oil price contain seven features .*

Output: *Prediction of oil price .*

```
        // Preprocessing
1:      For  each record in original  dataset
2:          For  each feture in original dataset
3:              Call checking outlier
4:              Call checking missing value
5:              Call normalization
6:          End for
7:      End for
        // Select feature based multivarient analysis
8:      For each  record in original dataset
9:          For  each feature in original dataset
10:             Call correlation
11:             Call entropy
12:             Call information gain
13:         End for
14:     End for
        // Split dataset
15:     For each  record in clean dataset
16:         For each  features  in clean dataset
17:             Split dataset into training and testing
18:         End for
19:     End for
        //Bulid predictor
20:     For each  record in training dataset
21:         For each  features  in training dataset
22:             Training GRU
23:             Compare if max less than max error generation
24:                 Jump to 29
25:             Else
26:                 Jump to 16
27:             End if
```

```
    End for
  End for
Call Evaluation Measure
End HOMP-NCT
```

3.2 Main Stages of HMOP-NCT

This section will show the main stages of the HMOP-NCT and explain the specific details for each stage.

3.2.1 Preprocessing Stage

The pre-processing stage is begin after in this stage checking both outlier and missing data ,in general ; missing data meaning the record have nan values while outliers meaning record have values not on the specific rang of that features .The proposed model handle both the above states by dropping the record have these values .Also ; in this stage checking if the feature lay in the normal distribution or not through apply the normalization on it . as explain algorithm 3.2.

Algorithm #3.2: Pre-Processing

```
Input: Dataset of oil price contain seven features .
Output: Clean dataset (features selection) based on multivariate analysis
// Treatment outlier values
For each row in oil price dataset
    For each column in oil price dataset
        IF value of column not in rang(Maxf,Minf)
            droping the row contains that columns
        End IF
    End for
End for
// Treatment missing values
For each row in oil price dataset
    For each column in oil price dataset
        IF value of column =Nan
            droping the row contains that columns
        End IF
```

| ***End** for*
***End** for*
// Apply normale Distribtion
***For** each row in oil price dataset*
| ***For** each column in oil price dataset*
| | *Compute MinMaxScaler* *// MinMaxScaler =* $\frac{k-\min k}{\max k-\min k}$
| ***End** for*
***End** for*
End Pre-processing

3.2.2 Building predictor stage

After the pre-processing stage of the data is completed begin feature selection stage through apply three types on multivariate analysis are (correlation ,Entropy and information gain) to determined which features more affect in oil price.Then determined the main parameters of the deep neurocomputing Network called GRU this parameters are size of windows, number of hidden layers , number of node in each layer ,activation function and number of epoch .In general ;this network, which maintains the state of the cell and the values of the weights. GRUs help make it easier to capture dependencies without ignoring previous information from huge chunks of a sequential database, as a result of this network having two gates (update gate and reset gate) to save important information from the previous step (i.e., especially long ago), or ignoring it. GRU is superior to other networks in terms of its relatively simple modification due to the presence of memory modules, which makes the process of training and learning faster compared to other networks. As shown in algorithm 3.3.

Algorithm #3.3: Essential components of GRU

***Input**: X : Dataset of oil price*
***Output**: Prediction values of oil price*
// The forward components
***For** each time(t) in dataset of oil price*

// Compute: h˜ , M, N ,h

h~ = tanh (Wh . Kt + r .Uh . ht-1 + bh) *// Memory Cell*

Mt = σ (Wm . Kt + Um . ht-1 + bm) *// Update gate*

Nt = σ (Wn . Kt + Un . h(t-1) + bn) *// Reset gate*

h= m ht-1 + (1-m) * h˜* *// Output Cell*

Y=h

End for

// The backward components.

***For** each time(t) update*

// Update Output Cell , loss function

Yt = σ (Uh . ht) *//Update Output Cell*

$E = \frac{1}{2}(y - yt)^2$ *//Loss function*

End for

// The Final updates to the internal parameters is compute

$\delta Un = \sum_{t=0}^{T} \delta\, Unt \times Xt$ *//Update the weight of reset gate*

$\delta Um = \sum_{t=0}^{T} \delta\, Umt \times Xt$ *// Update the weight of update gate*

$\delta U\tilde{h} = \sum_{t=0}^{T} \delta\, U\tilde{h}t \times Xt$ *//Update the weight of intermediate memory*

$\delta b = \sum_{t=0}^{T} \delta b$ *// Update bias*

End Essential components of GRU

We can implementation this algorithm 3.3 as explain in 3.4.

Algorithm #3.4: Execution of HMOP-NCT MODEL

Input: *Dataset split into part training and testing*

Output: *Average values of oil price for next year*

***For** each iteration*

***For** all samples in the training dataset*

Unroll a set of num_unrollings batches

Train the GRU with the unrolled batches

Compute the average training loss

***End** for*

***For** each recod in the testing dataset*

Update the HMOP-NCT state

Found num_unrolling data points
Make forecasting for forecasting_once steps continuosly
Current input is Previous prediction
End for
End for
End Execution of HMOP-NCT MODEL

3.2.3 Evaluation stage

After completing the model design phase and implementing it on a data set to predict oil prices. This The stage shown how can determined the efficiency of proposal model through usage the final predictor result from the previous stage to estimation the oil price of testing dataset and compute three measures (R2 ,MSE and MAE).

Algorithm #3.5: Evaluation Stage

Input: *Testing dataset*
Output: *Average error rate based on R2 ,MSE,MAE*
For *each record in testing dataset*
For *each column in testing dataset* *// m= number of feature*
$R2=1-\frac{\sum_{i=1}^{m}(Ki-Yi)^2}{\sum_{i=1}^{m}(\hat{Y}-Yi)^2}$ *// k= predict value , y= actual value*
$MSE=\frac{1}{m}\sum_{i=1}^{m}(Ki-Yi)^{2}$
$MAE=\frac{1}{m}\sum_{i=1}^{m}|Ki-Yi|^{2}$
End for
End for
End Evaluation stage

Chapter Four:

Implementation of HMOP-NCT

Chapter Four: Implementation of HMOP-NCT

4.1 Introduction

This chapter explains the implementation of model present in chapter three and analysis the result of each stage. Also, it shown the specific description of the dataset used to testing the behavers of proposed model and focus the light on the main features of that dataset and important of each feature.

4.2 Implement of HMOP-NCT

The proposed model will be applied, and their results will be evaluated based on the three measures as follow:

4.2.1 Description of Dataset

In this section, shown the dataset used and described its .In general ; dataset collected over the past 10 years with 7 base features and 4947 numeric records, excluding the date feature, which is of string type .The dataset used to train the model contain the following features ,while table (4.1) shown samples of dataset to implement the HMOP-NCT.

- ***Date*** is the trading date.
- ***Crude Oil WTI Futures Historical (WTI)***is a contract that determines the price of oil at a particular future date.
- ***Gold Futures Historical Data(GOLD)*** is one of the basic characteristics that effectively affect the determination of oil prices.
- **Standard and Poor's 500 Futures Historical Data-INVESTOR(SP 500)** serves as a measure for both market and national economic growth in the United States of America. Its name is a result of the name Standard Company and Pros Publishing Company coming together. This indicator's name comes from the fact that it was created

to include 500 companies. This index derives its value from the market value of the companies that come under it.

- ***US Dollar Index Futures Historical Data(US DOLLAR INDEX)*** refers to the rate at which the US dollar is exchanged for other currencies. The most well-known dollar index, or USDX, which shows how the dollar compares to other currencies, is produced by the New York Chamber of Commerce.
- ***Historical Bond Yield Information for the United States (US 10YR BOND)***, is one of the features used to anticipate oil prices.
- ***The Dow Jones Utility Average (^DJU)***
  is a stock exchange index in the United States. It is the world's oldest index, tracking the performance of the top publicly traded companies. It is one of the key features used to anticipate oil prices.

Table (4.1): Sample of the Dataset

Date	WTI	GOLD	SP 500	US DOLLAR INDEX	US 10YR BOND	DJU
4/1/2010	81.51	1117.7	1128.75	77.83	3.823	399.49
5/1/2010	81.77	1118.1	1132.25	77.85	3.763	395.37
6/1/2010	83.18	1135.9	1133	77.655	3.829	398.36
7/1/2010	82.66	1133.1	1137.5	78.105	3.827	396.61
8/1/2010	82.75	1138.2	1141.5	77.655	3.836	396.31
-------------	-------------	-------------	-------------	-------------	-------------	-------------
-------------	-------------	-------------	-------------	-------------	-------------	-------------
3/1/2011	91.55	1422.6	1265.25	79.384	3.334	406.58
4/1/2011	89.38	1378.5	1265.25	79.699	3.338	408.52
5/1/2011	90.3	1373.4	1271.75	80.538	3.463	406.2
6/1/2011	88.38	1371.4	1270.25	81.093	3.399	406.43
7/1/2011	88.03	1368.5	1267.5	81.321	3.328	407.72
-------------	-------------	-------------	-------------	-------------	-------------	-------------

-------------	-------------	-------------	-------------	-------------	-------------	-------------
3/1/2012	102.96	1599.7	1272	79.912	1.956	455.72
4/1/2012	103.22	1611.9	1273	80.425	1.984	453.02
5/1/2012	101.81	1619.4	1273	81.25	1.996	453.88
6/1/2012	101.56	1616.1	1274.25	81.597	1.958	451.2
9/1/2012	101.31	1607.5	1275.5	81.388	1.956	452.58
-------------	-------------	-------------	-------------	-------------	-------------	-------------
-------------	-------------	-------------	-------------	-------------	-------------	-------------
2/1/2013	93.12	1687.9	1457	79.948	1.835	461.46
3/1/2013	92.92	1673.7	1453.5	80.515	1.911	461.57
4/1/2013	93.09	1648.1	1457.75	80.61	1.903	464.62
7/1/2013	93.19	1645.5	1455.75	80.369	1.901	459.37
8/1/2013	93.15	1661.5	1452.25	80.445	1.864	458.46
-------------	-------------	-------------	-------------	-------------	-------------	-------------
-------------	-------------	-------------	-------------	-------------	-------------	-------------
2/1/2014	95.44	1362.7	1826.5	80.788	2.991	482.88
3/1/2014	93.96	1376.5	1825.5	80.955	2.998	481.4
6/1/2014	93.43	1376	1820.75	80.806	2.958	482.3
7/1/2014	93.67	1366.5	1830.75	80.98	2.941	486.36
8/1/2014	92.33	1363.6	1832.5	81.162	2.991	484.17
-------------	-------------	-------------	-------------	-------------	-------------	-------------
-------------	-------------	-------------	-------------	-------------	-------------	-------------
2/1/2015	52.69	1253.6	2046.25	91.383	2.114	621.61
5/1/2015	50.04	1269.4	2016	91.622	2.034	614.08
6/1/2015	47.93	1279.2	1994.5	91.738	1.938	615.33
7/1/2015	48.65	1270.6	2019.5	92.115	1.969	621.55
8/1/2015	48.79	1268.6	2055	92.602	2.018	626.61
-------------	-------------	-------------	-------------	-------------	-------------	-------------
-------------	-------------	-------------	-------------	-------------	-------------	-------------

4/1/2016	36.76	1110.5	2009	98.968	2.243	577.48
5/1/2016	35.97	1113.4	2011.75	99.496	2.239	580.97
6/1/2016	33.97	1126.9	1986	99.266	2.17	580.5
7/1/2016	33.27	1138.9	1933	98.272	2.149	577.51
8/1/2016	33.16	1128.5	1911.5	98.599	2.116	578.82
-------------	-------------	-------------	-------------	-------------	-------------	-------------
-------------	-------------	-------------	-------------	-------------	-------------	-------------
3/1/2017	52.33	1209.7	2252.5	103.201	2.448	657.32
4/1/2017	53.26	1211.9	2264.25	102.709	2.441	659.28
5/1/2017	53.76	1227.1	2264.25	101.528	2.348	659.85
6/1/2017	53.99	1220.5	2271.5	102.209	2.421	661.56
9/1/2017	51.96	1232.4	2265	101.909	2.368	653.19
-------------	-------------	-------------	-------------	-------------	-------------	-------------
-------------	-------------	-------------	-------------	-------------	-------------	-------------
2/1/2018	60.37	1354.8	2693	91.573	2.462	715.72
3/1/2018	61.63	1357.9	2711	91.896	2.445	706.52
4/1/2018	62.01	1361.1	2723.75	91.581	2.453	700.52
5/1/2018	61.44	1362.1	2742.5	91.673	2.476	700.74
8/1/2018	61.73	1360.5	2746.75	92.1	2.48	707.87
--------------	--------------	--------------	--------------	--------------	--------------	--------------
--------------	--------------	--------------	--------------	--------------	--------------	--------------
2/1/2019	46.54	1296.9	2511	96.423	2.633	701.04
3/1/2019	47.09	1307.7	2447.75	95.878	2.552	701.55
4/1/2019	47.96	1298.7	2531.25	95.75	2.668	711.91
7/1/2019	48.52	1302.9	2550.5	95.228	2.698	704.45
8/1/2019	49.78	1298.8	2572.5	95.479	2.73	711.43
--------------	--------------	--------------	--------------	--------------	--------------	--------------
--------------	--------------	--------------	--------------	--------------	--------------	--------------

4.2.2 Results of Preprocessing

This stage consists of several steps, each of which includes solving a specific problem within the dataset. The steps are as follows:

- **Outlier**

In this section, the dataset is examined. Does it include external data? This means that there is data that is unusual for the specific data set of oil prices. This data is processed by testing whether it is close to the specified range and can be rounded to the closest possible value or close to it within the specified value range. Otherwise, this data will be deleted. In this data set, there is no external data because this data is regular.

- **Missing Values**

In this section, the dataset is examined. Does it contain missing data? This means that if any of the features related to oil prices lose a certain value, then the error in entering a value leads to the loss of this value, and as a result, the record of these values will be deleted. The total number of rows in the data set (4947), the number of rows that contain missing data (152), and the total number of samples in the data set after dropping the missing values (4849). The result of this step is shown in Table 4.2:

Table (4.2): Missing values

Name of feature	# Record contain missing value	Rate of remained data	Rate of missing data
Date	0	100%	0
WTI	0	100%	0
GOLD	6	99.87	0.13
SP 500	1	99.97	0.03
US DOLLAR INDEX	30	99.39	0.61
US 10YR BOND	27	99.45	0.55
DJU	88	98.22	1.78

Table 4.2 shows how to process the missing data by deleting the row containing the missing value. We note that the rate of missing values are: Date = 0.0, WTI =0.0, Gold = 0.13, SP500 = 0.03, USDOLAR = 0.61, US

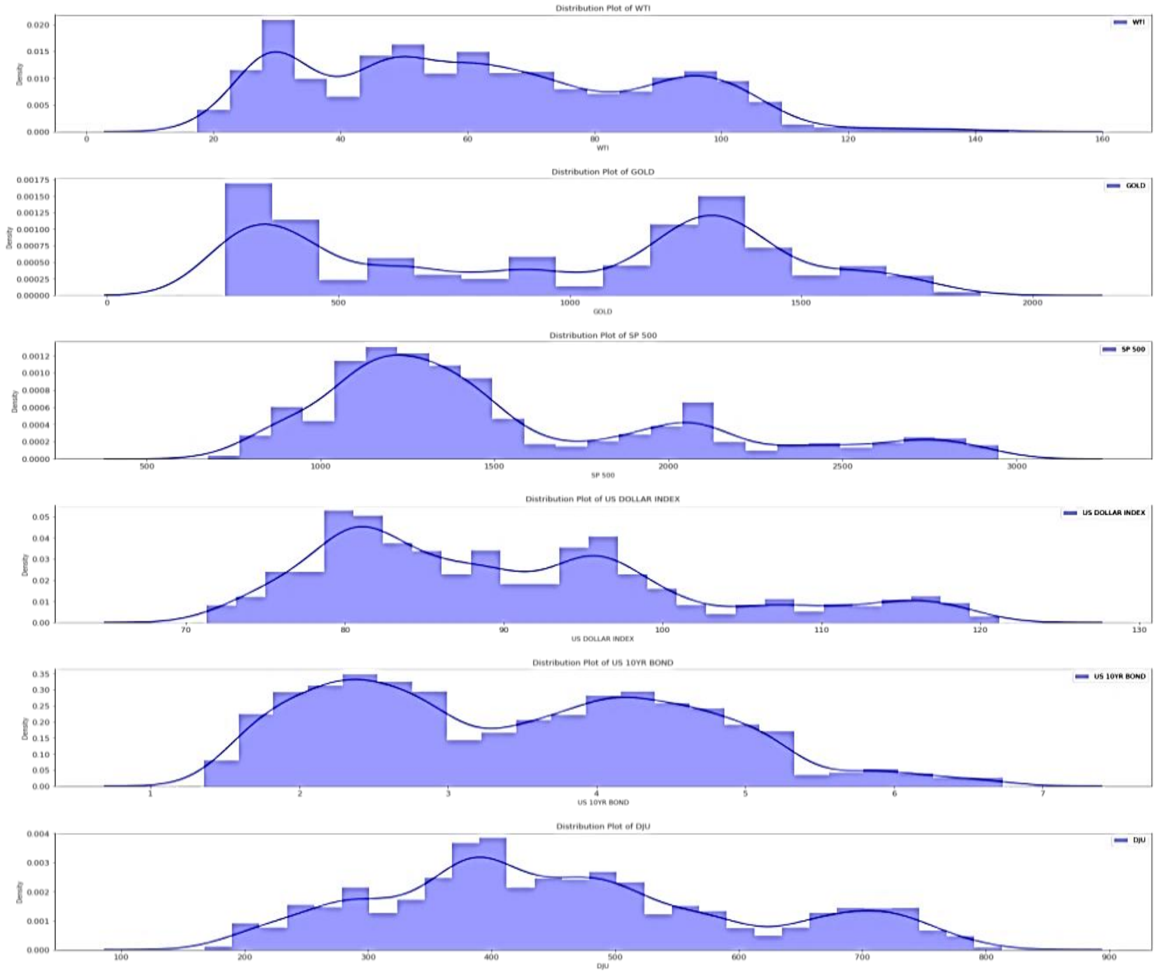
10 YR = 0.55, DJU =1.78 .Table(4.3), shown the statistical measurements for each features, after deleting the missing data.

Table (4.3): Statistical Measures

	count	mean	std	min	25%	50%	75%	max
WTI	4849.0	62.233	26.540	17.450	40.150	59.630	84.050	145.29
GOLD	4849.0	940.95	474.51	255.10	423.50	1046.4	1331.2	1888.7
SP 500	4849.0	1534.8	540.35	676.00	1145.7	1349.2	1923.7	2948.5
US DOLLAR	4849.0	90.438	11.706	71.304	80.911	88.240	96.881	121.21
US 10YR	4849.0	3.4540	1.2140	1.3580	2.3860	3.4080	4.4120	6.7300
DJU	4849.0	459.36	147.32	167.57	359.82	438.14	553.40	812.90

- **Normalization**

In this step compute also the normal distribution to known the relationship between the values of each feature through different 10 years with the price as shown in figure 4.1 as shown in figure 4.1 .



Figure(4.1): Normalization

4.2.3 Results of Multivariate Analysis

- **Person Correlation**

After pre-processing complete , begin the multivariate analysis of features to determine the important of each feature and degree of effect on the target (i.e. ,oil price) .Therefore; we will compute person correlation ; In general ;persons correlation which is used to find the relationship or the degree of correlation between the two variables whose value is between {-1, +1}, and the direction of this relationship, by correlation can conclusion four types of relationship among the variables:

- All wise the correlation of feature with themselves equal 1 .
- Positive relationship meaning both features inverse together and that inverse direct the +1.
- Negative relationship meaning both features inverse together and that inverse direct the -1.
- Zero relationship meaning there are no relation between both features.

Table (4.4):Results of Person Correlation

Features	WTI	GOLD	SP500	USDOLAR	US10YR	DJU
WTI	1.000000	0.668837	0.086988	-0.822593	-0.404814	0.286392
GOLD	0.668837	0.000000	0.510900	-0.531831	-0.871586	0.656323
SP500	**0**.086988	0.510900	1.000000	0.099090	-0.473430	0.926697
USDOLAR	-0.822593	-0.531831	0.099090	1.000000	0.376463	-0.115449
US10YR	-0.404814	-0.871586	-0.473430	0.376463	1.000000	-0.603977
DJU	0.286392	0.656323	0.926697	-0.115449	-0.603977	1.000000

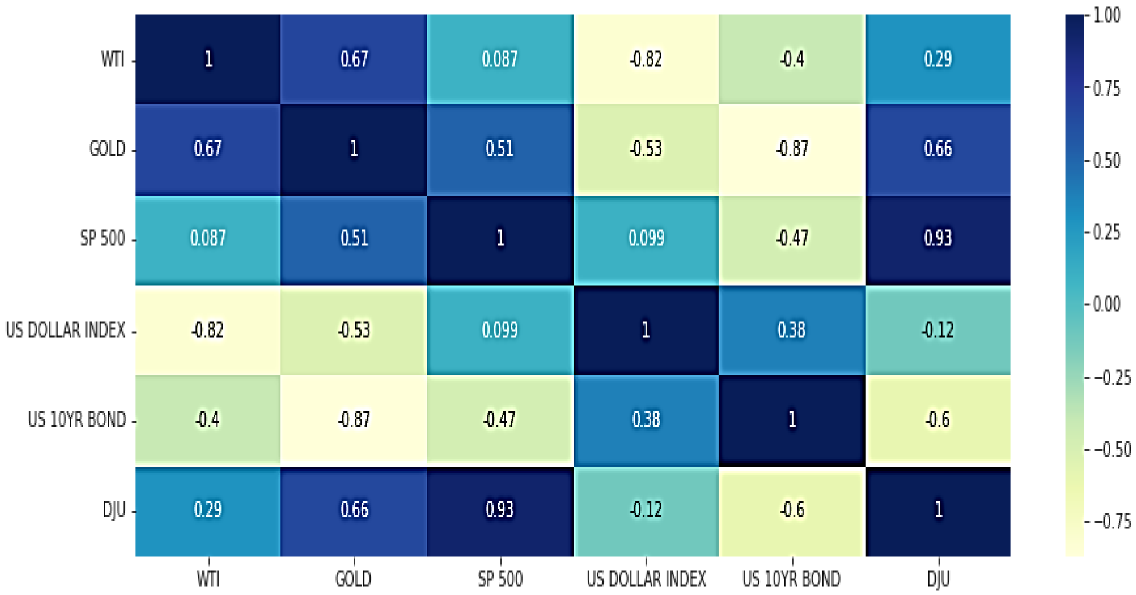


Figure (4.2): Results of Person Correlation

Table 4.4 and figure 4.2 shown the relationship among the features:

- Highest positive correlation WTI with gold = 0.67.
- We also note that SP500 and WTI have less correlation = 0.087 .
- High negative correlation between WTI and the US dollar index = -0.82.
- Medium positive correlation between DJU and WTI =0.29 .
- US 10 years and WTI are inversely correlated = -0 .4 .
- We note that there is a very high positive correlation between DJU and SP 500 = 0.93 .
- We note the relationship between SP500 and US DOLARINDEX is very weak because it is close to zero = 0.099.Likewise for other values.

The important of characteristics are examined and processed using an entropy scale and an information gain scale. Where the term "entropy" refers to the measure of uncertainty. In other words, the more random a variable is, the higher its entropy value is; Hence, when the entropy value is 0, it shows that the data is bad for learning, while the entropy value of 1 shows that the data is good for learning. In order to measure the degree of randomness, we will use the information gain metric, where the lower the entropy, the more information is obtained for decision-making and choosing the best partition to obtain properties with pure data that it can in the typical training phase of the algorithm used to predict oil prices. As a result, Entropy and information gain are inversely related; the lower the entropy, the greater the information acquisition, and consequently, the better predictability, and ultimately, the desired results. To choose the best evaluation of the data's features, information gain is utilized. As seen in the Table 4.5:

Table (4.5): MVA base on Entropy and Information Gain

Name of feature	Entropy of features	Information Gain
WTI	11.72219	11.72219
GOLD	11.74697	11.22737
SP 500	11.47777	10.95874
US DOLLAR INDEX	11.86626	11.34671
US 10YR BOND	11.24566	10.72665
DJU	12.13483	11.61412

4.2.4 Results of prediction(GRU)

At this stage, the pure data is entered into the model designed to predict oil prices through the (GRU) algorithm used to train the characteristics data resulting from the preprocessing. Table 4.6 shows the results from 50 Epoch to calculate the error rate based on three main measures (R2 ,MSE, MAE), where the total implementation time was (43.796) , while number of hidden layer =3 , activation function is tansh , while windows size =10 and total number of epoch = 50 .

Table (4.6): Results of GRU

# Iterations	MSE	MAE
Epoch 1/50	0.071550049	0.226995587
Epoch 2/50	0.05047923	0.195226431
Epoch 3/50	0.050553888	0.195025399
Epoch 4/50	0.050536904	0.195037976
Epoch 5/50	0.050507609	0.195116967
Epoch 6/50	0.050560009	0.195175931
Epoch 7/50	0.050520163	0.195246294
Epoch 8/50	0.050574478	0.194959551
Epoch 9/50	0.050525077	0.195000052
Epoch 10/50	0.050553054	0.195270523
Epoch 11/50	0.050592862	0.194887325
Epoch 12/50	0.050523791	0.195016965
Epoch 13/50	0.050608497	0.195157528
Epoch 14/50	0.050624907	0.195070922
Epoch 15/50	0.050545406	0.195120275
Epoch 16/50	0.050549649	0.195141643
Epoch 17/50	0.050485484	0.195048869
Epoch 18/50	0.0505859	0.195091203
Epoch 19/50	0.050498731	0.195062712
Epoch 20/50	0.050513774	0.195147857
Epoch 21/50	0.050562486	0.195265546
Epoch 22/50	0.050572738	0.194986105
Epoch 23/50	0.050522637	0.195187658
Epoch 24/50	0.050570983	0.19492352
Epoch 25/50	0.050574798	0.195028767
Epoch 26/50	0.05049159	0.194986135
Epoch 27/50	0.05052375	0.195039347

Epoch 28/50	0.050485361	0.195089012
Epoch 29/50	0.050545663	0.194996879
Epoch 30/50	0.050544661	0.195151016
Epoch 31/50	0.050547466	0.195161283
Epoch 32/50	0.05055977	0.195048094
Epoch 33/50	0.050519958	0.19513756
Epoch 34/50	0.050512973	0.195099309
Epoch 35/50	0.050503902	0.194896117
Epoch 36/50	0.050505612	0.195073172
Epoch 37/50	0.050649598	0.195155561
Epoch 38/50	0.050533637	0.195063382
Epoch 39/50	0.050504424	0.195067212
Epoch 40/50	0.050476044	0.194865659
Epoch 41/50	0.050513979	0.195024297
Epoch 42/50	0.050518908	0.195127293
Epoch 43/50	0.050537635	0.195164934
Epoch 44/50	0.050595324	0.194961667
Epoch 45/50	0.050592467	0.19504638
Epoch 46/50	0.050557058	0.194740504
Epoch 47/50	0.050530098	0.195137724
Epoch 48/50	0.050513219	0.195086613
Epoch 49/50	0.05058201	0.19524844
Epoch 50/50	0.050560113	0.194801405

4.2.5 Results of Evaluations

After, complete the building of predictor model based on combination between multivariate analysis and Gate Recurrent unit based on using the five cross validation as measure to split of dataset into training and dataset and compute the three evaluation measures for each spilt of 5.cross validation to choose the best split that generated less error . In general ;in this project used the split 80 % of dataset as training data to build the model . while 20% of dataset to testing the effectiveness of model. The result ,for training shown table 4.7 . on other hand the result of evaluation measure to testing dataset shown in 4.8 .

Table (4.7) : Evaluations Measures for Training Dataset

Five score validations	R2	MSE	MAE
80 % -20 %	0.97	0.0505	0.0304
60 % -40 %	0.91	0.1005	0.0805
50 % -50 %	0.87	0.1306	0.1002
40 % -60 %	0.80	0.1808	0.1204
20 % -80 %	0.72	0.2205	0.1608

Table (4.8) : Evaluations measures for Testing Dataset

Five score validations	R2	MSE	MAE
20 % -80 %	**0.77**	0.2504	0.1406
40 % -60 %	0.81	0.2006	0.1308
50 % -50 %	0.85	0.1601	0.1109
60 % -40 %	0.91	0.1403	0.0909
80 % -20 %	0.95	0.0905	0.0508

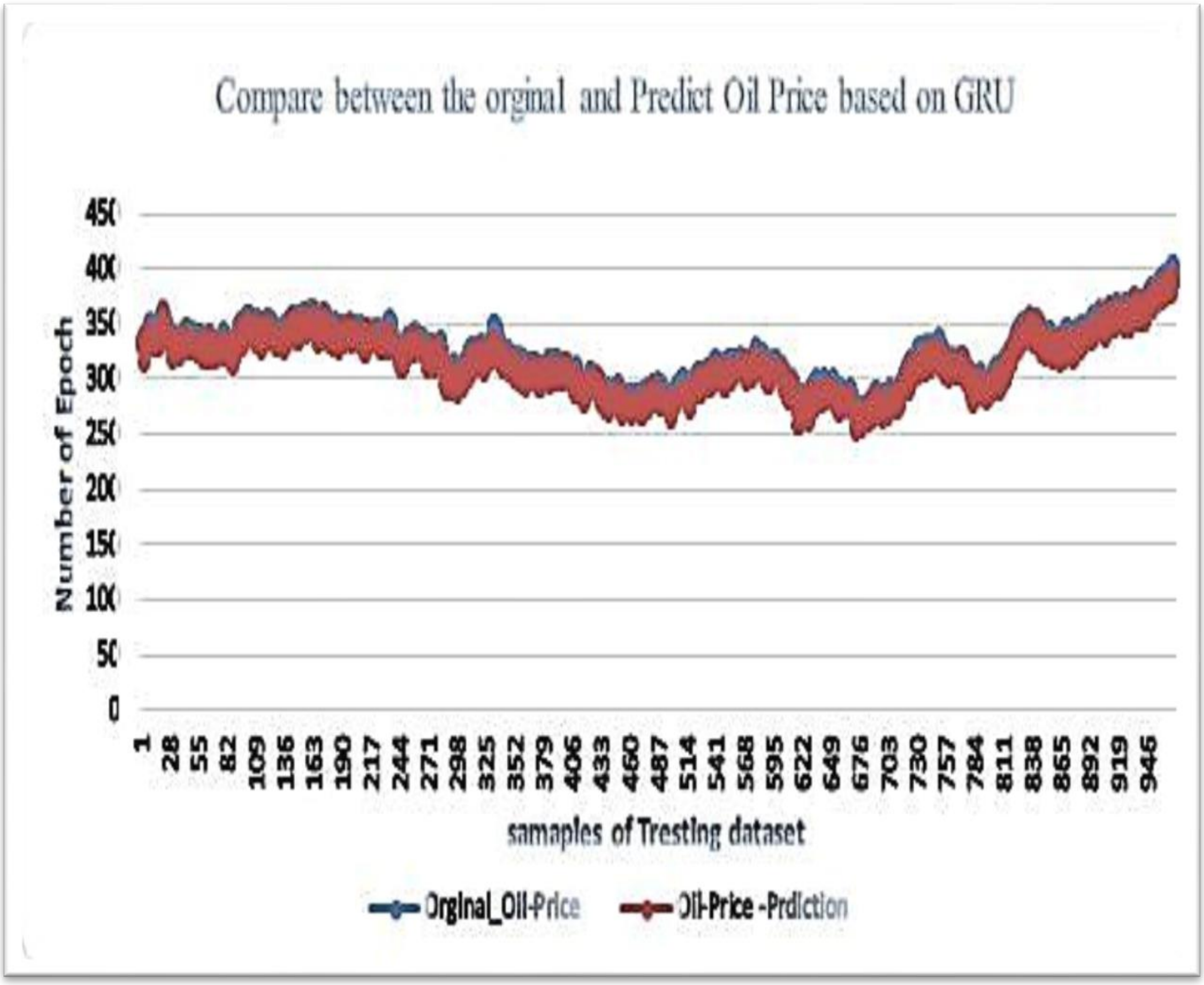


Figure (4.3): Compare between original and predict Oil Price Based on GRU

Chapter Five:

Conclusions and Future Work

Chapter Five: Conclusions And Future Works

5.1 Introduction

This chapter reviews the most important conclusions reached through applying the HMOP-NCT to the dataset and focuses on how to address the challenges raised in the previous chapters, which include (programming challenges and application challenges). In addition, we will suggest a set of recommendations for researchers to work on it in the future.

5.2 Conclusions

A. The problem of prediction the oil prices needs highly efficient techniques. Therefore, the challenge for this project is how to avoid these limitations which are believed to adopt an effective oil price prediction technique based on many features and capture data from several years prior.

B. The data used in this project is distinguished by contain missing values in multi features while the goal of this project build predictor that based on the law the results of prediction become true if predictor build from the fact otherwise the results become virtual therefore; we dropping all the row that contain missing values; in general number of these data is 152 from total dataset rows that include 4947 and remained only 4849 as fact used in build that predictor.

C. In this work, a prediction model will be designed and implemented that determines oil prices based on seven basic characteristics. The proposed system is based on knowledge discovery techniques through data analysis (which includes 7 characteristics to ten years). The proposed model consists of three main stages From the correlation we found; The highest positive correlation with target variable WTI is with Gold(ρ: 0.67); We also observe that SP500 and WTI have less correlation.(ρ : 0.087); US 10 YR Bond and WTI are moderately inversely correlated (ρ : -0 .4); DJU & WTI positive moderate correlation.(ρ : 0.29); US Dollar Index & WTI have high correlation .(ρ : -0.82), We observe high correlation between DJU & SP 500 (ρ : 0.93) observed the SP 500 and DJU also have very high positive correlation. This is kind of Expected since both are indexes. While; As shown in the Normal Distribution The range of 25 to 35 is where WTI values were mostly frequently found. Less frequently, the WTI oil price rises above 120 , All Features don't follow Normal Distribution. Finally; the predictor characteristic by it need Less memory consumption, faster execution, less training time.

D. The GRU network is considered one of the important neuro computing techniques because it is characterized by high speed and accuracy in extracting results. But on the other hand it needs to specify a large number of parameters. This work handle this problem through using multivariate analysis.

E. Give the approximation prices of oil can’t base on the small data or momentum therefore; this project analysis the data related to multi several years(2010-2019) to build accuracy predictor of oil price. On other hand the build of these types of model need to expended knowledge in mathematical and programming.

5.3 Recommendations:

A. It is possible to find an optimal method to determine the original parameters of the GRU to reduce its execution time, such as using (Genetic Algorithm(GA), practical swarm optimization PSO, and whale optimization algorithm (WOA) algorithms).

B. The same project can be implemented using another prediction techniques related to Data Mining rather than of neurcomputing, such as MARS data mining algorithm that based on mathematical principle .

C. Can test the accuracy of model based on another types of evaluation measures such as Confusion Matrices that include five measures(i.e., Accuracy, Precision, Recall , F, Fb).

References

References

(Al-Janabi, S., Alkaim, A. F., & Adel, Z. (2020)). An Innovative synthesis of deep learning techniques (DCapsNet & DCOM) for generation electrical renewable energy from wind energy. Soft Computing, 24(14), 10943-10962.https://doi.org/10.1007/s00500-020-04905-9

(Al-Janabi, S., Alkaim, A., Al-Janabi, E., Aljeboree, A., & Mustafa, M. (2021)). Intelligent forecaster of concentrations (PM2. 5, PM10, NO2, CO, O3, SO2) caused air pollution (IFCsAP). Neural Computing and Applications, 33(21), 14199-14229.https://doi.org/10.1007/s00521-021-06067-7

(Zabezhailo, M. I., & Trunin, Y. (2019)). On the problem of medical diagnostic evidence: intelligent analysis of empirical data on patients in samples of limited size. Automatic Documentation and Mathematical Linguistics, 53(6), 322-328.https://doi.org/10.3103/S0005105519060086

(Peres, R. S., Rocha, A. D., Leitao, P., & Barata, J. (2018)). IDARTS–Towards intelligent data analysis and real-time supervision for industry 4.0. Computers in industry, 101, 138-146.https://doi.org/10.1016/j.compind.2018.07.004

(Li, X., Shang, W., & Wang, S. (2019)). Text-based crude oil price forecasting: A deep learning approach. International Journal of Forecasting, 35(4), 1548-1560.https://doi.org/10.1016/j.ijforecast.2018.07.006

(Alkaim A.F., Al_Janabi S. (2020)).Multi Objectives Optimization to Gas Flaring Reduction from Oil Production. In: Farhaoui Y. (eds) Big Data and Networks Technologies. BDNT 2019. Lecture Notes in Networks and Systems, vol 81. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-23672-4_10

(Hair, J.F. (2011)) Multivariate Data Analysis: An Overview. In: Lovric, M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04898-2_395

(Vighnesh D, 2021),” Multivariate analysis: an overview”, https://s4be.cochrane.org/blog/2021/09/09/multivariate-analysis-an-overview

(Al-Janabi, S. & Alkaim, A.F(2020). A nifty collaborative analysis to predicting a novel tool (DRFLLS) for missing values estimation, Springer, Soft Comput (2020), Volume 24, Issue 1, pp 555–569. DOI 10.1007/s00500-019-03972-x

(Xu, Q., Fu, B., & Wang, B. (2022)) ,The effects of oil price uncertainty on China’s economy. *Energy Economics*, *107*, 105840. https://doi.org/10.1016/j.eneco.2022.105840

(Abbass, K., Sharif, A., Song, H., Ali, M. T., Khan, F., & Amin, N. (2022)) ,Do geopolitical oil price risk, global macroeconomic fundamentals relate Islamic and conventional stock market? Empirical evidence from QARDL approach. Resources Policy, 77, 102730. https://doi.org/10.1016/j.resourpol.2022.102730

(Sun, C., Zhan, Y., Peng, Y., & Cai, W. (2022)) ,Crude oil price and exchange rate: Evidence from the period before and after the launch of China's crude oil futures. Energy Economics, *105*, 105707. https://doi.org/10.1016/j.eneco.2021.105707

(He, H., Sun, M., Li, X., & Mensah, I. A. (2022)),A novel crude oil price trend prediction method: Machine learning classification algorithm based on multi-modal data features. Energy, 244, 122706 https://doi.org/10.1016/j.energy.2021.122706

(Gupta, N., & Nigam, S. (2020)). Crude oil price prediction using artificial neural network. *Procedia Computer Science*, *170*, 642-647. https://doi.org/10.1016/j.procs.2020.03.136

(Carpio, L. G. T. (2019)),The effects of oil price volatility on ethanol, gasoline, and sugar price forecasts. *Energy*, *181*, 1012-1022. https://doi.org/10.1016/j.energy.2019.05.067

(Li, T., Qian, Z., Deng, W., Zhang, D., Lu, H., & Wang, S. (2021)). Forecasting crude oil prices based on variational mode decomposition and random sparse Bayesian learning. Applied Soft Computing, 113, 108032,https://doi.org/10.1016/j.asoc.2021.108032

(Good, P.(2005)),Multivariate analysis. Permutation, Parametric and Bootstrap Tests of Hypotheses, 169-188 , https://doi.org/10.1007/0-387-27158-9_9

(Wang, M., Tian, L., & Zhou, P. (2018)). A novel approach for oil price forecasting based on data fluctuation network. Energy Economics, 71, 201-212, https://doi.org/10.1016/j.eneco.2018.02.021

(Grace, G. (2018, October 30)). What is the difference between univariate and multivariate data analysis. Ask Data Science , https://askdatascience.com/491/what-difference-between-univariate-multivariate-analysis .

(Bhumika Dutta ,(2021)),” What is Multivariate Data Analysis?”, https://www.analyticssteps.com/blogs/what-multivariate-data-analysis .

(Sateesh Gottumukkala ,(2018)) ,” How to Select Multivariate Data Analysis Techniques ?-Part 1” , https://www.linkedin.com/pulse/multivariate-data-analysis-techniques-part-1-sateesh-gottumukkala

(Hua-Kai Chiou (2005)),"Multivariate Statistical Data Analysis with Its Applications" , https://slideplayer.com/slide/6661200

(Ray Poynter ,(2017)), "Some thoughts on Metric and Nonmetric Data in Market Research" , https://www.linkedin.com/pulse/some-thoughts-metric-nonmetric-data-market-research-ray-poynter

(Harsha Nimkar & Great Learning Team ,(Jul 29, 2020)), " Overview of Multivariate Analysis | What is Multivariate Analysis and Model Building Process?" , https://www.mygreatlearning.com/blog/introduction-to-multivariate-analysis .

(Ganya Erickson, (2014)),"Other Multivariate Techniques", https://www.slideserve.com/ganya/other-multivariate-techniques .

(Zeiler, M.D., Fergus, R. (2014, September)),Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53

(Russakovsky, O., Deng, J., Su, H. et al,(2015)). ImageNet Large Scale Visual Recognition Challenge. Int J Comput Vis 115, 211–252 (2015). https://doi.org/10.1007/s11263-015-0816-y

(Divya Shree, B., Brunda, R., Shobha Rani, N. (2019)),Fruit Detection from Images and Displaying Its Nutrition Value Using Deep Alex Network. In: Wang, J., Reddy, G., Prasad, V., Reddy, V. (eds) Soft Computing and Signal Processing . Advances in Intelligent Systems and Computing, vol 898. Springer, Singapore. https://doi.org/10.1007/978-981-13-3393-4

(Sam, S. M., Kamardin, K., Sjarif, N. N. A., & Mohamed, N.(2019)) ,Offline signature verification using deep learning convolutional neural network (CNN) architectures GoogLeNet inception-v1 and inception-v3. Procedia Computer Science, 161, 475-483 .https://doi.org/10.1016/j.procs.2019.11.147

(Szegedy.et.al ,(2015)) "Comparing Local Descriptors and Bags of Visual Words to Deep Convolutional Neural Networks for Plant Recognition - Scientific Figure on ResearchGate " https://www.researchgate.net/figure/The-illustration-of-the-GoogleNet-architecture-Szegedy-et-al-2015-All-convolutional_fig1_314119821

(Arora S, Bhaskara A, Ge R, and Ma T (2013)), "Provable bounds for learning some deep representations," arXiv:abs/1310.6343

(Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. (2018)). Brain intelligence: go beyond artificial intelligence. Mobile Networks and Applications, 23(2), 368-375. https://doi.org/10.1007/s11036-017-0932-8

(Shao, W., Zhang, Y., Guo, B., Qin, K., Chan, J., Salim, F.D. (2019)).Parking Availability Prediction with Long Short Term Memory Model. In: Li, S. (eds) Green, Pervasive, and Cloud Computing. GPC 2018. Lecture Notes in Computer Science(), vol 11204. Springer, Cham. https://doi.org/10.1007/978-3-030-15093-8_9

(Rahman, Md & Siddiqui,. (2019)),An Optimized Abstractive Text Summarization Model Using Peephole Convolutional LSTM. Symmetry. 11. 1290. DOI: 10.3390/sym11101290

Abedin, M.Z., Moon, M.H., Hassan, M.K. et al. (2021),Deep learning-based exchange rate prediction during the COVID-19 pandemic. Ann Oper Res. https://doi.org/10.1007/s10479-021-04420-6

(Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M (2020)). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. Applied soft computing, 90, 106181.https://doi.org/10.1016/j.asoc.2020.106181

(Datta, R.K., Sajid, S.W., Moon, M.H., Abedin, M.Z. (2021)). Foreign Currency Exchange Rate Prediction Using Bidirectional Long Short Term Memory. In: Musleh Al-Sartawi, A.M.A. (eds) The Big Data-Driven Digital Economy: Artificial and Computational Intelligence. Studies in Computational Intelligence, vol 974. Springer, Cham. https://doi.org/10.1007/978-3-030-73057-4_17

(Liu, P., Qiu, X., Chen, X., Wu, S., & Huang, X. J. (2015)), September). Multi-timescale long short-term memory neural network for modelling sentences and documents. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 2326-2335), https://aclanthology.org/D15-1280.pdf .

(Jiao, M., Wang, D., & Qiu, J. (2020)). A GRU-RNN based momentum optimized algorithm for SOC estimation. Journal of Power Sources, 459, 228051. https://doi.org/10.1016/j.jpowsour.2020.228051

(Ganya Erickson (2014)),"Other Multivariate Techniques " , https://www.slideserve.com/ganya/other-multivariate-techniques

(Sparkle Russell-Puleri and Dorian Puleri , (Feb 25, 2019)),” Towards Data Science,Gated Recurrent Units explained using matrices: Part 1 ", https://towardsdatascience.com/gate-recurrent-units-explained-using-matrices-part-1-3c781469fc18

(Ambika Choudhury ,(october 16, 2020)) ,” IN DEVELOPERS CORNER,Gated Recurrent Unit – What Is It And How To Learn “, https://analyticsindiamag.com/gated-recurrent-unit-what-is-it-and-how-to-learn .

(Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014)). Learning phrase representations using RNN encoder-decoder for statistical machine translation. https://doi.org/10.48550/arXiv.1406.1078

(DENNY BRITZ, OCTOBER 27, 2015), “Recurrent Neural Network Tutorial, Part 4 – Implementing a GRU/LSTM RNN with Python and Theano” , http://www.wildml.com/2015/10/recurrent-neural-network-tutorial-part-4-implementing-a-grulstm-rnn-with-python-and-theano

(Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014)). Empirical evaluation of gated recurrent neural networks on sequence modeling. https://arxiv.org/abs/1412.3555v1

Xin Wang et .al 2019 “An Optimized Gated Recurrent Unit Neural Network”, https://iopscience.iop.org/article/10.1088/1742-6596/1325/1/012089/pdf

(Wang, Yusen, Wenlong Liao, and Yuqing Chang. (2018)). "Gated Recurrent Unit Network-Based Short-Term Photovoltaic Forecasting" Energies 11, no. 8: 2163. https://doi.org/10.3390/en11082163

(Gabriel Loye ,(Jul 22, 2019)) ,” Gated Recurrent Unit (GRU) With PyTorch” , https://blog.floydhub.com/gru-with-pytorch

(Zhou, GB., Wu, J., Zhang, CL. et al. (2016)) .Minimal gated unit for recurrent neural networks. Int. J. Autom. Comput. 13, 226–234 . https://doi.org/10.1007/s11633-016-1006-2

Dey, Rahul; Salem, Fathi M., 2017. "Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks", https://doi.org/10.48550/arXiv.1701.05923

(Chicco, D., Warrens, M. J., & Jurman, G. (2021)). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*, e623. https://doi.org/10.7717/peerj-cs.623

الخلاصة

يعتبر التنبؤ بأسعار النفط وتحديد نسبة الزيادة او الانخفاض في اسعارها واحد من اهم العوامل المؤثرة في اقتصاد الدول وموثر فعال في ميزانية اي دولة نفطية ومنها العراق وعامل اساسي في سوق العمل بشكل عام. تمثل تقنيات التنبؤ اداة فعالة لاكتشاف المعرفة من قواعد البيانات الضخمة و المعقدة في مجالات عدة منها اسعار النفط. في هذا العمل، سيتم تصميم و تنفيذ نموذج تنبؤ يقوم بتحديد اسعار النفط بالاعتماد على سبع خصائص اساسية وهي (Date, WTI, GOLD, SP 500, US DOLLAR INDEX, US 10YR BOND, DJU). و يستند النظام المقترح على تقنيات الشبكات المعرفة من خلال تحليل 7 خصائص إلى عشر سنوات. يتكون النموذج المقترح من ثلاث مراحل اساسية : المرحلة الاولى تسمى المعالجة الاولية وتتضمن تدقيق البيانات الخارجة عن النطاق والمفقودة واجراء عملية تنظيم للبيانات لتحضيرها الى المرحلة التالية التي تم فيها تحديد الخصائص الاكثر اهمية من خلال حساب(الترابط ما بين الخصائص , الانتروبي , information gain). بعد ذلك تم تقسيم البيانات الى جزئيين بيانات تدريب واختبار. حيث استخدم الجزء الاول لبناء متنبئ سمي HMOP-NCT بينما القسم الثاني استخدم لتقييم HMOP-NCT باعتماد ثلاث مقاييس هي (R2 , MSE ,MAE) . اثبت النموذج المصمم HMOP-NCT قابليته على تجهيزنا بنتائج جيدة ومقبولة بأسعار النفط مع نسبة خطأ قليلة . حيث تحليل (multivariate) أثبت ان العناصر الاكثر تأثير بأسعار النفط هي (WTI ,GOLD , US DOLAR) , فكانت قيم information gain لهم هي (WTI=11.272 ,GOLD =11.227 , DJU=11.614) , بينما تقنية الشبكة العصبية المعمقة وهي (GRU) اوضحت قدرتها على معالجة قاعدة بيانات ذات خصائص مختلفة من حيث السلوك لعدة سنوات . واعطت نتائج مقبولة مع نسبة خطأ قليلة فكانت (R2=0.945 MSE=0.0505,MAE=0.1948) بوقت تنفيذ قصير نسبيا".

جمهورية العراق
وزارة التعليم العالي والبحث العلمي
جامعة بابل كلية العلوم للبنات
قسم علوم الحاسوب

التنبؤ بأسعار النفط باستخدام تقنية الشبكات العصبية المعمقة (GRU) وتحليل متعدد المتغيرات

بحث

مقدم الى مجلس كلية العلوم للبنات في جامعة بابل كجزء من متطلبات الحصول على درجة الدبلوم العالي في العلوم / علوم الحاسوب

من قبل
حنان غانم عبد الواحد

بأشراف
أ. د سماهر الجنابي

2022 ميلادي **1444 هجري**