Discover Artificial Intelligence

Creating a cutting-edge neurocomputing model with high precision

Mahdi Abed Salman[1] · Samaher Al-Janabi[1] ⬤

Abstract

The prediction of oil prices has a significant impact on the economies of countries, particularly in oil-rich nations like Iraq, and affects the labor market. Prediction techniques are vital tools for extracting knowledge from complex databases, such as oil prices. This study aims to develop a prediction model that accurately determines oil prices based on seven fundamental characteristics, including Date, WTI, GOLD, SP 500, US DOLLAR INDEX, US 10YR BOND, and DJU. The proposed model utilizes advanced neurocomputing techniques that analyze the seven features over a ten-year period. The model comprises three main stages: preprocessing, determining feature importance through computing correlation, entropy, and information gain, and splitting the dataset into training and testing. The first part of the dataset builds the predictor called Hybrid Model to Oil Price based on Neurocomputing Techniques, while the second part evaluates model using three error measures: R2, MSE, and MAE. The model proves its ability to provide accurate predictions with low error rates. Multivariate analysis shows that WTI, GOLD, and US DOLLAR INDEX have a more significant impact on oil prices, with information gain values of WTI = 11.272, GOLD = 11.227, and DJU = 11.614. The Gate Recurrent Unit neurocomputing technique demonstrates its ability to handle datasets with features that behave differently over multiple years and provides accurate predictions with low errors in a short time, withR2 = 0.945, MSE = 0.0505, and MAE = 0.1948. This study provides valuable insights into the prediction of oil prices and highlights the efficacy of advanced neurocomputing techniques for extracting knowledge from complex databases.

Keywords Network of oil · GRU network · Forecasting · Advanced neurocomputing

1 Introduction

Data is a valuable asset that serves as the foundation for various fields in computer science. It can be generated from different sources and analyzed using various techniques [1]. Data science is an interdisciplinary field that combines data, intelligence, and statistics to extract useful knowledge from data. Intelligent Data Analysis (IDA) [2] involves preparing, mining, and validating data to obtain accurate results. The advantage of IDA is that it can work on real-world problems and produce high-accuracy results. However, the increasing size of data sets in the real world leads to complex calculations that require a lot of time to reach the desired results. Deep learning is a subset of artificial intelligence that combines three types of learning: supervised, unsupervised, and reinforcement. Prediction is a vital principle in designing models that can forecast future events or find unknown variables based on existing data or evidence. Multivariate Analysis (MVA) [3] is a statistical procedure that analyzes data with multiple variables to obtain more accurate conclusions. MVA techniques can be divided into two parts: dependence and interdependence techniques. The fundamental benefit of

✉ Samaher Al-Janabi, samaher@itnet.uobabylon.edu.iq | [1]Department of Computer Science, Faculty of Science for Women (SCIW), University of Babylon, Babylon, Iraq.

Discover

MVA [4] is that it considers multiple independent variables that can affect the variability of dependent variables, leading to more accurate conclusions. Accurately predicting oil prices is a challenging task that requires efficient techniques [5]. This study aims to build a reliable prediction model for oil prices by defining the variables or characteristics that influence oil prices, collecting and analyzing data, and studying the factors affecting oil. The challenges of this work include avoiding the limitations of the GRU algorithm, suchas high computation and trial-and-error in determining parameters, and building an efficient technique to predict oil prices based on various features over ten years. The benefits of this study can be summarized as follows:

- The study aims to build a reliable prediction model for oil prices by analyzing the variables or characteristics that influence oil prices. This will help in accurately forecasting oil prices, which is crucial for the global economy.
- The study uses MVA techniques, which take into account multiple independent variables that can affect the variability of dependent variables, resulting in more accurate conclusions.
- The study utilizes neuro computing techniques, such as the GRU algorithm, which gives high-accuracy results and works with real data. By combining MVA and neuro computing techniques, the study aims to build an efficient technique to predict oil prices.
- The study is focused on a real-world problem, which makes it highly relevant and useful. The results obtained from this study can be applied to real-world scenarios, such as the global economy.
- The study is an example of how the advancement in technology, such as deep learning and MVA, can be utilized to solve complex problems and obtain accurate results.
- Accurate forecasting of oil prices can lead to better decision-making, such as in the energy sector, where it can help in planning and investment decisions.
- The study contributes to the field of artificial intelligence and data analysis by exploring the application of MVA and neuro computing techniques in predicting oil prices. The findings of this study can serve as a guide for future research in this area.

2 Related works

The prediction of oil prices and determining the percentage of increase or decline is a crucial factor that affects the economy of countries, particularly oil-producing nations like Iraq. This section of the study aims to review previous research in the same field and compare it based on five points, including author(s) name, database/dataset used, pre-processing, and methodology measure. Table 1 illustrates the comparison.

Xu et al. [8, 14] investigated the relationship between China's economic activity and the volatility of the world oil price using a vector auto-regression model with stochastic volatility in the mean. Their research is similar to ours in terms of working with the same dataset, but we use different pre-processing and methodology.

Abbas et al. [9, 15] used the QARDL approach to calculate the monthly average empirically return of the traditional stock market index. The study focused on the impact of geopolitical oil price risk on assets in a bullish or bearish scenario. Our work differs from theirs in terms of data processing method and the use of deep neurocomputing techniques with evaluation measures.

Sun et al. [10, 16] compared the crude oil price with the current price in China to determine any correlation. Their study used a model that revealed the change in crude oil prices over time, showing the clear effects of the introduction of crude oil futures contracts. Our work, on the other hand, determines the oil price through analysis of seven different features based on multivariate analysis.

He al. [11, 17] proposed a new hybrid forecasting model that uses VMD and ML algorithms to anticipate the trend in crude oil prices. Their study found that the hybrid prediction model with a support vector machine classifier has greater predictive power than other classifiers. Additionally, their model is more accurate at forecasting high volatility than low volatility in crude oil prices. Our work differs from theirs in terms of the prediction model used to obtain the desired results.

Gupta et al. [12, 18] proposed a novel approach to predict oil prices using an artificial neural network that considers the shifting pattern of oil price variations by determining the appropriate delay and the number of delay effects that regulate oil prices. Their study found that the proposed model outperformed the control group in terms of accuracy.

Table 1 Compare among previous works

Name	Dataset/database	Preprocessing	Methodology
Xu and Wang [14]	The effects of oil price uncertainty on China's economy https://doi.org/10.1016/j.eneco.2022.105840	Autoregression model (AR-model)	New-Keynesian model
Abbass et.al. [15]	Geopolitical oil price risk https://doi.org/10.1016/j.resourpol.2022.102730	Autoregressive distributed lag (ARDL) models	Quantile Autoregressive Distributed Lags Errors Correction Model (QARDL)
Sun et.al. [16]	Crude oil price and exchange rate https://doi.org/10.1016/j.eneco.2021.105707	MS-VAR model	Markov-switching vector autoregressive model
He et.al. [17]	Crude oil price https://doi.org/10.1016/j.energy.2021.122706	Multi-modal data	Variational mode decomposition (VMD), (ML)
Gupta and Nigam [18]	Crude Oil Price Prediction https://doi.org/10.1016/j.procs.2020.03.136	Artificial neural network (ANN) model	Oil Price Prediction Model
Carpio [19]	THE effects of oil price volatility on ethanol, gasoline, and sugar price forecasts https://doi.org/10.1016/j.energy.2019.05.067	Long-term and short-term model	(VECX) model

Our work is similar to theirs in terms of using intelligent methods of processing to obtain results, but we differ in terms of the prediction model used.

Carpio et al. [13, 19] analyzed the long-term and short-term implications of oil prices on ethanol, gasoline, and sugar price forecasts using the vector error correction with exogenous variable (VECX) model. Their study found that the oil price has long-term implications for the other three price estimates, and forecasts for ethanol and gasoline costs are more vulnerable to changes in upcoming oil prices than forecasts for sugar prices. Our work is similar to theirs in terms of using the same dataset, but we use a different prediction technique called Gate recurrent unit.The accurate prediction of oil prices and the determination of the percentage of increase or decrease is a critical factor that impacts the economy of countries, especially oil-producing nations like Iraq. In this section, we review previous research in the same field and compare it based on five key points: author(s) name, database/dataset used, pre-processing, methodology measure, and present the findings in Table 1.

Xu et al. [8, 14] examined the relationship between China's economic activity and the volatility of the world oil price using a vector auto-regression model with stochastic volatility in the mean. Though they worked on a similar dataset, our study used different pre-processing and methodology measures.

Abbas et al. [9, 15] utilized the QARDL approach to calculate the monthly average empirically return of the traditional stock market index. The study focused on the impact of geopolitical oil price risk on assets in both bullish and bearish scenarios. Our study differs in terms of the data processing method and the use of deep neurocomputing techniques with evaluation measures (Table 1).

Sun et al. [10, 16] compared the crude oil price with the current price in China to identify any correlation. Their model revealed the change in crude oil prices over time and the effects of introducing crude oil futures contracts. Our study, on the other hand, analyzes seven different features using multivariate analysis to determine the oil price.

He et al. [11, 17] proposed a new hybrid forecasting model that combines VMD and ML algorithms to predict the trend in crude oil prices. Their study found that the hybrid prediction model with a support vector machine classifier was more effective at predicting high volatility than low volatility in crude oil prices. Our study differs in terms of the prediction model used to obtain the desired results.

Gupta et al. [12, 18] proposed a novel approach to predict oil prices using an artificial neural network that considers the changing pattern of oil price variations by determining the appropriate delay and number of delay effects that regulate oil prices. Their study found that the proposed model outperformed the control group in terms of accuracy. Our study is similar in terms of using intelligent processing methods to obtain results, but we differ in terms of the prediction model used.

Carpio et al. [13, 19] analyzed the long-term and short-term implications of oil prices on ethanol, gasoline, and sugar price forecasts using the vector error correction with exogenous variable (VECX) model. Their study found that the oil price has long-term implications for the other three price estimates, and forecasts for ethanol and gasoline costs are more vulnerable to changes in upcoming oil prices than forecasts for sugar prices. Our study is similar in terms of using the same dataset, but we use a different prediction technique called Gate recurrent unit.

3 Multivariate analysis

Multivariate Analysis (MVA) is a statistical technique used for analyzing data sets with more than one variable. This approach is complex but essential, as it considers multiple variables to minimize bias and provide accurate results [13]. MVA involves examining and analyzing several statistical variables simultaneously, considering the impact of all factors on the desired outcomes across multiple dimensions, especially when working with correlated data [20]. Since multivariate equations govern everything in the world, many issues arise from multiple reasons [21].

MVA techniques can be divided into two categories: dependence and interdependence techniques. The dependence approach involves identifying one or more dependent variables to be predicted or explained by other variables known as independent variables. Machine learning is an example of a dependence technique used to build predictive models based on input data and the independent variables used for prediction. Multiple regression is an example of a dependence technique [22].

Dependence techniques can be classified based on the number of dependent variables and the measurement scales used by the variables [12]. Dependence approaches can be further divided into three categories: techniques with a single dependent variable, multiple dependent variables, or multiple dependent/independent connections [22]. The second

category of dependence techniques can be classified based on whether the dependent variables are metric (quantitative/numeric) or non-metric (qualitative/categorical) [22]. If a study involves only one metric-dependent variable, the best approach to use is: [missing information].

Metrics data, also known as quantitative, interval, or ratio data, are measurements used to describe individuals or things based on the extent to which a characteristic can define the subject, not just the presence of an attribute. Examples of metrics data include age and weight [23].

Non-metric data, on the other hand, encompasses all structured data that market researchers use that is not metrics data. This includes ordinal data, which is information that is rated, and nominal or categorical data, which is information without a clear linear trend [24].

When using interdependence approaches, variables cannot be classified as dependent or independent. Instead, these approaches aim to identify correlations between variables without assuming any particular distribution for the variables [25]. There are several ways to classify interdependence approaches, including:

A. Factor analysis, which combines multiple variables into a small number of components to make data interpretation easier [26].
B. Cluster analysis, which groups items such as respondents, goods, businesses, or variables based on their similarity and distinctiveness within a cluster [26].
C. Multidimensional scaling (MDS), which generates a map that represents the relative locations of several objects based on a distance table. The map can be one, two, three, or even four dimensions, and the answer can be metricor non-metric. The proximity matrix refers to the distance table, which can be derived directly from trials or indirectly using a correlation matrix [25].
D. Correspondence analysis is a method that uses points on a map to represent the locations of the rows and columns of a table of non-negative data [25].

The primary advantage of multivariate analysis is that it considers more independent factors, leading to more accurate inferences and more reasonable results that reflect the actual situation. However, MVA requires complex calculations to achieve a decent result, which is its primary drawback.

4 Prediction techniques

Prediction techniques can reveal the unknown values of a target variable. These techniques are utilized to create plans by extending assumptions about future conditions and possibilities over time using specific methods. In essence, prediction involves forecasting future events while considering all potential influencing factors. Prediction techniques are extensively used in various fields such as marketing, finance, telecommunications, healthcare, and medical diagnosis to improve future decision-making. There are several types of prediction techniques available, but we will focus on the following two:

4.1 Prediction techniques of data mining

Prediction\forecasting is utilized to make plans using special techniques, and also explain the comparison among the main prediction techniques to determine the algorithm more suitable to solve the problem of that study (Table 2).

4.1.1 Random forest regression and classification (RFRC)

The RFC model, as described by Jieyu Li et al., exhibits high accuracy in classification, minimal sensitivity to flood samples, and a high level of stability. Compared to other models, it is better at dynamically identifying effective reservoirs. This model uses a decision-making algorithm that employs various algorithms to reach a final decision by using a forest of trees and the higher voting principle. The RF model uses two types of random selection, one when constructing a tree by selecting a random number of samples from a large set and the other when building each subtree by selecting a random number of features from a large set of features. Furthermore, this model is simple, easy to parallelize, and effective.

Table 2 Comparison among main prediction techniques related to data mining

Prediction techniques related to data mining	Advantage	Disadvantage
RFRC	■ -Excellent precision Unaffected by outliers and noise ■ More effective than boosting and bagging ■ simple and easy to paralyze	■ Computation time might range from hours to days, depending on the size of the data set
BTCR	■ Limit the number of times you've iterated ■ deal with a wide range of target variables—- be able to work with a diverse set of target variables outlier removal	■ Poor prediction due to an incorrect tree size limit and a small sample size
CART	■ Automatic balancing of classes. Missing Values are handled automatically ■ Handle outliers with ease—with justifications ■ Multiple splits are possible ■ Deals with market segmentation	■ Only one variable is split. one at a time
CHASE	■ Multiple splits are possible ■ Deals with market segmentation	■ Transformation of variables ■ Merge steps must be completed within a certain amount of time
E- CHASE	■ Multi-panhandlemarket segmentation is possible	■ Transformation of variables ■ There are two user-specific parameters ■ Merge steps must be completed within a certain amount of time
MARS	■ There are no parameters that can be set by the user ■ It's more flexible because there are no variable transformations ■ The variable selection process is fully automated	■ Discontinuity in sub-region boundaries has an impact on precision. need backward fix
XGBoost	■ Learning on a Schedule ■ Gradient Tree Boosting, Column Subsampling, and Shrinkage (prevent overfitting.) ■ a distributed gradient boosting library that is optimized for efficiency, flexibility, and portability	■ If the training data is large, we can take samples of it; if the data is small, we can use the entire training data to determine the gradient
ETC	■ a type of ensemble learning a type of ensemble learning of ensemble learning ■ combines the results of several de-correlated decision trees that have been collected in a forest	■ a lackluster performance ■ Over-fitting is a difficult problem to solve

4.1.2 Boosted tree classifiers and regression (BTCR)

BTCR is a forward-looking algorithm for developing binary tree sequences, similar to RFRC, but without relying on random sampling or variable selection. This finite loop CART algorithm is designed to improve time prediction accuracy (Iniyan et al. 2021). Each iteration tree is unique and predicts the residuals from the previous tree. The algorithm stops the tree from growing beyond a certain limit, and the tree's size is measured and compared to a predetermined value in each epoch to avoid over-fitting. BTCR outperforms RFRC and can handle a variety of samples without requiring data processing or outlier eradication. It can also deal with predictor interaction effects automatically (Iniyan et al. 2021).

However, the maximum tree size has an impact on BTCR prediction since it does not consider the interactions between variables. If the BTCR is too low, the tree size may be selected incorrectly, leading to poor prediction, especially when there are only a few samples available for generalization.

4.1.3 Classification and regression tree (CART)

CART is one of DT's methods for classifying data in a more comprehensible way. It uses an interesting variable (X) to classify a data problem and discover the value of the target variable (Y). The tree is constructed by recursively dividing the data from top to bottom, and each branch represents a test for one of the X variables, determining whether the child nodes should go right or left. The goal of CART is to identify predictors of outcome and subgroup. If there are no more questions about which direction to expand, the tree will terminate into a terminal node. CART makes splits dependent on only one variable at each level, resulting in a more precise split when dealing with a small number of variables. However, if the number of variables is too large, the precise split from a combination of variables may be lost, and the tree may have too many levels, requiring more computation time.

4.1.4 Chi-squared automatic interaction detection (CHAID)

One of the decision tree strategies involves methods with more than two children per parent node. CHAID is used to convert continuous predictors into categorical ones by allowing multi-splitting of each node, which increases the chances of each variable appearing. However, CHAID is only suitable for categorical variables with nominal or ordinal values. Continuous variables require more pre-processing time since they must be transformed into ordinal variables. Additionally, CHAID requires the user to make decisions about numerous parameters, according to Al-Janabi et al. Prediction data mining techniques are widely used in various fields to aid in making better decisions in the future [7].

4.1.5 Exhaustive chi-squared automatic interaction detection (E-CHAID)

Unlike CHAID, ECHAID allows the parent node to divide into more than two children, resulting in a general tree rather than a binary tree. ECHAID differs from CHAID primarily in that it combines a larger number of steps by merging comparable idea pairs until only one pair remains. This approach requires the user to determine fewer parameters, which is a benefit of the strategy over CHAID. ECHAID also has several advantages over CHAID, including the absence of an alpha-level merge or an alpha-level split-merge, resulting in more automated processes [4]. When selecting the best grouping factor, E-CHAID employs the group-to-the-end technique, which involves retaining the preprocessing results of the input variables and using them as decision tree branches.

4.1.6 Multivariate adaptive regression splines (MARS)

The MARS model uses a divide and conquer approach as one of its prediction strategies. It divides the input variables into multiple regions and then reduces the number of mathematical equations by determining bias equations and coefficients. Despite being slower to build than recursive partitioning models, MARS is capable of making quick predictions based on previously unknown data. However, it has a sub-regional boundary discontinuity on occasion, which can undermine its accuracy [1, 2]. Nevertheless, MARS has been found to produce the best results and can be used to make accurate predictions.

4.1.7 EXtreme gradient boosting (XGBoost)

The XGBoost model is a classifier that operates similarly to Gradient Boosting but adds the ability to assign weight to each sample, similar to Ad boost. This tree-based model has gained considerable attention recently and differs from gradient boosting by fitting multiple weak learners (decision trees) in parallel, leading to faster computation times. Additionally, XGBoost features scalability, allowing it to run on distributed systems and handle larger datasets. It employs a Log Loss function to reduce loss and increase accuracy. In Ref. [5] aimed to use machine learning methods to develop a strategy for accurately predicting levels.

4.1.8 Extra tree classifier (ETC)

The ETC model utilizes a meta-estimator that improves prediction accuracy by training numerous weak learners (randomized decision trees) on diverse dataset samples. It is a classification ensemble learning model, similar to RF, with the primary difference being the construction of the forest trees. ETC builds decision trees using the original training sample, while RF uses bootstrap samples from the original dataset to create decision trees. Each decision tree receives a random sample of k features from the feature collection at each test node, and it must select the best feature to split.

4.2 Prediction techniques of neurocomputing

There are multi techniques of Neurocomputing can be describe in this section while compare among them shown in Table 3:

4.2.1 Convolutional neural networks (CNN)

It have achieved significant success in image processing and data analysis. The filters in the convolutional layer extract the input data's properties, and the pooling layer extracts local properties, making CNN a powerful image recognition tool. CNN is one of the most popular deep learning approaches in medical imaging, as it can automatically learn domain-specific image features, which explains its success. CNN training often involves transferring knowledge from a previously trained network to a new task, which is preferred in medical imaging due to its speed and ease of implementation without requiring a large annotated dataset for training. Alex Nets, GoogLeNet, and ResNet50 are the most popular CNNs for object and object detection classification of images.

4.2.2 Recurrent neural networks (RNN)

It are a flexible tool for modeling sequences, including those with contextual and specialized information. As the time steps increase, the unit is affected by a larger neighborhood. RNN-based techniques include connectionist temporal classification (CTC), Gated recurrent unit (GRU), and long short-term memory (LSTM).

4.2.3 Multitask neural networks

Can perform multiple predictive outputs using a single neural network, outperforming competitors in terms of performance, speed, and time spent on a single task. While these networks are difficult to train, convergence of strong common features that benefit all is a simplified concept of deep learning with the benefit of parallelism. It helps in resource and parameter sharing across tasks, reducing training time for training twomodels separately.

Table 3 Comparison among main prediction techniques related to neurocomputing

Prediction techniques related to neurocomputing	Advantage	Disadvantage
CNN	▪ Speed ▪ Perfect when orthology matters ▪ Excellent for brief texts (e.g., headlines)	▪ difficult with long texts ▪ does not have some of these larger transformer models' advantages
RNN	▪ Classification of recurrent neural networks as "feedback" networks. It can contain signals that travel forward and backward, may contain different "loops" in the grid as numbers or values are fed back into the grid	▪ The reason why it is difficult to train them is because they are advanced neural networks
Multi NN	▪ Data gathers the massive amount of the data ▪ Algorithms: popular due to the advancement made in the algorithms itself ▪ Ability to work with incomplete information ▪ Fault tolerance ▪ Dynamic: it is advantageous to model nonlinear data with several inputs ▪ processing in paralle ▪ Computational power	▪ Black box ▪ Amount of data: ▪ Computationally expensive ▪ Determination of proper network structure ▪ The duration of the network is unknown
TCM	▪ Learning temporal sequences In the bias terminology, a convolution function is utilized to produce a trainable envelope of time sensitivity ▪ Provides a flexiblframework for building gradient-based temporal generative model training algorithms	▪ The requirement to boost the learning rate derives from the fact that the learning rules are dependent on prediction errors, and as these errors are decreased through training, it becomes helpful to amplify the error signal ▪ Phoneme identification, real-time visual processing, and signal processing are all key issues
Stacked Autoencoders	▪ Boost performance ▪ High accuracy ▪ Dealing with both the conventional supervised classification and the pretrained ▪ Backpropagation can be used to adjust neural networks, (Vařeka et.al.,2017)	▪ further training
Generative deep learning	▪ Educating Machines in the Arts of Painting, Writing, Composing, and Playing ▪ Recent breakthroughs have taught a computer to excel at human tasks like sketching, making music, and completing tasks by generating an awareness of how its actions affect its surroundings	▪ Learn how to create generative models that can learn to paint, write, and complex in a reinforcement learning environment to complete tasks
Recursive Deep Learning	▪ Jointly learn classifiers and continuous feature representations for variable-sized ▪ Accurate, large scale natural language processing inputs ▪ In order to improve a specific prediction on the training data, the cost function is optimized	▪ Additional analysis

Table 3 (continued)

Prediction techniques related to neurocomputing	Advantage	Disadvantage
Long Short-Term Memory Networks	▪ A particular variety of recurrent neural network (RNN),the best results have been achieved, Cotfas et,al.,(2021) ▪ Because time series data can contain unknown lags, LSTM networks are well suited to classifying, evaluating, and making predictions based on this data ▪ LSTMs were developed to solve the exploding and disappearing gradient issues that might arise when training conventional RNNs.	▪ LSTMs are easy to overfit need more memory ▪ Take longer to train ▪ Difficult to perform dropout in ▪ While working with enormous volumes of data, use a lot of parameters.
Gated Recurrent Unit(GRU)	▪ GRU is similar to LSTM, but it has fewer gates ▪ It relies solely on a hidden state for memory transfer between recurrent units ▪ There is no separate cell state ▪ GRU runs its trains a little bit quicker than the LSTM(J Noh et.al.,2020) ▪ Overcome the vanishing gradient and inflating gradient issues that have impacted back-propagation through time in early attempts at RNNs(MS Al-Rakhamii et.al.,2021)	▪ GRU updates weights using the stochastic gradient descent optimization approach runs the danger of slipping into local optimization(B) ▪ The neural network cannot be fully trained if there are not enough training samples, which will have an impact on the model's ability to make accurate predictions(B)

4.2.4 Temporal convolutional machines (TCMs)

TCMs are a type of temporal sequence learning architecture that uses convolutions for statistical modeling. They are particularly useful for representing noisy temporal sequences and making decisions based on them.

4.2.5 Stacked autoencoders

Autoencoder are a type of neural network that uses backpropagation for unsupervised learning. They are made up of a series of sparse autoencoders, with one layer's outputs feeding into the next. The number of nodes decreases as you move from input to output, and sparsity measures the number of active neurons for a particular activation function. Evolutionary algorithms are a popular method for optimizing hyperparameters in deep learning algorithms, and they have been found to be more effective than traditional optimization algorithms in some cases.

4.2.6 Generative deep learning networks

It can learn patterns and create new content, including essays, paintings, images, and other media. They are being used in computational biology to model molecular structure and motion, which is critical for understanding biological events in living cells.

4.2.7 Recursive deep learning (RNN)

Is an extension of RNNs that recursively applies the same set of weights for neural nodes over a structured input. Instead of batch-processing all inputs, it provides the most accurate estimate available at any given moment. This approach is efficient and reduces latency.

4.2.8 Long short-term memory networks (LSTM)

LSTM are designed to avoid relying on outdated data. They have a repeating structure like conventional RNNs, but each element has four levels. The layers of an LSTM determine which old data should be passed on to the following layer, and they have the ability to forget information. There are different types of LSTMs, but they all have the potential to forget things, which is essential for modeling long-term dependencies in data.

4.2.9 Gated recurrent units (GRU)

GRU are a modified version of LSTMs, but with less complexity. GRUs use the hidden state to transfer information instead of the cell state, and they only have two gates: a reset gate and an update gate. The update gate selects information that needs to be added and/or ignored, while the reset gate decides how much past information to forget. GRUs are smaller in operation than LSTM networks, which makes them faster. Both LSTMs and GRUs are successful because of their gating mechanism, which preserves contextual information and long-term sequences while avoiding gradient issues. The gating functions allow the network to modulate how much the gradient vanishes, taking different values at each time step, which are learned functions of the current input and hidden state.

5 Gate recurrent unit (GRU)

Gated Recurrent Units (GRUs) offer a solution to the problem of storing "memory" from past time steps to inform future predictions. While GRUs are commonly used in machine translation, they are also a simpler version of Long Short-Term Memory (LSTM) units, which were introduced in [14]. The GRU is similar to an LSTM with a forget gate, but it has fewer parameters and lacks an output gate. In some cases, GRUs have been found to outperform LSTMs on smaller and less frequent datasets [6].

GRUs have an internal mechanism known as gates, which include update and reset gates. These gates control the information to be retained or discarded at each time step and regulate the flow of information. GRUs only have one hidden state that is passed between time steps, which can hold both long-term and short-term conditions simultaneously due to the gating mechanisms and algorithms that the state and information pass through [6]. The architecture of GRU

is illustrated in Fig. 1, where the update gate is calculated in the first step by combining the previous hidden state and the current input data using the following equation:

$$Zt = \sigma(Wz.Kt + Uz.ht - 1 + bz) \tag{1}$$

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:

$$Rt = \sigma(Wr.Kt + Ur.ht - 1 + br) \tag{2}$$

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

$$h \sim= tanh(Wh.Kt + r.Uh.h(t - 1) + br) \tag{3}$$

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

$$h = z.ht - 1 + (1 - z).h \sim \tag{4}$$

$$y = h \tag{5}$$

Primary parameters: Z: update gate at t. R: reset gate at t. h˜: current memory.

While; secondary parameters: K: input vector. h,y: output vector T: timestep. Ht-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 $\sigma(x) \in \{0,1\}$, tanh $\in \{-1,1\}$.
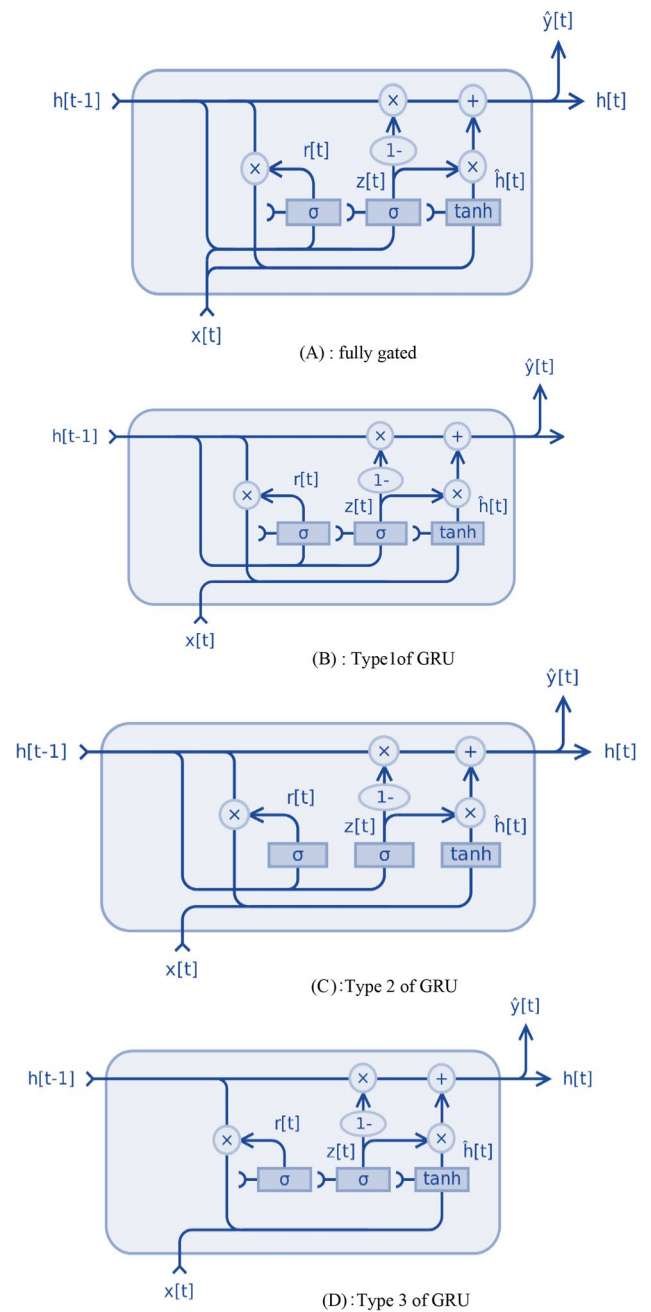
6 Intelligent neurocomputing model

The Intelligent Neurocomputing Model for Predicting Oil Price is a proposed model that comprises three stages. The first step, known as preprocessing, involves identifying outliers, missing values, and normalizing the data to prepare it for the next stage. In the second stage, the importance of each feature is determined by computing the correlation, entropy, and information gain. The dataset is then split into two parts: training and testing. The first part of the dataset is used to build the predictor, while the second part is used to evaluate it using three error measures: R2, MSE, and MAE, as shown in Fig. 2 and Algorithm #1.

In the preprocessing stage, the data is carefully examined to ensure that it is free from outliers and missing values. Any data points that do not fit within the expected range are marked as outliers and removed from the dataset. Missing values are also identified and replaced with appropriate values. After this, the data is normalized to ensure that all features have the same scale and range, which is necessary for accurate predictions.

In the second stage, the importance of each feature is determined by computing the correlation, entropy, and information gain. The correlation measures the strength and direction of the linear relationship between two variables. The entropy calculates the uncertainty of a variable, and information gain measures the reduction in entropy when a variable is used to split the data. These measures help to identify the most important features to include in the prediction model.

The dataset is then split into two parts: training and testing. The training set is used to build the predictor, which is a mathematical model that can predict future oil prices based on historical data. The testing set is used to evaluate the performance of the predictor using three error measures: R2, MSE, and MAE. R2, also known as the coefficient of determination, measures how well the predictor fits the data. MSE, or mean squared error, measures the average squared difference between the predicted and actual values. MAE, or mean absolute error, measures the average absolute difference between the predicted and actual values.

Fig. 1 Gru architecture [5]

(A) : fully gated

(B) : Type1 of GRU

(C) : Type 2 of GRU

(D) : Type 3 of GRU

The performance of the predictor is evaluated on the testing set using these three error measures, as shown in Algorithm #1. The results of the evaluation can then be used to fine-tune the predictor and improve its accuracy. The Intelligent Neurocomputing Model for Predicting Oil Price is a sophisticated model that can produce accurate predictions of future oil prices, and its three-stage approach ensures that the data is carefully prepared, the most important features are identified, and the predictor is thoroughly evaluated.

Discover

Algorithm 1 Intelligent Neuro Computing Model For Prediction Oil Price

Input: *Dataset of oil*
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
28:   **End** for
29:   **End** for
29: **Call Evaluation Measure**
End Algorithm

Benefit   **Name of Layer**

- Check Missing Values
- Handle Outliers   MVA
- Transformation
- Find Important Features

- Best Features
- Developing GRU   Normal Distribution (ND)
- Compute Distance

- Filter Distinct Values
- Predict Oil Prices   Advanced GRU
- Indicate of future

- Find accuracy of model
- Reduce Computation   Evaluations
- Enhance the Performance
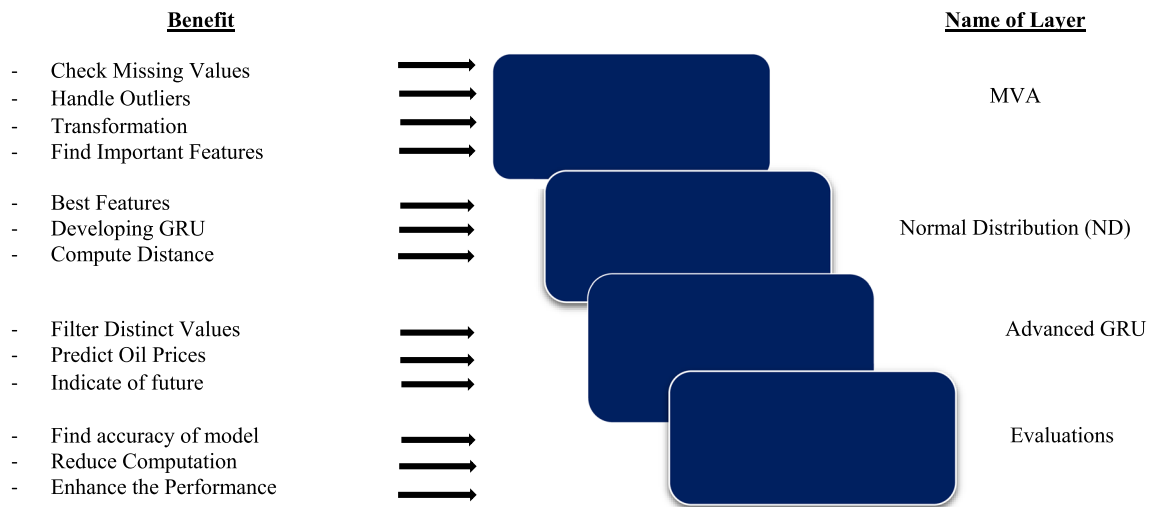
Fig.2 Layers of Proposal Model

7 Main stages of intelligent neurocomputing model

This section will show the main stages of the intelligent neuro computing model and explain the specific details for each stage.

7.1 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 #2.

Algorithm 2 Pre-Processing

Input: Oil price dataset
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 <> 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**

7.2 Building predictor stage

Once the pre-processing stage of the data is complete, the feature selection stage begins. This involves applying three types of multivariate analysis, namely correlation, entropy, and information gain, to determine which features have the greatest impact on oil prices. The results of this analysis are then used to select the most important features to be included in the deep neurocomputing network called GRU. The main parameters of the GRU network include the size of the windows, the number of hidden layers, the number of nodes in each layer, the activation function, and the number of epochs.

The GRU network is able to maintain the state of the cell and the values of the weights, making it easier to capture dependencies without ignoring previous information from large sequential databases. This is achieved through the use of two gates, namely the update gate and the reset gate, which allow for the retention or ignoring of important information from previous steps. The GRU network is superior to other networks in terms of its relatively simple modification due to the presence of memory modules. This makes the process of training and learning faster compared to other networks.

The implementation of the model is shown in Algorithm #3, which outlines the steps involved in building and training the GRU network. The first step involves selecting the most important features based on the results of the feature selection stage. The data is then split into training and testing sets, with the training set used to train the network and the testing set used to evaluate its performance. The network is initialized with random weights and biases, and the training process involves iteratively adjusting these values to minimize the error between the predicted and actual oil prices.

During the training process, the network is exposed to the training dataset multiple times, with each iteration known as an epoch. The number of epochs is a parameter that needs to be tuned to achieve optimal performance. The network is trained using backpropagation, which involves calculating the gradient of the error with respect to the weights and biases and using this to update their values.

The GRU network is a powerful tool for predicting oil prices, as it is able to capture dependencies and retain important information from previous time steps. Its relatively simple modification and fast training process make it an efficient choice for large-scale datasets. The implementation of the model shown in Algorithm #3 provides a clear and concise roadmap for building and training a GRU network for oil price prediction.

Algorithm 3 GRU

Input: X : Dataset of oil price

Output: Prediction values of oil price

// The forward components

  1: **For** each time(t) in dataset of oil price

//Compute: $h\tilde{}$, M, N ,h

  2: $h\sim = tanh\ (Wh . Kt\ + r .Uh . ht\text{-}1 + bh\ )$ // Memory Cell

  3: $Mt = \sigma\ (Wm\ . Kt\ + Um . ht\text{-}1 + bm\ )$ // Update gate

  4: $Nt = \sigma\ (Wn\ . Kt\ + Un . h(t\text{-}1) + bn\ )$ // Reset gate

  5: $h= m^*\ ht\text{-}1 + (1\text{-}m)\ ^* h\tilde{}$ // Output Cell

  6: $Y=h$

  7: **End for**

// The backward components.

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

     // Update Output Cell , loss function

  9: $Yt = \sigma\ (Uh . ht\ )$ //Update Output Cell

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

  11: **End for**

// The Final updates to the internal parameters is compute

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

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

  14: $\delta Uh\tilde{} = \sum_{t=0}^{T} \delta\ Uh\tilde{}t \times Xt$ //Update the weight of intermediate memory

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

End GRU

7.3 Evaluation stage

Once the model design phase is complete and the model has been implemented on a dataset to predict oil prices, the next step is to determine the efficiency of the proposed model. This is done by using the final predictor result from the previous stage to estimate the oil price of the testing dataset. The accuracy of the model is then evaluated using three measures: R2, MSE, and MAE, as shown in Algorithm #4.

By computing these three measures, the efficiency of the proposed model can be determined. If the R2 value is high and the MSE and MAE values are low, then the model is considered to be efficient in predicting oil prices. However, if the R2 value is low and the MSE and MAE values are high, then the model may need to be adjusted orimproved to achieve better results.

It is important to note that the efficiency of the model should be evaluated on a testing dataset that is separate from the training dataset. This ensures that the model is not overfitting to the training data and can generalize well to new data.

In Algorithm #4, the final predictor result is used to estimate the oil price of the testing dataset, and the R2, MSE, and MAE measures are computed. These measures provide a quantitative evaluation of the model's performance and can be used to compare the efficiency of different models.

Algorithm 4 Evaluation

> **Input:** *Testing dataset*
>
> **Output:** *Average error rate based on R2 ,MSE,MAE*
>
> **1:**   **For** *each record in testing dataset*
>
> **2:**   **For** *each column in testing dataset*   // *m= number of feature*
>
> **3:** $R2 = 1 - \frac{\sum_{i=1}^{m}(Ki - Yi)^2}{\sum_{i=1}^{m}(\hat{Y} - Yi)^2}$ // *k= predict value , y= actual value*
>
> **4:** $MSE = \frac{1}{m}\sum_{i=1}^{m}(Ki - Yi)^2$
>
> **5:** $MAE = \frac{1}{m}\sum_{i=1}^{m}|Ki - Yi|^2$
>
> **6:**   **End for**
>
> **7:**   **End for**
>
> **End Evaluation**

8 Results of intelligent neurocomputing model

The dataset used in this paper consists of 7 base features and 4947 numeric records collected over the past 10 years, with the date feature being of string type and excluded from the analysis. The dataset is used to train the model and contains the following features:

- Date: The trading date.
- Crude Oil WTI Futures Historical (WTI): A contract that determines the price of oil at a particular future date.
- Gold Futures Historical Data (GOLD): One of the fundamental characteristics that significantly impact 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. The name of the index is a combination of Standard Company and Pros Publishing Company. It includes 500 companies, and its value is derived from the market value of the companies under it.
- US Dollar Index Futures Historical Data (US DOLLAR INDEX): Refers to the exchange rate of the US dollar with respect to other currencies. The most well-known index, the USDX, is produced by the New York Chamber of Commerce and shows how the dollar compares to other currencies.
- Historical Bond Yield Information for the United States (US 10YR BOND): One of the features used to anticipate oil prices.
- The Dow Jones Utility Average (^DJU): A stock exchange index in the United States that is the world's oldest index, tracking the performanceof top publicly traded companies. It is also one of the key features used to anticipate oil prices.

The dataset used in this study is comprehensive, containing crucial features that significantly impact the prediction of oil prices. The inclusion of features such as WTI, GOLD, SP 500, US DOLLAR INDEX, US 10YR BOND, and ^DJU ensures

that the model is built on a solid foundation of relevant and informative data. The availability of a large dataset with thousands of records spanning a decade provides ample data for the training and evaluation of the model.

The date feature, although excluded from the analysis, remains a critical aspect of the dataset as it provides a temporal dimension to the data. It enables the model to take into account the changes and fluctuations in oil prices over time and make predictions based on historical trends and patterns.

The use of a diverse range of features from different sectors, such as the stock market, commodity markets, and foreign exchange markets, ensures that the model is not biased towards a single sector and can capture the complex and dynamic relationships between different factors that impact oil prices.

8.1 Results of checking outlier

During the data examination process, outliers are identified which are data points that are unusual or different from the rest of the data in the specific oil price dataset. The presence of external data is also checked to ensure that the dataset only contains relevant and valid information. If an outlier is identified, it undergoes a processing step where it is tested to see if it falls within the specified value range. If it does, it is rounded to the nearest possible value within that range. Otherwise, it is removed from the dataset.

In this particular dataset, no external data was found, and the data was regular. Therefore, there was no need to remove any outliers or perform any further processing steps. The absence of external data is an advantage as it ensures that the dataset contains only relevant and reliable information that can be used to build an accurate model for predicting oil prices.

Identifying and removing outliers is a crucial step in the data preparation process as it can significantly impact the accuracy and reliability of the model. Outliers can skew the results and introduce errors, making it difficult to predict oil prices accurately. Therefore, it is essential to carefully examine the dataset to identify and handle outliers appropriately.

8.2 Results of checking missing values

The data set used in this study contains a total of 4947 rows, out of which 152 rows contain missing data. After removing these missing values, the total number of samples in the data set reduces to 4849. The outcome of this data cleaning process is presented in Table 4.

The removal of missing data is a critical step in the data preparation process as it ensures that the dataset contains only complete and accurate information. Missing data can introduce errors and bias into the analysis, leading to inaccurate predictions of oil prices. Therefore, it is crucial to handle missing data appropriately to ensure the reliability and validity of the dataset.

The total number of rows in the dataset provides an indication of the size of the dataset and the amount of information available for analysis. The reduction in the number of rows after removing missing data highlights the impact of missing data on the dataset's size and completeness.

Table 3 presents a clear and concise summary of the data cleaning process, providing a visual representation of the number of rows with missing data and the resulting number of samples after removing these values. This information is essential for understanding the dataset's quality and completeness and for making informed decisions regarding the use of the data for building and evaluating the model.

8.3 Normal distribution

The next step involves calculating the normal distribution to understand the relationship between the values of each feature and the oil price over the past 10 years. This is depicted in Fig. 3, which illustrates the distribution of the different features and their correlation with oil prices.

The normal distribution is a statistical technique used to model the distribution of a set of data. It helps to understand the central tendency, dispersion, and shape of the data, as well as the relationship between different variables. By calculating the normal distribution for each feature, it is possible to gain insights into how each feature is related to oil prices and evaluate their impact on the prediction of oil prices.

Figure 4 provides a visual representation of the normal distribution of each feature and its relationship with oil prices. By examining the distribution of each feature, it is possible to identify trends and patterns that can help to build a more accurate model for predicting oil prices.

Table 4 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
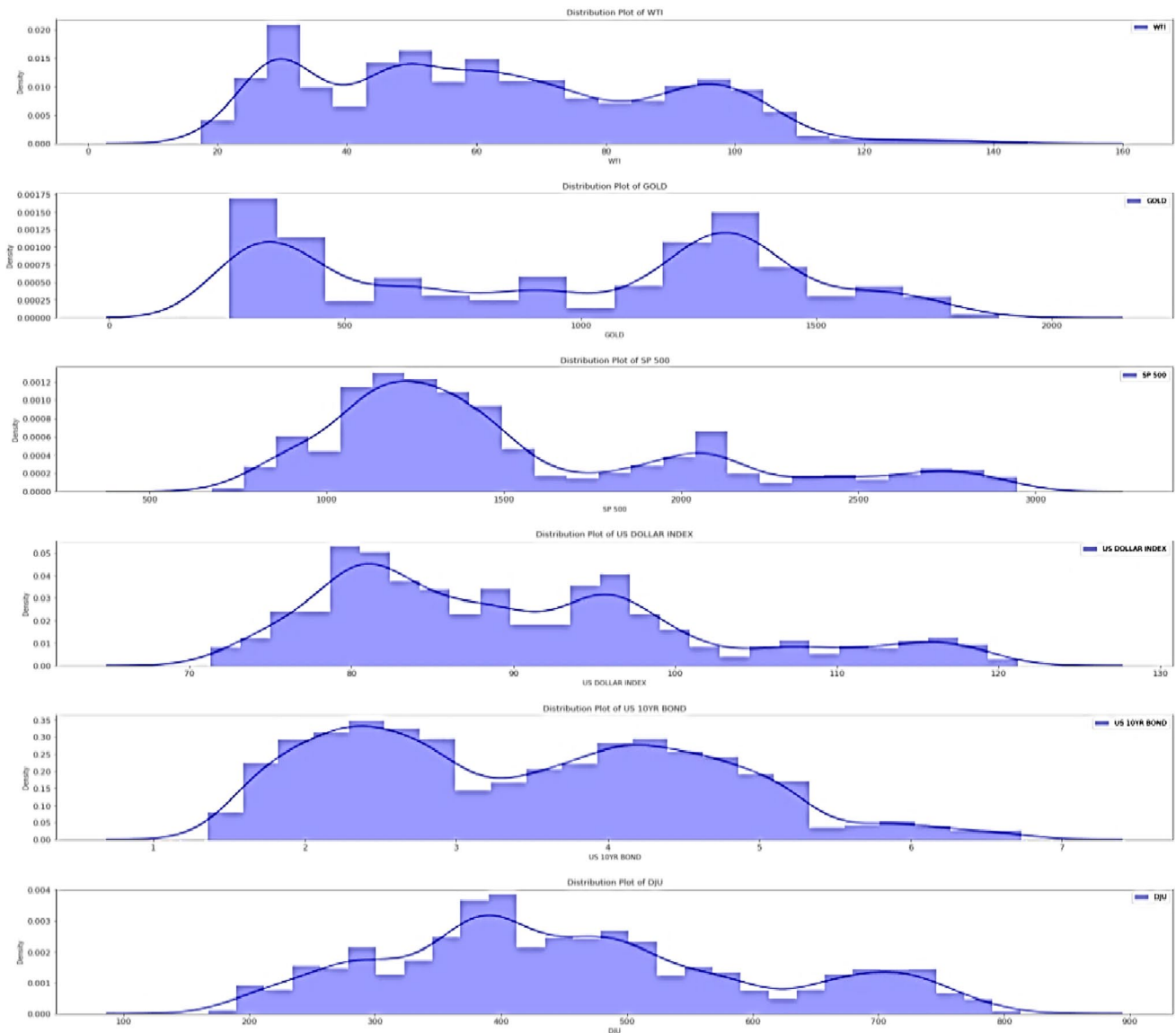
Fig. 3 Analysis of Oil Price Distribution with Multiple Features: Historical Data of Crude Oil WTI Futures, Gold Futures, S&P 500 Futures (INVESTOR), US Dollar Index Futures, United States 10-Year Bond Yield, and ^DJU

Understanding the relationship between different features and oil prices is crucial for building an accurate model that can capture the complex and dynamic relationships between different factors that impact oil prices. By using the normal distribution technique, it is possible to gain insights into these relationships and develop a more robust and reliable model.

8.4 Results of multivariate analysis

Once the pre-processing is complete, the next step is to perform multivariate analysis of the features to determine their importance and degree of effect on the target variable, which is the oil price. This is done by computing the Pearson correlation, which measures the degree of correlation between two variables and ranges between -1 and $+1$. The direction and strength of the relationship can be inferred from the correlation coefficient. There are four types of relationships that can be concluded from the correlation coefficient: positive, negative, zero, and perfect.
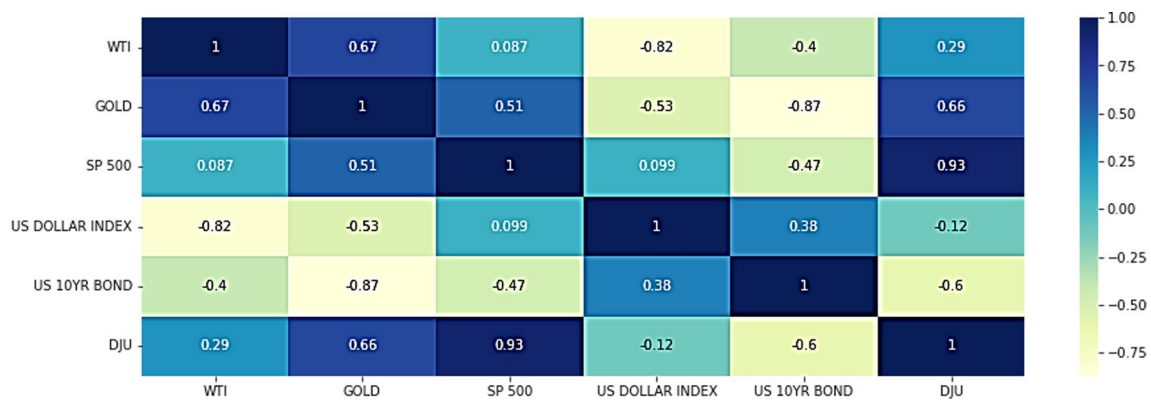
Fig. 4 Correlation among all the features

The results of this analysis are presented in Table 5 and Fig. 4, while the results of the entropy and information gain analysis are shown in Table 6. It is important to note that the correlation of each feature with itself is always equal to 1, as this represents a perfect correlation.

A positive correlation indicates that both features move in the same direction, while a negative correlation indicates that they move in opposite directions. A correlation coefficient of + 1 or − 1 represents a perfect correlation, while a coefficient of 0 indicates no correlation between the features (Table 6).

The multivariate analysis of features is a critical step in the data analysis process as it helps to identify the most important features that significantly impact the prediction of oil prices. By understanding the relationship between different features and oil prices, it is possible to build a more accurate and reliable model.

8.5 Results of prediction (GRU)

In this stage, the preprocessed data is fed into a model designed to predict oil prices using the GRU algorithm. The characteristics data resulting from the preprocessing are used to train the model. The results of the model shows in Fig. 5 the error rate calculated based on three main measures, namely R2, MSE, and MAE. The total implementation time for the model was 43.796, with a number of hidden layers of 3 and a tansh activation function. The window size used was 10, and the total number of epochs was set to 50.

Figure 5 provides a comparison between the original and predicted oil prices based on the GRU model. This comparison helps to evaluate the accuracy and reliability of the model and identify any trends or patterns that may have been missed during the data analysis process.

The use of the GRU algorithm is a state-of-the-art technique for predicting time series data, such as oil prices. By using this algorithm, it is possible to capture the complex and dynamic relationships between different variables and build a more accurate and reliable model.

The results presented in Fig. 6 provide insights into the performance of the model and its ability to predict oil prices accurately. The low error rates indicate that the model is robust and can capture the complex relationships between different factors that impact oil prices.

8.6 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 paper 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 7. on other hand the result of evaluation measure to testing dataset shown in Table 8.

Table 5 Results of person correlation

Features	WTI	GOLD	SP500	USDOLAR	US10YR	DJU
WTI	1.00	0.67	0.08	− 0.82	− 0.40	0.29
Gold	0.67	1.00	0.51	− 0.53	− 0.87	0.67
SP500	**0.09**	0.51	1.00	0.09	− 0.47	0.93
US dolar	− 0.822	− 0.53	0.09	1.00	0.38	− 0.12
US10yr	− 0.40	− 0.87	− 0.47	0.38	1.00	− 0.60
DJU	0.29	0.66	0.92	− 0.12	− 0.60	1.00

Table 6 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

Fig. 5 Compare between original and predict Oil Price Based on GRU
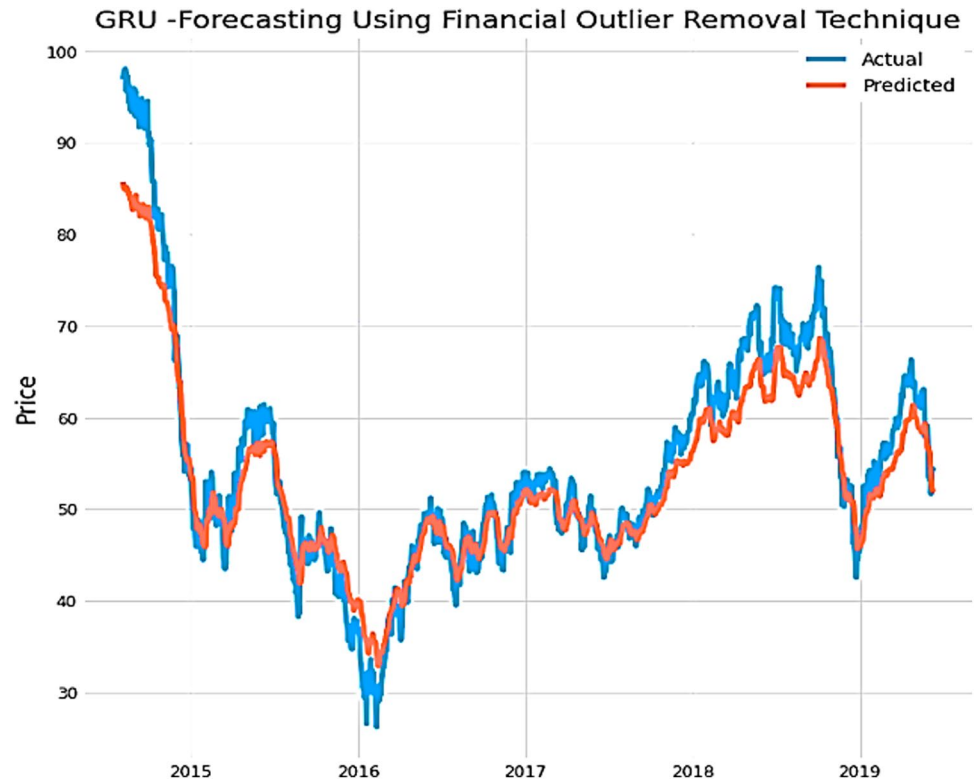
9 Conclustion

The prediction of oil prices requires the use of highly efficient techniques. The challenge of this study was to overcome the limitations of existing techniques by adopting an effective oil price prediction technique that is based on multiple features and captures data from several years prior. Figure 6 illustrates the data used in this study, which contains missing values in multiple features. The goal of this study was to build a predictor based on facts, which means that the results of the prediction will be accurate if the predictor is built on factual data. Therefore, all rows that contained missing values were dropped, resulting in a total of 4849 rows from the original dataset of 4947.

The predictor characteristic of this study was designed to be memory-efficient, fast, and requiring less training time. The GRU network was chosen as one of the important neurocomputing techniques due to its high speed and accuracy in extracting results. However, it requires the specification of a large number of parameters. This study addressed this problem by using multivariate analysis.

To achieve accurate predictions of oil prices, it is necessary to analyze data related to several years. Therefore, this study analyzed data from 2010 to 2019 to build an accurate predictor of oil prices. However, building such models requires extensive knowledge in mathematics and programming.
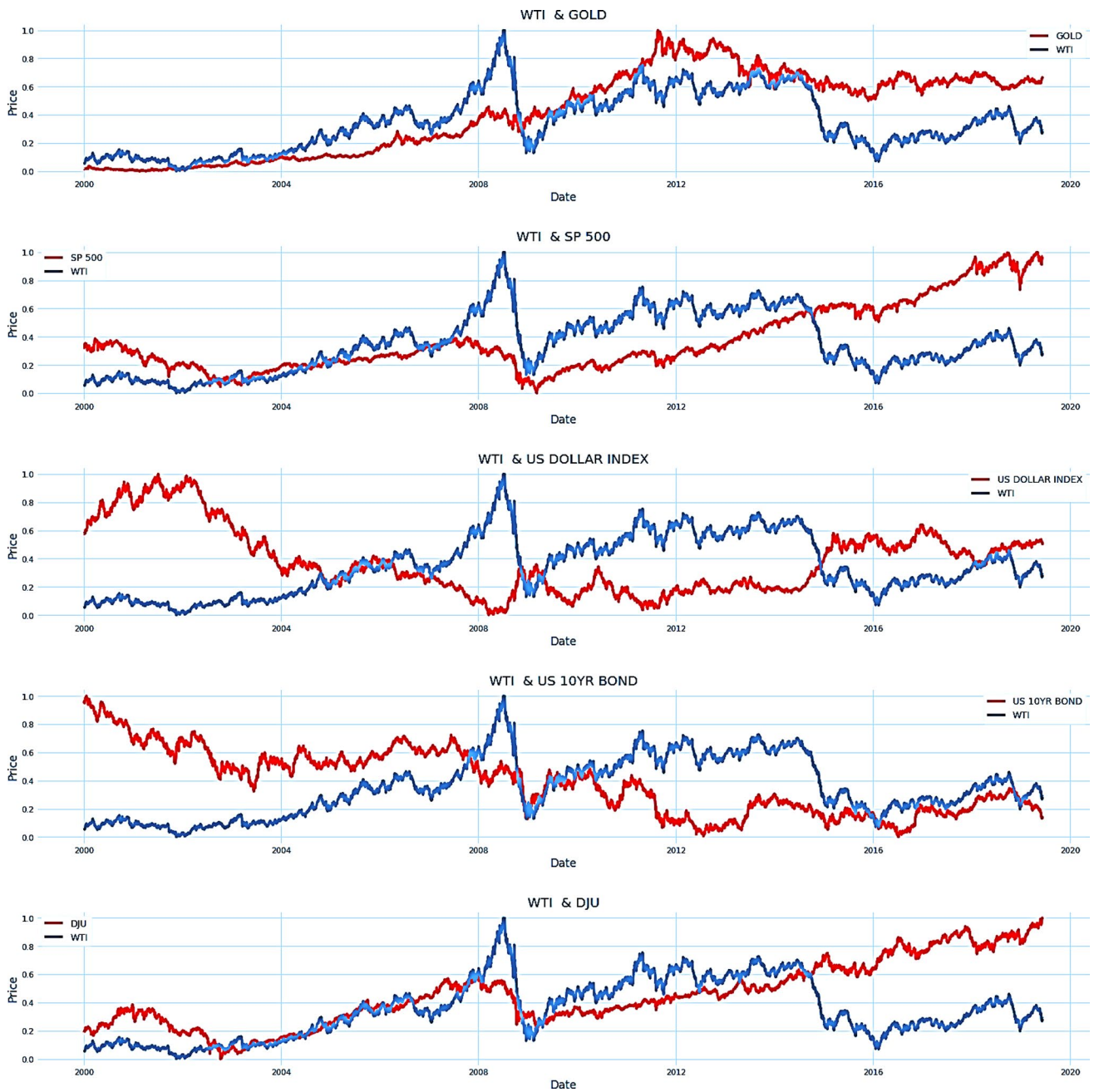
Fig.6 Compare between each both feathers

In future works, it is recommended to find an optimal method for determining the original parameters of the GRU network to reduce its execution time, such as using Genetic Algorithm (GA), Practical Swarm Optimization (PSO), or Whale Optimization Algorithm (WOA) algorithms. Additionally, this study can be implemented using other prediction techniques related to Data Mining, such as the MARS data mining algorithm based on mathematical principles. The accuracy of the model can also be tested using other types of evaluation measures, such as Confusion Matrices that include five measures, namely Accuracy, Precision, Recall, F, and Fb. Table 9 presents a comparison between this study and previous works.

Table 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 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

Table 9 Compare this study with the previous works

Name	Preprocessing	Methodology	Evaluation measures	Advantages/disadvantage
This Study	▪ Checking outlier ▪ Checking missing value ▪ Normalization ▪ Find most important featuers (Correlation, entropy and Information Gain	Intelligent Neurocomputing Model is hybrid model based on cooperative between Multivariant analysis and GRU	▪ MSE ▪ MAE ▪ R2	▪ Reduce the implantation time ▪ Find Most Important Feature ▪ Reduce computation ▪ Give the results very near of real
Xu et al. [14]	Autoregression model (AR-model)	New-Keynesian model	Loss	
Abbass et al. [15]	Normalization and remove duplication	Autoregressive distributed lag (ARDL) Models QARDL	Loss	▪ Give speed results because it based on distributed principle ▪ Select some of random numbers ▪ Loss rate of original data
Sun te al. [16]	MS-VAR model	Markov-switching vector autoregressive model	Accuracy	▪ Need high experience of mathematical principles ▪ High computation ▪ High accuracy
He et al. [17]	▪ Handle the noise ▪ Check missing vales	Variational mode decomposition (VMD), (ML) Algorithm	Accuracy	▪ High computation ▪ High accuracy
Gupta et al. [18]	Artificial neural network (ANN) model	Oil Price Prediction Model	MSE	▪ Select parameters of ANN based on try and error principle ▪ High implementation time
Carpio [19]	long-term and short-term Model	(VECX) Model	Loss	▪ High accuracy ▪ High implementation time ▪ Need high experience of mathematical principles

Declarations

Ethics approval and consent to participate This article does not contain any studies with human participants or animals performed by any of the author.

Competing interests The authors declare that they have no competing interests.

References

1. Al-Janabi S. Overcoming the main challenges of knowledge discovery through tendency to the intelligent data analysis. Int Conf Data Anal Business Ind. 2021;2021:286–94. https://doi.org/10.1109/ICDABI53623.2021.9655916.
2. Al_Janabi, S., Salman, M.A, Mohammad, M. (2019). Multi-level Network Construction Based on Intelligent Big Data Analysis. In: Farhaoui, Y., Moussaid, L. (eds) Big Data and Smart Digital Environment. ICBDSDE 2018. Studies in Big Data, vol 53. Springer, Cham. https://doi.org/10.1007/978-3-030-12048-1_13
3. Al-Janabi S, Alkaim A, Al-Janabi E, et al. Intelligent forecaster of concentrations (PM2.5, PM10, NO2, CO, O3, SO2) caused air pollution (IFCsAP). Neural Comput Applic. 2021;33:14199–229. https://doi.org/10.1007/s00521-021-06067-7.
4. Al-Janabi S, Alkaim AF, Adel Z. An Innovative synthesis of deep learning techniques (DCapsNet & DCOM) for generation electrical renewable energy from wind energy. Soft Comput. 2020;24:10943–62. https://doi.org/10.1007/s00500-020-04905-9.
5. Alkaim AF, Al_Janabi S. 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 SystemsSpringer, Cham. 2020; 81. https://doi.org/10.1007/978-3-030-23672-4_10
6. Al-Janabi S, Alkaim AF. A nifty collaborative analysis to predicting a novel tool (DRFLLS) for missing values estimation. Springer Soft Comput. 2020;24(1):555–69. https://doi.org/10.1007/s00500-019-03972-x.
7. Al-Janabi S, Mahdi MA. Evaluation prediction techniques to achievement an optimal biomedical analysis. Int J Grid and Utility Comput. 2019;10(5):512–27.
8. Zabezhailo MI, Trunin Y. On the problem of medical diagnostic evidence: intelligent analysis of empirical data on patients in samples of limited size. Autom Doc Math Linguist. 2019;53(6):322–8. https://doi.org/10.3103/S0005105519060086.
9. Peres RS, Rocha AD, Leitao P, Barata J. IDARTS–Towards intelligent data analysis and real-time supervision for industry 4.0. Comput Ind. 2018;101:138–46. https://doi.org/10.1016/j.compind.2018.07.004.
10. Li X, Shang W, Wang S. Text-based crude oil price forecasting: a deep learning approach. Int J Forecasting. 2019;35(4):1548–60. https://doi.org/10.1016/j.ijforecast.2018.07.006.
11. Alkaim A.F, Al_Janabi S. 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. 2020; 81. https://doi.org/10.1007/978-3-030-23672-4_10.
12. Hair JF. 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. 2011.
13. Vighnesh D. Multivariate analysis: an overview. https://s4be.cochrane.org/blog/2021/09/09/multivariate-analysis-an-overview. 2021.
14. Xu Q, Fu B, Wang B. The effects of oil price uncertainty on China's economy. Energy Econ. 2022;107:105840. https://doi.org/10.1016/j.eneco.2022.105840.
15. Abbass K, Sharif A, Song H, Ali MT, Khan F, Amin N. Do geopolitical oil price risk, global macroeconomic fundamentals relate Islamic and conventional stock market? Empirical evidence from QARDL approach. Resour Policy. 2022;77:102730. https://doi.org/10.1016/j.resourpol.2022.102730.
16. Sun C, Zhan Y, Peng Y, Cai W. Crude oil price and exchange rate: Evidence from the period before and after the launch of China's crude oil futures. Energy Econ. 2022;105:105707. https://doi.org/10.1016/j.eneco.2021.105707.
17. He H, Sun M, Li X, Mensah IA. A novel crude oil price trend prediction method: machine learning classification algorithm based on multi-modal data features. Energy. 2022;244:122706. https://doi.org/10.1016/j.energy.2021.122706.
18. Gupta N, Nigam S. Crude oil price prediction using artificial neural network. Procedia Comput Sci. 2020;170:642–7. https://doi.org/10.1016/j.procs.2020.03.136.

19. Carpio LGT. The effects of oil price volatility on ethanol, gasoline, and sugar price forecasts. Energy. 2019;181:1012–22. https://doi.org/10.1016/j.energy.2019.05.067.
20. Grace G. What is the difference between univariate and multivariate data analysis. Ask Data Science. https://askdatascience.com/491/what-difference-between-univariate-multivariate-analysis. 2018.
21. Bhumika Dutta. What is Multivariate Data Analysis?. https://www.analyticssteps.com/blogs/what-multivariate-data-analysis . 2021.
22. Sateesh Gottumukkala. How to Select Multivariate Data Analysis Techniques ?-Part 1. https://www.linkedin.com/pulse/multivariate-data-analysis-techniques-part-1-sateesh-gottumukkala. 2018.
23. Hua-Kai Chiou. Multivariate Statistical Data Analysis with Its Applications. https://slideplayer.com/slide/6661200. 2005.
24. Ray Poynter. Some thoughts on Metric and Nonmetric Data in Market Research" , https://www.linkedin.com/pulse/some-thoughts-metric-nonmetric-data-market-research-ray-poynter. 2017.
25. Harsha Nimkar, Great Learning Team. Overview of Multivariate Analysis | What is Multivariate Analysis and Model Building Process?. https://www.mygreatlearning.com/blog/introduction-to-multivariate-analysis .2020.
26. Ganya Erickson. Other Multivariate Techniques. https://www.slideserve.com/ganya/other-multivariate-techniques . 2014.