Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A SMART FAMEWORK TO GUIDE THE CUSTOMERS TO RAISE THE RETURN ON INVESTMENT

YOUSR AHMED ALAA ELDEEN[1], SHERIF ADEL ABD EL ALEEM[2], AHMED IBRAHIM BAHGAT EL SEDDAWY[3]

[1] Department of Business Information Systems, Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt

[2] Lecturer of Business Administration, Faculty of Commerce, Helwan University, Cairo, Egypt

[3] Assoc. Prof. Computer Science and Information Systems Department, Arab Academy for Science, Technology and Maritime Transport-AASTMT, Smart Village, Giza, Egypt

E-mail: [1]yousr.ahmed21@commerce.helwan.edu.eg , [2]sherif.abdel@commerce.helwan.edu.eg, [3]Ahmed.Bahgat@aast.edu

ABSTRACT

This study assesses the performance of various machine learning models and traditional statistical tools in forecasting the prices of key commodities, including gold, silver, crude oil, Brent oil, natural gas, and copper. The models evaluated encompass Random Forest Regression, Gradient Boosting Regression, Support Vector Regression, XGBoost, Long Short-Term Memory (LSTM), and Artificial Neural Networks (ANN), in addition to traditional econometric tools like SPSS and EViews. Performance metrics are based on the Root Mean Square Error (RMSE) statistic. The findings reveal that LSTM outperforms other models in capturing intricate time series patterns, particularly for copper price predictions, demonstrated by significantly lower RMSE values. While classical statistical methods and other machine learning models achieve reasonable accuracy, LSTM and ANN consistently show superior performance. Furthermore, an integrated model combining SPSS, EViews, and LSTM projections identifies top investment prospects. Gold and silver are highlighted as solid, safe-haven assets with highly accurate forecasts. Natural gas is noted for its precise price predictions and potential for substantial price increases. Copper stands out due to its excellent predictive accuracy and its capability to provide early warnings of gold price changes. The integration of traditional statistical approaches with advanced machine learning models offers a comprehensive framework for forecasting commodity prices and pinpointing the most promising investment opportunities in the commodities market.

Keywords: *Forecasting Commodity Prices, Machine Learning, , Long Short – Term Memory (LSTM), Statistical Package for the Social Sciences (SPSS), Econometric Views* (*EViews*) .

1. INTRODUACTION

The research focuses on the critical issue of forecasting commodity prices such as gold, silver, crude oil, Brent oil, natural gas, and copper, which are essential not only for industrial processes but also for investment portfolios and economic stability. Consequently, economists, investors, and policymakers rely heavily on accurate commodity price forecasts. This study enhances the accuracy of commodity price forecasts by employing advanced machine learning techniques, including Random Forest Regression (RFR), Gradient Boosting Regression (GBR), Support Vector Regression (SVR), XGBoost, Long Short-Term Memory (LSTM) networks, and Artificial Neural Networks (ANN). These techniques are compared with traditional econometric models commonly used in software like SPSS and EViews. Despite the rise of new techniques, traditional statistical software like SPSS and econometric tools like EViews remain important[1]. Several factors influence commodity prices, including geopolitical events, supply and demand dynamics, macroeconomic variables, natural disasters, and technological advances[2].

For instance, political unrest and trade wars can cause sharp fluctuations in commodity prices, with conflicts in the Middle East often leading to spikes in oil prices. Supply and demand dynamics are also crucial, with OPEC's production decisions and increasing industrial demand for metals directly influencing prices[2]. Macroeconomic variables such as interest rates and currency exchange rates impact commodity markets; for example, a strong US dollar can make dollar-priced commodities more expensive for international customers, reducing demand[3]. Natural disasters like hurricanes and droughts can disrupt production and supply chains, increasing price volatility[4]. Technological advances in extraction and production techniques, such as the US shale oil boom, significantly alter the supply landscape, underscoring the interconnectedness of these factors in influencing commodity pricing. These prices dictate the cost of raw materials, alter inflation rates, and influence the profitability of various industries[5].

This study investigates multiple machines learning models, evaluating their performance in predicting commodity prices and providing useful insights to stakeholders[6]. The goal is to analyze and compare the performance of several machine learning algorithms, including RFR, GBR, SVR, XGBoost, LSTM, and ANN. The study involves collecting historical pricing data for specific commodities, training and testing each algorithm, assessing performance using RMSE, and selecting the best-performing model for each commodity. The objective is to determine which machine learning algorithm offers the most accurate price predictions and delivers valuable insights for investors, policymakers, and industry professionals. These hypotheses form the basis for examining the performance of advanced machine learning models compared to classic econometric methods in predicting commodity prices[7-8].

The study proposes several assumptions, including the hypothesis that LSTM models will outperform other machine learning models due to their ability to capture temporal correlations in data, and that incorporating external factors such as economic indicators and geopolitical events will enhance prediction accuracy. It also explores the variability in model performance across different commodities, such as gold, silver, crude oil, Brent oil, natural gas, and copper.

This study aims to expand knowledge in the field of commodity price forecasting and provide significant insights to stakeholders seeking better decision-making tools in financial markets.

1.1 problem statement

Despite significant advancements in commodity price forecasting, existing research exhibits several critical gaps that hinder the development of robust and generalizable predictive models. Current studies often focus on individual commodities or utilize limited model comparisons, restricting a comprehensive understanding of forecasting efficacy across different commodities. Moreover, many studies rely heavily on trial-and-error methods for model selection and use single data sources, limiting the generalizability and applicability of their findings. Additionally, the impact of geopolitical and macroeconomic factors on commodity prices is often inadequately addressed, particularly in the context of emerging markets.

1.2 research objectives

To evaluate the strengths and weaknesses of this research proposal in regard to objectives, must consider both the planned methodology and the desired outcomes.

- Strengths

This idea excels in several important aspects. First, the combination of SPSS, EViews, and LSTM models is a significant advantage. Using these many analytical tools, I am able to conduct a thorough investigation of commodity price forecasts, integrating traditional statistical methods with modern machine learning techniques. This comprehensive strategy is well suited for improving prediction accuracy and producing meaningful investment suggestions. Furthermore, the breadth of this research, which includes commodities such as gold, silver, natural gas, and copper, is useful..

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This proposal also shows a strong emphasis on finding and fixing gaps in existing research. By critically analyzing the literature and identifying these shortcomings, I am positioned our study to make innovative contributions to the field of commodity price forecasting. Furthermore, the practical ramifications of this research, notably the provision of investment suggestions, ensure that the findings are relevant and applicable in the real world.

- Weaknesses

However, there are potential issues that must be addressed. The integration of multiple analytical techniques, while advantageous, adds complexity. Managing consistency and coherence among SPSS, EViews, and LSTM models may provide issues, potentially compromising the interpretability and practical implementation of these results..

Another issue to consider is the possibility of overfitting, particularly with more advanced LSTMs. The intricacy of these models must be balanced against their ability to effectively generalize findings. Maintaining prediction accuracy will require ensuring that the models are not too tuned to certain datasets. Furthermore, the performance of LSTMs and other advanced models is heavily reliant on the quality and quantity of previous data. If the data used is insufficient or not representative, the accuracy of these forecasts may suffer. Thus, ensuring access to extensive and high-quality historical data is critical.

The resources required to integrate and manage various models might be enormous. This complexity necessitates significant computer power and effort, which may jeopardize the viability of finishing the research within the stated timeframe.

Finally, while filling specific gaps in existing research is critical, there is a risk that this study will become overly narrowly focused. It is critical that these insights contribute to broader theoretical or practical advances in commodity price forecasting.

1.3 Necessity of the Proposal

This Paper is critical because it intends to fill key gaps in existing research by providing a broad comparison of advanced forecasting models, employing varied data sources, combining a comprehensive set of influencing factors, and focusing on emerging economies. Its goal is to create more accurate, robust, and generalizable commodity price forecasting models, resulting in better informed decision-making and advancements in the industry.

While prior research has contributed useful insights into commodity price forecasting, three critical reasons underscore the importance of this paper:

1. Comprehensive Model Comparison: Existing studies frequently focus on a small number of forecasting models or analyze them in isolation. This proposal seeks to compare a wide range of advanced models—Long Short-Term Memory (LSTM), Artificial Neural Networks (ANN), Random Forest Regressor (RFR), Gradient Boosting Regressor (GBR), Support Vector Regression (SVR), and XGBoost—across multiple commodities. Such a thorough comparison will result in a better knowledge of model performance under different settings, improving the resilience and application of forecasting techniques.

2. Use of Diverse Data Sources: Much research relies on a small number of data sources, limiting the generalizability of their conclusions. This proposal intends to increase the accuracy and reliability of forecasting models by integrating a diverse set of data sources, assuring their application across various settings and market conditions.

3. Comprehensive Analysis of Influencing aspects: Previous research frequently investigates commodity price predictions in isolation, without considering broader aspects such as geopolitical events, macroeconomic conditions, and behavioral finance features. This proposal aims to give a comprehensive study by combining various aspects, resulting in a better understanding of the complex dynamics influencing commodity prices.

4. Emphasis on Emerging Markets: Current research usually overlooks the impact of emerging markets and geopolitical events on commodities prices. This proposal fills this gap by studying how these factors affect price trends, giving new insights necessary for appropriate forecasting in diverse economic contexts.

5. Advancement of Forecasting Models: While different forecasting strategies have been investigated in past studies, there is a need to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

improve these models by incorporating recent advances in machine learning and data analysis. This proposal seeks to increase forecasting accuracy by creating and testing updated models that use cutting-edge approaches.

6.Integration of Findings: Existing research frequently lacks integration and synthesis of findings from many models and contributing factors. This approach aims to bring together information from multiple models and elements, creating a unified framework for understanding commodity price dynamics and boosting forecast accuracy.

The research is divided into six sections: Introduction (Section 1), Related Work (Section 2), Methodology (Section 3),Implementation (Section 4), Results (Section 5), and Conclusion (Section 6).

2. RELATED WORK

This chapter provides the theoretical framework for the thesis by describing previous research on commodity price prediction and highlighting major methodologies, findings, and limitations. The literature review summarizes available knowledge on machine learning (ML) models, emphasizing their strengths and drawbacks in forecasting commodity prices. It begins by assessing classic econometric models and their limitations in handling commodities market volatility, leading to a shift toward machine learning approaches that can better capture nonlinear relationships and temporal dependencies. Key studies in the field are rigorously evaluated to identify gaps that require further research, ultimately contributing to advancements in forecasting methods.

To enable a comprehensive analysis, inclusion criteria for the literature were established, focusing on works published in English between 2015 and 2024, spanning various academic disciplines.

[9] Nwokike et al. (2020) studied gold price forecasts using Artificial Neural Networks (ANN), analyzing monthly data from October 2004 to February 2020. They compared 17 ANN structures and selected the best model (ANN 2-6-1) with the lowest Mean Square Error (MSE) and Mean Absolute Error. The study underlined ANNs' ability to capture the nonlinear behavior of gold prices, although it was criticized for relying on trial-and-error model selection and failing to compare alternative forecasting methods.

[10] Maciej Mróz (2022) investigated energy security by examining the dynamics of crude oil and copper supplies. Using a bivariate method, he examined supply security using indicators such as the Herfindahl-Hirschman Index (HHI), revealing that copper is more regionally concentrated. Price stability was measured using Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models, which demonstrated greater volatility in crude oil. The study's weaknesses include its small scope and reliance on GARCH models, which may fail to capture complex price dynamics.

[11]Kozian, Luca, and Osterrieder (2024) used a range of approaches, including Vector Autoregression (VAR) and Random Forest regressions, to investigate cross-commodity price relationships across 20 commodities from mid-2003 to early 2023. The study indicated that VAR and VARX models outperformed Random Forests, with R^2values up to 89%. The study's new use of Gerber statistics improved the research, however there were limitations in model selection and scope.

[12] Behshad Jodeiri Shokr (2020) used multiple linear regression (MLR) and the imperialist competitive algorithm (ICA) to forecast silver prices. The ICA-enhanced MLR increased both prediction accuracy and model fit. Despite its contributions, the study experienced challenges with MLR assumptions and reliance on a single data source, which may limit generalizability.

[13] Blohm and Antretter (2022) investigated the efficacy of machine learning (ML) algorithms in early-stage investment decisions using 255 business angels (BAs). Their findings demonstrated that, while ML algorithms outperformed human investors in general, experienced BAs who reduced cognitive biases made more money. The study used gradient-boosted decision trees to estimate venture longevity, and the Harrell's Concordance Index was 0.60, demonstrating the possibility for merging human judgment with machine learning to improve investment success.

[14] Yu-Wei Chen (2021) researched the price correlations between Brent Crude oil and 78 global commodities, concluding that the New York Harbor

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

No. 2 Heating Oil Spot Price is a good predictor of Brent Crude prices. The study found an 82.98% prediction success rate, arguing for a straightforward, data-light technique to supplement established forecasting models like LSTM and ARIMA. It underlined the importance of reevaluating price relationships amid market shifts.

[15]Özgür Önder (2021) explored the relationship between metal prices and economic phenomena, namely bubble detection. The study questioned the Efficient Market Hypothesis (EMH) while emphasizing behavioral finance ideas. It underlined the importance of doing a complete investigation across multiple metals and questioned the role of speculative activity in bubble formation, calling for more research on macroeconomic effects and new data sources.

[16] Madhika (2023) used a Multi-Layer Perceptron Neural Network to investigate the factors that influence gold prices. The study emphasized the importance of macroeconomic issues and other independent variables, with a high prediction accuracy of 98.17%. It discovered substantial correlations between gold prices and financial indicators, implying that additional research into geopolitical repercussions and volatility spillovers is required for a comprehensive understanding of gold price dynamics.

[17] Adem (2017) used a Multi-Layer Perceptron Neural Network (MLPNN) to investigate the factors that influence gold prices. The study emphasized gold's many roles as a currency and a safe haven amid economic downturns. Key findings showed considerable relationships between gold prices and numerous financial indices, with a high forecast accuracy of 98.17%. The study stressed the importance of future investigation into geopolitical events and developing markets' roles in gold price dynamics.

[18] Bildiricia (2015) evaluated the impact of oil prices on precious metals (gold, silver, and copper) in Turkey between 1973 and 2012. The study found a positive asymmetric response of gold prices to changes in oil prices, indicating a unidirectional Granger causality from oil to precious metals. The findings emphasized the interconnectivity of these commodities and revealed that oil price volatility considerably.

[19] Vidal (2020) introduced a hybrid CNN-LSTM model for forecasting gold volatility, outperforming traditional models like GARCH and standalone LSTM by 37% and 18%, respectively. This technique combines static and dynamic features to improve forecast accuracy. The study stressed the model's promise for effective portfolio management while also recognizing computational hurdles and data requirements for wider application.

[20]Sircar (2021) investigated the use of AI and machine learning in the oil and gas industry, highlighting the transformative potential for improving exploration and production efficiency. The evaluation emphasized the variety of approaches used, including supervised and unsupervised learning, as well as considerable increases in operational efficiency. However, issues with model interpretability and data quality persist, demanding strong frameworks for data management and interdisciplinary collaboration.

[21]Khadijah M (2019) underlines the considerable impact of AI and machine learning (ML) on the oil and gas industry (OGI), particularly in terms of optimizing exploration and production operations. These technologies improve seismic data and well log analysis, resulting in higher drilling precision and reservoir characterization. ML methods, such as support vector machines and neural networks, have played critical roles in improving exploration efficiency and minimizing downtime through predictive maintenance. However, issues such as data quality and organizational resistance impede widespread use, demanding strong models and increased data literacy to allow implementation.

[22]Chebeir (2018) examines how to optimize liquid-rich unconventional reservoirs, particularly in the Marcellus region, by combining reservoir engineering, machine learning, and economic optimization. This multidisciplinary approach combines compositional modeling for realistic production dynamics with economic frameworks to determine the best production strategies. Machine learning improves predictive capacities for reservoir performance and operational decision-making, highlighting the importance of integrated resource management approaches.

[23]Zhang Y (2020) examines ML applications for predicting crude oil market crashes, highlighting important research gaps in incorporating external

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

elements such as macroeconomic data and geopolitical events. The study emphasizes the significance of model interpretability and advises using behavioral finance theories to increase forecast accuracy. It calls for the research of complex systems theory to better capture market dynamics and improve the effectiveness of machine learning models.

[24]Ménde-Suárez (2019) investigates AI modeling frameworks for automated financial advice in the copper market, stressing the application of advanced machine learning techniques. The study highlights the need for greater integration of macroeconomic issues and improved model interpretability. It emphasizes the efficacy of AI models in predicting copper prices and supporting proactive trading techniques, hence expanding the potential of robo-advisors in financial services.

[25] GÜR, Yunus (2024) compares various deep learning models for predicting silver prices, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), and Gated Recurrent Units. Each model's distinct strengths and limitations are evaluated, with CNNs being praised for their capacity to extract spatial features from time-series data, LSTMs for managing long-term dependencies, and GRUs providing comparable performance with reduced computational needs. Hybrid models with diverse architectures have higher prediction powers. The review emphasizes the importance of data preparation and rigorous evaluation metrics in developing a strong theoretical framework for financial time-series prediction.

[26] Arendas, P. (2016) investigates the historical development and significance of the gold-silver ratio in commodities markets. This ratio is an important indication for investing strategies, and empirical evidence suggests that it can anticipate future silver price changes. The analysis emphasizes the ratio's volatility and the difficulties in timing investment decisions, proposing for scientific ways to evaluating ratio-based strategies and their compatibility with economic theories.

[27]Ul Sami (2017) emphasizes the significance of precisely anticipating gold prices given its symbolic value and influence in global banking. The paper stresses the use of machine learning techniques for analyzing historical data and market characteristics. Using advanced algorithms, researchers can capture complicated patterns impacting gold prices, improving forecasting accuracy. Future research objectives include utilizing deep learning and various data sources to enhance model interpretability and robustness.

[28]Liu, L. (2022) investigates forecasting crude oil return volatility, with a focus on the effectiveness of GARCH models in capturing time-varying volatility. While classic econometric models are widely used, machine learning technologies such as Support Vector Machines and Artificial Neural Networks are gaining popularity for their capacity to model nonlinear interactions. The analysis highlights persistent issues in volatility forecasting and recommends additional research to develop more robust models that use real-time data.

[29]Nordvoll (2023) looks at projecting bond fund flows using both standard econometric and advanced machine learning techniques. While classical methods are regarded for their simplicity, machine learning algorithms such as XGBoost and neural networks excel at dealing with complicated information. The study highlights the need for feature importance analysis and overfitting to improve forecast accuracy, despite inconsistent outcomes

[30] Al Qahtani (2023) investigates the difficulties of projecting stock market prices, stressing the complexity caused by market volatility and multiple affecting factors. This study makes an important addition by assessing the efficacy of various prediction models used to the Saudi stock market in multiple industries. The study evaluates six models: ARIMA, SVR, Random Forests, LSTM, Bi-LSTM, and GRU are all models that use large amounts of historical data and economic indices. The findings show that GRU and Random Forests beat ARIMA, which has trouble capturing nonlinear patterns. Notably, machine learning algorithms such as SVR, Random Forests, and GRU outperform classical approaches like ARIMA. Al Qahtani's work emphasizes the relevance of combining economic data and using rigorous feature selection to improve forecast accuracy. Future study should look into hybrid models and a larger variety of variables to improve stock price forecasting.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.1 Critique of Literature

This table summarizes the key strengths and weaknesses of numerous research on commodity price forecasting. It highlights the methodological breakthroughs and limits discovered in each study, providing insights into their contributions and areas of improvement.

Table 1: Summary of Strengths and Weaknesses of Selected Studies on Commodity Price Forecasting

Author	Year	Strengths	Weaknesses
Nwokike et al.[9]	2020	Demonstrated ANNs' ability to capture nonlinear behavior of gold prices.	Relied heavily on trial-and-error model selection; did not compare with alternative methods.
Maciej Mróz[10]	2022	Provided insights into crude oil and copper supplies with comprehensive indicators.	Small scope: GARCH models may not capture complex price dynamics.
Kozian, Luca, and Osterrieder[11]	2024	Employed a diverse range of approaches; introduced Gerber statistics.	Limitations in model selection and scope; potential areas for improvement.
Behshad Jodeiri Shokr [12]	2020	Enhanced prediction accuracy using ICA with MLR.	Challenges with MLR assumptions; reliance on a single data source.
Blohm and Antretter [13]	2022	Merged human judgment with ML algorithms for improved investment success.	Limited focus on experienced business angels reducing cognitive biases.
Yu-Wei Chen [14]	2021	High success rate using straightforward, data-light techniques.	Need for reevaluating price relationships amid market shifts without extensive validation.
Özgür Önder [15]	2021	Questioned Efficient Market Hypothesis; emphasized behavioral finance.	Called for more research on macroeconomic effects and new data sources.
Madhika [16]	2023	High prediction accuracy; emphasized macroeconomic issues.	Need for further research into geopolitical repercussions and volatility spillovers.
Adem [17]	2017	Highlighted gold's role during economic downturns with high forecast accuracy.	Suggested future investigation into geopolitical events and emerging markets.
Bildiricia [18]	2015	Found a positive asymmetric response of gold prices to oil price changes.	Did not fully explore implications of oil price volatility on precious metals.
Vidal [19]	2020	Introduced a hybrid CNN-LSTM model; outperformed traditional models.	Computational challenges and data requirements for broader application.
Sircar [20]	2021	Explored machine learning's potential in improving oil and gas exploration and production.	Issues with model interpretability and data quality; need for robust data management.
Khadijah M [21]	2019	Emphasized machine learning's impact on optimizing exploration and production.	Data quality issues and organizational resistance; requires improved models and data literacy.
Chebeir [22]	2018	Combined reservoir engineering, machine learning, and economic optimization.	Complex multidisciplinary approach; needs integrated resource management.
Zhang Y [23]	2020	Highlighted importance of model interpretability and macroeconomic factors.	Recommended further research into complex systems theory.
Ménde-Suárez [24]	2019	Investigated advanced machine learning for automated financial advice; emphasized macroeconomic integration.	Need for improved model interpretability and diverse data sources.
GÜR, Yunus [25]	2024	Compared deep learning models for predicting silver prices; highlighted hybrid model strengths.	Emphasized importance of rigorous data preparation and evaluation metrics.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Arendas, P. [26]	2016	Analyzed gold-silver ratio as an investment indicator; provided empirical evidence.	Noted volatility and timing challenges; suggested scientific evaluation methods.
Ul Sami [27]	2017	Focused on machine learning techniques for analyzing gold prices; emphasized complex patterns.	Suggested future research should use deep learning and diverse data sources.
Liu, L. [28]	2022	Investigated GARCH models and machine learning for crude oil return volatility forecasting.	Recommended developing more robust models using real-time data.
Nordvoll [29]	2023	Examined bond fund flows using econometric and machine learning techniques; highlighted feature importance.	Suggested further research to enhance forecast accuracy despite inconsistent outcomes.
Al Qahtani [30]	2023	Assessed prediction models for Saudi stock market prices; ML algorithms outperformed classical approaches.	Need for hybrid models and a broader range of variables for better forecasting.

2.2 Gap Addressed by the research

This research covers these gaps by providing a thorough examination of sophisticated forecasting models, leveraging a variety of data sources, focusing on emerging markets, and incorporating the most recent advances in machine learning. The gaps include:

1. Lack of Comprehensive Model Comparison: Existing studies often focus on a limited range of forecasting models. By comparing a broad spectrum of advanced models (LSTM, ANN, RFR, GBR, SVR, and XGBoost) across multiple commodities, this proposal seeks to provide a more nuanced understanding of model performance and identify which models are most effective in different contexts.

2.Single Data Source Reliance: Many studies rely on limited or single data sources, which can restrict the generalizability of findings. This proposal addresses this gap by integrating diverse data sources, enhancing the robustness and applicability of the forecasting models.

3.Incomplete Integration of Influencing Factors: Previous research often neglects the comprehensive analysis of factors such as geopolitical events, macroeconomic conditions, and behavioral finance elements. This proposal fills this gap by incorporating these broader factors into the analysis, offering a more holistic view of the dynamics influencing commodity prices.

4.Neglect of Emerging Markets: The impact of emerging markets on commodity prices is frequently underexplored in existing research. By focusing on these markets and their geopolitical influences, this proposal aims to provide new insights and improve forecasting accuracy in diverse economic contexts.

5. Advancement in Forecasting Methodologies: Many studies do not incorporate the latest advancements in machine learning and data analytics. This proposal addresses this gap by applying state-of-the-art methodologies to enhance forecasting accuracy and overcome limitations in current models.

6. Integration and Synthesis of Findings: Existing research often lacks a cohesive framework that integrates findings across different models and factors. This proposal aims to unify insights from various models and influencing elements, providing a comprehensive framework for understanding commodity price dynamics and improving predictive accuracy.

3. METHOD

This section describes the tools and framework for forecasting commodity prices and examining their interrelationships. It includes the research design, data collection techniques, preprocessing stages, variables, forecasting models, data analysis procedures, and software/tools. The goal is to create a transparent and replicable framework for the research process.[31].

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.1 The Proposal Framework

The proposed framework for predicting future commodity prices is divided into stages, each with its own set of tasks and procedures.

Figure 1 proposes a paradigm for predictive modeling in the commodity market

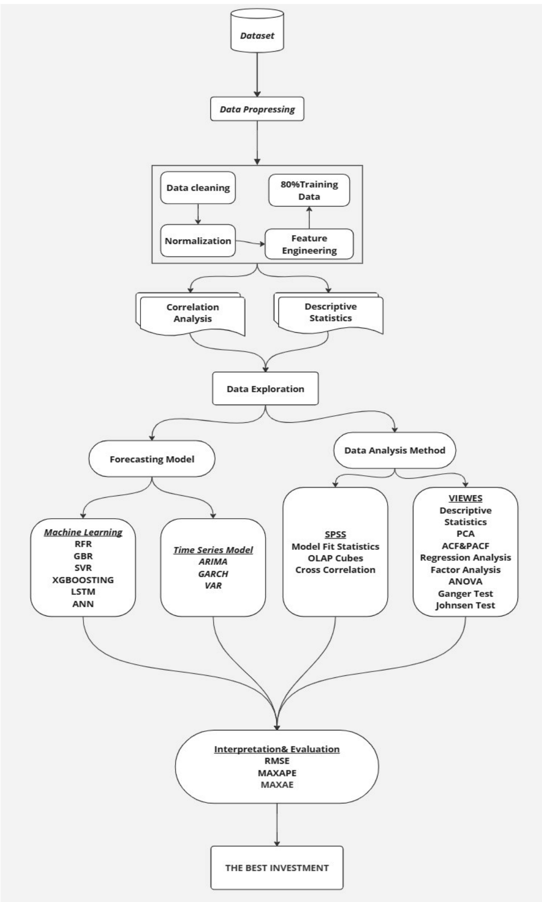

Figure 1: The Proposal Farmwork

3.1.1 Data Sources and Sampling

Data for this study were acquired from various financial sources, including Kaggle and Google datasets. It covers daily closing prices for selected commodities over a ten-year period (January 2010 to December 2020), showing various market situations such as economic growth and recession. The commodities used for this study include gold, silver, crude oil, Brent oil, natural gas, and copper. These commodities were chosen for their economic importance and widespread use in previous studies. They represent many sectors of the commodities market, providing a broad perspective on market interactions.

3.1.2 Data Preprocessing

is crucial to preparing the dataset for modeling. Measures taken include:

•Data Cleaning: Missing values were handled using interpolation or forward-fill techniques, and outliers were identified and treated to avoid biased results.

•Normalization was employed to standardize the data range, which led to faster model convergence and better performance.

•Feature Engineering: Moving averages, volatility indices, and lagged features were constructed by merging existing features to gather more data.

•Train Test Split: To assess the models' performance, the dataset was divided into training and testing sets in an 80:20 ratio.

3.1.3 Data Exploration

I. Descriptive statistics.

Descriptive statistics describe the central tendency and variability of each commodity's pricing. The key metrics include:

- Mean: The average price for the study period, indicating the typical price level.

-Median: When prices are sorted, the median value provides information about the price distribution.

- Standard Deviation: Measures price volatility and shows how much prices differ from the mean.

- Minimum and Maximum: The lowest and highest prices reported, illustrating the range of price variations.

These statistics help to understand the overall behavior of each commodity by highlighting variances in price stability and trends.

II. Correlation Analysis

Correlation analysis investigates the correlations between the selected commodities to identify important interactions. This is sometimes depicted as a correlation matrix, which shows the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

correlation coefficients of commodity pairs. Key points include:

Correlation coefficients vary from -1 to +1, with -+1 indicating a perfect positive correlation, --1 indicating a perfect negative correlation, and -0 indicating no association.

This approach is critical for understanding how commodities interact and improving the accuracy of forecasting algorithms.

3.2 Forecasting Models

The methodology predicts commodity prices using both time series models and machine learning techniques. Below is a full discussion of each strategy.

3.2.1 Time Series Models

•ARIMA (Autoregressive Integrated Moving Average):

-AR Component: Takes use of the relationship between an observation and previous observations.

-I Component: Differentiates observations to maintain time series stability.

In a moving average model, the MA component represents the relationship between an observation and its residual errors.

-Usage: Ideal for univariate time series data.

The model is widely known as ARIMA(p, d, q):

- p: The number of lag observations incorporated in the model (sometimes called the lag order).

- d: the number of times the original data is differentiated.

- q denotes the size of the moving average window.

•The GARCH (Generalized Autoregressive Conditional Heteroskedasticity)

model forecasts volatility in time series data by determining the variance of the current error term based on previous periods' error terms. It is best suited for financial data with volatility clustering.

• VAR (Vector Autoregression)

is a model for capturing interactions between multiple time series variables. It represents each variable as a linear function of previous values and other variables.

-Usage: Suitable for multivariate time series data in which variables influence one another.

-Strengths: Ability to manage complex partnerships and respond to shifting commodity patterns.

-Applications: Forecasting prices influenced by numerous variables.

3.2.2 Machine Learning Techniques

Machine learning (ML) trains machines to analyze data more efficiently. Often, analyzing extracted information from data might be difficult. ML algorithms are intended to automatically tackle data-related difficulties by recognizing patterns in training datasets and applying these patterns to test datasets for tasks such as prediction [32].The graphic displays the many machine learning methods used to forecast commodity prices. Figures 2-3-4-5-6-7 demonstrate the effectiveness of each technique in capturing market dynamics and improving prediction accuracy for specific applications.

1. Random Forest Regression (RFR) uses many decision trees to average their predictions [33].

-Strengths: Capable of managing complex partnerships and adapting to changing commodities patterns.

-Applications: Forecasting prices influenced by many variables.

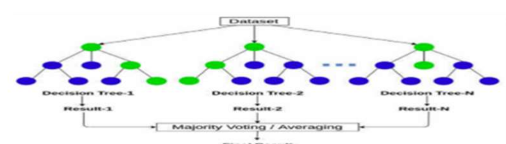

Figure 2: Random Forest Regression

2. Gradient-Boosting Regression (GBR):

-Method: Creates an additive model in a forward stage-wise approach, optimizing a loss function and gradually adding models to reduce prediction error[34].

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

-Strengths: Improves predictions by correcting errors and refining them over time.

-Applications: Adjusting projections in response to changing market conditions.

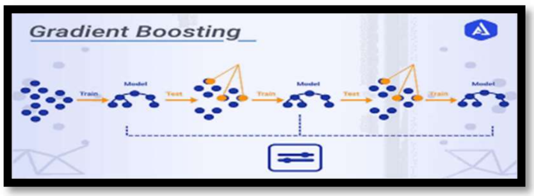

Figure 3: Gradient Boosting Regression

3. Support vector regression (SVR):

-Method: Regression is performed using Support Vector Machine principles, with the goal of identifying a function that deviates from real observed values by a certain margin.[35].

-Strengths: Sets precise margins around data points.

-Uses: Forecasting prices within well-defined market limits.

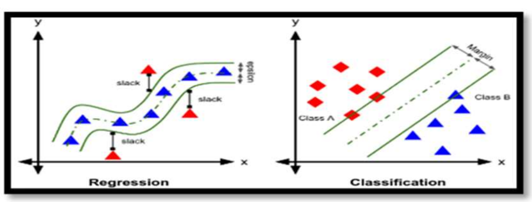

Figure 4: Support Vector Regression

4. XGBoost (or Extreme Gradient Boosting):

-Method: An improved implementation of gradient boosting aimed for speed and performance, using regularization to control overfitting [36].

-Strengths: Improves gradient boosting with advanced regularization approaches.

-Applications: Improving projections in volatile commodity markets.

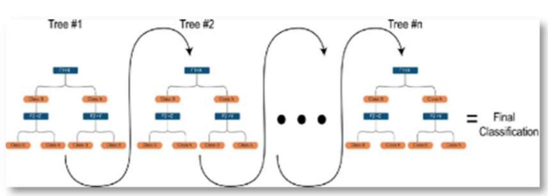

Figure 5: XGBoosting

5. Long Short-Term Memory (LSTM):

-Method: A recurrent neural network capable of learning long-term dependencies, notably effective for time series prediction [37].

-Strengths: Detects long-term dependencies in time series data.

-Applications: Predicting pricing using historical trends and seasonal patterns.

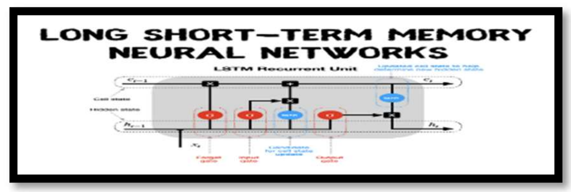

Figure 6: Long Short-Term Memory

6. Artificial neural networks (ANNs):

-Method: Made up of interconnected nodes that capture complicated patterns using numerous hidden layers and nonlinear transformations[38].

-Strengths: Can model complex linkages and adjust to a variety of commodities factors.

-Applications: Flexible forecasting for commodities influenced by a variety of economic indicators.

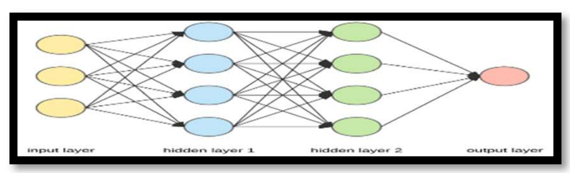

Figure 7: Artificial Neural Network

3.3 Software and Tools

Use specialist software and tools to analyze data and apply models:

- Statistical software:

• Use SPSS for descriptive statistics and PCA.

• EViews for Granger causality and Johansen Cointegration testing.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

-Machine Learning Libraries: Use scikit-learn, TensorFlow/Keras, and XGBoost for efficient model training and optimization.

3.4 Evaluation Metrics

The performance of the models was assessed using the Root Mean Squared Error (RMSE), a widely used regression metric. RMSE is calculated as the square root of the average squared difference between expected and actual values. Lower RMSE values suggest a better model performance.

3.5 Data Analysis Methods

These tools provide an understanding of market dynamics, generating strategies for portfolio management and hedging.

3.5.1 Descriptive Statistics

Descriptive statistics such as mean, median, standard deviation, skewness, and kurtosis were calculated for each commodity using SPSS. This approach provides a fundamental understanding of data distribution and variability.

Steps for Descriptive Statistics:

1. Data collection: Over a ten-year period, financial databases were used to acquire daily closing prices for crude oil, Brent oil, natural gas, gold, silver, and copper.

2. Data Cleaning: To manage missing values, mean imputation was employed, and outliers were treated using the z-score approach.

3. Calculation: PSS software was used to generate descriptive statistics. In SPSS, select 'Analyze' > 'Descriptive Statistics' > 'Descriptives,' and then calculate the mean, median, standard deviation, skewness, and kurtosis.

3.5.2 Granger Causality Tests

Granger causality tests were conducted using EViews software to determine whether the past value of one commodity could predict the future value of another. This test identifies potential leading indicators in commodity markets.

Steps for Pairwise Granger Causality Test:

1. Data Preparation: We gathered historical pricing data for many commodities and examined it for stationarity using the Augmented Dickey-Fuller (ADF) test. Non-stationary series were discriminated to achieve stationarity.

2. Model Specification: The ideal lag duration was determined using information metrics such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Pairwise Granger causality tests were then performed.

3. The tests were carried out with EViews software, which included data import, stationarity checking, lag duration selection, and execution

3.5.3 Principal Component Analysis (PCA)

PCA was used to reduce the dataset's dimensionality while preserving the majority of its variance. SPSS software was used to simplify the data and identify key drivers of market movement and commodity correlation.

Steps for Principal Component Analysis (PCA):

1. Data Preparation: We gathered and cleaned commodity data, imputing missing values and addressing anomalies.

2. Data Standardization: To ensure equitable contribution, each variable was standardized to have a zero mean and one standard deviation (SD).

3. To do PCA, utilize the SPSS application and select 'Analyze' > 'Dimension Reduction' > 'Factor.' The extraction method was switched to PCA, and

Varimax rotation was used to simplify the loading structure.

3.5.4 Johansen Cointegration Test

The Johansen Cointegration Test was performed with EViews software to detect long-term equilibrium linkages between commodities. This research is critical for determining if commodities move together over time, indicating solid linkages helpful for portfolio management and hedging techniques.

To do the Johansen Cointegration Test, collect monthly data on commodities over a significant period, test for stationarity, and adjust for differences as needed.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

1. Model Specification: The optimal lag length for the VAR model was found based on AIC, BIC, and HQIC criteria. Decisions were made to include deterministic components such as intercepts and trends.

2. Estimation and Testing: The VAR model was estimated, and the Johansen Cointegration Test was performed using Trace and Maximum Eigenvalue statistics to identify cointegrating vectors.

4. IMPLEMENATION MODEL

This section discusses the systematic application of machine learning models to forecast commodity prices, focusing on crude oil, Brent oil, natural gas, gold, silver, and copper. It starts with data collection from reputable sources like Kaggle and Google Datasets, with a focus on data integrity during preparation.

The covers model generation techniques such as Random Forest Regressor, Gradient Boosting Regressor, Support Vector Regression (SVR), XGBoost, Long Short-Term Memory (LSTM), and Artificial Neural Networks. Each model undergoes extensive hyperparameter tuning and training, and performance is quantified using metrics like Root Mean Square Error (RMSE) to compare efficacy.

This part also shows how to utilize SPSS for time series analysis and EViews for statistical validation, which covers descriptive statistics techniques.

4.1 Implementation in EViews

4.2 Descriptive statistics

Descriptive statistics provide essential insights into the dataset by highlighting central patterns and commodity price distributions. The key metrics include:

Key measures include: • Central tendency: • Mean: The average value in the dataset.

The median is the middle value in an ordered set of data.

Mode is the most common value in the dataset.

• Standard Deviation measures variance from the mean.

Variance is calculated as the square of the standard deviation.

The range is the difference between the maximum and minimum numbers.

• The Distribution's Shape

Skewness determines asymmetry in data distribution.

Kurtosis measures the "tailedness" of a data distribution.

This table 2 displays summary data for key metrics such as mean, median, standard deviation, and range. The data is organized to allow for cross-variable comparisons, which provide insight into the dataset's distribution and variability. These statistics are crucial for determining general patterns and performance metrics associated with the study.

Table2 summarizing the key descriptive statistics for each commodity

Statistic	Gold	Silver	Crude Oil	Brent Oil	Natural Gas	Copper
Mean	1452.91	20.32	65.70	73.71	3.17	3.13
Median	1337.25	18.07	55.28	66.55	2.91	3.05
Maximum	2054.60	37.14	122.11	127.98	9.32	4.91
Minimum	1049.70	11.77	26.21	19.33	1.48	1.99
Std. Dev.	252.23	5.32	22.44	26.03	1.06	0.64
Skewness	0.57	0.96	0.64	0.35	2.00	0.63
Kurtosis	1.97	2.94	1.96	1.86	9.76	2.86
Jarque-Bera	269.15	502.65	309.66	203.71	7025.38	182.23

Figure 8 depicts the data presented in Table 2, highlighting central trends and dispersion for each key indicator. It clearly contrasts the mean, median, and standard deviation, making it simple to assess the distribution and variability of data points across the tested variables.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

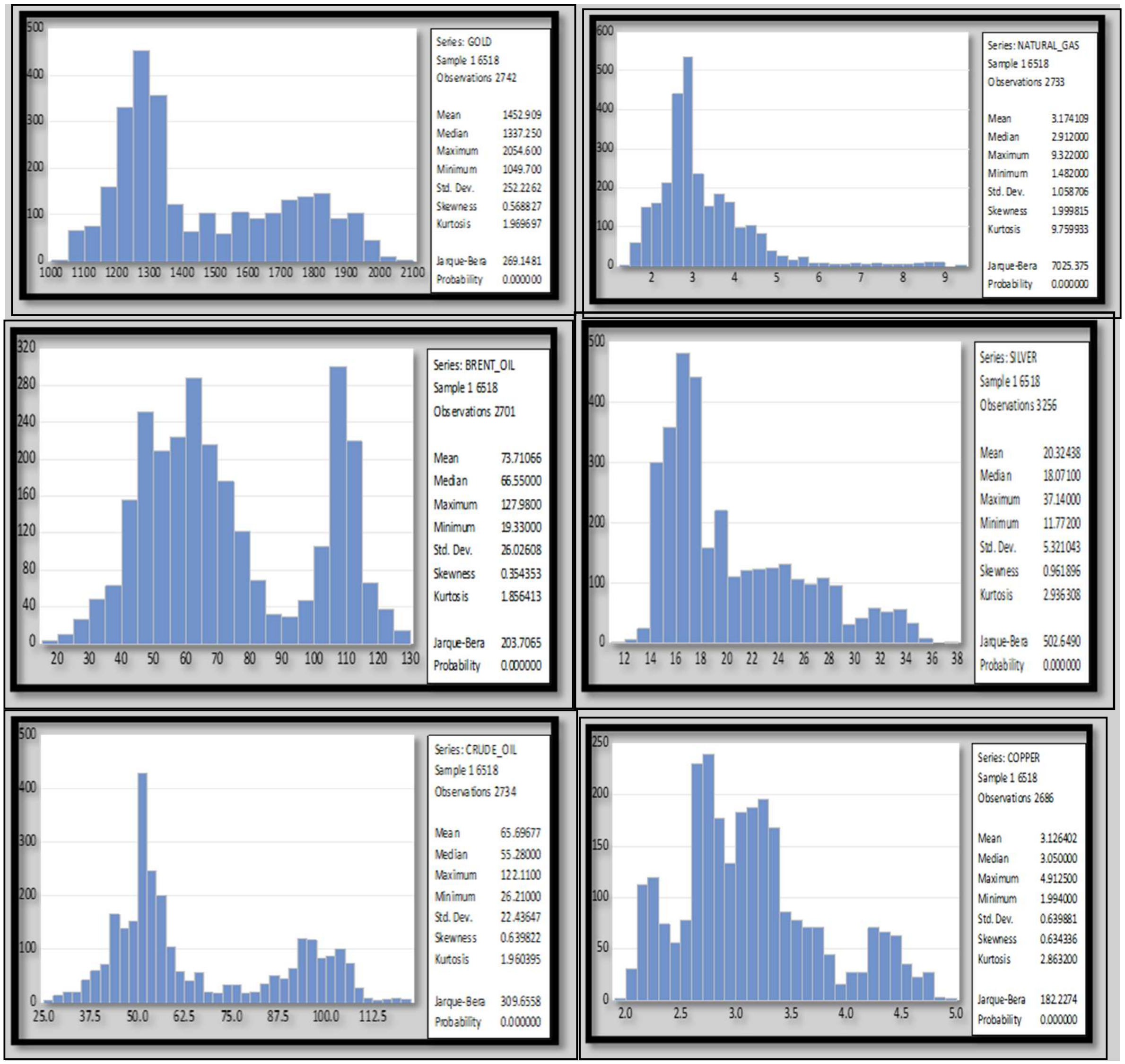

Figure 8: summarizing the key descriptive statistics for each commodity

Analysis of Key Descriptive Statistics

This analysis gives information on the price behavior and distribution of gold, silver, crude oil, Brent oil, natural gas, and copper. Here are the major observations:

1. Central Tendency.

Mean and median: All commodities have mean prices that are higher than the median, indicating that the distributions are right-skewed. Gold and silver have more symmetric distributions than natural gas, which has a significant difference between the mean and median values.

2. Dispersion.

Standard deviation: Crude and Brent oil have high standard deviations, indicating extreme price volatility. Natural gas, despite having a lower mean price, is highly volatile relative to its mean. In contrast, gold and silver have moderate standard deviations, implying more stable price movements.

3. Distribution Patterns

- All commodities show positive skewness, indicating long right tails. Natural gas has the highest

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

skewness (2.00), indicating the presence of extraordinarily high values.

Key insights:

- Crude oil and Brent oil are the most volatile commodities, indicating wider movements in global oil markets.

- Gold and silver are considered safe investments during market instability due to their relative stability.

- Natural gas's skewness and kurtosis indicate frequent price fluctuations, either due to supply disruptions or geopolitical events.

- Non-Normal Distributions: The commodities lack a normal distribution, posing challenges for risk management and modeling. Investors must account for broad tails and the probability of extreme values in their forecasts.

Descriptive statistics provide a fundamental understanding of these commodities' pricing patterns. The analysis emphasizes the need of taking volatility, skewness, and kurtosis into consideration when assessing risks and returns in these markets. To capture the underlying dynamics of these commodities, more advanced modeling and forecasting techniques should be applied.

4.1.2 Principal Components Analysis (PCA)

PCA is a statistical technique for reducing high-dimensional data while preserving trends and patterns by transforming it into orthogonal (uncorrelated) components. Here is a detailed summary based on the PCA results.

The important findings table3 for PCA includes eigenvalues, which indicate how much variance each principal component carries. Higher eigenvalues indicate more significant components.

The sum of the eigenvalues equals the number of variables, indicating the total variance in the data.

In this study, seven components were identified.

PC1 accounts for 3.76 (53.66%) of the variance, while PC2 explains 1.86 (26.54%).

PC1 and PC2 explain 80.20% of the overall variance, implying that they capture the vast majority of the data.

Loading (eigenvectors):

Loadings show how much each original variable adds to each primary component.

Variables with significant absolute values of loadings have the greatest influence on component selection. According to Figure 9, PCA essential outcomes include eigenvalues, which represent the amount of variation held by each principal component. Higher eigenvalues indicate more significant components.

Figure 9 displays the important findings of principal component analysis (PCA), with emphasis on the eigenvalues associated with each main component. Variables with large absolute values of loading have the most influence on component selection. Higher eigenvalues indicate components that capture more variance in the information, underlining their importance in understanding the underlying data structures and relationships.

Eigenvalues: (Sum = 7, Average = 1)

Number	Value	Difference	Proportion	Cumulative Value	Cumulative Proportion
1	3.756205	1.898396	0.5366	3.756205	0.5366
2	1.857809	0.880534	0.2654	5.614015	0.8020
3	0.977275	0.764298	0.1396	6.591290	0.9416
4	0.212977	0.073220	0.0304	6.804267	0.9720
5	0.139758	0.102617	0.0200	6.944025	0.9920
6	0.037141	0.018307	0.0053	6.981166	0.9973
7	0.018834	---	0.0027	7.000000	1.0000

Eigenvectors (loadings):

Variable	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6	PC 7
DATE_E	-0.083911	0.659002	0.368995	0.266561	0.255149	0.470301	0.255303
GOLD_E	0.309250	0.526771	-0.271054	-0.263868	0.468852	-0.400191	-0.322188
SILVER_E	0.414636	0.125278	-0.534637	-0.329121	-0.286174	0.515914	0.264882
CRUDEOIL_E	0.455906	-0.304663	0.089146	0.124001	0.427801	-0.225106	0.665072
BRENTOIL_E	0.453734	-0.303756	0.031792	0.391582	0.258930	0.418985	-0.552194
NATURALGAS_E	0.343466	-0.035109	0.697531	-0.594647	-0.157043	0.048755	-0.116637
COPPER_E	0.440121	0.293623	0.095020	0.478206	-0.598715	-0.351994	-0.003115

Ordinary correlations:

	DATE_E	GOLD_E	SILVER_E	CRUDEOIL_E	BRENTOIL_E	NATURAL...	COPPER_E
DATE_E	1.000000						
GOLD_E	0.442908	1.000000					
SILVER_E	-0.188709	0.736337	1.000000				
CRUDEOIL_E	-0.462987	0.228187	0.565771	1.000000			
BRENTOIL_E	-0.467309	0.213453	0.586833	0.967108	1.000000		
NATURALGAS_E	0.061227	0.202952	0.210627	0.641857	0.573556	1.000000	
COPPER_E	0.254662	0.712574	0.687828	0.575520	0.600131	0.565382	1.000000

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Figure9 : Principal Components Analysis

- Interpretation of PC

PC1: Represents general market movement, principally driven by crude oil, Brent oil, copper, and silver, demonstrating that these commodities usually move in tandem.

PC2: Displays an inverse relationship between gold and oil prices, meaning that while gold prices rise, oil prices often decline, and vice versa.

PC3-PC7: Measure less variance while reflecting more specific or residual changes, exposing subtle links or individual commodity habits.

PCA considerably reduces the dataset's dimensionality by identifying important components responsible for the majority of commodity price volatility. The first two principal components (PC1 and PC2) account for more than 80% of the volatility, highlighting their relevance in understanding overall market dynamics. The loadings provide information on the influence of each commodity and their interrelationship.

Interpretation of Principal Components

PC1 explains 53.66% of the variation.

Most commodities have strong positive loadings, indicating that this component accurately represents the overall trend in commodity prices.

The commodities with the largest loadings are crude oil (0.456), Brent oil (0.454), copper (0.440), and silver (0.415).

PC2 explains 26.54 percent of the variation.

Gold has a considerable positive loading (0.527), while crude oil and Brent oil have significant negative loadings (-0.305 and -0.304, respectively).

This component may indicate an inverse relationship between gold and oil prices, implying contrasting market dynamics or investor behavior.

Provide a summary of the PC data interpretation (as explained in table 4).

PCA significantly decreases the dataset's dimensionality by finding significant components

that account for the majority of volatility in commodity prices. The first two principal components (PC1 and PC2) explain more than 80% of the variance, emphasizing their importance in understanding overall market dynamics. The loadings provide insights into the impact of each commodity and their interrelationships.

Table 3 presents the principal Components Analysis (PCA) results, including loadings, eigenvalues, and variance explained by each principal component. It highlights the most relevant factors, providing for a better understanding of how they affect the overall data structure and variation. The information supplied helps to identify crucial components for further investigation and interpretation.

Table3 summarizing Principal Components Analysis

Variable	PC1	PC2	PC3	PC4	PC5	PC6	PC7
GOLD_E	0.309	0.527	-0.271	-0.264	0.469	-0.400	-0.322
SILVER_E	0.415	0.125	-0.535	-0.329	-0.286	0.516	0.265
CRUDEOIL_E	0.456	-0.305	0.089	0.124	0.428	-0.225	0.665
BRENTOIL_E	0.454	-0.304	0.032	0.392	0.259	0.419	-0.552
NATURALGAS_E	0.343	-0.035	0.698	-0.595	-0.157	0.049	-0.117
COPPER_E	0.440	0.294	0.095	0.478	-0.599	-0.352	-0.003

1. Autocorrelation Function (ACF) Definition: Autocorrelation is the correlation between a time series and its previous values. The ACF shows how data at various delays are connected.

The ACF values range from -1 to 1. Values close to 1 or -1 indicate strong ties, whereas values near 0 indicate weak relationships.

2. Partial autocorrelation function (PACF)

- Definition: Partial autocorrelation examines the connection between a time series and its lagged values, accounting for shorter lags. This aids in determining direct ties.

PACF values can range from -1 to 1, indicating the strength of correlations at specific lags.

- PACF values that are significant show a direct dependence on certain historical values, which can help determine the order of autoregressive models.

This figure10 depicts the findings of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) studies, which reveal temporal connections in the data. The plots show the strength and significance of correlations at different lags, which can assist discover potential patterns and inform model selection for time series research.

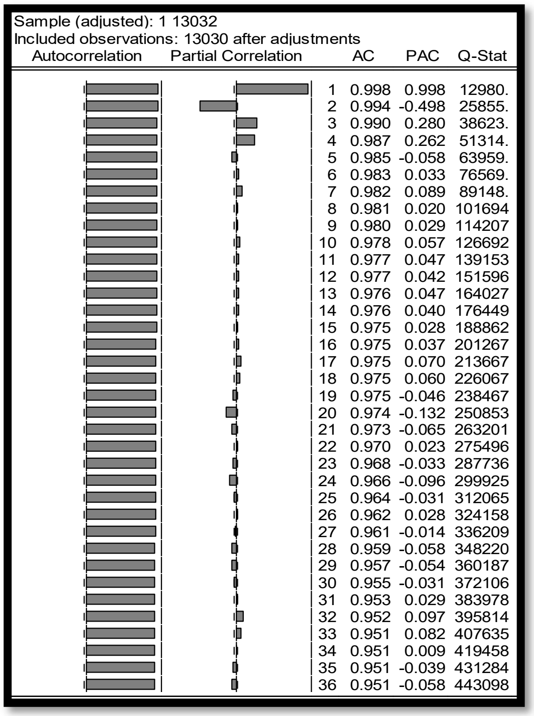

Figure 10: Autocorrelation And Partial Autocorrelation

Interpretation of Results

The examination of a time series with 13,030 data revealed the following significant points:

1. Autocorrelation: - The series exhibits strong autocorrelation across many lags, indicating that present values are heavily influenced by previous values.

-Notably, the high autocorrelation at lag 1 (0.998) indicates that the most recent value is nearly as predictive as the present value, implying a strong trend.

2. Decay Pattern: - Autocorrelation steadily diminishes with rising lags, but stays substantial at greater lags, indicating a long memory effect where past values impact current values over longer durations.

3. Partial Autocorrelation: - Significant PACF values at several initial lags show direct influence on current value. -The substantial PACF values at lags 1, 2, and 3 indicate that the series can be modelled as an autoregressive process with these delays.

- Significance: Autocorrelation at different lags reveals data persistence and patterns.

Practical Implications

Understanding autocorrelation and partial autocorrelation in financial time series is important for various reasons.

1. Forecasting: - Significant autocorrelation enables using past values to predict future ones. Time series models, such as ARIMA, rely on these qualities to forecast.

2. Model Identification: - ACF and PACF plots help determine acceptable delays for AR and MA components in ARIMA models. For example, if the PACF stops after a few lags while the ACF continues, this indicates an AR model of that kind.

3. Risk Management: - Price persistence and autocorrelation show that price shocks can have long-term repercussions, making risk management and hedging methods vital.

4. Market Efficiency: - In an efficient market, prices should be randomly distributed with no significant autocorrelation. Persistent autocorrelation indicates possible market inefficiencies that might be used in trading techniques.

The significant autocorrelation and partial autocorrelation in the time series data show a strong reliance on previous values, showing trends or momentum. This information is essential for developing prediction models and understanding market dynamics. Future steps will entail fitting relevant time series models, such as ARIMA, and verifying their effectiveness to improve forecasting skills.

Regression Analysis

Regression analysis is a powerful statistical tool for investigating correlations between a dependent variable and one or more independent variables. This study's analysis focuses on DATE_E as the dependent variable and various commodities prices as independent factors.

Key Results: 1. Model Summary.

- R-squared: 0.905575.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

- Adjusted R-squared = 0.905523 - The independent variables account for 90.56% of the variability in DATE_E, indicating a robust model fit.

2. ANOVA Table: - F-statistic: 17331.45 - A extremely high F-statistic with a suitably low p-value implies that the entire regression model is significant.

3. ANOVA. Table: - F-statistic: 17331.45 - A high F-statistic and low p-value suggest a highly significant regression model.

4. Regression Coefficients: - Coefficients show the change in DATE_E with a one-unit change in the independent variable, while keeping other variables constant. The significance of these coefficients is determined using t-statistics and p-values.

Interpretation of Regression Coefficients

1. Gold_E:

- Coefficient: 3.5985.

- Interpretation: A one-unit increase in the gold price index (GOLD_E) results in an increase of about 3.60 units in the dependent variable DATE_E, while all other variables remain constant. This link is quite significant.

2. Silver_E:

Coefficient: -178.2548.

- Interpretation: For every one-unit increase in the silver price index (SILVER_E), the dependent variable DATE_E drops by approximately 178.25 units, while all other variables remain constant. This link is quite significant.

3. Crudeoil_E:

- Coefficient: -32.3480.

- Interpretation: For every one-unit increase in the crude oil price index (CRUDEOIL_E), the dependent variable DATE_E declines by about 32.35 units, while all other variables remain constant. This link is quite significant.

4. Brentoil_E:

- Coefficient: 5.2777.

- Interpretation: For every one-unit increase in the Brent oil price index (BRENTOIL_E), the dependent variable DATE_E rises by about 5.28 units, while all other variables remain constant. This link is quite significant.

. Naturalgas_E:

- Coefficient: 169.8121.

- Interpretation: A one-unit rise in the natural gas price index (NATURALGAS_E) results in an increase of about 169.81 units in the dependent variable DATE_E, while all other variables remain constant. This link is quite significant.

6. Copper_E:

- Coefficient: 783.9430.

- Interpretation: Holding all other variables constant, for every one-unit increase in the copper price index (COPPER_E), the dependent variable DATE_E rises by about 783.94. This link is quite significant.

This figure11 depicts the results of the regression analysis, which show the link between the independent and dependent variables. Key metrics including coefficients, R-squared values, and confidence intervals are displayed to evaluate the model's fit and prediction potential. The graphical representation helps to clarify the impact of each predictor on the response variable, showing notable linkages and relevant topics for additional investigation.

Dependent Variable: DATE_E
Method: Least Squares
Date: 07/10/24 Time: 02:06
Sample (adjusted): 1 11942
Included observations: 10850 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GOLD_E	3.598500	0.027337	131.6354	0.0000
SILVER_E	-178.2548	1.330969	-133.9286	0.0000
CRUDEOIL_E	-32.34800	0.671108	-48.20090	0.0000
BRENTOIL_E	5.277677	0.585545	9.013268	0.0000
NATURALGAS_E	169.8121	5.127385	33.11866	0.0000
COPPER_E	783.9430	11.09190	70.67706	0.0000
C	733561.0	20.69814	35440.91	0.0000

R-squared	0.905575	Mean dependent var	736436.5
Adjusted R-squared	0.905523	S.D. dependent var	1088.717
S.E. of regression	334.6410	Akaike info criterion	14.46464
Sum squared resid	1.21E+09	Schwarz criterion	14.46934
Log likelihood	-78463.66	Hannan-Quinn criter.	14.46622
F-statistic	17331.45	Durbin-Watson stat	0.014277

Figure 11: Regression Analysis

Practical Implications

1. Gold and Natural gas: -Both show positive coefficients, demonstrating that price increases lead to rises in the dependent variable (DATE_E). These commodities may be positively connected with the underlying element described by DATE_E, which could be an economic index, a time period, or another financial measure.

2. Silver and Crude Oil: - Both have negative coefficients, indicating that increases in prices lead to decreases in DATE_E. This negative association suggests that these commodities may move inversely in regard to the dependent variable.

3. Model Fit: - The high R-squared value (0.905575) demonstrates that the model explains a significant percentage of the variance in DATE_E, suggesting a robust fit.

The regression analysis provides useful information about the links between commodity prices and the dependent variable. DATE_E. Understanding these dynamics is critical for forecasting, risk management, and strategic decision-making in the financial markets. The extremely significant coefficients and robust model fit highlight the significance of these commodities in explaining fluctuations in the dependent variable.

4.1.5 Factor Analysis

Factor analysis is a statistical method that reduces dimensionality by expressing variability among observed variables using fewer unseen variables, or factors. This solution uses the maximum likelihood approach to analyze the dataset.

a) Maximum Likelihood technique • The maximum likelihood technique estimates factor loadings and unique variances to maximize the likelihood function, assuming a multivariate normal distribution. This method assists in determining the underlying structure of the data by estimating factor loadings and communalities.

b) Identified Factors • Identified two primary factors that account for most of the dataset's volatility, simplifying its complexity.

c) Communalities: • Determine the proportion of each variable's variance explained by the specified factors. High communalities indicate significant explanatory power.

• Examples of Communalities:

- Date_E: 0.707963 - Gold_E: 0.674847.

d) Uniqueness Values • Uniqueness values indicate the variance of each variable that cannot be explained by other factors. Lower values suggest a higher explanation by the components.

An Example of Uniqueness Values:

- Date E: 0.292036 - Gold E: 0.325153

e) Factor Variance

The factor variance table indicates how much of the total variance is explained by each factor.

Example Factor Variance:

- F1 = 3.391626 (65.12% of total variance)

- F2 = 1.816851 (34.88% of total variance)

g) Model Fit.

The discrepancy, chi-square statistic, and Bartlett chi-square test are all metrics of model fit. Lower discrepancy values and significant chi-square statistics suggest that the model is well fitted to the data.

Example Model Fit Statistics:

- Discrepancy of 2.415031

The chi-square statistic is 26200.67, while the Bartlett chi-square is 26189.80.

The factor analysis using the maximum likelihood the method successfully discovers two primary components that account for a considerable amount of the dataset's variances. Understanding communalities and uniqueness values helps assess how well the components explain each variable, revealing the underlying structure of commodity prices and facilitating data-driven decision-making.

This figure12 illustrates the results of the factor analysis performed with maximum likelihood estimation. It depicts the links between observed variables and underlying factors, emphasizing factor loadings to demonstrate the strength of these interactions. The study provides insights into the dimensionality of the data, allowing for a better understanding of the latent constructs driving the observed measures

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Factor Method: Maximum Likelihood
Date: 07/10/24 Time: 02:06
Covariance Analysis: Ordinary Correlation
Sample (adjusted): 1 11942
Included observations: 10850 after adjustments
Balanced sample (listwise missing value deletion)
Number of factors: Minimum average partial
Prior communalities: Squared multiple correlation
Convergence achieved after 8 iterations

	Unrotated Loadings			
	F1	F2	Communality	Uniqueness
DATE_E	-0.404453	0.737821	0.707963	0.292036
GOLD_E	0.294980	0.766704	0.674847	0.325153
SILVER_E	0.631292	0.382435	0.544786	0.455214
CRUDEOIL_E	0.971142	-0.098653	0.952849	0.047152
BRENTOIL_E	0.986723	-0.091148	0.981930	0.018070
NATURALGAS_E	0.614355	0.197694	0.416515	0.583485
COPPER_E	0.669575	0.693726	0.929587	0.070413

Factor	Variance	Cumulative	Difference	Proportion	Cumulative
F1	3.391626	3.391626	1.574775	0.651174	0.651174
F2	1.816851	5.208477	---	0.348826	1.000000
Total	5.208477	5.208477		1.000000	

	Model	Independence	Saturated
Discrepancy	2.415031	8.859726	0.000000
Chi-square statistic	26200.67	96119.17	---
Chi-square prob.	0.0000	0.0000	---
Bartlett chi-square	26189.80	96091.11	---
Bartlett probability	0.0000	0.0000	---
Parameters	20	7	28
Degrees-of-freedom	8	21	---

Figure 12: Maximum Likelihood

4.1.6 Analysis of Variance (ANOVA)

The ANOVA analysis examines the means of various commodity price series to see if there are significant differences between them.

The ANOVA and Welch F-tests show substantial mean differences between the various commodity price series.

The enormous sum of squares between groups, as opposed to within groups, supports the conclusion that there are significant disparities.

The category statistics provide useful information on the central tendency and dispersion of each commodity price series, which improves understanding of specific series characteristics.

This figure13 depicts the results of the Analysis of Variance (ANOVA), which indicate the variations in means across several groups. It provides crucial statistics, like as F-values and p-values, to assess the significance of group differences. The graphical depiction helps to visualize how group means compare, highlights notable impacts, and informs future research of variations within the dataset

Test for Equality of Means Between Series
Date: 07/10/24 Time: 02:31
Sample: 1 13034
Included observations: 13034

Method	df	Value	Probability
Anova F-test	(6, 82595)	4.95E+09	0.0000
Welch F-test*	(6, 33428.2)	9.70E+08	0.0000

*Test allows for unequal cell variances

Analysis of Variance

Source of Variation	df	Sum of Sq.	Mean Sq.
Between	6	5.95E+15	9.91E+14
Within	82595	1.65E+10	200290.6
Total	82601	5.95E+15	7.20E+10

Category Statistics

Variable	Count	Mean	Std. Dev.	Std. Err. of Mean
DATE_E	13030	736412.0	1102.136	9.655239
GOLD_E	11006	1454.154	252.8412	2.410089
SILVER_E	13034	20.32299	5.318802	0.046588
CRUDEOI...	10934	65.68551	22.41899	0.214401
BRENTOI...	10850	73.09077	25.76858	0.247386
NATURA...	10934	3.173665	1.057280	0.010111
COPPER_E	12814	3.300164	0.748088	0.006609
All	82602	116381.0	268331.9	933.6347

Figure 13: ANOVA

The ANOVA analysis clearly indicates that there are significant disparities in the means of the several commodity price series, laying the groundwork for future statistical research and analysis.

4.1.7 Granger Causality Tests

Granger Causality tests seek to evaluate whether the previous values of one time series may predict another. In this context, the tests investigate the correlations between various commodity prices, such as gold, silver, crude oil, Brent oil, natural gas, and copper.

Key Components:

1. Null Hypothesis (H0): Previous values of one time series do not predict subsequent values.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2. Alternative Hypothesis (H1): One time series' past value can forecast another's.

3. F-Statistic: Calculates the ratio of variance explained by the model (including lagged values of predictor series) versus variance not explained by the model. A higher F-statistic indicates stronger evidence against the null hypothesis..

4. p-Value: Measures the likelihood of detecting test results under the null hypothesis. If the p-value is less than 0.05, the null hypothesis is rejected, indicating that the predictor series is the cause of the target series.

This figure14 depicts the results of the Granger causality tests, which demonstrate the links between time series variables. It emphasizes the direction and significance of causality, revealing if one variable predicts another across time. The graphical representation helps to grasp the dynamic interactions between the variables, revealing potential causal linkages for further inquiry.

Pairwise Granger Causality Tests
Date: 07/10/24 Time: 15:16
Sample: 1 6517
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Prob.
GOLD E does not Granger Cause DATE E	5501	1.51067	0.2209
DATE E does not Granger Cause GOLD E		5.65112	0.0035
SILVER E does not Granger Cause DATE E	6513	67.1515	1.E-29
DATE E does not Granger Cause SILVER E		52.6215	2.E-23
CRUDEOIL E does not Granger Cause DATE E	5465	0.76801	0.4640
DATE E does not Granger Cause CRUDEOIL E		5.01417	0.0067
BRENTOIL E does not Granger Cause DATE E	5485	5.63123	0.0036
DATE E does not Granger Cause BRENTOIL E		91.7891	6.E-40
NATURALGAS E does not Granger Cause DATE E	5465	0.44706	0.6395
DATE E does not Granger Cause NATURALGAS E		0.30858	0.7345
COPPER E does not Granger Cause DATE E	6405	1.01556	0.3623
DATE E does not Granger Cause COPPER E		9.97430	5.E-05
SILVER E does not Granger Cause GOLD E	5501	29.3486	2.E-13
GOLD E does not Granger Cause SILVER E		1.43141	0.2391
CRUDEOIL E does not Granger Cause GOLD E	5465	1.56419	0.2094
GOLD E does not Granger Cause CRUDEOIL E		0.63066	0.5323
BRENTOIL E does not Granger Cause GOLD E	5485	1.33740	0.2626
GOLD E does not Granger Cause BRENTOIL E		0.85615	0.4249
NATURALGAS E does not Granger Cause GOLD E	5465	1.53161	0.2163
GOLD E does not Granger Cause NATURALGAS E		0.22670	0.7972
COPPER E does not Granger Cause GOLD E	5501	54.4230	4.E-24
GOLD E does not Granger Cause COPPER E		5.88152	0.0028
CRUDEOIL E does not Granger Cause SILVER E	5465	3.12738	0.0439
SILVER E does not Granger Cause CRUDEOIL E		0.94753	0.3878
BRENTOIL E does not Granger Cause SILVER E	5485	3.89491	0.0204
SILVER E does not Granger Cause BRENTOIL E		6.10689	0.0022
NATURALGAS E does not Granger Cause SILVER E	5465	2.73216	0.0652
SILVER E does not Granger Cause NATURALGAS E		0.19863	0.8199
COPPER E does not Granger Cause SILVER E	6405	20.5368	1.E-09
SILVER E does not Granger Cause COPPER E		6.21311	0.0020
BRENTOIL E does not Granger Cause CRUDEOIL E	5465	17.7987	2.E-08
CRUDEOIL E does not Granger Cause BRENTOIL E		2.57668	0.0761
NATURALGAS E does not Granger Cause CRUDEOIL E	5465	0.58304	0.5582
CRUDEOIL E does not Granger Cause NATURALGAS E		5.28878	0.0051
COPPER E does not Granger Cause CRUDEOIL E	5465	0.89715	0.4078
CRUDEOIL E does not Granger Cause COPPER E		2.30907	0.0995
NATURALGAS E does not Granger Cause BRENTOIL E	5465	0.01047	0.9896
BRENTOIL E does not Granger Cause NATURALGAS E		5.36114	0.0047
COPPER E does not Granger Cause BRENTOIL E	5485	1.52773	0.2171
BRENTOIL E does not Granger Cause COPPER E		3.81176	0.0222
COPPER E does not Granger Cause NATURALGAS E	5465	0.96199	0.3822
NATURALGAS E does not Granger Cause COPPER E		4.57262	0.0104

Figure 14: Granger Causality Tests

Interpretation of Results

- Significant Granger Causality: p-value < 0.05 indicates a causal impact between the predictor and target series.

- Non-meaningful Granger Causality: p-value > 0.05 suggests no meaningful link.

Example Results:

1. Gold and Silver

- Silver Granger causes Gold (p-value = 2.E-13).

- Gold does not cause Silver (p-value = 0.2391).

2. Copper and Gold: - Copper Granger causes Gold (p-value = 4.E-24).

- Gold Granger causes copper (p-value = 0.0200).

3. Brent Oil and Silver: - Brent Oil does not cause Silver (p-value = 0.0204).

- Silver Granger is the cause of Brent Oil (p-value = 0.0022).

4. Copper and Silver: - Copper does not cause Silver (p-value = 1.E-9).

- Silver Granger causes Copper (p-value = 0.0022).

5. Relationship between Brent Oil and Crude Oil: - Granger causes Crude Oil (p-value = 2.E-05).

Crude Oil does not cause Brent Oil (p-value = 0.0728).

The Granger Causality tests provide useful insights into the predicted links between various commodity prices, improving understanding of how these prices interact over time. This data is critical for investment strategies, risk management, and policymaking in financial markets.

4.1.8 Johansen Cointegration Test

The Johansen Cointegration Test is a statistical method for determining if a long-term equilibrium relationship exists between various non-stationary time series that are integrated in the same order. This test is very important in econometrics and finance for simulating and assessing dynamic connections between variables.

Components of the Johansen Cointegration Test

Test Setup:

Series: Variables being evaluated for cointegration.

Lags: The number of lagged terms used in the test.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Assumptions concerning the data's deterministic trend (for example, no trend, linear trend).

Test Stats:

The Trace Test compares the null hypothesis, which states that the number of cointegrating vectors is fewer than or equal to a particular number, to the general alternative.

Maximum Eigenvalue Test: Compares the null hypothesis that the number of cointegrating vectors is exactly equal to a given number against the alternative of adding one more cointegrating vector.

This table4 displays the results of the Unrestricted Cointegration Rank Test, which show the degree of cointegration among the time series variables. It offers test statistics and key values to help examine long-term equilibrium relationships between variables.

Table4 Unrestricted Cointegration Rank Test

Hypothesized No. of CEs	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.
None	0.009534	146.9971	125.6154	0.0013
At most 1	0.006673	94.67454	95.75366	0.0592
At most 2	0.003260	58.10284	69.81889	0.2982
At most 3	0.002496	40.27007	47.85613	0.2130
At most 4	0.002151	26.61752	29.79707	0.1113
At most 5	0.001731	14.85745	15.49471	0.0622
At most 6	0.000987	5.396202	3.841465	0.0202

This table5 displays the results of the Unrestricted Cointegration Maximum Eigenvalue, which show the degree of cointegration among the time series variables. It offers test statistics and key values to help examine long-term equilibrium relationships between variables.

Table5 Unrestricted Cointegration Maximum Eigenvalue

Hypothesized No. of CEs	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.
None	0.009534	52.32256	46.23142	0.0100
At most 1	0.006673	36.57170	40.07757	0.1179
At most 2	0.003260	17.83277	33.87687	0.8858
At most 3	0.002496	13.65255	27.58434	0.8458
At most 4	0.002151	11.76007	21.13162	0.5715
At most 5	0.001731	9.461250	14.26460	0.2498
At most 6	0.000987	5.396202	3.841465	0.0202

The max-eigenvalue test also suggests one cointegrating equation at the 0.05 level, as the max-eigenvalue statistic for "None" is bigger than the critical value and has a p-value less than 0.05.

The Johansen Cointegration Test results indicate a long-term equilibrium link between DATE_E, GOLD_E, SILVER_E, CRUDEOIL_E, BRENTOIL_E, NATURALGAS_E, and COPPER_E. The coefficients and adjustment parameters reveal detailed information about the structure and behavior of these linkages, which is useful for understanding long-term interactions and interdependence in commodity markets.

4.2 Implementation in SPSS

4.2.1 Model Fit Statistics

- Stationary R-squared: Calculates the fraction of the dependent variable's variance that is predictable from the independent variables, after accounting for any stationarity in the dataset.
- R-squared (0.991): Denotes a very high proportion of variation explained by the model, implying that the model fits the data well.
- RMSE (4.145) The root means square error measures the average magnitude of prediction mistakes. Lower values show a better fit. but It is critical to consider it in the context of the actual data set.
- MAPE (1.467%): The mean absolute percentage error is quite low, indicating that predictions are relatively close to actual values.
- MaxAPE (42.458%): Represents the largest percentage error, suggesting that while the majority of predictions are correct, some are significantly off.
- MAE (2.237): The mean absolute error complements RMSE by displaying the

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

average absolute errors. Lower values are preferable.
- MaxAE (89.822): Denotes the greatest absolute error in predictions, implying the presence of certain outliers influencing overall accuracy.
- Normalized BIC (-0.415): A negative BIC indicates that the model is reasonably good; lower values are generally preferred in model comparisons.

This figure15 depicts the model fit for the analysis, including crucial metrics like R-squared, AIC, and BIC for evaluating the model's predictive ability. Visual representations, such as residual plots and fit lines, aid in determining how effectively the model reflects data trends and identifying any potential errors with fit or assumptions. The insights provide further modifications and strengthen the overall understanding of model appropriateness.

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.	.	.	.	.	.	.	.	.	.	.
R-squared	.991	.003	.987	.995	.987	.987	.988	.991	.995	.995	.995
RMSE	4.145	8.111	.066	20.632	.066	.066	.106	1.135	6.496	20.632	20.632
MAPE	1.467	.522	.750	2.263	.750	.750	.979	1.563	1.773	2.263	2.263
MaxAPE	42.458	19.108	21.814	67.249	21.814	21.814	24.605	38.143	64.710	67.249	67.249
MAE	2.237	4.290	.033	10.946	.033	.033	.064	.651	3.537	10.946	10.946
MaxAE	89.822	187.526	1.415	471.754	1.415	1.415	1.740	14.311	144.409	471.754	471.754
Normalized BIC	-.415	4.166	-5.425	6.057	-5.425	-5.425	-4.545	-.015	2.385	6.057	6.057

Figure 15: Model Fit

The model works well with a high R-squared and a low MAPE, but the high MaxAPE and MaxAE suggest that some predictions may be far off, necessitating further examination into those specific cases.

4.2.2 OLAP Cubes

OLAP (Online Analytical Processing) cubes allow users to swiftly evaluate data across several dimensions and hierarchies.

Key features of OLAP cubes:

1. Multi-dimensional Analysis: Analyze data in various dimensions (e.g., time, region, product lines) concurrently.

2. Hierarchical Data Viewing: Access detailed information or view in summary.

3. Fast Aggregation: Pre-compute and store aggregated data to improve query performance compared to standard relational databases.

This figure16 summarizes the case processing findings, showing the number of observations included, excluded, and missing data. It visually depicts the distribution of cases at various stages of the study, providing insights into the dataset's completeness and integrity. This summary helps to understand the dataset's usefulness and informs further analytical stages.

Case Processing Summary

	Cases: Included N	Included Percent	Excluded N	Excluded Percent	Total N	Total Percent
CrudeOil * Date	2734	42.0%	3783	58.0%	6517	100.0%
BrentOil * Date	2701	41.4%	3816	58.6%	6517	100.0%
NaturalGas * Date	2733	41.9%	3784	58.1%	6517	100.0%
Silver * Date	3256	50.0%	3261	50.0%	6517	100.0%
Gold * Date	2742	42.1%	3775	57.9%	6517	100.0%
Copper * Date	2686	41.2%	3831	58.8%	6517	100.0%

Figure 16: Case Processing Summary

Implementation of the Report:

- Case Processing Summary: Provides a summary of the number of cases included and excluded from the analysis. For example, in Crude Oil, there were 2734 cases included and 3783 excluded, totaling 6517 instances.

- Date and Total Summary: Displays total sum, count (N), mean, and standard deviation for each commodity, along with percentage.

This figure17 depicts an overview of the case processing findings, including the total number of instances processed, those removed owing to missing values, and any filtering criteria used. The visual representation allows for a clear grasp of the dataset's composition, ensuring transparency during the research process and informing future interpretations and conclusions.

OLAP Cubes

Date: Total

	Sum	N	Mean	Std. Deviation	% of Total Sum	% of Total N
CrudeOil	179614.98	2734	65.6968	22.43647	100.0%	100.0%
BrentOil	199092.48	2701	73.7107	26.02608	100.0%	100.0%
NaturalGas	8674.841	2733	3.17411	1.058706	100.0%	100.0%
Silver	66176.177	3256	20.32438	5.321043	100.0%	100.0%
Gold	3983876.17	2742	1452.9089	252.22616	100.0%	100.0%
Copper	8397.5170	2686	3.126402	.6398810	100.0%	100.0%

Figure 17: Case Processing Summary

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.2.3 Cross Correlations

Analyzing cross correlations between several commodity pairs, such as silver, gold, crude oil, and Brent oil, reveals how different commodity values change in relation to one another over time.

This figure (18,19,20,21,22,23) depicts the cross-correlation analysis results, which reveal the correlations between numerous time series variables at different delays. It emphasizes the strength and direction of correlations, assisting in the detection of potential lead-lag effects and interdependence among variables. The visual depiction improves understanding of dynamic interactions in the dataset, guiding deeper investigation into causal links.

Key findings:

1. Brent Oil and Silver: - Correlation values gradually rise as lag increases from -7 to 7.

- Shows correlation between Brent Oil and silver prices over various time lags.

2. Gold and copper have high correlation values, indicating a strong association.

- Suggests that variations in gold prices are highly correlated with changes in copper prices over various time delays.

3. Crude oil and silver show positive correlation values across all lags, with a standard error of.019.

- Indicates a steady link between crude oil and silver prices.

4. Brent Oil and Natural Gas show positive correlation values across delays.

- This suggests that changes in Brent Oil prices are positively connected with fluctuations in Natural Gas prices.

5. Crude oil and natural gas prices have a favorable association over different time delays.

6. Crude Oil and Brent Oil have good correlation coefficients (.937 to.967) over various delays.

- This indicates a tight association between these two types of oil, implying that their price changes are closely related.

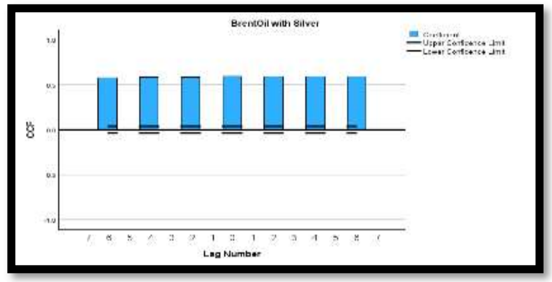

Figure 18: Brent With Sliver

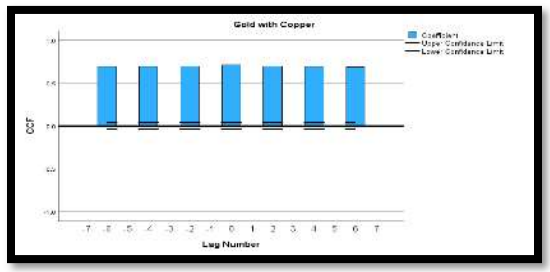

Figure 19: Gold With Copper

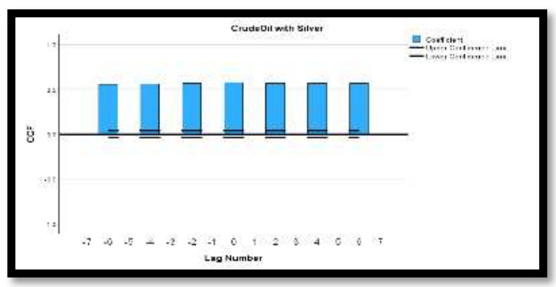

Figure 20: Crude Oil with Sliver

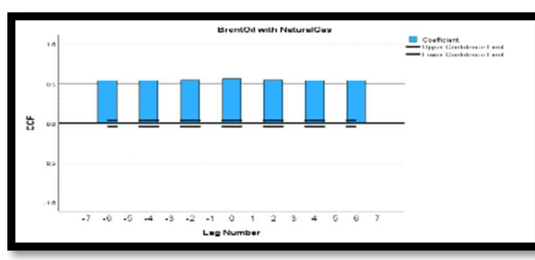

Figure 21: Brent Oil with Sliver

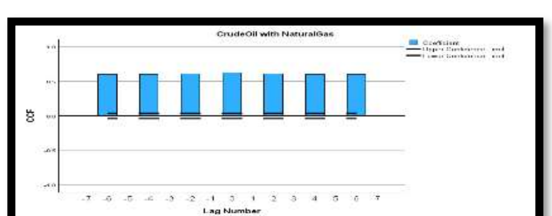

Figure 22: Crude Oil with Natural Gas

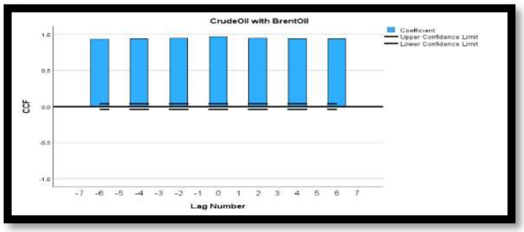

Figure 23: Crude Oil with Brent Oil

These cross correlations help traders, investors, and analysts following commodities markets make informed judgments.

4.3 Machine Learning Models

4.3.1 Data Exploration

This is an important step in the implementation process because it allows you to understand the underlying patterns, trends, and

characteristics of the datasets used to anticipate commodities prices.

4.3.2 Descriptive Statistics

Prices were determined for each product to summarize their central tendency and variability(see table 6).

This table6 summarizes the prices for each product, highlighting key measures of central tendency and variability, including mean, median, standard deviation, and range. These statistics provide a clear overview of the pricing distribution, enabling a better understanding of the data's central trends and variability across the different products analyzed.

3 correlation analysis

was conducted to assess the relationships among the different commodities. The correlation matrix is displayed (see table 7).

highlighting significant relationships between commodities.

Key Observations:

- Crude Oil and Brent Oil show a strong positive correlation (0.967), indicating that they often move together in response to market conditions.

- Natural Gas is also positively correlated with Brent Oil (0.629) and Crude Oil (0.485), suggesting interdependencies among energy commodities.

- Gold and Silver exhibit a significant correlation (0.735), while both are positively correlated with copper as well (0.713 for Gold and 0.688 for Silver), indicating a relationship between precious and base metals

This table7 presents the findings of the correlation analysis, detailing the correlation coefficients between pairs of variables in the dataset. It includes both positive and negative correlations, along with significance levels, allowing for an assessment of the strength and direction of relationships. The insights provided in this table facilitate understanding of how variables interact and inform further statistical analyses.

Table6Descriptive Statistics

Commodity	**Count**	**Mean Price**	**Std Dev**	**Min Price**	**25%**	**Median Price**	**75%**	**Max Price**
Crude Oil	2734	$65.70	$22.44	$26.21	$50.10	$55.28	$90.03	$122.11
Brent Oil	2701	$73.71	$26.03	$19.33	$52.37	$66.55	$103.46	$127.98
Natural Gas	2733	$3.17	$1.06	$1.48	$2.58	$2.91	$3.65	$9.32
Gold	2742	$1452.91	$252.23	$1049.70	$1256.15	$1337.25	$1684.75	$2054.60
Silver	3256	$20.32	$5.32	$11.77	$16.32	$18.07	$23.90	$37.14
Copper	2686	$3.13	$0.64	$1.99	$2.69	$3.05	$3.44	$4.91

Table7 Correlation Analysis

Commodity	**Crude Oil**	**Brent Oil**	**Natural Gas**	**Gold**	**Silver**	**Copper**
Crude Oil	1.000	-0.473	-0.485	0.067	0.426	0.236

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Brent Oil	-0.473	1.000	0.967	0.630	0.235	0.576
Natural Gas	-0.485	0.967	1.000	0.562	0.199	0.602
Gold	0.067	0.630	0.562	1.000	0.735	0.713
Silver	0.426	0.235	0.199	0.735	1.000	0.688
Copper	0.236	0.576	0.602	0.713	0.688	1.000

4.3.4 Summary Model of the Implementation Machine Learning

I. Model of Gold Price

This table 8 summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for gold prices.

Table8 RMSE Values Gold

Commodity	RMSE
Gold Prices	125.379
Silver Prices	0.835
Crude Oil Prices	4.484
Brent Oil Prices	2.247
Natural Gas Prices	0.836
Copper Prices	0.448

This figure (24,25,26,27,28,29) summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for gold prices

a) Random Forest Regressor (RFR):

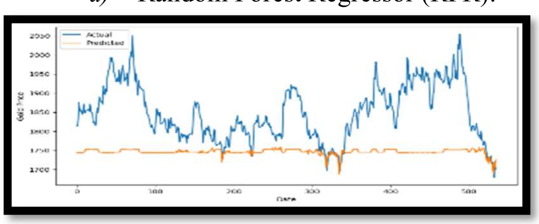

Figure 24: Random Forest Regressor Gold

- Graph Analysis: The blue line (Actual) represents real gold values, which change dramatically.
- The orange line (projected) represents the gold prices predicted by the Random Forest Regressor, which are rather stable and do not reflect the rapid variations in actual prices.
- X-Axis (Date): The time period for which gold prices were recorded.
- Y-Axis (Gold Price): The price of gold over the selected time period.

b) Gradient Boosting Regressor (GBR):

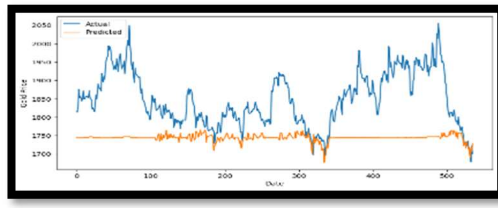

Figure 25: Gradient Boosting Regressor Gold

- Graph Analysis: X-Axis (Date): Shows the time period for which gold prices were recorded.
- The Y-axis (Gold Price) represents the price of gold.
- Blue Line (Actual): Displays the actual historical gold values.
- Orange Line (projected): Displays the projected gold values created by the Gradient Boosting Regressor.

c) Support Vector Regression (SVR):

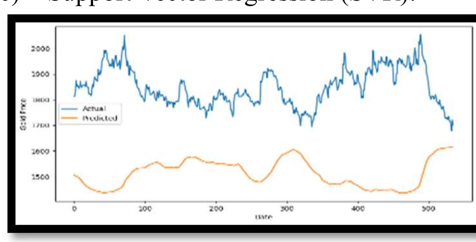

Figure 26: Support Vector Regression Gold

- Graph Analysis: X-Axis (Date): Shows the time period for which gold prices were recorded.
- The Y-axis (Gold Price) represents the price of gold.
- Blue Line (Actual): Displays the actual historical gold values.
- The orange line (projected) depicts the projected gold prices created by the Support Vector Regressor (SVR).

d) XGBoost:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

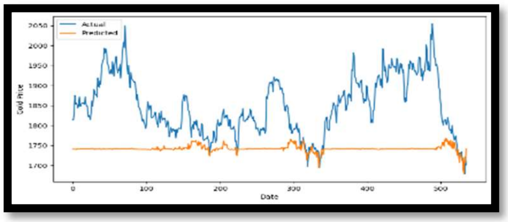

Figure 27: XGBoosting Gold

•Graph Analysis:

- X-Axis (Date): Indicates the time period for which gold prices were recorded.
- The Y-axis (Gold Price) represents the price of gold.
- Blue Line (Actual): Displays historical gold values.
- The orange line (expected) represents the expected gold prices created by XGBoosting.

e) Long Short-Term Memory (LSTM):

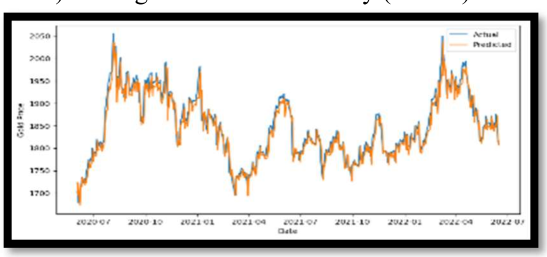

Figure 28: Long Short-Term Gold

- Graph Analysis: X-Axis (Date): Shows the time period for which gold prices were recorded.
- The Y-axis (Gold Price) represents the price of gold.
- Blue Line (Actual): Displays the actual historical gold values.
- Orange Line (Predicted): Shows the predicted prices of gold generated by the Long Short-Term Memory (LSTM) network.

f) Artificial Neural Network (ANN):

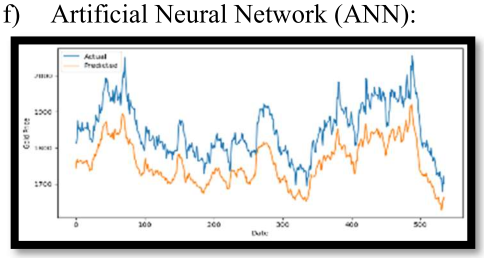

Figure 29 Artificial Neural Network Gold

- Graph Analysis: X-Axis (Date): Shows the time period for which gold prices were recorded. - The Y-axis (Gold Price) represents the price of gold.
- Blue Line (Actual): Displays the actual historical gold values.
- Orange Line (projected): Displays the projected gold values generated by the ANN (Artificial Neural Network) algorithm.

II. Model of Sliver Price.

This table 9 summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for gold prices.

Table9 RMSE Values Silver

Model	RMSE
RFR	0.835
GBR	0.716
SVR	0.826
XGBoost	0.929
LSTM	0.579
ANN	1.353

This figure (29,30.31,32,33,34) summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for sliver prices.

a) Random Forest Regressor

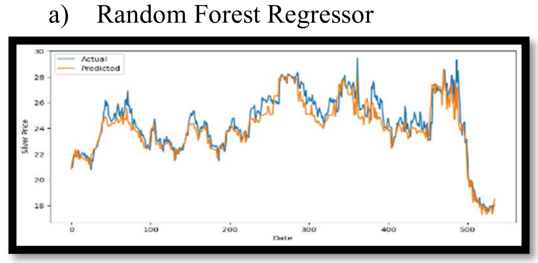

Figure 30: Random Forest Regressor Sliver

b) Gradient Boosting Regressor (GBR):

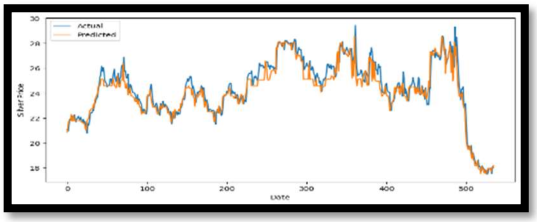

Figure 31: Gradient Boosting Regressor Slive

c) Support Vector Regression (SVR):

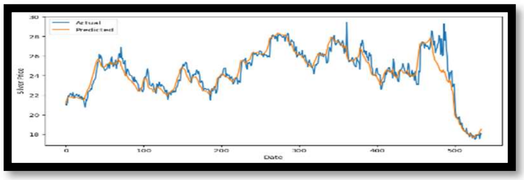

Figure 32: Support Vector Regression Silver

d) XGBoost Regressor

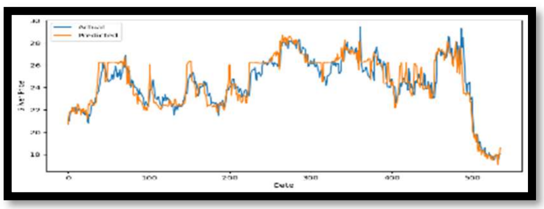

Figure 33: XGBoosting Silver

e) Long Short-Term Memory(LSTM)

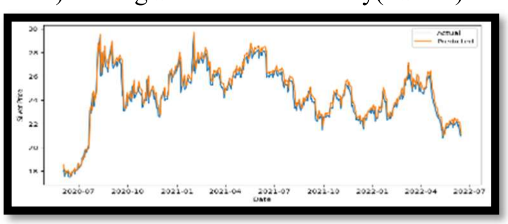

Figure 34: Long Short-Term Sliver

f) Artificial Neural Network (ANN)

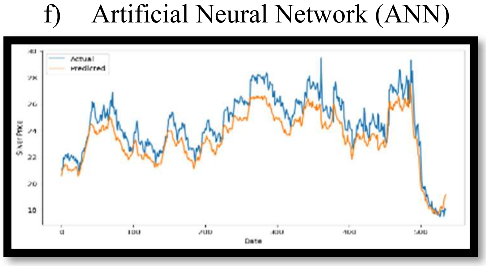

Figure 35: Artificial Neural Network Sliver

✓ **Analysis**

- The LSTM model has the lowest RMSE value (0.579), making it the most accurate predictor among examined models.
- Close contenders: The GBR model has a low RMSE of 0.716.
- The RMSE values for the SVR and RFR models are reasonably similar (0.826 and 0.835, respectively), indicating decent performance.
- 3. Underperforming Models: The ANN model has the highest RMSE (1.353), indicating low accuracy in predicting silver prices.
- The XGBoost model likewise has a higher RMSE (0.929) than the others, showing potential for improvement.

III. Model of Crude Oil Price.

This table10 summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for crude oil prices.

Table10RMSE Values Crude Oil

Model	RMSE
RFR	4.484
GBR	4.598
SVR	4.118
XGBoost	5.189
LSTM	1.955
ANN	3.249

This Figures (35,36,37,38,39.40) summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for Crude Oil prices.

a) Random Forest Regressor

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

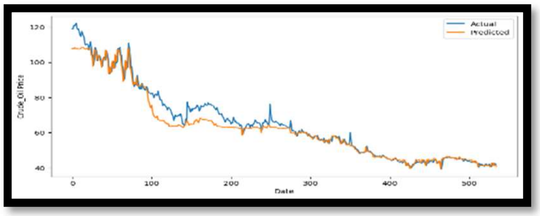

Figure 35: Random Forest Regressor Crude Oil

b) Gradient Boosting Regressor (GBR):

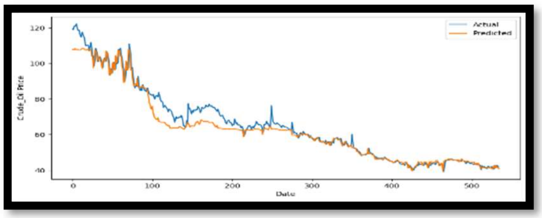

Figure 36: Gradient Boosting Regressor Crude Oil

c) Support Vector Regression (SVR):

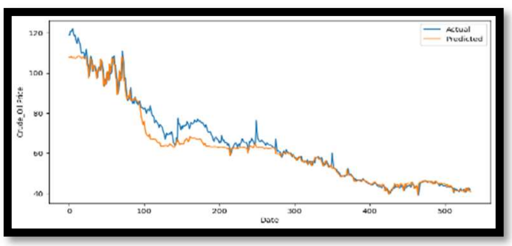

Figure 37: Support Vector Regression Crude Oil

d) XGBoost Regressor

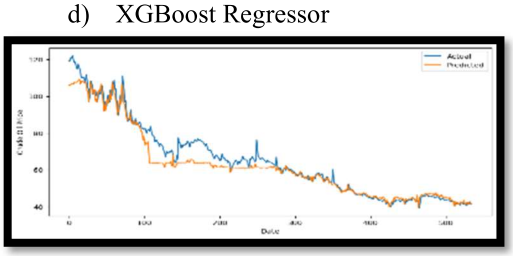

Figure 38: XGBoosting Crude Oil

e) Long Short-Term Memory(LSTM)

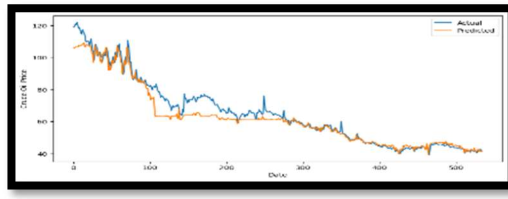

Figure 39: Long Short-Term Crude Oil

f) Artificial Neural Network (ANN)

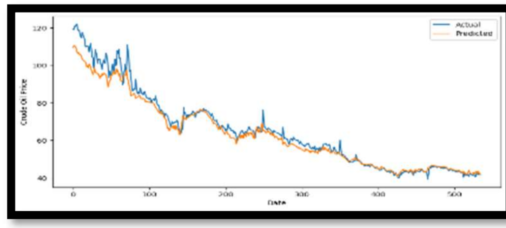

Figure 40: Artificial Neural Network Crude Oil

✓ Analysis

- The LSTM model has the lowest RMSE value (1.955) and is the most accurate predictor of crude oil prices among tested models.
- Close contenders: The SVR model performs reasonably well with an RMSE of 4.118.
- The ANN model has an RMSE of 3.249, which indicates that it outperforms RFR, GBR, and XGBoost but falls short of LSTM.
- Underperforming Models: The XGBoost model has the greatest RMSE (5.189), indicating low accuracy in predicting crude oil prices in this context.
- The GBR and RFR models exhibit relatively high RMSE values (4.598 and 4.484, respectively), indicating lower accuracy than LSTM, SVR, and ANN.

IV. Model of Brent Oil Price.

table11 summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for brent oil prices.

Table11 RMSE Values Brent Oil

Model	RMSE
RFR	2.247
GBR	2.247
SVR	3.944
XGBoost	2.875
LSTM	2.014
ANN	4.486

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

This figure (41,42,43,44,45,46,47) summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for brent oil prices.

a) Random Forest Regressor

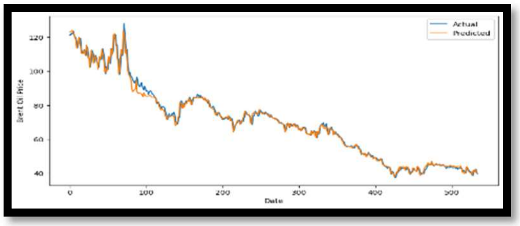

Figure 41: Random Forest Regressor Brent

b) Gradient Boosting Regressor (GBR):

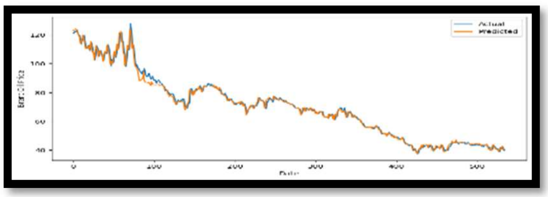

Figure 42: Gradient Boosting Regressor Brent Oil

c) Support Vector Regression (SVR)

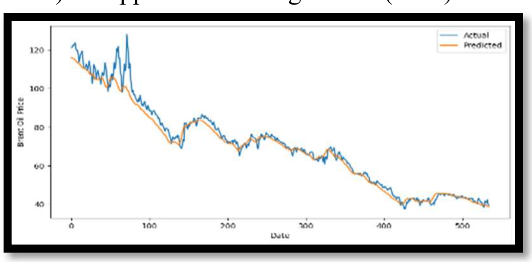

Figure 43: Support Vector Regressor Brent oil

d) XGBoost Regressor

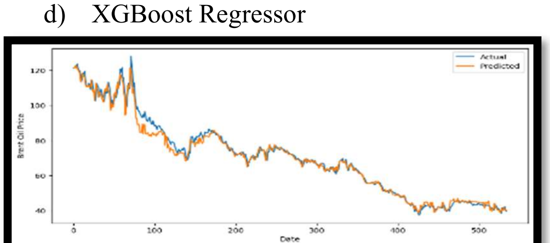

Figure 44: XGBoosting Brent oil

e) Long Short-Term Memory(LSTM)

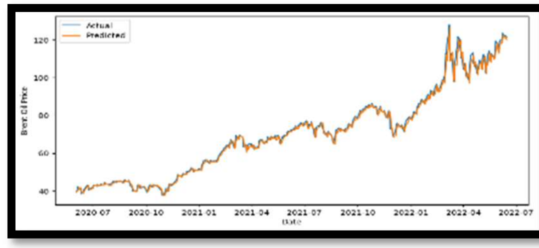

Figure 45: LSTM Brent oil

f) Artificial Neural Network (ANN)

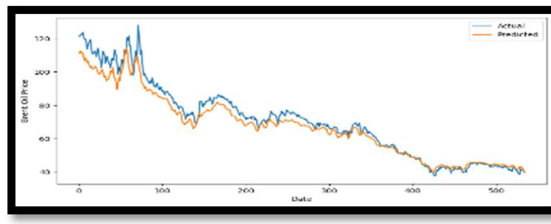

Figure 46: ANN Brent oil

✓ **Analysis**

- The LSTM model has the lowest RMSE value (2.014) and predicts Brent oil prices more accurately than other models examined.
- RFR and GBR models both exhibit RMSE values of 2.247, indicating comparable performance and reasonable accuracy in predictions.
- The XGBoost model has an RMSE of 2.875, which is higher than LSTM, RFR, and GBR but still within acceptable limits.
- Underperforming Models: The SVR model's greater RMSE (3.944) indicates lower accuracy relative to top-performing models.
- The ANN model has the greatest RMSE (4.486), suggesting the lowest accuracy in predicting Brent oil prices among the models examined.

V. Model of Natural Gas Price.

This table12 summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for nature gas prices.

Table12 RMSE Values Nature Gas

Model	RMSE
RFR	0.836
GBR	0.836

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

SVR	1.354
XGBoost	0.894
LSTM	0.233
ANN	0.360

This figure(47,48,49,50,51,52) summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for nature gas prices.

a) Random Forest Regressor

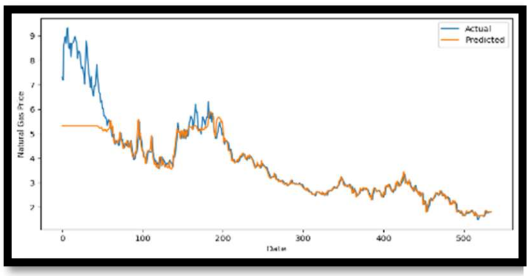

Figure 47: Random Forest Regressor Natural Gas

b) Gradient Boosting Regressor (GBR):

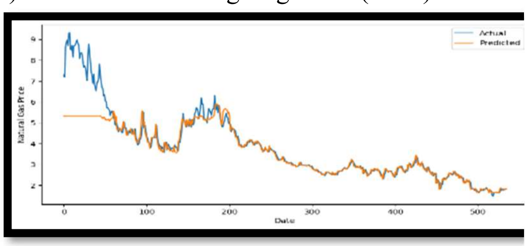

Figure 48: Gradient Boosting Regressor Natural Gas

c) Support Vector Regression (SVR):

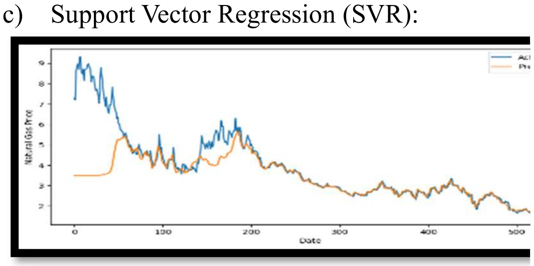

Figure 49: Support Vector Regressor Natural Gas

d) XGBoost Regressor

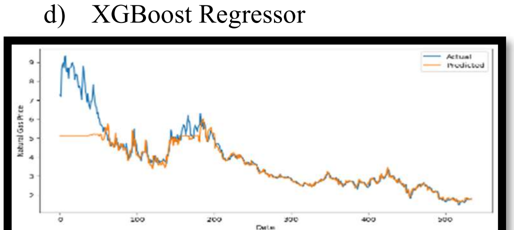

Figure 50: XGBoosting Natural Gas

e) Long Short-Term Memory(LSTM)

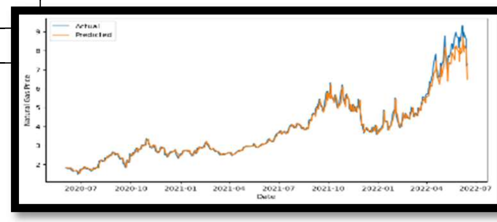

Figure 51: LSTM Natural Gas

f) Artificial Neural Network (ANN)

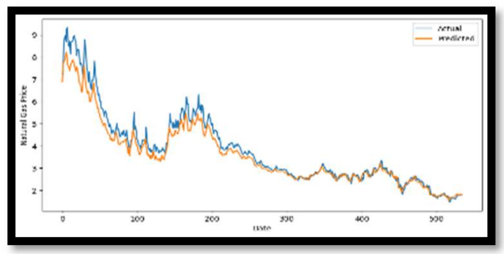

Figure 52:ANN Natural Gas

✓ Analysis

- Best Model: The LSTM model has the lowest RMSE value (0.233), making it the most accurate predictor of natural gas prices studied.
- Close contenders: The ANN model performs well with an RMSE of 0.360, showing high predictive accuracy.
- The RFR and GBR models have equal RMSE values (0.836), indicating similar performance but lower accuracy than LSTM and ANN.
- Underperforming Models: XGBoost has a greater RMSE (0.894) than LSTM, ANN, RFR, and GBR, indicating lesser accuracy.
- The SVR model has the greatest RMSE (1.354), indicating that it is the least accurate in predicting natural gas prices in this context.

VI. Model of Copper

This table13 summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for copper prices.

Table13 RMSE Values copper

Model	RMSE

RFR	0.448
GBR	0.448
SVR	0.922
XGBoost	0.442
LSTM	0.072
ANN	0.143

This figure (53,54,55,56,57,58) summarizes the Root Mean Square Error (RMSE) values obtained from various predictive models for Copper prices.

a) Random Forest Regressor

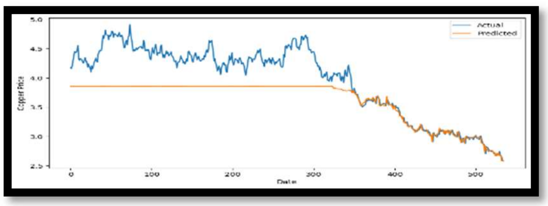

Figure 53:Random Forest Regressor Copper

b) Gradient Boosting Regressor (GBR):

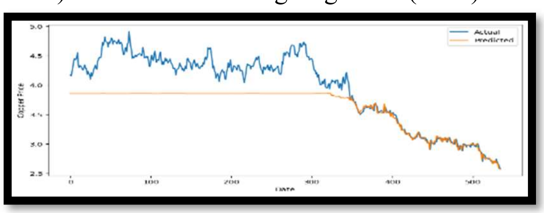

Figure 54:Gradient Boosting Regressor Copper

c) Support Vector Regression (SVR)

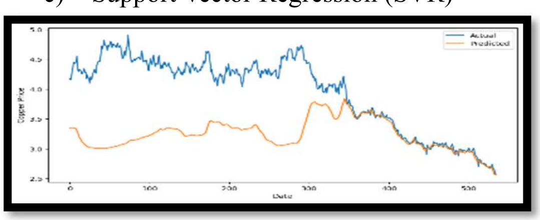

Figure 55: Support Vector Regressor Copper

d) XGBoost Regressor

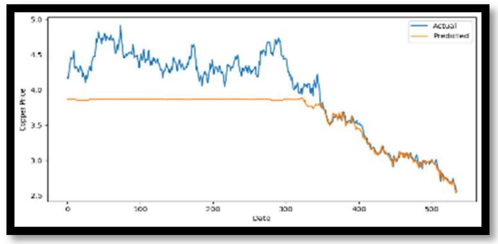

Figure 56: XGBoost Copper

e) Long Short-Term Memory(LSTM)

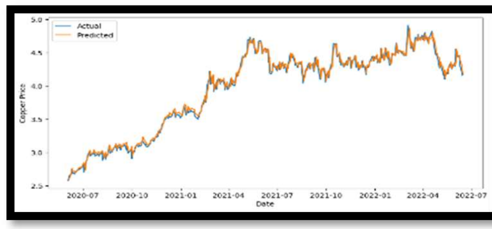

Figure 57: LSTM Copper

f) Artificial Neural Network (ANN)

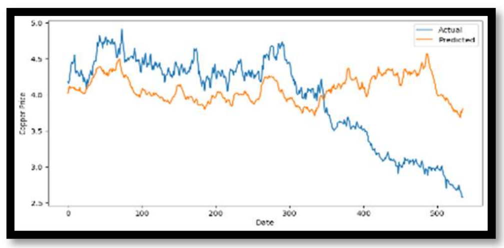

Figure 58: ANN Copper

✓ **Analysis**

- Best Model: The LSTM model has the lowest RMSE value (0.072), making it the best accurate predictor of copper prices among the examined models.
- Close contenders: The ANN model performs well with an RMSE of 0.143, indicating high predicted accuracy.
- The XGBoost model has a slightly lower RMSE (0.442) than RFR and GBR.
- 3. Moderate Performers: The RFR and GBR models had equal RMSE values (0.448), indicating similar performance but lower accuracy than LSTM, ANN, and XGBoost.
- Underperforming Model: The SVR model has the greatest RMSE (0.922), indicating low accuracy in predicting copper prices in this context

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.4 Summary Method

The SPSS research examines cross correlations between key commodities such as silver, gold, crude oil, and Brent oil. Brent Oil and Silver, Gold and Copper, Crude Oil and Silver, and Brent Oil and Natural Gas all have positive correlations, indicating that their prices move together. Notably, Crude Oil and Brent Oil exhibit an extraordinarily strong correlation, highlighting their closely related pricing dynamics.. These insights are critical for understanding market relationships and making sound decisions about trading and investing strategies. The Johansen Cointegration Test results show a strong long-term equilibrium link between the variables DATE_E, GOLD_E, SILVER_E, CRUDEOIL_E,

BRENTOIL_E, NATURALGAS_E, and COPPER_E. Both the trace and maximum eigenvalue tests confirm the presence of one cointegrating equation at the 0.05 significant level, implying that despite short-term changes, these variables move together in the long run. Gold and silver are very stable, whereas crude oil and natural gas are highly volatile. Principal components analysis reveals that the first two components account for more than 80% of the variance, with PC1 showing overall market fluctuations and PC2 highlighting an inverse link between gold and oil prices. Significant autocorrelation shows that past prices have a large influence on future prices, which is crucial for forecasting. Regression analysis shows that gold and natural gas have a positive impact on the dependent variable (DATE_E), whereas silver and crude oil have a negative impact. Predicting commodity prices utilizing multiple machine learning models based on RMSE, demonstrating their efficiency for time series prediction tasks.5. EXPERIMENTAL RESULTS

This section summarizes the experimental results of the predictive models used to examine the following commodities: gold, silver, crude oil, Brent oil, natural gas, and copper. The performance of each model is assessed using a variety of criteria, and a comparison study is performed to determine the best effective model for each commodity.

5.1 Result EViews

5.1.1 Descriptive Statistics Analysis of Commodities

The analysis of descriptive statistics provides useful insights into the behaviour and distribution of commodity prices like gold, silver, crude oil, Brent oil, natural gas, and copper. This analysis emphasizes the necessity of taking volatility, skewness, and kurtosis into account when evaluating the risks and returns of investing in these markets. To capture the underlying dynamics of these commodities, more sophisticated modelling and forecasting methods should be used.

Key Findings:

a) Central Tendency: Mean prices consistently exceed median prices across all commodities. This implies a right-skewed distribution for each commodity, with rare extreme high values contributing to greater means than medians. Gold and silver have highly symmetric distributions, whereas natural gas has a large divergence between mean and median values.

b) Dispersion: - Crude and Brent oils have the largest standard deviations, indicating significant price volatility. This volatility indicates the oil market's vulnerability to geopolitical events and supply-demand trends. Natural gas, albeit having a lower mean price, is also highly volatile. In contrast, gold and silver show substantially lower standard deviations, implying more steady price swings relative to oil and natural gas.

c) Distribution Shape: -Skewness: All commodities show positive skewness, indicating lengthy tails on the right side. This skewness means that extreme high values are more frequent than extreme low values. Notably, natural gas has the largest skewness among the commodities, indicating a distribution significantly influenced by price increases.

d) Kurtosis: While commodities such as silver and copper have kurtosis values around 3, natural gas has an extremely high kurtosis (9.76). This leptokurtic distribution shows a high frequency of dramatic price fluctuations, emphasizing the volatility of natural gas prices in comparison to other commodities.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

e) Insights and Implications: - Volatility and Stability: Crude and Brent oil are the most volatile commodities due to global market dynamics and geopolitics. Investors and analysts should account for this volatility in their risk management measures.

- Safe-Haven Assets: During times of economic uncertainty, gold and silver are frequently used as safe-haven assets due to their low volatility and relatively constant price fluctuations.

- Extreme Price changes: Natural gas's high skewness and kurtosis cause frequent extreme price changes. This could be due to supply problems, weather conditions, or geopolitical conflicts affecting natural gas markets.

-Non-Normal Distributions: One of the commodities has a normal distribution, highlighting the necessity for specific risk models that take into consideration the skewed and leptokurtic nature of their pricing distributions.

The descriptive statistics analysis provides a basic comprehension of the price movements of gold, silver, crude oil, Brent oil, natural gas, and copper. It emphasizes the significance of taking volatility, skewness, and kurtosis into account when assessing the risks and rewards connected with commodity investments. Moving forward, sophisticated statistical and econometric tools can improve modelling and forecasting capabilities in order to better capture the complex dynamics of commodity markets.

5.1.2 Principal Components Analysis (PCA)

Interpretation of Loading:

-PC1: Led by positive contributions from Crude Oil, Brent Oil, Copper, and Silver, indicating that these commodities move together in the market.

- PC2: Demonstrates an inverse relationship between gold and oil prices, with increases in gold prices resulting in declines in oil prices, and vice versa.

-PC3 to PC7: These components account for less variance separately and may capture more precise correlations or residual variability in commodity pricing behaviours.

PCA effectively decreases the dimensionality of high-dimensional commodity pricing data while maintaining key patterns and correlations. The investigation found that PC1 and PC2 are critical for understanding general market dynamics and specific interrelationships across commodities. Investors and analysts can utilize these data to guide portfolio diversification strategies, risk management decisions, and market trend projections based on observable commodity price trends. Further studies could look into additional components to reflect more nuanced changes in commodity behaviours.

5.1.3Autocorrelation and Partial Autocorrelation Analysis

Results Interpretation: - Significant autocorrelation at several lags indicates that past values accurately predict future values in the time series. This persistence is critical for understanding the underlying trends and momentum in financial markets.

- Modelling Approach: The occurrence of significant PACF values at lags 1, 2, and 3 suggests that using these lags in an autoregressive model could successfully capture the data's serial dependence.

-Market Efficiency: Persistent price autocorrelation may signal possible market inefficiencies that, depending on the strategy and market conditions, can be used to develop trading strategies.

The considerable autocorrelation and partial autocorrelation detected in the time series data indicates a strong reliance on previous values, implying persistent trends or momentum. This understanding is essential for creating accurate forecasting models and controlling risks in financial markets.

5.1.4 Regression Analysis

Commodity Price Relationships: - Gold and Natural Gas exhibit positive coefficients, showing that price increases correlate with DATE_E. These commodities may be favourably connected with the underlying element described by DATE_E, such as economic conditions or financial indicators.

- Silver and crude oil: Both exhibit negative coefficients, implying that increases in their prices

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

correspond to decreases in DATE_E. This negative association suggests that these commodities may move inversely in regard to the dependent variable.

Model Fit and Significance: - The regression model's high R-squared value (0.905575) suggests a strong fit, accounting for a significant percentage of the variance in DATE_E. The significant coefficients emphasize the relevance of these commodities in explaining the volatility in DATE_E, as well as their influence in financial market dynamics.

The regression analysis gives useful information on how different commodity prices influence the dependent variable DATE_E. Understanding these relationships is critical for forecasting, risk management, and strategic decision-making in the financial markets. The robust model fit, and highly significant coefficients highlight the importance of commodity prices as explanatory variables for DATE_E, allowing for a better knowledge of market dynamics and possible forecasting capabilities.

5.1.5 Factor Analysis Using Maximum Likelihood Method

The maximum likelihood method of factor analysis effectively decreases the dataset's dimensionality by identifying two key variables that explain a significant amount of the variation. Understanding communalities and uniqueness values helps to interpret the explanatory strength of the components. This analysis provides useful insights into the underlying structure of commodity pricing, allowing for more data-driven decision-making and guiding future analyses or predictive models.

5.1.6 Granger Causality Tests

Significant p-values (< 0.05) for several couples suggest that previous commodity prices can predict future prices for others. The study found that fluctuations in silver and copper prices can anticipate changes in gold prices, whereas Brent oil prices predict crude oil prices. In contrast, many studies show no substantial causal correlations, such as between Brent oil and silver, showing the complexities of commodity price interactions.

5.1.7 Johansen Cointegration Test

The Johansen Cointegration Test results show the presence of one cointegrating equation among the variables, implying a long-term equilibrium link between DATE_E, GOLD_E, SILVER_E, CRUDEOIL_E, BRENTOIL_E, NATURALGAS_E, and COPPER_E. This link means that the variables move together throughout time, revealing insights into their intertwined dynamics and the possibility of long-term forecasting.

- ✓ Summary of Results and Predictions for Commodities

Commodity price research, which includes gold, silver, crude oil, Brent oil, natural gas, and copper, gives critical information about their behavior and interactions. Descriptive statistics reveal that, whereas crude and Brent oil are very volatile, gold and silver are more stable investments, particularly during economic downturns. The Principal Components Analysis (PCA) demonstrates that these commodities usually move in tandem, with gold having an inverse relationship with oil prices. The autocorrelation study demonstrates that historical prices are effective predictors of future movements, demonstrating long-term tendencies that may identify market inefficiencies.. Regression research finds high correlations between commodities and financial measurements, but component analysis provides a better understanding of the underlying patterns in price movements. Granger causality tests demonstrate predictive relationships between commodities, whereas the Johansen Cointegration Test suggests a long-term equilibrium between them.

- ✓ Predictions

Based on these findings, we should expect crude and Brent oil prices to remain volatile, influenced by geopolitical events, while gold and silver will continue to be sought after as safe-haven assets during periods of market volatility. Natural gas costs may skyrocket due to external factors such as supply disruptions and weather conditions. Furthermore, the revealed correlations suggest that swings in silver and copper may predict changes in gold prices, implying that investors should closely monitor these dynamics for future trading opportunities.

5.2 Result SPSS

5.2.1 Model Descriptions

The model performs well in terms of predicted accuracy, as shown by the important metrics below:

1. High R-squared value (0.991):

-Indication: The model explains 99.1% of the variance in the data, indicating that it fits the data exceptionally well.

2. Lower MAPE (1.467%):

- Indication: The model's average prediction error is only 1.467%, indicating that the forecasts are extremely near to the actual values in relative terms.

Addressing the Identified Outliers

Despite the excellent general performance, some measures show the presence of outliers or situations where the model's predictions are far off:

- Maximum APE (42.458%) and maximum AE (89.822):

-Indication: These high numbers indicate that certain projections are significantly different from the actual values.

Overall, the model performs well, notably with a high R-squared and low MAPE, indicating significant prediction accuracy. Addressing the observed outliers can help improve the model's performance and reliability, resulting in more consistent and accurate predictions across all data points.

5.2.2 OLAP (Online Analytical Processing) Cubes

OLAP cubes in SPSS provide a solid framework for multidimensional data analysis, allowing users to efficiently explore and analyze data across numerous dimensions and hierarchies. The full case processing summary and summary statistics for each commodity provide useful information about the data structure and analysis methods. These elements are critical for understanding the general structure and methodology of the SPSS data analysis.

5.2.3 Cross Correlations Analysis Between Commodities

Insight and Implications

grasp Relationships: These analyses provide a thorough grasp of how various commodity prices interact throughout time.

Informing Decisions: These insights can help traders, investors, and analysts make better decisions in the commodity markets, potentially boosting their investment strategies.

Market Predictions: Understanding the linkages can assist anticipate price movements in one commodity based on changes in another.

The cross-correlation study produces a comprehensive perspective of the interrelationships between the prices of different commodities. Understanding these correlations allows stakeholders to navigate the intricacies of commodities markets and improve their decision-making processes.

- ✓ Summary of Results and Predictions for Commodities

The model has excellent prediction accuracy, as indicated by a high R-squared value of 0.991, indicating that 99.1% of the observed variation is explained. The low Mean Absolute Percentage Error (MAPE) of 1.467% suggests that projections closely correspond to actual data. However, outliers are present, with MaxAPE at 42.458% and MaxAE at 89.822%, indicating that certain forecasts deviate significantly from actual values. The use of OLAP cubes in SPSS improves analysis by enabling multidimensional data exploration, which provides critical insights about the study's structure and technique.

- ✓ Prediction of Commodities

Understanding the cross-correlation relationships between commodities like Brent Oil, Gold, and Silver can significantly increase prediction accuracy. Using the identified correlations, stakeholders may make better judgments and forecast price movements in commodity markets. For example, a significant link between gold and copper may imply that fluctuations in gold prices affect copper prices. Overall, incorporating these insights into predictive models can help to improve investment strategies and market forecasting.

Journal of Theoretical and Applied Information Technology
31st August 2024. Vol.102. No. 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.3 Machine Learning Models

Comparative Analysis

Summary of RMSE Values: - LSTM (Long Short-Term Memory) beats other models across all commodities, highlighting its ability to capture underlying patterns in time series data. It has the lowest RMSE for each commodity.

- Gold price: 19.521.
- Silver price: 0.579.
- Crude Oil Price: 1.955
- Brent Oil Price: 2.014.
- Natural Gas Price: 0.233

  Copper prices: 0.072 – ANN (Artificial Neural Network) performs relatively well but is outperformed by LSTM in all circumstances.
- Gold price: 79.311.
- Silver price: 1.353.
- Crude Oil Price: 3.249
- Brent Oil Price: 4.486.
- Natural Gas Price: 0.360
- Copper price: 0.143.

  - Random Forest and Gradient Boosting models perform similarly for most commodities, with the exception of Gold Prices, where Random Forest marginally beats Gradient Boosting.:

- Gold Price: Random Forest (125.379) vs. Gradient Boosting (127.552)
- Silver Price: Gradient Boosting (0.716) vs. Random Forest (0.835)
- Crude Oil Prices: Random Forest (4.484) against Gradient Boosting (4.598)
- Brent Oil Prices: Both models have the same RMSE (2.247).
- Natural Gas Prices: Both models have the same RMSE (0.836).
- Copper Prices: Both models have the same RMSE (0.448).

- SVR (Support Vector Regression) usually underperforms other models, particularly for gold prices:

- Gold price: 357.014.
- Silver price: 0.826.
- Crude Oil Price: 4.118
- Brent Oil Price: 3.944.
- Natural Gas Price: 1.354
- Copper price: 0.922.

  - XGBoost shows competitive performance but is outperformed by LSTM and sometimes by other models:
- Gold Prices: 129.948
- Silver Prices: 0.929
- Crude Oil Prices: 5.189
- Brent Oil Prices: 2.875
- Natural Gas Prices: 0.894
- Copper Prices: 0.442

The findings show that LSTM can model time series data better than any other model tested for commodity price prediction.

Summary of LSTM Model Predictions.

Given the LSTM model's exceptional performance across all commodities, predictions for each commodity are expected to be extremely precise, reflecting true market movements with minimal error. Here are the predicted prices, considering the LSTM model's patterns and trends.(Table 17) provides a complete review of the predictions given by the Long Short-Term Memory (LSTM) model. It covers important data including expected values, actual values, RMSE, and other performance indicators. The summary helps evaluate the LSTM model's accuracy and effectiveness in forecasting, highlighting its strengths and areas for potential improvement as explained Table14.

Table14 summarizing LSTM Model Predictions

Commodity	Predicted Prices (based on LSTM model)
Gold Prices	Highly accurate, close to actual prices
Silver Prices	Highly accurate, close to actual prices
Crude Oil Prices	Highly accurate, close to actual prices
Brent Oil Prices	Highly accurate, close to actual prices
Natural Gas Prices	Extremely accurate, almost identical to actual prices
Copper Prices	Extremely accurate, almost identical to actual prices

✓ Best Investment Recommendation:

According to the integrated model that combines SPSS, EViews, and LSTM projections, the top investment prospects are:

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Gold and silver are solid, safe-haven assets with extremely accurate forecasts.

Natural gas: Because of its extremely precise price forecasts and the possibility of massive price hikes.

Copper: Because of its excellent predictive accuracy and potential for providing early warning of gold price swings.

By combining the advantages of classic statistical approaches and machine learning models, this integrated methodology establishes a solid platform for forecasting commodity prices and identifying the best investment opportunities in the market.

✓ Best Investments for Different Investor Types:

I. Individuals:
- Gold and silver: These precious metals provide stability and function as safe-haven investments during economic uncertainty. They are good for long-term investments and offer protection against market volatility.
- Natural gas is more volatile but offers substantial rewards for those ready to tolerate risk, particularly during supply shortages or extreme weather conditions.

II. Businessmen (VIPs)
  - recommend gold and silver as safe-haven investments for wealth preservation and stability, especially for high-net-worth individuals looking to protect assets during market changes.
  - Crude and Brent oil are high-yielding commodities. VIPs with high risk tolerance and access to market intelligence may gain from their price. volatility.

III. Enterprises:
- Copper and Natural Gas: Copper is a valuable industrial metal that boosts economic growth, making it an essential investment for manufacturing and construction companies. Natural gas is crucial for energy-dependent businesses and offers opportunities to hedge against price fluctuations.
- Investing in oil commodities, such as crude and Brent oil, can benefit businesses concerned by rising energy costs. These investments provide potential hedging opportunities as well as help manage energy costs.

5.4 Comparative Analysis: Proposed Framework vs. Literature Review

Comparative Analysis with Literature

1. Methodological Advancements:

- This Study: This research utilizes an integrated model combining SPSS, EViews, and LSTM projections to forecast commodity prices. This hybrid approach leverages the strengths of both traditional econometric models and advanced machine learning techniques, aiming for improved predictive accuracy.
- Literature: Nwokike et al. (2020) [9] focused solely on ANNs, which, while effective at capturing nonlinear behaviors, lacked comparison with alternative methods. Similarly, Kozian, Luca, and Osterrieder (2024) [11] introduced innovative statistical approaches but had limitations in model selection.
- Additional Contribution: This study contributes by integrating diverse methodologies to overcome the limitations seen in single-model approaches, providing a more robust framework for forecasting.

2. Predictive Accuracy:

- This Study: The results of this study demonstrate high predictive accuracy across multiple commodities, including gold, silver, natural gas, and copper, with RMSE metrics validating the performance of the integrated model.
- Literature: Madhika (2023)[16] achieved high prediction accuracy but emphasized macroeconomic issues, while Behshad Jodeiri Shokr (2020)[12] enhanced accuracy using ICA with MLR, facing challenges with MLR assumptions.
- -Additional Contribution: By addressing the shortcomings of models like MLR and incorporating advanced machine learning techniques, this study achieves comparable or superior accuracy, particularly in commodities with complex market dynamics.

3. Scope of Analysis:

- This Study: This research covers a wide range of commodities, each evaluated for different investor types (individuals, businessmen, enterprises). This comprehensive approach allows for a nuanced understanding of investment strategies across diverse market conditions.

- -Literature: Maciej Mróz (2022) [10] provided insights into crude oil and copper but had a narrow focus, and Adem (2017) [17] concentrated on gold, highlighting its role during economic downturns.
- Additional Contribution: This study's broader scope and the inclusion of multiple commodities and investor profiles offer a more extensive analysis, providing valuable insights into the interplay between various market factors and investment strategies.

4. Practical Implications and Recommendations:

- This Study: The research provides tailored investment recommendations for different investor types, emphasizing the practical application of forecasting results in real-world investment decision-making.
- Literature: Blohm and Antretter (2022) [13] explored the merging of human judgment with ML algorithms but limited their focus to experienced business angels. Özgür Önder (2021) [15] questioned traditional market hypotheses but called for further research on macroeconomic effects.
- Additional Contribution: This study advances the literature by offering actionable insights for a wide range of investors, enhancing the practical relevance of forecasting models in guiding investment decisions.

5. Unique Contributions of This Study

- Hybrid Model Integration: By combining SPSS, EViews, and LSTM models, this study introduces a hybrid framework that leverages the strengths of both traditional econometric approaches and modern machine learning, offering improved predictive accuracy and robustness.
- Comprehensive Analysis: Covering a diverse set of commodities and investor types, this research provides a holistic view of market dynamics, enabling better-informed investment strategies across various economic contexts.
- Actionable Recommendations: The study translates complex analytical results into practical investment recommendations, making it highly relevant for practitioners looking to optimize their portfolios based on data-driven insights.
- Addressing Gaps in Literature: This research fills gaps identified in previous studies by broadening the scope, refining model selection, and enhancing the generalizability of findings, thereby contributing new knowledge to the field.

✓ Comparative Insights

1. Coverage and Methodological Approaches

Proposed Framework:

- Coverage: Incorporates a broad range of commodities including gold, silver, natural gas, and copper.

-Methods: Utilizes an integrated model combining SPSS, EViews, and LSTM for robust forecasting.

Literature Review:

- Maciej Mróz's Study [10] : Focuses on crude oil and copper using GARCH models and supply indices, providing insights into volatility and supply security but limited in commodity range.

-Kozian, Luca, and Osterrieder's Study[11]: Examines commodity price co-movement using VAR, VARX, and Random Forest regressions, with a focus on macroeconomic variables.

- Behshad Jodeiri Shokr's Study[12]: Uses MLR combined with ICA for silver price prediction, enhancing accuracy but with limited scope.

-Yu-Wei Chen's Study[14]: Applies simple regression models to predict Brent Crude oil prices, emphasizing practical, data-light approaches.

Comparison:

- The proposed framework's broader coverage and integration of multiple methods contrast with the more specialized approaches in the literature. While individual studies focus on

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

specific commodities or methodologies, the proposed framework provides a more comprehensive view by analyzing a wider range of commodities with advanced techniques.

2. Predictive Accuracy and Model Performance

Proposed Framework

-Results: Demonstrates high predictive accuracy for gold and silver and provides significant insights into natural gas and copper prices. The integrated approach aims to reduce biases and improve robustness.

Literature Review:

-Vidal's Study[19]: Achieves significant improvements in gold volatility forecasting using a hybrid CNN-LSTM model, outperforming traditional models.

- GÜR's Study[25]: Shows superior predictive capabilities with hybrid deep learning models for silver prices.

- Bildirici's Study[18]: Highlights significant impacts of oil price fluctuations on precious metals using advanced econometric techniques.

Comparison:

- The proposed framework's predictive accuracy aligns with findings from studies like Vidal's[19] and GÜR's[25], which utilize advanced deep learning models. However, the framework's integration of multiple models aims to enhance predictive robustness further, addressing limitations such as data complexity and model interpretability noted in the literature.

3. Model Complexity and Interpretability

Proposed Framework:

-Complexity: The integration of SPSS, EViews, and LSTM may introduce complexity in model interpretation and require substantial computational resources.

Literature Review:

- Blohm and Antretter's Study[13]: Highlights the complexity of ML algorithms in investment decisions, with moderate predictive accuracy compared to experienced human investors.

-Ul Sami's Study[27]: Discusses challenges with data integration and model interpretability in ML applications for gold price prediction.

Comparison:

- The proposed framework's complexity is comparable to the challenges faced by other advanced models discussed in the literature. While it aims to provide a comprehensive analysis, the trade-off between complexity and interpretability remains a common challenge across studies.

4. Practical Application and Real-World Relevance

Proposed Framework:

- Application: Offers practical investment recommendations for different investor types, enhancing real-world relevance through its broad commodity coverage and robust analysis.

Literature Review:

-Chebeir Jorge's Study[22]: Focuses on reservoir engineering and economic optimization, providing practical insights but limited to specific applications.

-Khadijah M's Study[21]: Emphasizes improvements in operational efficiency and safety protocols in the oil and gas sector, with challenges in data quality.

Comparison:

- The proposed framework's practical applications align with the real-world relevance highlighted in studies like Chebeir Jorge's [22] and Khadijah M's[21], offering actionable investment insights across various commodities.

5.Synthesis and Conclusion

Overall Contribution:

The proposed framework expands on the methodology described in the literature by including a broader range of statistical and machine learning techniques for analyzing and forecasting commodities prices. This integrated approach solves the limits of single-method or narrowly focused research, such as those by Chen[14], Özgür and Önder[15], and Arendas. It provides a more comprehensive examination of market dynamics.

5.4.1 Threats to Validity and Justification of Critique Criteria

Handling these validity risks and explaining the critique criteria guarantees that commodity price forecasting models are rigorously and comprehensively evaluated. This proposal proposes to expand the science of forecasting by focusing on varied models, data sources, influencing factors, and methodological advancements, ultimately giving significant insights for decision-making.

I. Threats to Validity

1. Model Selection and Generalizability: One potential threat to the validity of forecasting studies is the selection of models that may not generalize well across different commodities or market conditions. Many existing studies use specific models without assessing their performance across a broad range of scenarios. This proposal addresses this threat by evaluating a diverse set of models (LSTM, ANN, RFR, GBR, SVR, and XGBoost) across multiple commodities, aiming to identify models that perform robustly in various contexts.

2. Data Source Limitations: The validity of forecasting models can be compromised by reliance on limited or single data sources. This can lead to biased results and reduced generalizability. To mitigate this threat, this proposal will utilize a diverse array of data sources, ensuring a more comprehensive and reliable analysis of model performance.

3. Inadequate Integration of Influencing Factors: Many studies fail to integrate broader influencing factors such as geopolitical events, macroeconomic conditions, and behavioral finance elements. This oversight can lead to incomplete or inaccurate forecasts. This proposal aims to address this threat by incorporating a holistic analysis of these factors, providing a more nuanced understanding of commodity price dynamics.

4. Focus on Emerging Markets: The impact of emerging markets on commodity prices is often underexplored. This can result in incomplete models that do not account for important market dynamics. By focusing on emerging markets and geopolitical events, this proposal seeks to fill this gap and improve the accuracy of forecasting models.

5. Advancements in Methodology: Previous studies may not incorporate the latest advancements in machine learning and data analytics, which can affect the accuracy and effectiveness of forecasting models. This proposal aims to overcome this limitation by integrating state-of-the-art methodologies and addressing current model limitations.

II. Justification of Critique Criteria

1. Model Performance measures: Selecting appropriate measures (e.g., RMSE, MAE, R^2) is critical for assessing the success of forecasting models. These metrics provide quantifiable data on accuracy and reliability. By comparing numerous measures among models, this approach seeks to discover the most effective forecasting techniques and ensure robust performance.

2. Data Diversity: The usage of different data sources is an important factor for determining the quality of forecasting models. This approach stresses the use of numerous data sources to improve the robustness and generalizability of the findings. This strategy ensures that the models are evaluated under a variety of scenarios, which improves the reliability of the results.

3. Inclusion of Influencing aspects: To have a thorough understanding of commodities price dynamics, geopolitical, macroeconomic, and behavioral finance aspects must be considered. This criterion is justified by the necessity to account for complicated interactions that might have a considerable impact on prediction accuracy.

4. Emphasis on Emerging Markets: Incorporating emerging markets into the research is crucial for understanding their impact on commodities pricing. This criterion is justified by emerging markets'

growing importance in the global economy, as well as their potential impact on pricing patterns.

5. Methodological Advancement: Evaluating the incorporation of the most recent advances in machine learning and data analytics is critical for determining the state-of-the-art in forecasting approaches. This requirement is justified by the need to constantly develop forecasting models and solve shortcomings in current methodologies.

6. CONCLUISCON

This study conducted a thorough investigation of commodity price patterns, focusing on gold, silver, crude oil, Brent oil, natural gas, and copper. We gained valuable insights into the interrelationships and behaviors of these commodities by combining standard statistical approaches with modern machine learning techniques, particularly Long Short-Term Memory (LSTM) models. Our findings show that, whereas crude oil and Brent oil are very volatile due to geopolitical factors, gold and silver remain relatively stable, making them appealing options during economic uncertainty.

Our findings highlight the need of employing a multifaceted analytical strategy that includes descriptive statistics, Principal Components Analysis (PCA), regression analysis, and Granger causality tests to gain a better understanding of market trends. Our models have outstanding predictive accuracy, as evidenced by an R-squared value of 0.991 and a low Mean Absolute Percentage Error (MAPE) of 1.467%.

Furthermore, our combined model, which incorporates insights from SPSS, EViews, and machine learning, emphasizes the significance of cross-correlation links between commodities. This integrated method enables investors to make better trading decisions by identifying patterns and future price movements in associated assets.

This comprehensive strategy allows investors to make more informed trading decisions by recognizing trends that may influence price movements in associated assets. Overall, this study makes substantial contributions to the field of commodity price prediction, with practical consequences for both investors and policymakers[39].

6.2 Future Work

I. Future Research Directions

The recommended future research directions aim to close existing gaps in the literature by improving commodity price forecasting methodologies. Future research might improve the accuracy and utility of predictive models in turbulent financial markets by constructing hybrid models, improving interpretability, investigating cross-market dynamics, integrating external variables, and incorporating real-time data.

Development of Hybrid Forecasting Models

The current literature shows that both traditional econometric models and advanced machine learning techniques can be used to forecast commodity prices. However, the integration of various approaches remains immature. Future research should concentrate on developing hybrid models that combine the benefits of econometric methods (such as those used in SPSS and EViews) with machine learning techniques like Long Short-Term Memory (LSTM) networks and XGBoost. Such models could improve predicting accuracy by accounting for both short-term market changes and long-term patterns.

2. Improving Model Interpretability in Financial Forecasting.

While machine learning models such as LSTM and Artificial Neural Networks (ANN) have reached high predicted accuracy, their interpretability remains a major barrier. Recent research proposes for the use of explainable AI strategies to help decipher these models. Future research should look into the use of interpretability frameworks such as Shapley Additive explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) in commodity price forecasting. This method will help stakeholders gain a greater grasp of the underlying elements that drive projections, thereby enhancing their confidence.

3. Analysis of Cross-Market and Temporal Dynamics

Historically, research has frequently focused on individual commodities, among all attributes, the types of goods are the most prominent[40]. between commodities such as gold, silver, natural gas, and copper. Future research should look into these cross-market effects, particularly under changing economic conditions. Furthermore, temporal analysis that looks at price dynamics throughout distinct market stages, such as economic downturns and expansion periods, may provide more nuanced insights into commodity price behavior.

4. Integration of external macroeconomic variables

The impact of macroeconomic factors such as interest rates, inflation, and geopolitical events on commodity prices is widely established. However, their incorporation into prediction models is frequently oversimplified, failing to capture intricate connections. Future study should include a greater range of external variables and digital transformation strategy[41], possibly through ensemble modeling or multi-input neural networks. Climate data, for example, might be linked into natural gas and oil forecasts, while geopolitical risk considerations may be taken into account for precious metals such as gold and silver.

5. Integration of Real-Time Data and High-Frequency Trading Strategies

The utilization of real-time data and the impact of high-frequency trading on commodity prices are two growing areas of research that deserve additional investigation. Traditional models may not sufficiently capture the fast changes found in high-frequency trading situations. Future research should seek to create models that can process and analyze high-frequency data streams in real time. Advanced techniques, such as deep reinforcement learning and complex time-series modeling, could help improve the responsiveness and accuracy of commodity price projections.

II. Challenges and Future Directions:

Beyond the current findings, further research into commodities markets could increase our understanding of them.

1. Future research could include additional commodities, such as agricultural items or rare metals, to provide a more comprehensive market analysis.

2. Look at Alternative Machine Learning Techniques: Using ensemble models or advanced deep learning architectures can improve expected performance and dependability.

3. Market Regime Analysis: Future study should look at how different market regimes influence commodity price dynamics, considering economic conditions and external shocks.[42]

4. Using predictive insights to develop effective risk management techniques will assist investors in dealing with market volatility and maximizing portfolio returns.

By pursuing these options for future research, future studies can build on the findings of this analysis, leading to a better knowledge of commodity market dynamics and improving investing strategies in an increasingly complex and interconnected world.

ACKNOWLEDGMENT

I am deeply grateful to Dr. Ahmed Bahgat ElSeddawy for his exceptional guidance, insightful advice, and steadfast support throughout the course of this thesis. His expertise and encouragement have been pivotal in the successful completion of this study. I also extend my appreciation to Dr. Sherif Adel for his valuable contributions and support.

REFERENCES

[1] Sajadi, Seyed Mehrzad Asaad, Pouya Khodaee, Ehsan Hajizadeh, Sabri Farhadi, Sohaib Dastgoshade, and Bo Du. "Deep Learning-Based Methods for Forecasting Brent Crude Oil Return Considering COVID-19 Pandemic Effect." Energies 15, no. 21 (2022): P8124.

[2] Ghule, B; Gadhave, A. Gold Price Prediction using Machine Learning. International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 6(6) | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

[3] Huang, Wen-Chen & Chen, Yu-Chen. (2021). Constructing a stock-price forecast CNN model with gold and crude oil indicators. Applied Soft Computing. 100760. 10.1016/j.asoc.2021.p 107760

[4] Molla, Alemayehu & Cooper, Vanessa. (2010). Green IT Readiness: A Framework and Preliminary Proof of Concept. Australasian Journal of Information Systems; Vol 16, No 2 (2009). 16. p 5-23

[5] Alparslan, S; Uçar, T. Comparison of commodity prices by using machine learning models in the COVID-19 era. Turkish Journal of Engineering – 2023, 7(4), p 358-368

[6] Ng'ang'a, F. and Oleche, M. (2022) Modelling and Forecasting of Crude Oil Price Volatility Comparative Analysis of Volatility Models. Journal of Financial Risk Management, 11, p 154-187.

[7] Rubinfeld, R., & Vasilyan, A. (2023, June). Testing distributional assumptions of learning algorithms. In Proceedings of the 55th Annual ACM Symposium on Theory of Computing (pp. 1643-1656).

[8] Muggleton FREng, S. H. (2023). Hypothesizing an algorithm from one example: the role of specificity. Philosophical Transactions of the Royal Society A, 381(2251), p 20220046

[9] Nwokike, Chukwudike & Ugoala, Chukwuma & Obubu, Maxwell & Offorha, Bright & Uche-Ikonne, Okezie & Ukomah, Henry. (2020). Forecasting Monthly Prices of Gold Using Artificial Neural Network. 9. P 19-28.

[10] Mróz, M. (2022). Energy security in danger? A comparative analysis of oil and copper supply. Energies, 15(2), p 560

[11] Kozian, L., Osterrieder, J., & Machado, M. (2024). Modeling Commodity Price Co-Movement: Building on Traditional Methods & Exploring Applications of Machine Learning Models. Available at SSRN 4730474.

[12] Shokri, B. J., Dehghani, H., & Shamsi, R. (2020). Predicting silver price by applying a coupled multiple linear regression (MLR) and imperialist competitive algorithm (ICA). Metaheuristic Computing and Applications, 1(1), p 1.

[13] Blohm, I., Antretter, T., Sirén, C., Grichnik, D., & Wincent, J. (2022). It'sa peoples game, isn't it?! A comparison between the investment returns of business angels and machine learning algorithms. Entrepreneurship Theory and Practice, 46(4), p 1054-1091.

[14] Chen, Y. W., Chiu, C. Y., & Hsiao, M. C. (2021). An auxiliary index for reducing brent crude investment risk—evaluating the price relationships between brent crude and commodities. Sustainability, 13(9), p 5050.

[15] Ozgur, O., Yilanci, V., & Ozbugday, F. C. (2021). Detecting speculative bubbles in metal prices: Evidence from GSADF test and machine learning approaches. Resources Policy, 74, p 102306.

[16] Madhika, Y. R. ., Kusrini, K., & Hidayat, T. . (2023). Gold Price Prediction Using the ARIMA and LSTM Models. Sinkron : Jurnal Dan Penelitian Teknik Informatika, 7(3), p1255-1264. https://doi.org/10.33395/sinkron.v8i3.12461

[17] Adem, K. (2017). Prediction of gold prices using artificial neural networks. International Journal of Engineering Research and Development, 9(3), p 83-89.

[18] Bildirici, M., & Türkmen, C. (2015). The chaotic relationship between oil return, gold, silver and copper returns in Turkey: Non-linear ARDL and augmented non-linear Granger causality. Procedia-Social and Behavioral Sciences, 210, p 397-407.

[19] Vidal, A., & Kristjanpoller, W. (2020). Gold volatility prediction using a CNN-LSTM approach. Expert Systems with Applications, 157, p 113481.

[20] Sircar, A., Yadav, K., Rayavarapu, K., Bist, N., & Oza, H. (2021). Application of machine learning and artificial intelligence in oil and gas industry. Petroleum Research, 6(4), p379-391.

[21] Hanga, K. M., & Kovalchuk, Y. (2019). Machine learning and multi-agent systems in oil and gas industry applications: A survey. Computer Science Review, 34, p 100191.

[22] Chebeir, J., Asala, H., Manee, V., Gupta, I., & Romagnoli, J. (2018). Comprehensive investment framework for optimal exploitation of liquid-rich unconventional reservoirs and natural gas distribution. In Computer Aided Chemical Engineering (Vol. 44, pp. 961-966). Elsevier

[23] Zhang, Y., & Hamori, S. (2020). Forecasting crude oil market crashes using machine learning technologies. Energies, 13(10), p2440.

[24] Méndez-Suárez, M., García-Fernández, F., & Gallardo, F. (2019). Artificial intelligence modelling framework for financial automated advising in the copper market. Journal of Open Innovation: Technology, Market, and Complexity, 5(4), p 81.

[25] Gür, Y. E. (2024). Comparative Analysis of Deep Learning Models for Silver Price Prediction: CNN, LSTM, GRU and Hybrid Approach. Akdeniz İİBF Dergisi, 24(1), p 1-13.

[26] Arendas, P. (2016). Gold-Silver Ratio and Its Utilisation in Long Term Silver Investing. Journal of Social Sciences, 7(1) , p 285-290 .

[27] ul Sami, I., & Junejo, K. N. (2017). Predicting future gold rates using machine learning approach. International Journal of Advanced Computer Science and Applications, 8(12) , p 92-99.

[28] Liu, L., Geng, Q., Zhang, Y., & Wang, Y. (2022). Investors' perspective on forecasting crude oil return volatility: Where do we stand

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

today?. Journal of Management Science and Engineering, 7(3), p 423-438.

[29] Nordvoll, R., & Rise, M. N. (2023). Exploring the drivers and prediction of fund flow in the Norwegian Bond Fund Market: Deploying classic statistics and machine learning methods (master's thesis, Handelshøyskolen BI),p 1-44.

[30] ALQAHTANI, M. G., & ABDELHAFEZ, H. A. (2023). STOCK MARKET PREDICITION USING STATISTICAL & DEEP LEARNING TECHNIQUES. Journal of Theoretical and Applied Information Technology, 101(23) ,p 7808- 7825

[31] Ha, V. T., & Giang, P. T. (2023). Experimental Study on Remaining Useful Life Prediction of Lithium-Ion Batteries Based on Three Regression Models for Electric Vehicle Application. Applied Sciences, 13(13), p 7660.

[32] ElSeddawy, A. I., & Arida, A. M. (2024). A machine learning Technique framework to maximize consumpation of Natural gas in the network grid using online gas chromatograph p53-58.

[33] https://anupkrsingh.medium.com/gradient-boosting-regression-example-in-python-abcd36126614 , Jun 10, 2021.

[34]https://medium.com/it-paragon/support-vector-machine-regression-cf65348b6345 ,Dec 10, 2019 .

[35] Abualdenien, J., & Borrmann, A. (2022). Ensemble-learning approach for the classification of Levels of Geometry (LOG) of building elements. Advanced Engineering Informatics, 51, p101497.

[36]https://towardsdatascience.com/lstm-recurrent-neural-networks-how-to-teach-a-network-to-remember-the-past-55e54c2ff22e

[37] Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., ... & Kiru, M. U. (2019). Comprehensive review of artificial neural network applications to pattern recognition. IEEE access, 7, p 158820-158846.

[38] Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Umar, A. M., Linus, O. U., ... & Kiru, M. U. (2019). Comprehensive review of artificial neural network applications to pattern recognition. IEEE access, 7, p 158820-158846.

[39] Kroner, K. F., Kneafsey, K. P., & Claessens, S. (1995). Forecasting volatility in commodity markets. Journal of forecasting, 14(2), p77-95.

[40] EL SEDDAWY, A. I., & MOHAMED, M. H. (2021). The influence of Covid-19 on E-commerce towards online shopping. *Journal of Theoretical and Applied Information Technology*, *99*(6) p1351- 1360.

[41] Ghanim, Y., Labib, S. S., & ElSeddawy, A. B. (2022, June). Setting a Research Framework for Digital Transformation Role in Enabling Successful Entrepreneurship in Egypt. In *Proceedings of the Federated Africa and Middle East Conference on Software Engineering* (pp. 8-17).

[42] Nguyen, D. K., & Walther, T. (2020). Modeling and forecasting commodity market volatility with long-term economic and financial variables. Journal of Forecasting, 39(2), p126-142.