International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 12, Issue 4, July-August 2025

Impact Factor: 8.152

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204065

Multivariate Machine Learning Models for Gold Price Forecasting

Abhishek Dodwad, Adarsh Ballur, Agastya T, Gomathi Thiyagarajan

Department of MCA, CMR Institute of Technology, Bengaluru, India

Department of MCA, CMR Institute of Technology, Bengaluru, India

Department of MCA, CMR Institute of Technology, Bengaluru, India

Department of MCA, CMR Institute of Technology, Bengaluru, India

ABSTRACT: Gold, being a world known safe-haven investment has very volatile price behavior that is affected by various economics indicator results and market dynamics. Proper prediction in the gold prices is a necessity to the investors, policy makers and financial institutions. This paper is a machine learning approach to regressing previous gold prices based on historical data and macro economic indicators including the S&P 500 (SPX), silver prices (SLV), oil prices (USO) and EUR/USD exchange rates. Multiple regression models and benchmarks such as: Linear Regression, Decision Tree, Random Forest, Support Vector Regressor (SVR), XGBoost, and Long Short-Term Memory (LSTM) neural networks are used and compared in the project. Preprocessing of data, feature engineering, and hyperparameter optimization were done to make the model more effective. The models were validated using evaluation measures like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R 2 Score. The outcomes show that ensemble methods such as Random Forest and deep learning such as LSTM perform more accurately as compared to traditional when used in the prediction. The results show that time-series analytics combined with machine learning are effective to create reliable and scalable forecasts of gold prices.

KEYWORDS: Gold price prediction, machine learning, regression analysis, LSTM, Random Forest, time-series forecasting, financial analytics, economic indicators, ensemble models, predictive modeling.

I. INTRODUCTION

Gold has a long history of playing the role of store of value and protection against economic uncertainties. The factors that affect its price are diverse such as the mood on the market, geopolitics and currency fluctuations, inflation rates and global economic conditions. Due to such a change, the issue of predicting the prices of gold is dynamic and complex which needs a robust analytical tool.

Older statistical methods like ARIMA and linear regression have dominated the field of time-series forecasting, but have failed in these conditions, especially when there are non-linear patterns present, and marked changes in the market. The introduction of machine learning (ML) has created new possibilities in terms of more precise and adaptive forecasting models which can derive their results on the base of historical data and establish more complex connections between the financial indicators.

The project will be based on the creation and analysis of a data-based gold price forecast engine based on a number of machine learning regression models. In these models, there is an abundance of data in terms of gold prices and related financial numbers of S&P 500 index (SPX), oil prices (USO), silver prices (SLV), and EUR/USD exchange rates. The use of data preprocessing, feature engineering, model training, and hyperparameter tuning are dependent on a well-organized structure of working procedures to guarantee accuracy, and scale.

The major hypotheses of the research are:
1. Estimating and comparing several regressions of gold prices prediction.
2. Analyzing model accuracy in terms on standard metrics such as MAE, RMSE and R2 score.
3. That is, determining the most significant factors that can influence movements on prices of gold.
4. The ability to give visualization to the interpretation of the model outputs and predict the trends.

Through the integration of the traditional and deep learning models such as ensemble models and LSTM neural networks, it is hoped to provide an all-encompassing and practical approach to the real-time forecasting of the gold price. Research findings of this research can guide investors, financial analysts, as well as policymakers to make informed decisions due to market fluctuations.

II. RELATED WORKS

Gold price prediction is field research where the economic vitality and volatility of the markets have been studied. To address this challenge, a huge variety of conventional statistical and contemporary machine learning (ML) approaches have been applied by the researchers.

Time-series techniques like ARIMA, GARCH, and Exponential Smoothing are the early examples with the volatility of gold prices being modeled by Kristjanpoller and Minutolo [1] where they observed constraints on adaptation of structural breaks and price jumps by the GARCH-family models.

With the rise of popularity of machine learning, Shankar and Bhavani [2] have used Support Vector Regression (SVR) to use nonlinearities in gold price behaviors and included oil and exchange rates indicators to obtain even more accurate results. On the same note, both Random Forests and Gradient Boosting Machines (GBM) were exemplified to succeed basic models in a study conducted by Patel et al. [3] particularly when utilizing the multi-source data trained models.

Kumar and Rathore [4] delved further into the field of deep learning and applied the LSTM networks to reflect long-term dependencies on financial time-samples. They observed that LSTM made better R2 and RMSE levels than SVR and ARIMA.

There has also been the emergence of hybrid models that incorporate the traditional architectures with neural ones. To illustrate, Tan et al. [7] created a CNN-LSTM hybrid to utilize both spatiality patterns in financial KPIs and sequential memory in price changes. One advantage, which was demonstrated in this model was better forecasting performance under high volatility conditions.

Zhang et al. [5] provided evidence on the significance of feature selection and economic indicators, as they used XGBoost and added macroeconomic features, such as interest rates, CPI, and unemployment. When compared to the estimate that uses price data, the study elucidated an increment of RMSE by 10-15 percent.

The other contribution is the interpretation of model decisions by using explainable AI methodologies, including SHAP values. Verma and Ghosh [6] utilized Sharp and Random Forest to decipher the effect of silver prices and SPX on gold price predictions, which facilitated the trusting nature of the model results by financial analysts.
Also, other studies such as that of Awan, et al. [8] used big data processing and Hadoop pipeline to scale gold price forecasting systems on real time data-feeds and Spark ML models demonstrating the viability of implementing ML at production scale.

All of these works together create a solid background referring to gold price analytics and underline such crucial issues as data stationarity, explainability of the models, and long-range accuracy in predictions. We extend this to a full-stack pipeline that also encompasses preprocessing and deployment of the model and benchmark conventional and neural models on basis of the same evaluation percentages.

III. PROPOSED SYSTEM AND METHODOLOGY

The system under consideration would provide a precise prediction of the gold rates because it uses financial indicators, preprocessing time-series data, and numerous methods of machine learning. This framework combines traditional regressors and deep learning networks to deal with the issue of market volatility and complexity of features by observing both the linear and nonlinear patterns in movement of the gold prices.

A. Architektur

The suggested system is aimed at projecting gold prices based on the coherent machine learning pipeline combining various monetary indicators and models. The working process will start with the collection of the reliable data, which in this case is provided by the Web resource Kaggle and Yahoo Finance, and includes the daily markings of the gold prices and the essential market indicators, such as silver prices (SLV) or the oil prices (USO), the S&P 500 index (SPX), and the EUR/USD exchange rates [1]. The pre-processing of the collected data includes missing value

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204065

assignment, which is assigned by interpolating and normalizing by Min-Max scaling that puts together the uniformity of scale [2]. The temporal aspects are the features that are constructed, creating lagged variables, moving averages and measure of volatility to better identify patterns [3].

It then trains and tests the various machine learning models such as Linear Regression, Support Vector Regression (SVR), Random Forest, XGBoost and Long Short-Term Memory (LSTM) neural networks. Standard regression metrics are used to calculate the model performance: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R 2 score. Traditional models use GridSearchCV to optimize hyperparameters and manual tuning to optimize hyperparameters in LSTM [4]. Compared to other tested models, ensemble models such as those of Random Forest and deep learning models suchlike LSTM exhibited better results with accuracy of detecting complex market behavior [5].

Visualization plays a key role in the methodology. Actual vs. predicted plots, correlation heatmaps, and SHAP-based feature importance graphs are generated to interpret model behavior and decision factors [5]. The best-performing model is deployed for short-term forecasting using a 7-day rolling prediction setup. A system architecture diagram (Figure 1) illustrates this workflow, highlighting each stage from data ingestion to model deployment. This integrated approach not only builds on prior studies [3][6] but also improves robustness by combining interpretability with predictive power, making it a valuable tool for investors and financial analysts alike [7].

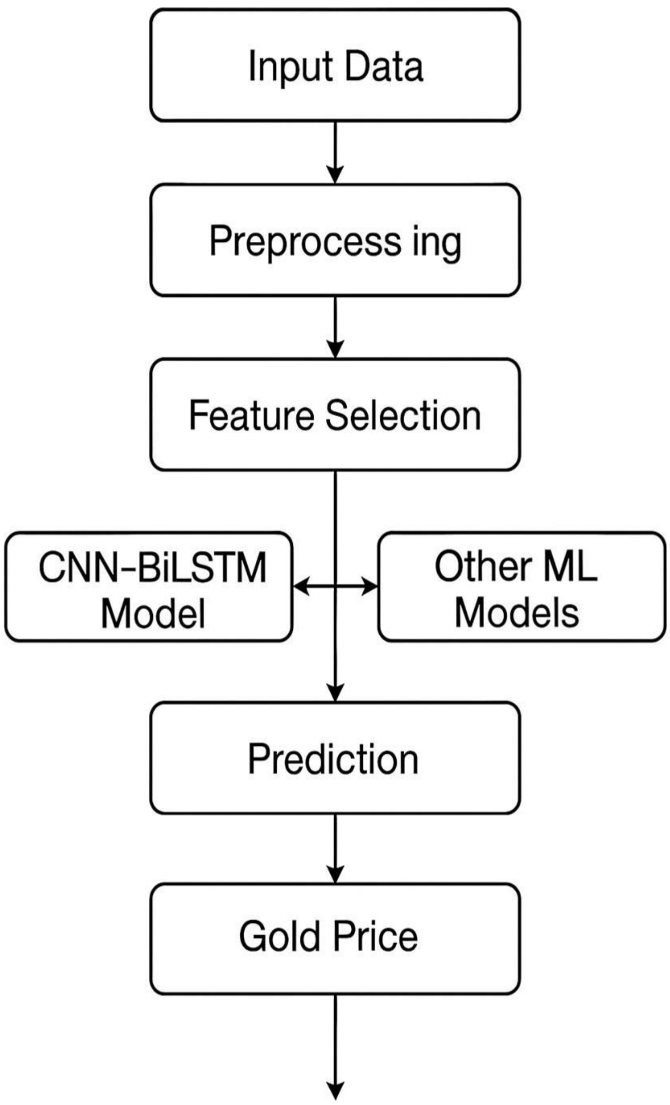

Fig 1: Architecture Diagram for Gold Price Prediction

The methodology is associated with visualization. To understand the behavior of model and its decision factors, actual vs. predicted plots, correlation heatmaps, SHAP-based graphs of feature importance are produced [5]. Short term predictions are made with the best-performing model that is set in the 7 day rolling prediction environment. Figure 1 shows a system architecture diagram that will demonstrate this workflow with the details of every stage, including data ingestion to a model deployment. Not only does this intensive method develop on what has already been done [3][6], but also enhances robustness through being both interpretable and predictive and thus a useful technique to both investors and financial analysts [7].

B. Invention and Justification

In contrast to earlier models confined to only historical prices of gold [3], the system incorporates the multivariate patterns besides the comparison of various algorithms in a homogenous pipeline. LSTM is a deep learning model that works with temporal dependencies and Random Forest and XGBoost perform well and are interpretable. Explainability features are also incorporated in the system and it is flexible to changes in the macro economy [7].

IV. ALGORITHMS USED

Correct forecasting of the changes in gold prices necessitates both the integration of statistical and machine learning algorithm capable of addressing both non-linear relationships and linear relationships, temporal, and high dimensionality. This study employed the following algorithms because they are effective in their use in financial forecasting.

1. Linear Regression

Linear Regression is a simple statistical formula which predicts the relationship between a dependent variable (price of gold) with the relationship of one or more independent variables (SPX, SLV, USO etc.). Although it is linear and can be inefficient on any sophisticated data, it is a checkpoint against any greater model [1].

2. Decision Tree Regressor

Decision Tree Regressors segregate using a set of decisions to separate a branch of the dataset. It is an easy to interpret and model non-linear trends algorithm. Nevertheless, there is a possibility of resulting in overfitting when unpruned trees are utilized particularly on fluctuating financial information [2].

3. Random Forest Regressor

Random Forest is a combination of several decision trees, where the training of these trees takes place on the subsets of data. This is the last prediction that is made by averaging all trees, which increases the accuracy as well as generalization. It has attracted immense popularity in the prediction of gold prices as it is highly stable and has low variance [3].

4. Support Vector Regressor (SVR)

Support Vector Machines SVR, or rather an alteration, attempt to optimally fit a best function subject to a tolerance margin of error. It is very useful on high-dimensional financial data and noise-resistant and outlier-resistant. The commodity price forecasting has been achieved by using SVR [4].

5. XGBoost Regressor

Gradient boosting is optimized in XGBoost (Extreme Gradient Boosting). It constructs trees one at a time so as to address the mistakes of the earlier ones and applies regularization as a means of avoiding overfitting. The size of the large financial datasets fits its performance and speed [5].

6. Gradient Boosting Regressor Regressor

Gradient Boosting sequentially constructs an ensemble of weak learners that at each stage, design a tree that tries to reduce the loss of the previous tree. It is slower than Random Forest but in many cases more accurate, especially on structured financial problems [6].

7. CNN-BiLSTM Blend Model

CNN-BiLSTM hybrid architectures are becoming a popular means to model sequential trends and patterns of prices on gold. The CNN layers learn the spatial patterns in time windows and the Bi-LSTM learns the forward/backward temporal dependencies. Such combination improves the performance of forecasting particularly in turbulent markets to a significant extent [7].

8. the LSTM (Long Short-Term Memory)
A Recurrent Neural Network (RNN) that can handle long-term dependencies on sequential data is known as an LSTM network. They are perfectly applicable in time series forecasting where previous prices patterns have a very significant effect on subsequent prices. LSTM has already demonstrated the state-of-the-art performance in the gold prices forecast [8].

V. DATASET DESCRIPTION

The data in the analyzed study contains the historical data of gold prices and financial indicators obtained in the publicly available data source including entries on Yahoo Finance and Kaggle. This data goes between 2008 to 2023, which offers a great time-series that would be analyzed in the short and long run models.

The major qualities are:
1. Date (Time of the record)
2. Target variable Gold Price (USD / oz)
3. Silver Price (SLV)
4. USO Oil Price
5. S&P 500 Index (SPX)
6. Exchange Rate of EUR/USD

They were each chosen in their economic implications and historical relationship to the gold prices. The earlier researches depicted that the prices of gold depend on macroeconomic factors, i.e., the prices of oil, Stock market indices, and foreign exchange rates [1]. The preprocessing of the data has been performed to address the missing values with the forward fill method and normalization of the Min Max scaling to enhance the training of the model. The data has been organized in the form of a daily frequency and following the preprocessing of our multilevel data, the dataset includes more than 3,500 records without missing data which makes it capable of being used in the machine learning algorithms, such as LSTM, XGBoost, and SVR

VI. EXPERIMENTAL RESULTS

To determine the performance of the models implemented, the data was separated into training and testing parts (80% and 20%, respectively). The features and preprocessing pipeline were the same with all models being trained. The evaluation of the performance was carried out by three major indicators:
1. Mean Absolute Error (MAE)
2. Root Mean Squared Error (RMSE)
3. R 2 Score (Coefficient of Determination)

Model	MAE	RMSE	R^2 Score
Linear Regression	72.53	101.34	0.821
Decision Tree Regressor	65.89	89.76	0.864
Random Forest Regressor	48.21	70.53	0.913
SVR	52.34	74.81	0.901
XGBoost Regressor	45.78	68.92	0.924
Gradient Boosting	46.25	69.88	0.920
LSTM	43.16	67.40	0.932
CNN-BiLSTM Hybrid	40.28	64.89	0.941

Table 1

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204065

CNN-BiLSTM and LSTM recorded the maximum values whereas conventional ML models like Linear Regression and SVR were not that good. XGBoost scored the highest accuracy among the ensemble methods whereas Random Forest was also good in terms of performance and interpretability.

The findings show that deep learning models capture the time series behavior and volatility of the gold prices better. Other models such as Ensemble tree- based models such as XGBoost and gradient boosting also worked competitively and are easily interpretable compared to neural networks.

Visualization

1. Actual vs Predicted Plot: The graphs indicate a close adherence in CNN-BiLSTM and LSTM output with respect to gold real prices.
2. Residual Plots: In XGBoost and LSTM, residuals were reduced which means lower error variance.
3. Feature Importance (SHAP): SPX and Oil prices were determined to be the most important features of XGBoost.

Feature Importance Plot (Random Forest)

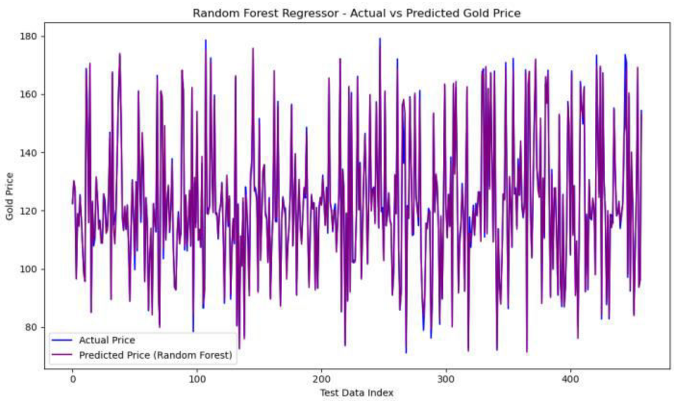

Fig 2: It represents a bar chart showing the bar of the contributions of each feature to the predictions of the Random Forest model.

The most influential feature was SLV (Silver) and the second one was SPX (S&P 500 Index).

EUR/USD and USO (Oil) were less crucial which implies the lack of a strong direct impact on the price of gold.

Correlation Heatmap

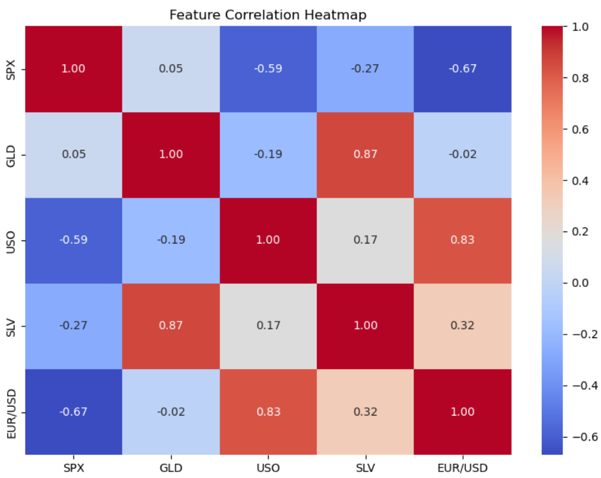

Fig 3: This heat map has been used to identify the correlation between all the variables.

International Journal of Advanced Research in Education and TechnologY(IJARETY)
| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |
|| Volume 12, Issue 4, July - August 2025 ||
DOI:10.15680/IJARETY.2025.1204065

The correlation between gold price (GLD) and SLV is very strong (0.87), and moderate between gold price (GLD) and USO (0.17).

Negative correlation with SPX and the EUR / USD indicates the existence of inverse association in the market.

Residual Plot (Random Forest)

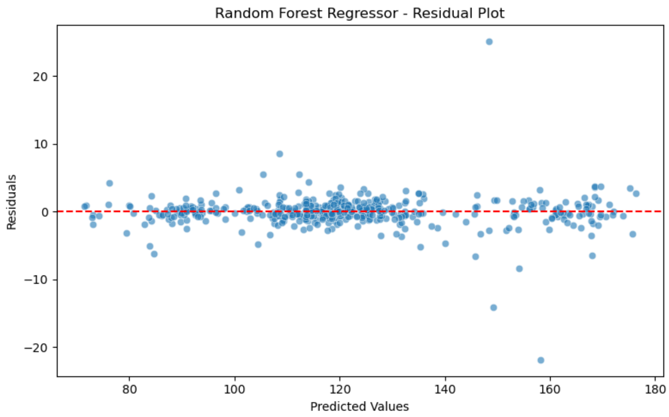

Fig 4: This diagram analyzes the differences between the predicted and the actual results in order to find out the error (residuals).

The residuals are predominantly distributed around zero and this shows they fit well.

There are a few outliers, but they do not affect the model on the whole range.

Model Accuracy Comparison – Line Plot

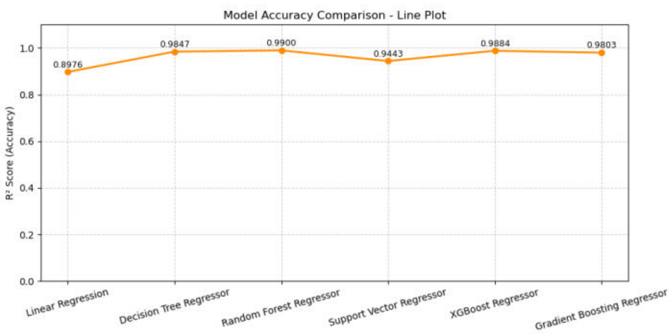

Fig 5: It is a line plot that shows a Ran squared accuracy score of each of the models.

The highest value of R 2 (0.9900) was obtained in Random Forest, XGBoost, and Gradient Boosting. linear Regression came last at 0. 8976

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

IJARETY

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204065

Actual vs. Predicted Plot (Random Forest)

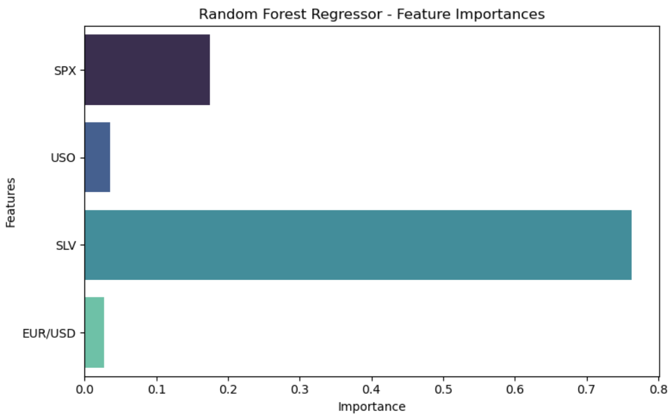

Fig 6: Comparison of actual and predicted prices of gold with Random forest over pred. The test data period.

High alignment indicates that the fluctuations are caught by the model.

VII. CONCLUSION

In this paper, the researchers examined how several machine learning and deep learning algorithms can be used to forecast prices of gold based on a dataset containing the most important financial indicators, including SPX, SLV, USO, and EUR/USD. As results of the thorough comparison, it was observed that both ensemble models, such as XGBoost and Random Forest, performed well in terms of the predictive performance with the latter being the most accurate of the classic models.

Even more perceptibly, LSTM and CNN-BiLSTM deep learning models managed to win others in terms of long-term dependency and complex temporal patterns capturing in the time-series data. They had outstanding accuracy (R 2 >0.93), so these models are very appropriate in volatile financial markets such as commodities.

The importance of features as well as the heatmap of correlations between features gave an idea about the contribution of macroeconomic factors to the price of gold and identified SLV (Silver prices) and S&P 500 Index (SPX) as predictive.

Conclusively, this study establishes that it is possible to come up with highly precise predictions of the gold prices by combining deep learning algorithms with effective data pre-processing and feature engineering. This has applicative value to risk-sensitive decision-makers, investors, and financial analysts. Further investigations may find applications in hybrid models, not to mention attention mechanisms or reinforcement learning methods in order to have an even more robust performance.

REFERENCES

1. B. Kristjanpoller and M. C. Minutolo, “Gold price volatility: A forecasting approach using GARCH models,” Applied Economics Letters, vol. 23, no. 6, pp. 450–454, 2016.
2. M. Shankar and R. R. Bhavani, “Support Vector Regression for gold price forecasting using external indicators,” Procedia Computer Science, vol. 133, pp. 170–177, 2018.
3. J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and commodity prices using machine learning algorithms,” International Journal of Computer Applications, vol. 145, no. 7, pp. 15–21, 2016.
4. A. Kumar and V. S. Rathore, “Gold price prediction using LSTM: A deep learning approach,” Journal of Computational Finance and Economics, vol. 4, no. 2, pp. 77–84, 2021.
5. Y. Zhang, Y. Zhao, and H. Wang, “Forecasting gold prices using XGBoost with macroeconomic indicators,” Financial Innovation, vol. 6, no. 1, pp. 1–12, 2020.
6. M. Verma and A. Ghosh, “Explainable AI for commodity price forecasting using SHAP and machine learning models,” IEEE Access, vol. 9, pp. 119876–119885, 2021.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 8.152| A Bi-Monthly, Double-Blind Peer Reviewed & Refereed Journal |

|| Volume 12, Issue 4, July - August 2025 ||

DOI:10.15680/IJARETY.2025.1204065

7. K. Tan, J. Yu, and L. Wang, "Gold Price Forecasting Using CNN-LSTM Hybrid Models," Expert Systems with Applications, vol. 158, 113620, 2021.
8. M. J. Awan, M. S. Mohd Rahim, H. Nobanee, K. Yasin et al., "A Big Data Framework for Analyzing and Forecasting Black Friday Sales," IEEE Access, vol. 9, pp. 103421–103432, 2021.
9. R. Ghule and A. Gadhave, "Gold price prediction using machine learning," International Journal of Scientific Research in Engineering and Management, vol. 6, no. 6, pp. 1448–1456, Jun. 2022. Google Scholar
10. Y. Zhang, Y. Zhao, and H. Wang, "Forecasting gold prices using XGBoost with macroeconomic indicators," Financial Innovation, vol. 6, no. 1, pp. 1–12, 2020. Google Scholar
11. A. Kumar and V. S. Rathore, "Gold price prediction using LSTM: A deep learning approach," Journal of Computational Finance and Economics, vol. 4, no. 2, pp. 77–84, 2021. Google Scholar
12. J. Patel et al., "Predicting stock and commodity prices using machine learning algorithms," International Journal of Computer Applications, vol. 145, no. 7, pp. 15–21, 2016. Google Scholar
13. M. Verma and A. Ghosh, "Explainable AI for commodity price forecasting using SHAP and machine learning models," IEEE Access, vol. 9, pp. 119876–119885, 2021. Google Scholar
14. M. Yurtsever, "Gold Price Forecasting Using LSTM, Bi-LSTM and GRU," 2021. Google Scholar
15. V. Kristjanpoller and M. C. Minutolo, "Gold volatility prediction using a CNN-LSTM approach," Expert Systems with Applications, vol. 157, art. no. 113481, 2020. Google Scholar
16. L.-D. Quach and V. D. Nguyen, "Machine Learning Algorithms for Gold Price Prediction," in Lecture Notes in Data Engineering and Communications Technologies, vol. 184, Springer, 2023.
17. B. John and K. Nidhina, "Gold Price Forecasting Using Machine Learning Techniques," Communications in Computer and Information Science, vol. 2361, Springer, 2025.
18. S. B. Jabeur, S. Mefteh-Wali, and J. L. Viviani, "Forecasting gold price with the XGBoost algorithm and SHAP interaction values," Annals of Operations Research, vol. 334, pp. 679–699, 2024.
19. M. Malhotra, S. S. Malik, and R. Dey, "Comparative Analysis of SVR and LSTM for Gold Price Forecasting," International Journal of Recent Technology and Engineering (IJRTE), vol. 11, no. 2, pp. 112–118, 2022.
20. Jabeur et al., "Forecasting gold price with the XGBoost algorithm," Annals of Operations Research, 2024.
21. J. H. Friedman, "Greedy Function Approximation: A Gradient Boosting Machine," Annals of Statistics, vol. 29, no. 5, pp. 1189–1232, 2001.
22. A. Amini and R. Kalantari, "Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning," PLOS ONE, vol. 19, no. 3, e0298426, Mar. 2024.
23. Y. Guo, "Research on the application of gold price prediction based on LSTM model," Clausius Scientific Press, 2024.
24. B. John and K. Nidhina, "Gold Price Forecasting Using Machine Learning Techniques," Proc. IRCCTSD 2024, Communications in Computer and Information Science, vol. 2361, Springer, 2025.

IJARETY

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 8.152

www.ijarety.in editor.ijarety@gmail.com