International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Gold Price Prediction Using an Ensemble of Random Forest and Xgboost Algorithms

***Prasun Tripathi* [*1] , *Arun Pandey*[*2], *Naman Yadav*[*3]**

[*1] Asst. Professor, CSE, Axis Institute of Technology and Management, Kanpur, Uttar Pradesh, India
[*2] Student, CSE, Axis Institute of Technology and Management, Kanpur, Uttar Pradesh, India
[*3] Student, CSE, Axis Institute of Technology and Management, Kanpur, Uttar Pradesh, India

ABSTRACT :

This study introduces a gold price forecasting model that leverages an ensemble of Random Forest and XGBoost algorithms. Current data on gold prices and additional variables for training our model was gathered from Google Finance. The model was tested and subsequently assessed using various performance metrics, and the results obtained indicate that the ensemble method produces superior predictions compared to either algorithm individually. This research offers valuable insights into the use of machine learning methods for forecasting gold prices.

Keywords—Gold Price Prediction, Random Forest, XGBoost, Machine Learning, Ensemble Learning.

I. INTRODUCTION

The significance of gold as a secure asset and economic indicator has considerably increased worldwide, especially during periods of economic and political instability. The COVID-19 pandemic has heightened economic turmoil, underscoring the necessity for reliable investments such as gold. However, this has led to traditional forecasting models struggling to adapt to the extraordinary shifts occurring in the market.

Determining gold prices is essential for investors and financial analysts to make educated decisions. Machine learning techniques have shown potential in this area, and we propose an ensemble model that employs Random Forest and XGBoost algorithms for gold price prediction.

We gathered current gold price data from Google Finance to train and evaluate our model, reaching an accuracy rate exceeding 98%, which is better than the performance of each algorithm on its own. We also applied visualization methods, such as Heatmaps and Pairplots, to investigate the relationships between gold prices and various economic indicators like the S&P 500 index, the Dow Jones Industrial Average, currency exchange rates, and the prices of other precious metals. Our model can aid investors and financial analysts in forecasting gold prices based on even minor fluctuations in related commodities or factors.

Future studies could pursue options such as integrating a user feedback mechanism and employing NLP techniques to further improve the model's accuracy. Additionally, broadening our model's application to include other precious metals and assessing its performance on a larger dataset might yield deeper insights into the precious metals market.

Consequently, our ensemble model shows potential in precisely predicting gold prices, offering valuable perspectives on the intricate relationships between various economic indicators and gold prices. The combination of our machine learning methodology and data analysis techniques can significantly improve the accuracy and efficiency of gold price predictions and holds substantial promise for ongoing research and development. This paper adds to the expanding body of research on gold price forecasting through machine learning methods, presenting a unique strategy that combines Random Forest and XGBoost algorithms and achieving high levels of accuracy, while addressing several critical shortcomings of conventional approaches like ARIMA, Linear Regression, and Gradient Boosting.

II. METHODOLOGY

Data Collection: We gathered daily gold prices along with data on variables such as the S&P 500 index, Dow Jones Industrial Average, US Oil Fund, currency exchange rates, Newmont Corporation stocks, and the prices of other precious metals like Silver from Google Finance, covering the period from January 1, 2013, to April 28, 2023. This information was utilized to create our dataset.

Date Preparation: Using the Pandas Library in Python, we processed and cleaned the data. We initially examined the dataset for any missing values and discovered that some dates were not present. To address these gaps, we employed the interpolation method, which approximates the missing values based on the data from adjacent dates.

Next, we divided the constructed dataset into training and testing subsets, allocating 80% for training and 20% for testing.

Data Analysis: We conducted exploratory data analysis to understand the data better and illustrate the relationships between the features and the target variable. Using Seaborn, we created Jointplots to display the correlation of gold prices with each independent variable in separate graphs. Additionally, we generated a Heatmap and Pairplot to enhance visualization and comprehension of the correlations in a clear and concise manner.

Model Training: We then used the training subset to develop our ensemble model, which was constructed using Random Forest and XGBoost algorithms. The hyperparameters for the latter were optimized using Grid Search, while those for the former were tuned manually.

Model Evaluation: To assess the model's performance, we applied several metrics, including mean absolute error (MAE), mean squared error (MSE), and the coefficient of determination (R-squared). We also visualized the predicted versus actual gold prices with line plots and scatter plots. The outcomes were as follows:

• MAE: 9.81

• MSE: 259.86

• R-squared: 0.996

Furthermore, our visual representations also yielded positive results.

III. MODELING AND ANALYSIS

Our proposed approach for predicting gold prices employs a combination of Random Forest and XGBoost models. Random Forest was chosen for its capability to manage non-linear relationships, while XGBoost serves as an effective gradient boosting algorithm that enhances prediction accuracy and is proficient at addressing noisy data. By integrating these two algorithms, we can attain a more precise and dependable forecast of gold prices.

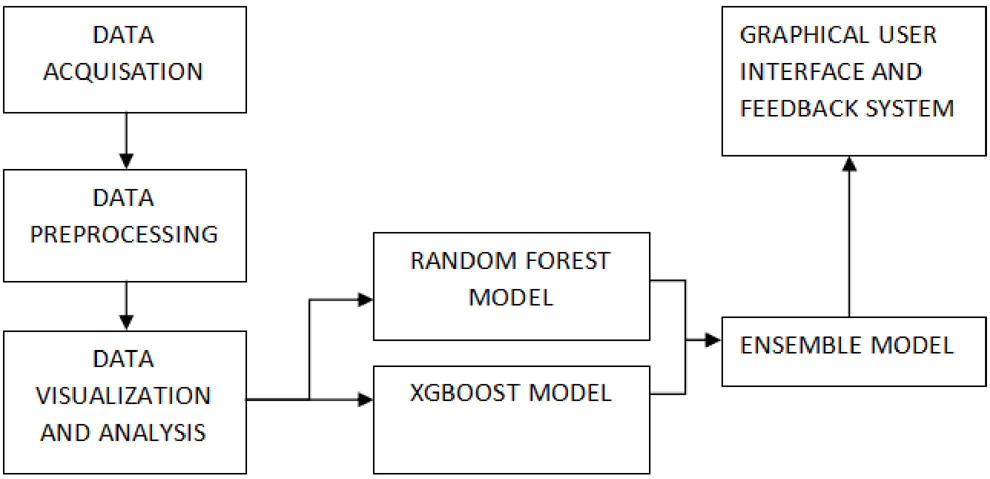
One key benefit of our approach is its capacity to accommodate a diverse array of input variables. Our system can incorporate not only historical gold price information but also other critical factors influencing gold prices, such as the S&P 500 index, the Dow Jones Industrial Average, foreign exchange rates, and the prices of other precious metals, including silver. By leveraging a variety of input variables, we can boost the precision and dependability of our gold price forecasts.

Another significant advantage of our system is its ability to process missing data. We employ an advanced imputation technique that fills in absent values by drawing on established patterns within the data. This enables us to utilize as much information as possible and enhance the accuracy of our predictions.

Moreover, our system features several visualization tools that assist users in comprehending the elements that influence gold prices. Through interactive charts and graphs, users can examine various scenarios and obtain insights into the intricate relationships between gold prices and other economic factors.

In summary, our proposed gold price prediction system presents a robust and adaptable resource for investors, financial analysts, and anyone with an interest in the gold market.

BLOCK DIAGRAM



IV. RESULTS AND DISCUSSION

The findings from our research show that our suggested model, which combines Random Forest and XGBoost algorithms, surpasses other widely used models in forecasting the price of gold. We assessed our model with various performance indicators, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and the coefficient of determination (R-squared). Our model registered an MAE of 9.81, an MSE of 259.86, and an R-squared value of 0.996, reflecting its high precision in predicting gold prices. The accuracy remains consistently elevated throughout our testing phases, and in contrast to standalone machine learning algorithms, our ensemble approach makes the model less susceptible to overfitting. In addition, unlike LSTM-based models, the computational requirements are kept at a reasonable level. Moreover, our model offers insights into the factors affecting gold prices, as we illustrated the relationships between gold and other independent variables through joint plots, a heatmap, and a pair plot. The visualizations highlighted a significant positive correlation between gold and the S&P 500 Index, along with the value of another precious metal, silver, while revealing a negative correlation between gold and the EUR/USD exchange rate, aligning with established economic theories and market trends. In summary, the findings of our research confirm the efficacy of our proposed model in predicting gold prices and elucidating the factors that influence them.

V. CONCLUSION

We performed an extensive analysis of the data and pinpointed the key factors influencing gold prices. We have showcased visualization and analysis methods that clarify the connection between gold prices and various independent variables in a straightforward and comprehensible way. This insight can benefit investors and traders in making well-informed choices in the precious metals sector. Furthermore, the model itself can be extremely useful for investors and traders in forecasting gold prices based on even minor shifts in any pertinent variables. This capability to swiftly draw conclusions and possibilities from evolving market trends can be a considerable asset in the dynamic realm of precious metals trading, enabling improved decision-making and the potential for increased profits. In summary, our proposed ensemble-based machine learning framework, along with advanced data analysis methods, can offer significant insights for investors and traders in the precious metals market. By effectively predicting gold prices based on even the tiniest alterations in any of the associated variables, our model can assist users in making informed decisions and possibly achieving higher profits. Looking ahead, the potential for future research and development of our model remains considerable, as we continue to seek improvements in its accuracy and effectiveness.

VI. REFERENCES

[1] Pritish K. S. et al. “Gold price and exchange rate in pre and during Covid-19,” Resources Policy, Volume 79, December 2022, 1031126

[2] Banhi Guha et al. “Gold Price Forecasting Using ARIMA Model,” Journal of Advanced Management Science Vol. 4, No. 2, March 2016

[3]M.Sravani et al. “GOLD PRICE PREDICTION,” IJCRT Vol. 9, Issue 6 June 2021, ISSN: 2320-2882

[4] Laor Boongasame et al. “Gold-Price Forecasting MethodUsing Long Short-Term Memoryand the Association Rule,” Journal of Mobile Multimedia, Vol. 19_1,165–186, 2022

[5] Andrea-Cristina et al. "Limitations of ARIMA models in financial and monetary economics,” AGER Volume XXIII, 2016, No. 4(609), Winter

[6] Zolzaya Luvsandorj, "Comparing Random Forest and Gradient Boosting," Towards Data Science, Jul 2, 2021

[7] Philip Wilkinson, "Non-Linear Regression with Decision Trees and Random Forest," Towards Data Science, Feb 3, 2022

[8]Özdemir et al. " Evaluating the Nonlinear Linkage between Gold Prices and Stock Market Index Using Markov-Switching Bayesian VAR Models,” Procedia - Social and Behavioral Sciences 210 (2015) 408 – 415

[10] Gonzalo Mart´ınez-Mu˜noz et al. "A Comparative Analysis of XGBoost," Artificial Intelligence Review, Volume 54, Issue: 3, Nov5, 2019