International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Study on Price Forecasting for Gold Commodities Using Tree-Based Customized Adaboost Algorithm

[1]Anandaraj SP, [2]Jyoti Shekhawat, [3]Hina Hashmi, [4]Apoorva Joshi

Submitted: 22/04/2023 **Revised**: 18/06/2023 **Accepted**: 27/06/2023

Abstract: For buyers and sellers to make wise choices, forecasting prices in the field of commodities is essential. The goal of this study is to create a reliable and accurate price forecasting model that is tailored to gold commodities. To do this, we suggest a brand-new tree-based AdaBoost (T-CA) technique that includes features specifically designed for the features of gold prices. The methodology of the report makes use of daily statistics from the World Gold Council between 1979 and 2019. We only utilize data from 2014 for model training and the rest for model validation. The outcomes show that our suggested algorithm performs better than the alternatives, with lower errors and higher forecasting abilities. The results of this study provide a trustworthy and effective method based on the T-CA for price forecasting gold commodities. The investors, sellers, and banks can make more precise predictions and improve how they invest in the gold sector thanks to the useful insights provided by the suggested T-CA model.

Keywords: *Gold commodity, price forecasting, tree-based customized AdaBoost (T-CA), accuracy*

1. Introduction

Gold is among the most valuable minerals on Earth. Gold is a reserve in every nation, even though it may be made into useful goods. The quantity of gold that is retained by the central bank of any country as an assurance that it can be used to buy or trade on the world market and, as a result, grow the economy of that nation is referred to as the gold reserve [1, 2]. Gold is by far the most well-liked option for financial investments when compared to any other material in the globe.Gold price is influenced by various factors, making price movement unpredictable. Among these are the inflation rate, supply and demand, and political issues. When inflation rises, a sign of a healthy economy, the price of gold rises naturally, and the opposite is true when supplies of a commodity are low and prices rise [3]. Gold prices tend to rise when there is widespread concern that the value of the dollar will decline because it is an international trading currency. Because of its significance, it has been called a refuge of stability amid economic downturns in other works of fiction.Gold Because of this, the gold price is quite volatile. Figure 1 plots the gold price over a 41-year period, from January 1982 to December 2018 [4, 5]. However, a possible future gold price may be predicted, allowing for future decisions to be made. Gold price fluctuations follow a time-series order, which simply means that they change over time; Until the development of machine learning and deep learning in the fields of banking and statistics, it was thus exceedingly impossible to foresee them [6, 7].

In this research article, we use ARIMA and tree-based customized Adaboost (T-CA) algorithm models to forecast gold prices. One of the oldest and most popular approaches is ARIMA. The T-CAmodel, an application of supervised machine learning, is used as a comparison standard.

The remainder of this paper is arranged as follows: Part 2-related work, part 3- metrology, Part 4- results, and Part 5-conclusion.

2. Related works

The financial sector has also begun to make good use of deep learning technologies. The basic explanation for this is that financial markets are extremely unstable and susceptible to being influenced by a wide range of factors. This uncertainty presents both a chance and a significant risk for participants in the financial markets. The hope is that future applications for deep learning created in this space will help investors both capitalize on market opportunities and hedge against potential market hazards [8]. Study [9] Proposed combining “LSTM and CNN neural networks” with the Attention Mechanism to forecast the daily gold price trend. In particular, they built an LSTM component, a Focus Mechanism, and a CNN component into their LSTM Focus-CNN model. The LSTM module created the new coding and stored the

[1]*Professor & HOD, Department of Computer Science and Engineering, Presidency University, Bangalore, India, Email Id: anandaraj@presidencyuniversity.in*
[2]*Assistant Professor, Department of Computer Science & Application, Vivekananda Global University, Jaipur, India, Email Id: jyoti.shekhawat@vgu.ac.in*
[3]*Assistant Professor, College of Computing Science and Information Technology, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh, India, Email id: hinahashmi170@gmail.com*
[4]*Assistant Professor, Department of Master of Computer Application, Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh, India, Email id: aproova.joshi@niet.co.in*

earlier processed information in long-term memory, while the Focus Mechanism enhanced the weights of every component in the encoding. The CNN module performed better since it was able to extract regional characteristics. To gauge the efficacy of the suggested methodology, they ran extensive tests using actual daily gold price information. The accuracy of the gold price forecast was crucial.

Article [10] Introduced a genetic algorithm and a regularized machine learning algorithm to calculate the ideal layer weight matrix and neurons in the hidden layer of a gold price forecasting model, mitigating the effect of the model's inherent randomness on subsequent sequential update stages of forecasting. Study [11] Proposed a NN forecasting method based on SSA smoothing to enhance prediction accuracy and trend detection. Using these reference models, they evaluate the proposed method on four different dimensions. Further, they investigated the interconnectedness of three distinct commodity futures and the factors that influence their prices. Researchers [12] Suggested a new forecasting framework called “CNN-LSTM for predicting gold price and motion." The suggested “CNN-LSTM model” was tested in two variations, each using two convolution layers with various amounts of filters, and compared against modern deep learning and machine learning forecasting techniques. Study [13] Examine six various machine learning models for their ability to forecast the price of gold. According to the results, XGBoost delivers superior performance compared to industry standard benchmark models and other leading methods.

To improve the reliability of gold price predictions in a major Indian commodities market, they designed and built many cutting-edge ANN algorithms. They recommended using multiple other optimization functions for this purpose rather than the standard quadratic error function based on the average square error. Using a combination of different conventional ANN algorithms and recently modified ANN algorithms, weekly gold price projections were made [14]. The ICA-GRUNN model is presented to predict the gold price using ICA and GRUNN. In this case, ICA is a statistical analysis method created specifically for identifying and resolving signal-to-noise ratio problems involving unidentified sources. Over the past four years since its introduction, GRUNN has gained widespread use [15].

3. Proposed Method

3.1 Dataset

The purpose of this investigation is to utilize deep learning in order to forecast the gold price. Daily gold price information from the World Gold Council is used in this analysis. The World Gold Council lists its prices in terms of a single troy ounce. The dataset contains daily prices beginning in January 1982 and ending in December 2018, for a total of 41 years of data collection. The next step is data cleansing, which involves verifying that all of the information is accurate and in the right places. The average number for the previous three days is used to fill in any gaps in the data. The data segmentation is completed after that. At this stage, the dataset is split into training and testing portions. The model is trained using the testing data, and then the model's performance on the data being tested is evaluated. Knowing the ARIMA method and T-CA is necessary since they will be used in this research to forecast the gold price.

3.2 ARIMA model

Time series forecasting has employed a variety of techniques to determine the most accurate forecasting approach, particularly in business and econometrics. The use of the ARIMA model in predicting is a highly effective strategy. Auto Retrograde Integrated Motion Average is an acronym for this method. The ARIMA model was developed a few years ago when statisticians and economists were analyzing time series trends without taking the impact of non-stationarities of data on forecast outcomes seriously.Forecasting and management were created by George Box and Gwilym Jenkins, who details how irregular data can be made stable. When dealing with time series data, ARIMA is a common econometric and statistical model. When a dataset exhibits irregularity, it is typically analyzed using this model.

A different name for this model is the Box- Jenkins approach. When there are seasons, the ARIMA mode is expressed as “ARIMA (p, d, q).ARIMA (p, d, q) (PDQ)” m is the abbreviation used to indicate the season. The number of autoregressive or lag order (p), number of irregular differences (d), number of average movement periods (q), and number of seasons (m) are all variables in autoregressive models. When two values of the (p, d, q) become zero, the ARIMA model becomes either AR, I, or MA. As an illustration, “ARIMA is identical to AR (p), where (p, 0, 0)” is the input. It is crucial to select the appropriate ARIMA model before making any predictions using an Arima model. Therefore, it is necessary to calculate the value of pdq.The model is trained with the help of the fit() tool. Remember that the time series must be stable before doing any calculations. There are a number of ways to test for a stationary state, but “the Augmented Dickey-Fuller” test will be employed here. Furthermore, The Box-Jenkins Model makes forecasts by combining the three theories of autoregression, differencing, and moving average. These three rules are denoted by the letters p, d, and q. The Box-Jenkins analysis makes use of each of these principles, and their combined representation as ARIMA (p, d, q) is very useful.

3.3 Tree-based customized adaboost algorithm

The term "AdaBoosting" refers to "Adaptive Boosting," a Type of machine learning approach. The AdaBoost strategy utilizes a large number of classifiers. A prior classifier is used to divide the sample for the purpose of training a subsequent classifier. The flexibility of the AdaBoost technique can be strengthened by using this model. AdaBoost is less susceptible to overfitting in some classification situations than certain other learning algorithms.It's possible that the AdaBoost method's classifier doesn't perform very well. The voting model suggests that Adaboost's classification capability can be enhanced by having all of the weak classifiers come to a collective decision about how to proceed. The AdaBoost approach is an iterative algorithm; after each round, this method will adjust a weak classifier until it reaches a preset level of classification accuracy that is considered to be suitably low. Every single training sample will serve as the weight. If the classifier has correctly assigned a label to this sample, then the classifier's probability of selecting this sample and its weight in the overall sample pool should decrease. On the other hand, if this sample is not identified correctly by this classifier, its weight needs to go up. In this approach, the AdaBoost algorithm will be able to concentrate on managing the more challenging samples.

In this study, AdaBoost is modified using a single model. Each sample's initial weight is equal in classic AdaBoost. A sample's starting weight in the modified AdaBoost model is established at a time point that is generated by the sample itself. The Equation is as follows:

$$u_1(j) = \frac{(1.01)^{x_i - min_{\mathcal{P}}(x_{\mathcal{P}})}}{\sum_{j=\mathcal{P}}^{\mathcal{S}} (1.01)^{x_{\mathcal{P}} - min_{\mathcal{P}}(x_{\mathcal{P}})}}$$

(1)

Where A2 is the year from which sample I was obtained. Because of this, a more recent sample will carry more weight than an older one.The most recent sample is to be given more consideration in stage 1. The improved version of AdaBoost has the benefit of being able to accurately categorize the most recent data.

4. Result

The machine learning techniques "Linear Regression (LR), Random Forest Regression (RFR), and Gradient Boosting Regression (GBR)" are implemented in this study by using Python. “Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE)” were considered to assess the predictive power of the regression techniques are used compared with our approach T-CA. Figure 5 shows the results of actual and predicted price of gold.

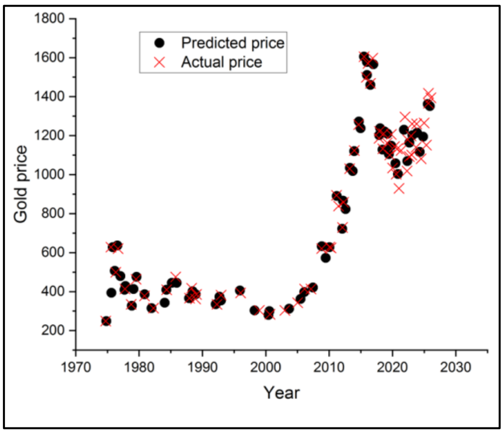

Fig 1: T-CA prediction and actual Price

The R2-score, also known as the coefficient of determination, measures the proportion of variance in the dependent variable that can be predicted by the independent variable(s) as in Equation 2. Figure 2 depicts the R2-score comparison for LR, RFR, and GBR is 0.957, 0.9802, and 0.9786 and our approach T-CA has 0.9987. It shows that our approach has a higher R^2-score when compared with other techniques.

$$R^2 = 1 - \frac{\sum (z_j - \hat{z})^2}{\sum (z_j - \bar{z})^2}$$

(2)

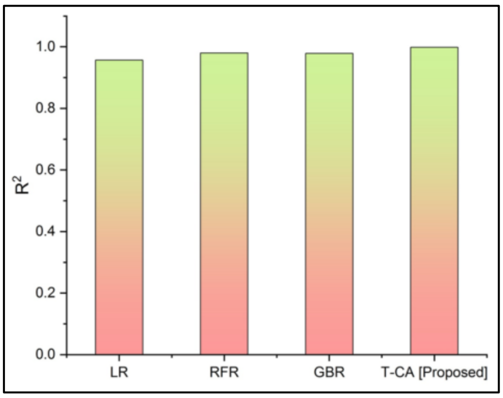

Fig 2: R2-score

The mean absolute error (MAE)” is the mean of the absolute disparities between projected and actual outcomes. It assigns equal importance to all errors, as in Equation 3. Figure 3 depicts the RMSE comparison for LR, RFR, and GBR as 4384.43, 2808.44, and 3007.8, and our technique T-CA has 2801.5. It illustrates that, when compared to other methods, our approach has a low MAE.

$$MAE = \frac{1}{M} \sum_{j=1}^{M} \left| z_j - \hat{z} \right|$$

(3)

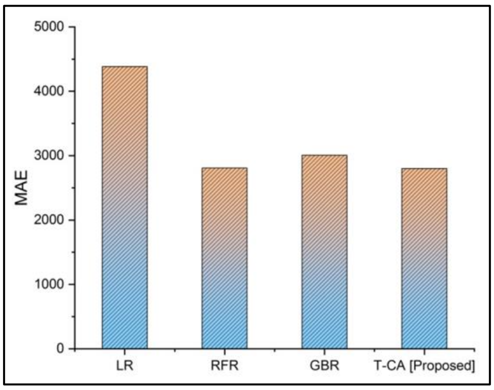

Fig 3: Mean Absolute Error

“The mean squared error (MSE)," which is the average of the squared discrepancies between the expected and actual values. It emphasizes huge discrepancies due to squaring, as in Equation 4. Figure 4 displays the MAE comparison for LR, RFR, and GBR is 35465098.85, 16294903.01, and 17609001.94, and our method T-CA has 13718103.86. It demonstrates that our approach has low MSE compared with others.

$$MSE = \frac{1}{M}\sum_{j=1}^{M}\left(z_j - \hat{z}\right)^2$$
(4)

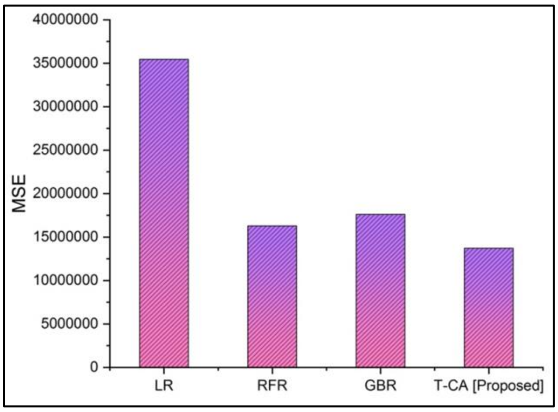

Fig 4: Mean Squared Error

“The square root of the MSE” is called the root mean squared error (RMSE). Because it is in the same units as the output, it is easier to understand as in Equation 5. Figure 5 shows the MSE comparison for LR, RFR, and GBR is 5955.25, 4036.69, and 4196.3, and our approach T-CA has 3869.5. It shows that our technique has a low RMSE when compared to other approaches. The values of MAPE, RMSE, and R^2 for ARIMA and T-CA are presented in Table 1.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(z_j - \hat{z}\right)^2} \quad (5)$$

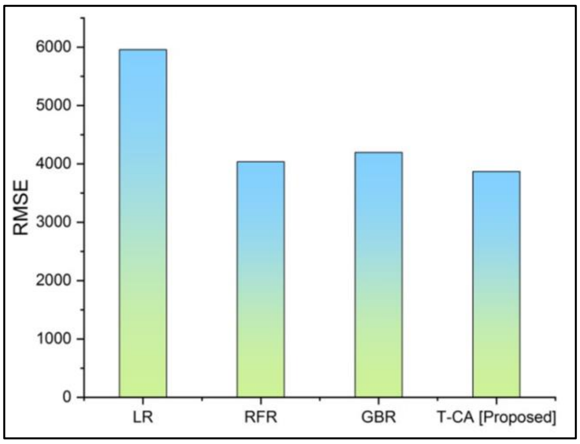

Fig 5: Root Mean Squared Error

Table 1: ARIMA vs T-CA

Methods	MAPE	RMSE	R2
ARIMA	2689.48	37.6262	0.8154
T-CA	2741.72	3869.5	0.9987

5. Conclusion

We suggested a brand-new tree-based AdaBoost (T-CA) technique that included features specifically designed for predicting the price of gold and related commodities. For the study, monthly price data from January 1982 to December 2018 were used. The study's findings show that T-CV performs more effectively than the other approaches. The Values of performance metrics for our proposed method were obtained in terms of R2-score, MAE, MSE, and RMSE of 0.9987, 3007.8, 13718103.86, and 3869.5, respectively. It has been shown experimentally that the proposed method was the most effective for forecasting and prediction analysis. For class predictions, Adaboost does not automatically offer probability estimates. In the future, models based on artificial neural networks will be constructed using the same financial metrics.

References

[1] ul Sami, I. and Junejo, K.N., 2017. Predicting future gold rates using a machine learning approach. International Journal of Advanced Computer Science and Applications, 8(12).

[2] Belasen, A.R. and Demirer, R., 2019. Commodity-currencies or currency-commodities: Evidence from causality tests. Resources Policy, 60, pp.162-168.

[3] Al-Dhuraibi, W.A. and Ali, J., 2018, May. Using classification techniques to predict gold price movement. In 2018 4th International Conference on Computer and Technology Applications (ICCTA) (pp. 127-130). IEEE.

[4] Raza, N., Ali, S., Shahzad, S.J.H. and Raza, S.A., 2018. Do commodities effectively hedge real estate risk? A multi-scale asymmetric DCC approach. Resources Policy, 57, pp.10-29.

[5] Dolatabadi, S., Narayan, P.K., Nielsen, M.Ø. and Xu, K., 2018. The economic significance of commodity return forecasts from the fractionally cointegrated VAR model. Journal of Futures Markets, 38(2), pp.219-242.

[6] Ayele, A.W., Gabreyohannes, E. and Tesfay, Y.Y., 2017. Macroeconomic determinants of volatility for the gold price in Ethiopia: the application of GARCH and EWMA volatility models. Global Business Review, 18(2), pp.308-326.

[7] Benlagha, N. and El Omari, S., 2022. Connectedness of stock markets with gold and oil: New evidence from COVID-19 pandemic. Finance Research Letters, 46, p.102373.

[8] Yurtsever, M., 2021. Gold price forecasting using LSTM, Bi-LSTM, and GRU. Avrupa Bilim ve Teknoloji Dergisi, (31), pp.341-347.

[9] He, Z., Zhou, J., Dai, H.N. and Wang, H., 2019, August. Gold price forecast based on the LSTM-CNN model. In 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech) (pp. 1046-1053). IEEE.

[10] Weng, F., Chen, Y., Wang, Z., Hou, M., Luo, J. and Tian, Z., 2020. Gold price forecasting research based on an improved online extreme learning machine algorithm. Journal of Ambient Intelligence and Humanized Computing, 11, pp.4101-4111.

[11] Wang, J. and Li, X., 2018. A combined neural network model for commodity price forecasting with SSA. Soft Computing, 22, pp.5323-5333.

[12] Livieris, I.E., Pintelas, E. and Pintelas, P., 2020. A CNN–LSTM model for gold price time-series forecasting. Neural computing and applications, 32, pp.17351-17360.

[13] Jabeur, S.B., Mefteh-Wali, S. and Viviani, J.L., 2021. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. Annals of Operations Research, pp.1-21.

[14] Verma, S., Thampi, G.T. and Rao, M., 2020. ANN-based method for improving gold price forecasting accuracy through modified gradient descent methods. IAES International Journal of Artificial Intelligence, 9(1), p.46.

[15] Jianwei, E., Ye, J. and Jin, H., 2019. A novel hybrid model on the prediction of time series and its application for the gold price analysis and forecasting. Physica A: Statistical Mechanics and its Applications, 527, p.121454.

[16] Dr. Antino Marelino. (2014). Customer Satisfaction Analysis based on Customer Relationship Management. International Journal of New Practices in Management and Engineering, 3(01), 07 - 12. Retrieved from http://ijnpme.org/index.php/IJNPME/article/view/26

[17] Ahire, P. G. ., & Patil, P. D. . (2023). Context-Aware Clustering and the Optimized Whale Optimization Algorithm: An Effective Predictive Model for the Smart Grid. International Journal on Recent and Innovation Trends in Computing and Communication, 11(1), 62–76. https://doi.org/10.17762/ijritcc.v11i1.5987