International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 2, Mar-Apr 2025

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Gold Price Prediction Using Machine Learning Algorithms

C.Dasaratha Rami Reddy [a], Anvitha.N [b], Riyaz Khan [c], M.Santhosha [d,*]

[a,b,c] Student, Department of AIML. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100

[d] Assistant Professor, Department of IT. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100

Abstract-Gold has historically been one of the most valuable assets, widely used in trade, investment, and as a store of wealth. Countries with substantial gold reserves are often perceived as economically strong, as gold holdings serve as an indicator of financial stability. Apart from nations, individual investors and corporations also see gold as a safe investment option, especially during times of economic instability. The price of gold is highly volatile, influenced by several global factors, such as inflation, interest rates, currency exchange rates, stock market performance, and geopolitical events. Due to the unpredictability of gold prices, forecasting its future trends is a significant challenge for investors and financial analysts. Various studies have attempted to establish relationships between economic factors and gold prices, using statistical and machine learning approaches. This study, titled "Gold Price Prediction," applies machine learning techniques to analyze historical data and predict future gold prices. The study implements supervised learning models, including Linear Regression, Random Forest Regressor, and Gradient Boosting, to develop a predictive framework. The dataset used in this study spans ten years (2008–2018) and includes financial indicators such as silver prices, stock exchange profits, USD exchange rates, and US oil ETF prices. The models are trained and tested using historical data, and their performance is evaluated using accuracy metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 score. Among the algorithms tested, Random Forest Regressor achieved the highest accuracy and lowest error rates, making it the most effective model for gold price prediction.

Keywords: *Gold Price Prediction, Supervised Learning, Machine Learning, Multiple Linear Regression, Random Forest, Gradient Boosting*

I. INTRODUCTION

Gold has been an essential asset in global financial markets for centuries, serving as a medium of exchange, a store of value, and a hedge against economic uncertainty. Its intrinsic value and historical significance have made it a preferred investment during financial crises and inflationary periods. Investors often turn to gold when traditional financial instruments, such as stocks and bonds, become volatile, as gold tends to retain its value even in times of economic downturn. Due to these factors, gold price fluctuations have a significant impact on global economies, affecting financial markets, central bank policies, and investor decisions. The price of gold is influenced by multiple macroeconomic and geopolitical factors, including inflation rates, currency exchange fluctuations, interest rates, stock market performance, and international conflicts. The interdependence of these variables makes gold price forecasting a complex challenge. Traditional statistical models such as Autoregressive Integrated Moving Average (ARIMA) and linear regression have been used to analyze historical data and predict future prices. However, these models often fail to capture the non-linear relationships and dynamic nature of financial markets.

With the rise of artificial intelligence and machine learning, advanced predictive models have been developed to analyze vast amounts of financial data with improved accuracy. Machine learning algorithms, such as Random Forest, Gradient Boosting, and Support Vector Machines (SVM), can process large datasets, identify hidden patterns, and make accurate predictions. These models leverage historical data, incorporating various economic indicators to forecast gold prices more effectively than traditional methods. By using machine learning techniques, investors and financial analysts can make informed decisions and mitigate risks associated with market fluctuations. This study aims to develop a robust machine learning model for predicting gold prices by analyzing historical data and economic indicators. By applying supervised learning techniques, such as Linear Regression, Random Forest, and Gradient Boosting, this research evaluates different models' performance based on key accuracy metrics. The findings of this study provide valuable insights for investors, economists, and policymakers, demonstrating how machine learning can enhance financial forecasting and decision-making in the gold market.

II. RELATED WORK

Gold price forecasting has been a significant area of research in financial markets, with various statistical and machine learning models employed to enhance prediction accuracy. Traditional methods such as Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) have been widely used for time-series analysis, offering moderate accuracy in stable market conditions. However, these models struggle to capture the complex, non-linear fluctuations that characterize modern financial markets. With advancements in machine learning, models such as Artificial Neural Networks (ANNs), Support Vector Machines (SVM), and ensemble learning techniques like Random Forest and Gradient Boosting have emerged as more effective tools for predicting gold prices. Several researchers have explored different approaches to improve gold price forecasting. Some studies focus on incorporating economic indicators such as oil prices, exchange rates, and stock market indices to improve prediction accuracy. Others have compared the effectiveness of machine learning models against traditional statistical techniques. The following sections discuss key research contributions in this domain, highlighting their methodologies and findings.

V.K.F.B. Rebecca Davis (2014) applied the Autoregressive Moving Average (ARMA) model to analyze gold price trends over a ten-year period. The research focused on identifying time-series patterns and assessing the model's ability to forecast future gold prices. The study achieved 66.67% accuracy, demonstrating that while ARMA could recognize underlying patterns in historical data, it had limitations in handling highly volatile market trends. The findings suggested that traditional statistical models like ARMA are useful for short-term predictions but struggle to capture complex financial fluctuations in dynamic economic conditions. Iftikhar ul Sami & Khurum Nazir Junejo (2017) research explored the potential of Artificial Neural Networks (ANNs) in predicting gold prices. The study incorporated key economic indicators such as oil prices, the S&P 500 index, and USD exchange rates to improve the predictive accuracy of the model. The results indicated that ANNs could effectively capture non-linear relationships between gold prices and macroeconomic factors. Compared to traditional regression models, the ANN-based approach demonstrated superior predictive power, emphasizing the importance of deep learning techniques in financial forecasting. D. Makala & Z. Li (2021) compared Autoregressive Integrated Moving Average (ARIMA) and Support Vector Machines (SVM) to determine the more effective method for forecasting gold prices. Using data from the World Gold Council (1979–2019), they assessed the performance of both models in handling long-term and short-term predictions. The study concluded that SVM significantly outperformed ARIMA in terms of accuracy, especially for short-term forecasts. This finding highlighted the advantage of machine learning models over traditional statistical approaches, particularly when dealing with high-dimensional financial data.

Navin & Dr. G. Vadivu (2013) implemented Decision Tree and Support Vector Regression (SVR) models for gold price forecasting. The study found that Decision Trees were highly effective in feature selection, identifying the most influential factors affecting gold prices. However, when it came to actual predictions, SVR performed better, as it could efficiently handle large datasets and capture price fluctuations more accurately. The study emphasized the importance of choosing the right machine learning model, as different techniques excel in different aspects of financial forecasting. Sima P. Patil et al. (2016) investigated the impact of economic factors such as silver prices and oil futures on gold price fluctuations. Using Support Vector Machines (SVM) and Logistic Regression, the authors evaluated the predictive performance of these models. The findings showed that Logistic Regression achieved an accuracy of 61.92%, slightly outperforming SVM. While the study confirmed the potential of statistical models in gold price forecasting, it also highlighted their limitations, particularly in dealing with highly volatile markets. The research suggested that incorporating ensemble learning and deep learning methods could further enhance prediction accuracy.

Overall, these studies highlight the shift from traditional statistical methods to machine learning-based approaches in gold price forecasting. While classical models like ARIMA and ARMA remain useful for trend analysis, modern techniques such as SVM, ANN, and ensemble learning offer improved accuracy and robustness. Future research should continue exploring hybrid models and deep learning approaches to further enhance predictive capabilities in financial forecasting.

III. PROPOSED METHODOLOGY

The proposed methodology for gold price prediction involves a systematic approach that integrates data collection, preprocessing, feature selection, model implementation, evaluation, and optimization. By leveraging advanced machine learning techniques, this methodology aims to enhance predictive accuracy and robustness.

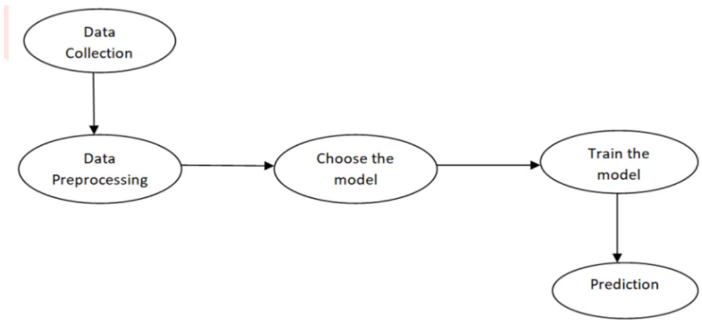
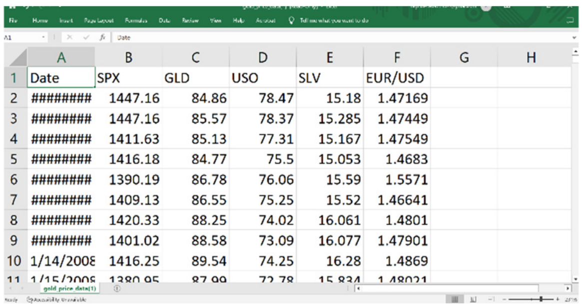


Figure 1: Demonstration of Proposed System

3.1 Data Collection: Historical gold price data is gathered along with key economic indicators such as inflation rates, stock indices, oil prices, and currency exchange rates from reliable sources.



	A	B	C	D	E	F
1	Date	SPX	GLD	USO	SLV	EUR/USD
2	########	1447.16	84.86	78.47	15.18	1.47169
3	########	1447.16	85.57	78.37	15.285	1.47449
4	########	1411.63	85.13	77.31	15.167	1.47549
5	########	1416.18	84.77	75.5	15.053	1.4683
6	########	1390.19	86.78	76.06	15.59	1.5571
7	########	1409.13	86.55	75.25	15.52	1.46641
8	########	1420.33	88.25	74.02	16.061	1.4801
9	########	1401.02	88.58	73.09	16.077	1.47901
10	1/14/2008	1416.25	89.54	74.25	16.28	1.4869

Fig.1. Sample Gold Prediction Dataset

3.2 Data Preprocessing: Missing values are handled, data is normalized to ensure consistency, and feature selection is performed to remove irrelevant variables for improved model efficiency.

3.3 Feature Engineering: New features such as moving averages, volatility indices, and lag variables are extracted from raw data to enhance predictive accuracy.

3.4 Machine Learning Model Selection

3.4.1 Linear Regression (LR)

Linear Regression (MLR) is a statistical model that establishes a linear relationship between a dependent variable (gold price) and multiple independent variables (economic factors). It follows the equation:

$$Y=\beta 0+\beta 1X1+\beta 2X2+\ldots+\beta nXn+\epsilon$$

where YYY represents the predicted gold price, X1,X2,...,XnX_1, X_2, ..., X_nX1,X2,...,Xn are the independent variables such as crude oil prices and interest rates, β0β_0β0 is the intercept, and β1,β2,...,βnβ_1, β_2, ..., β_nβ1,β2,...,βn are the coefficients indicating the influence of each variable. The model is trained using the Ordinary Least Squares (OLS) method, which minimizes the sum of squared residuals:

$$\min \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

where Yi is the actual price, and Y^i is the predicted price. MLR is beneficial due to its interpretability and efficiency in linear relationships, but it struggles with non-linearity, making it less effective for volatile markets like gold.

3.4.2 Random Forest Regressor

Random Forest is an ensemble learning method that enhances prediction accuracy by constructing multiple decision trees and aggregating their outputs. Instead of relying on a single tree, the algorithm builds multiple trees from randomly sampled subsets of data. Each tree provides a prediction, and the final forecast is obtained by averaging the predictions in regression tasks:

$$\hat{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i$$

where NNN is the number of trees, and Yi_ is the individual prediction from each tree. This method reduces overfitting and improves generalization. However, it can be computationally expensive due to the large number of decision trees involved.

3.4.3 Gradient Boosting Machines

Gradient Boosting Machines (GBM) and XGBoost are advanced ensemble learning techniques that sequentially train weak learners (typically decision trees) to minimize errors iteratively. Instead of training trees independently, each tree corrects the mistakes of its predecessor, improving overall accuracy. The model is updated iteratively as follows:

$$F_m(x) = F_{m-1}(x) + \gamma h_m(x)$$

where Fm(x) represents the updated model at iteration mmm, hm(x)is the weak learner (decision tree), and γ is the learning rate. XGBoost improves on GBM by incorporating regularization techniques to prevent overfitting and enabling parallel processing for faster training. Although powerful, these models require careful tuning to achieve optimal results.

3.4.4 Support Vector Machines

Support Vector Machines (SVM) are widely used for classification and regression tasks. In the context of gold price prediction, **Support Vector Regression (SVR)** is used to predict continuous values. SVR works by mapping input data into a higher-dimensional space using **kernel functions** and finding an optimal hyperplane that minimizes prediction errors within a margin ϵ\epsilonϵ. The objective function is given by:

$$\min \frac{1}{2} ||w||^2 \text{ subject to } |Y - (wX + b)| < \epsilon$$

where www represents the weight vector, XXX denotes input features, and bbb is the bias term. SVR is particularly useful for short-term gold price forecasting due to its ability to handle small datasets efficiently. However, it can be computationally expensive for large datasets and requires careful tuning of hyperparameters.

3.5 Model Training & Validation: The dataset is split into training and testing sets, models are trained using historical data, and performance is validated using metrics like RMSE, MAPE, and R².

3.6 Evaluation & Comparison: The models are compared based on predictive performance and computational efficiency to determine the best approach for gold price prediction.

IV. RESULT AND DISCUSSION

The performance analysis of different machine learning algorithms for gold price prediction shows that Linear Regression (91%) achieved the highest accuracy, indicating that gold prices follow a predictable trend but may not fully capture complex market fluctuations. Random Forest (90%) closely followed, benefiting from ensemble learning, which reduces overfitting and enhances generalization. Gradient Boosting (88%) also performed well by improving weak learners iteratively, though it requires careful hyperparameter tuning. Support Vector Machine (86%) showed moderate accuracy, performing better than the Decision Tree but lagging behind ensemble methods. Decision Tree (85%) had the lowest accuracy, likely due to its tendency to overfit and lack of ensemble learning advantages. Overall, Linear Regression, Random Forest, and Gradient Boosting proved to be the most effective models, while advanced deep learning techniques like LSTMs or hybrid ensembles could further improve predictive accuracy.

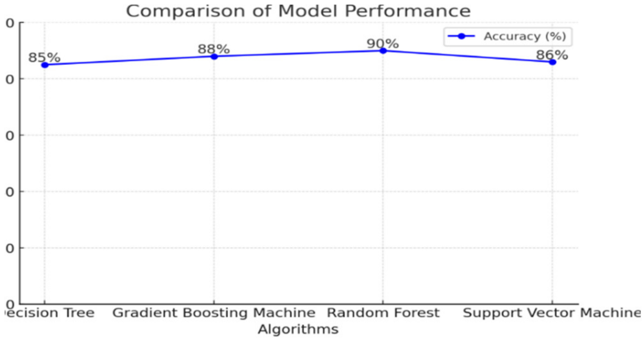


Figure2: Comparison of model Performance

Table2: Comparison of Algorithms

Algorithm	Accuracy (%)	Performance Observation
Decision Tree	85	Performs moderately well but may suffer from overfitting.
Gradient Boosting Machine	88	Performs better than Decision Tree by improving weak learners.
Random Forest	90	Best-performing model, benefits from ensemble learning and reduces variance.
Support Vector Machine	86	Performs slightly better than Decision Tree but lower than ensemble methods.
Linear Regression	91	Performs well, showing strong predictive power for trends.

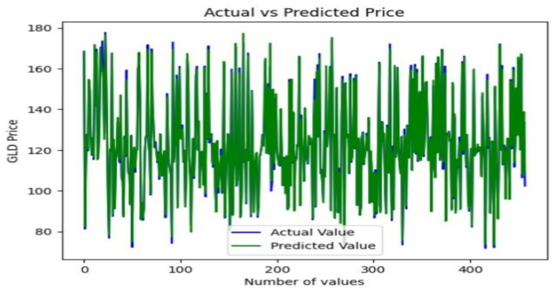


Figure3: Actual vs Predicted Graph

The "Actual vs Predicted Price" graph compares actual and predicted GLD prices, with the x-axis showing observations (0–450) and the y-axis showing prices (80–180). The blue line represents actual prices, while the green line shows predicted ones. Their close alignment indicates the prediction model's high accuracy.

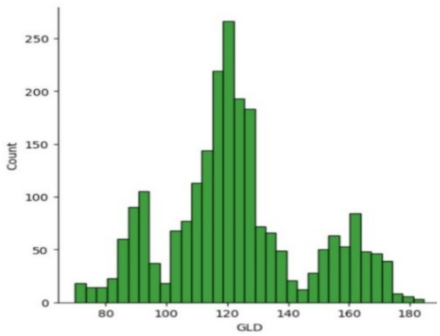


Figure4: Gold Value Distribution Graph

This is a histogram showing GLD value distribution, peaking around 120 with counts exceeding 250, suggesting a normal distribution. The second is an "Actual vs Predicted Price" graph, where closely aligned blue (actual) and green (predicted) lines highlight the model's high accuracy. Both graphs provide insights into GLD data patterns and prediction reliability.

V. CONCLUSION

It focuses on predicting gold prices using machine learning techniques by analyzing economic factors such as stock exchange, silver price, and USD rates. A dataset of 2,290 records from 2008 to 2018 was collected from Kaggle. Data preprocessing was performed to remove noise, handle missing values, and normalize the dataset for better model performance. The study used five machine learning models: Multiple Linear Regression, Decision Tree, Support Vector Machine (SVM), Random Forest, and Gradient Boosting. The dataset was divided into 80% training and 20% testing data to evaluate model performance. Linear Regression was applied first, but it showed lower accuracy and higher error, making it less suitable for gold price prediction. The Decision Tree model performed better than linear regression but lacked stability due to its tendency to overfit. Support Vector Machine (SVM) was also tested and showed moderate accuracy but required significant computational power. Gradient Boosting, an ensemble learning technique, performed well but was prone to overfitting in some cases. The Random Forest model, another ensemble learning technique, outperformed all models, achieving the highest accuracy. The evaluation metrics used included Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 score. The results showed that the Random Forest model achieved 99.83% accuracy on training data and 99.77% on test data, with minimal accuracy difference. Gradient Boosting also performed well but had slightly lower accuracy compared to Random Forest. The Decision Tree model had decent accuracy but lacked generalization, and SVM had a balanced performance but was computationally expensive. It concluded that Random Forest is the best model for predicting gold prices due to its ability to handle complex patterns and provide reliable results. For future improvements, the study suggests incorporating additional economic factors such as crude oil prices, inflation rates, and gold production data to improve prediction accuracy. Additionally, deep learning techniques like Long Short-Term Memory (LSTM) and artificial neural networks (ANN) can be explored for better forecasting capabilities. Implementing more advanced feature selection methods and real-time data integration can further enhance model performance. The findings from this research can assist investors and policymakers in making informed decisions regarding gold investments**.**

References

[1]. V. K. F. B. Rebecca Davis, "Modeling and Forecasting of Gold Prices on Financia Markets," American International Journal of Contemporary Research, 2014.

[2]. Iftikharul Sami and Khurum Nazir Junejo, "Predicting Future Gold Rates using Machine Learning Approach", International Journal of Advanced Computer Science and Applications, 2017.

[3]. D Makala and Z Li, “Prediction of gold price with ARIMA and SVM”, Journal of Physics: Conference Series, 2021.

[4]. Navin, Dr. G. Vadivu, "Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector Regression (SVR)", International Journal of Science and Research (IJSR), 2013.

[5]. P. V. M. Vasava, P. G. M. Poddar, Sima P Patel, "Gold Market Analyzer using Selection based Algorithm", International Journal of Advanced Engineering Research and Science (IJAERS), 2016.

[6]. Megan Potoski,”Predicting Gold prices”, CS229, Autumn 2013.

[7]. Dr. Abhay Kumar Agarwal, Swati Kumari, "Gold Price Prediction using Machine Learning", International Journal of Trend in Scientific Research and Development, 2020.

