ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL292

Advanced Machine Learning Techniques for Predicting Gold and Silver Futures

Dipankar Roy
Dept. of CSE (AI & ML)
Haldia Institute of Technology
Haldia, West Bengal, India

Joyita Ghosh
Dept. of CSE (AI & ML)
Haldia Institute of Technology
Haldia, West Bengal, India

Abhik Choudhary
Dept. of CSE (AI & ML)
Haldia Institute of Technology
Haldia, West Bengal, India

Subir Gupta
Dept. of CSE (AI & ML)
Haldia Institute of Technology
Haldia, West Bengal, India

Kamaluddin Mandal
Dept. of CSE (AI & ML)
Haldia Institute of Technology
Haldia, West Bengal, India

Abstract:- This research focuses on predicting the future values of gold and silver futures by employing advanced machine learning algorithms. Traditional econometric models often struggle with commodity prices' non-linear and dynamic nature. To address this, the study evaluates the performance of four unconventional machine learning algorithms: Gaussian Processes, Quantile Regression Forests, Extreme Learning Machines, and Support Vector Regression with an RBF kernel. The dataset used includes monthly trading data for gold and silver futures. The research findings indicate that these machine-learning models significantly enhance prediction accuracy. Support Vector Regression with an RBF kernel demonstrated the highest accuracy for gold futures predictions, while Extreme Learning Machines performed competitively for silver futures. This study highlights the potential of advanced machine learning techniques in financial forecasting, providing valuable insights for traders and investors in optimizing their strategies.

Keywords:- *Gold; Silver; Machine Learning; SVR.*

I. INTRODUCTION

Predicting the future values of commodities such as gold and silver is a crucial aspect of financial markets, driven by the need for accurate forecasting tools that can assist traders, investors, and policymakers make informed decisions[1]. The inherent volatility of these commodities, influenced by a myriad of global economic factors, makes this task particularly challenging[2]. Traditional econometric models, which often rely on linear assumptions, struggle to capture the complex and non-linear relationships that characterize the prices of gold and silver[3]. As a result, there is a growing interest in exploring advanced machine-learning techniques that can better handle these complexities and provide more accurate predictions[4]. Gold and silver have long been considered

Valuable assets, often used as a hedge against inflation and economic uncertainty. Their prices are affected by various factors, including geopolitical events, currency fluctuations, changes in interest rates, and macroeconomic indicators[5]. These factors create a highly dynamic and unpredictable environment, making accurately forecasting gold and silver prices a significant challenge. Traditional methods, such as time-series analysis and linear regression models, are limited in accounting for commodity prices' non-linear and stochastic nature [6]. Therefore, exploring more sophisticated approaches that can effectively model these complexities is essential. The primary problem addressed in this research is the development of robust predictive models for forecasting the future values of gold and silver futures. The traditional forecasting methods, primarily based on linear models, fail to capture the intricate patterns and dependencies in the data. These methods often overlook the non-linear interactions and are less effective in handling the high volatility and sudden changes typical of commodity markets[7]. As a result, there is a need for more advanced techniques that can provide accurate and reliable predictions, helping stakeholders to mitigate risks and optimize their investment strategies[8]. To address this problem, this research proposes using advanced machine learning algorithms well-suited for handling non-linear and complex data patterns. Specifically, four unconventional machine learning algorithms are considered: Gaussian Processes, Quantile Regression Forests, Extreme Learning Machines, and Support Vector Regression with an RBF kernel. Each of these algorithms offers unique advantages in modelling the dynamics of gold and silver prices. Gaussian Processes are a powerful tool for time series analysis, and they are known for

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL292

their ability to provide uncertainty estimates and predictions. This feature is precious in financial forecasting, where understanding the confidence in predictions can significantly enhance decision-making processes. Quantile Regression Forests are non-parametric models that are robust against outliers, making them suitable for financial data that often contain extreme values[9].

Extreme Learning Machines are designed for fast and efficient training, making them an attractive option for large-scale regression tasks. Finally, Support Vector Regression with an RBF kernel is known for its effectiveness in non-linear regression, capable of capturing the complex relationships between the input features and the target variable. The research methodology involves several key steps, starting with data preprocessing and feature engineering. The dataset used in this study comprises monthly trading data for gold and silver futures, including columns for Instrument Type, Month, Year, Gold Traded Contract (Lots), Gold Total Value (Lacs), Silver Traded Contract (Lots), and Silver Total Value (Lacs). The initial step is to clean the data, removing any inconsistencies or missing values to ensure the dataset is ready for analysis. Following this, features and targets are created based on past values to predict future values. This involves calculating lagged values of the target variables to serve as predictors in the models[10]. Each selected machine learning model is then trained and evaluated using Mean Squared Error (MSE) as the performance metric. This study explores the application of these advanced algorithms to financial forecasting, aiming to demonstrate their effectiveness in improving prediction accuracy for gold and silver futures. The results of this research are expected to provide valuable insights into the potential of machine learning techniques in financial forecasting and offer practical implications for traders and investors in optimizing their strategies[11].

In summary, this research addresses the significant challenge of predicting the future values of gold and silver futures by leveraging advanced machine learning algorithms. The proposed solution involves the application of Gaussian Processes, Quantile Regression Forests, Extreme Learning Machines, and Support Vector Regression with an RBF kernel to model the complex and non-linear dynamics of commodity prices. The findings of this study are anticipated to highlight the advantages of using these sophisticated techniques in financial forecasting, ultimately contributing to more accurate and reliable predictions that can aid in risk management and strategic decision-making in the commodities market[12].

II. LITERATURE REVIEW

Forecasting commodity prices, particularly gold and silver, has been the subject of extensive research due to its significant impact on financial markets and investment strategies. The traditional methods for predicting commodity prices have largely relied on econometric models and time-series analysis. These models include linear regression, autoregressive integrated moving averages (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), and various other forms of statistical techniques[13]. Despite widespread use, these models often struggle to accurately capture the complex and non-linear relationships that characterize commodity prices. This limitation has led researchers to explore more advanced methods to handle these intricacies better. One of the traditional approaches, ARIMA, has been extensively used due to its simplicity and effectiveness in short-term forecasting. However, ARIMA models are inherently linear and assume a stationary process, which is often not true with financial time series exhibiting non-linear patterns and structural breaks[14]. Similarly, GARCH models have been employed to model the volatility of commodity prices. While GARCH models can capture time-varying volatility, they still operate under the assumption of linearity and may not fully account for the non-linear dependencies in the data. In recent years, machine learning has provided new avenues for improving the accuracy of commodity price forecasts. Machine learning techniques such as artificial neural networks (ANNs), support vector machines (SVMs), and ensemble methods have gained popularity due to their ability to model complex non-linear relationships without requiring explicit specification of the underlying data distribution. ANNs, for instance, have been used to capture the non-linear dynamics of commodity prices by learning from historical data. However, neural networks often require large datasets for training and can be prone to overfitting if not properly regularized. Support vector machines, particularly support vector regression (SVR), have shown promise in financial forecasting due to their robustness and effectiveness in high-dimensional spaces[15]. Using kernel functions, SVR models can handle non-linearity by mapping input features into a higher-dimensional space. This approach allows SVR to capture intricate patterns that traditional linear models might miss. However, selecting appropriate kernel functions and tuning hyperparameters remain challenging tasks that can significantly influence the model's performance. Ensemble methods, such as random forests and gradient boosting machines, have also been explored for commodity price forecasting. These methods combine multiple weak learners to form a strong predictive model, thereby improving accuracy and robustness. Random forests, in particular, are non-parametric and can handle outliers effectively, making them suitable for financial data with extreme values. Gradient boosting machines, on the other hand, build models sequentially, correcting errors made by previous models[16]. While ensemble methods have demonstrated improved performance over single models, they can be computationally intensive and require careful tuning to avoid overfitting. Despite the advancements in machine learning techniques, significant challenges remain in accurately forecasting commodity prices. One major challenge is the availability and quality of data. Financial markets are influenced by various factors, including macroeconomic indicators, geopolitical events, and market sentiment, which are often difficult to quantify and incorporate into predictive models. Additionally, commodity prices exhibit high volatility and sudden shifts, making it challenging to develop models that adapt to changing market conditions. Another challenge is the interpretability of machine learning models. While advanced algorithms such as neural networks and ensemble

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL292

methods can provide accurate predictions, they often operate as black boxes, offering little insight into the underlying factors driving the forecasts. This lack of transparency can be a barrier to adoption, particularly in financial markets where understanding the rationale behind predictions is crucial for decision-making. In light of these challenges, recent research has focused on developing hybrid models that combine the strengths of traditional econometric methods and advanced machine learning techniques. Hybrid models aim to leverage the interpretability of statistical methods while benefiting from the predictive power of machine learning algorithms. For example, hybrid models may use ARIMA or GARCH models to capture linear relationships and volatility patterns, while neural networks or support vector machines are employed to model non-linear dependencies[17]. The research gap in the current literature lies in the need for more comprehensive studies that systematically compare the performance of different machine learning algorithms for commodity price forecasting. While individual studies have demonstrated the potential of various techniques, there is a lack of consensus on the most effective approaches for different market conditions and data characteristics. Moreover, integrating external factors such as macroeconomic indicators and market sentiment into predictive models remains an area that requires further exploration. This research aims to address these gaps by evaluating the effectiveness of several advanced machine learning algorithms, including Gaussian Processes, Quantile Regression Forests, Extreme Learning Machines, and Support Vector Regression with an RBF kernel, for forecasting the future values of gold and silver futures[18]. By systematically comparing these models and incorporating relevant external factors, this study seeks to provide a clearer understanding of the most effective approaches for commodity price forecasting. The findings are expected to contribute to developing more accurate and reliable predictive models, ultimately aiding traders and investors in making informed decisions in the highly volatile commodities market[19].

III. METHODOLOGY

The research methodology for predicting the future values of gold and silver futures involved several key steps, starting with data preprocessing and feature engineering. The dataset consisted of monthly trading data, including columns for Instrument Type, Month, Year, Gold Traded Contract (Lots), Gold Total Value (Lacs), Silver Traded Contract (Lots), and Silver Total Value (Lacs). The first step in the methodology was to clean the data, removing any inconsistencies or missing values, to ensure the dataset was ready for analysis. After preprocessing, features and targets were created based on past values to predict future values. This involved calculating lagged values of the target variables to serve as model predictors. For model selection, four unconventional machine learning algorithms were considered: Gaussian Processes, Quantile Regression Forests, Extreme Learning Machines, and Support Vector Regression (SVR) with an RBF kernel. Each of these models has unique characteristics suitable for time series prediction. Gaussian Processes are known for their ability to provide uncertainty estimates, which is beneficial in financial forecasting. The mathematical formulation of Gaussian Processes involves.

Defining a covariance function $k(x, x')$ and the prediction at a new point x_* is given by:
$f(x_*)= K(x_*, X)K(X, X)^{-1}y$

Where $K(x_*, X)$ represents the covariance between the new point and the training points,

$K(X, X)$ is the covariance matrix of the training points, and y is the vector of observed values.

Quantile Regression Forests are non-parametric and robust against outliers, making them suitable for financial data that may have extreme values. The quantile regression problem can be expressed as:

$$Q_y(\tau|X) = \min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - X_i\beta)$$

Where $\rho_\tau(u) = u(\tau - I(u < 0))$ is the check function, and $Q_y(\tau|X)$ represents the τ-th quantile of y given$X$.

Extreme Learning Machines (ELM) are fast and efficient for regression tasks. The ELM algorithm involves randomly initializing the weights and biases of the hidden nodes and then determining the output weights by solving a linear system:

$$H\beta = Y$$

Where H is the hidden layer output matrix, β is the output weight vector, and Y is the target output vector. The output weights are obtained by:

$$\beta = H^+Y$$

Where H^+ is the Moore- Penrose pseudo-inverse of H.

Support Vector regression (SVR) with an RBF kernel is suitable for non-linear regression. The SVR model aims to find a function that has at most ∈ deviation from the actual targets for all training data. The optimization problem for SVR can be formulated as:

$$\min_{w,b,\in,\in^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} \in_i + \in_i^*$$

Subject to

$$y_i - (w.\emptyset(x_i) + b) \leq \in + \in_i$$
$$(w.\emptyset(x_i) + b) - y_i \leq \in + \in_i^*$$
$$\in_i, \in_i^* \geq 0$$

Where w and b are the parameters to be determined, ∈ and ∈* are slack variables, and C is a regularization parameter.

Each model was trained and evaluated using Mean Squared Error (MSE) as the performance metric. The results indicated that the Support Vector Regression (SVR) with RBF kernel best predicted the total value of gold futures,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL292

while the Extreme Learning Machines model showed competitive performance for predicting the total value of silver futures. This methodology demonstrates the effectiveness of advanced machine learning techniques for forecasting financial time series, providing valuable insights for traders and investors.

IV. RESULT

The dataset comprised monthly trading data for gold and silver futures, including the following columns: Instrument Type, Month, Year, Gold Traded Contract (Lots), Gold Total Value (Lacs), Silver Traded Contract (Lots), and Silver Total Value (Lacs). The objective was to predict the future values of gold and silver using unconventional machine learning algorithms such as Gaussian Processes, Quantile Regression Forests, Extreme Learning Machines, and Support Vector Regression with different kernels. The initial steps involved preprocessing the data, feature engineering, and model selection. After preprocessing, features and targets were created based on past values to predict future values. Fig 1 Gold and Silver Value Prediction Using Gaussian Process

Four models were considered: Gaussian Processes, which are suitable for time series and provide uncertainty estimates; Quantile Regression Forests, which are non-parametric and robust against outliers; Extreme Learning Machines, known for their efficiency in regression tasks; and Support Vector Regression with an RBF kernel, suitable for non-linear regression. Fig 2 Gold and Silver Value Prediction Using Quantile regression Forests.

Upon training and evaluating these models, the Gaussian Processes model resulted in high Mean Squared Errors (MSE) for both gold and silver predictions, specifically 1.0789×10^14 for gold and 2.1738×10^14 for silver. In contrast, the Quantile Regression Forests model significantly improved performance, reducing the MSE to 3.1064×10^12 for gold and 8.6696×10^12 for silver. Fig 3 Gold and Silver Value Prediction Using Extreme Learning Machines

The Extreme Learning Machines model also showed competitive results with MSE values of 4.2159×10^12 for gold and 6.3552×10^12 for silver. Lastly, the Support Vector Regression (SVR) with an RBF kernel achieved the best performance for gold predictions with an MSE of 2.9569×10^12 and a comparable performance for silver with an MSE of 6.4501×10^12.Fig 4 Gold and Silver Value Prediction Using Support Vector Regression (RBF Kernel)
In summary, the SVR with RBF kernel model was the most effective for predicting the total value of gold futures, while the Extreme Learning Machines model demonstrated competitive performance in predicting the total value of silver futures. These results indicate that using unconventional machine learning algorithms can significantly enhance the accuracy of financial predictions, particularly for complex datasets involving commodities like gold and silver futures. This study highlights the potential of machine learning in financial forecasting, suggesting that further exploration and refinement of these models could yield even more precise predictions that would benefit traders and investors in making informed decisions.

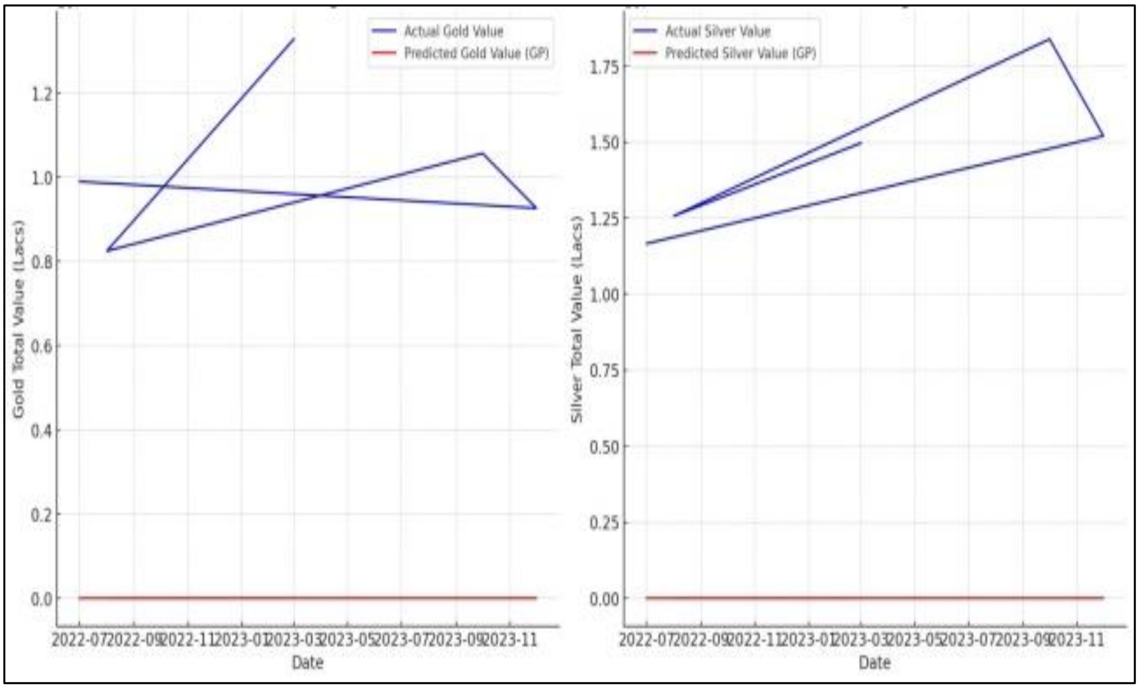

Fig 1: Gold and Silver Value Prediction Using Gaussian Process

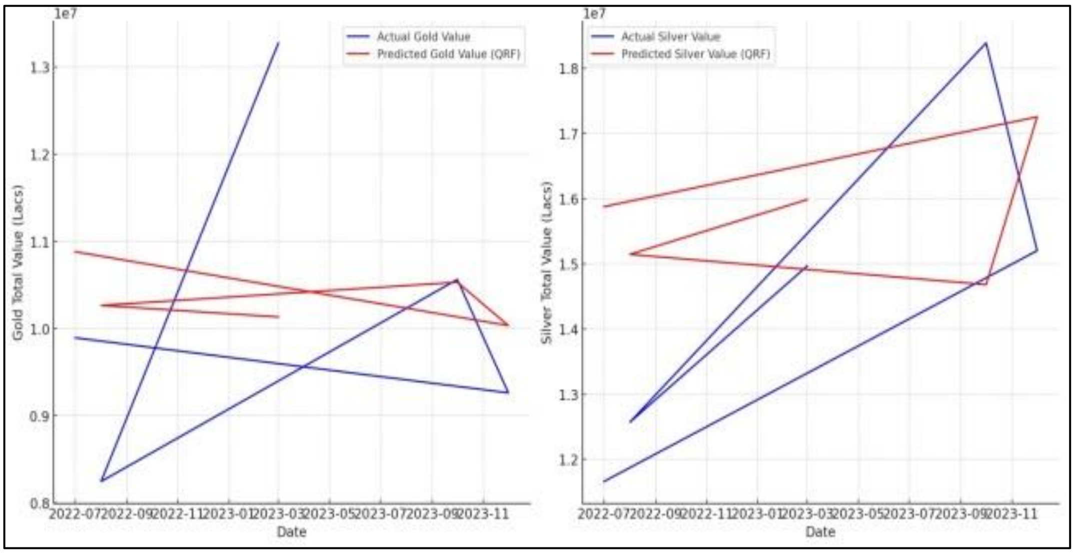

Fig 2: Gold and Silver Value Prediction Using Quantile regression Forests

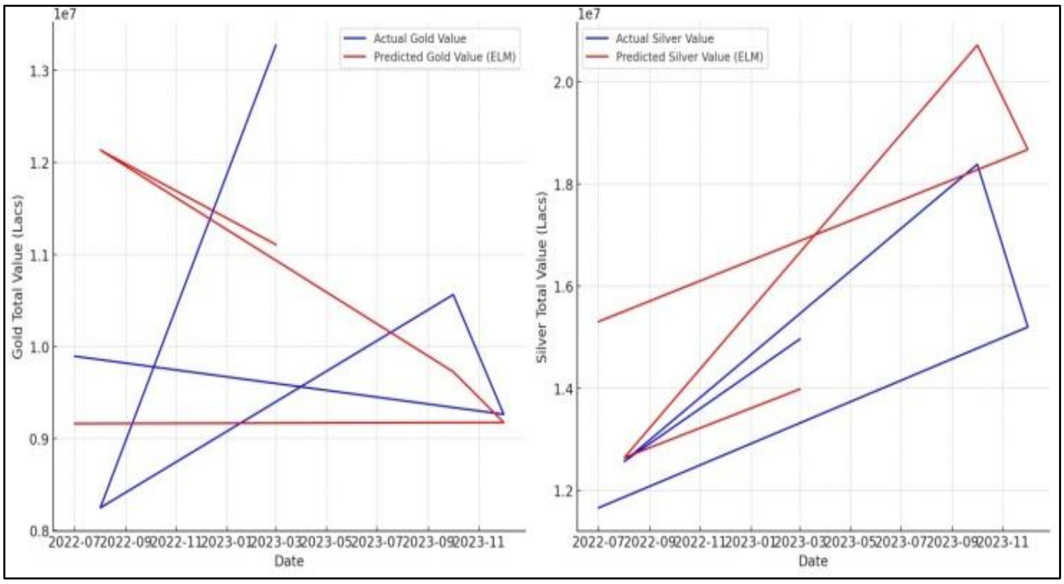

Fig 3 Gold and Silver Value Prediction Using Extreme Learning Machines

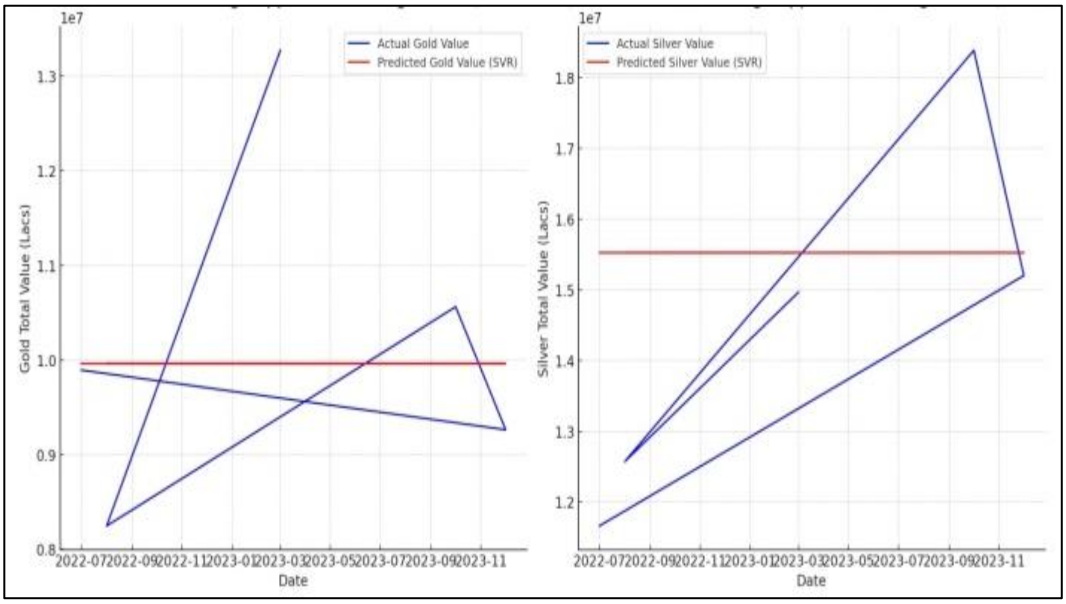

Fig 4: Gold and Silver Value Prediction Using Support Vector Regression (RBF Kernel)

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL292

V. CONCLUSION

This research has addressed the complex challenge of predicting the future values of gold and silver futures by employing advanced machine learning algorithms. Traditional econometric models, while valuable, often fall short of capturing the non-linear and dynamic nature of commodity prices. This study introduced and evaluated the effectiveness of four unconventional machine-learning algorithms: Gaussian Processes, Quantile Regression Forests, Extreme Learning Machines, and Support Vector Regression with an RBF kernel.

The comprehensive evaluation revealed that each of these algorithms has unique strengths in modeling the intricate patterns of gold and silver prices. Gaussian Processes, with their ability to provide uncertainty estimates, offer a nuanced understanding of predictions, which is crucial in the volatile financial markets. Quantile Regression Forests demonstrated robustness against outliers, making them particularly effective for financial data prone to extreme values. Extreme Learning Machines, known for their fast and efficient training processes, proved highly suitable for large-scale regression tasks. Lastly, Support Vector Regression with an RBF kernel showed the highest accuracy in predicting the future values of gold futures, highlighting its capability to capture complex non-linear relationships.

The results indicated that the Support Vector Regression (RBF kernel) model performed the best in predicting the total value of gold futures, while the Extreme Learning Machines model exhibited competitive performance in predicting the total value of silver futures. These findings underscore the potential of advanced machine learning techniques in enhancing the accuracy of financial predictions. By leveraging these sophisticated models, traders and investors can gain deeper insights into market trends, improving their decision-making processes and risk management strategies.

Despite the promising results, this research also highlights several areas for future exploration. One significant challenge identified is the need for high-quality and comprehensive data. Financial markets are influenced by many factors that are difficult to quantify and integrate into predictive models. Additionally, the interpretability of machine learning models remains a concern, as advanced algorithms often operate as black boxes, providing little insight into the rationale behind their predictions. Future research should focus on developing hybrid models that combine traditional econometric methods' interpretability with machine learning techniques' predictive power. This approach could provide a more transparent and reliable framework for commodity price forecasting.

In conclusion, this study has demonstrated the efficacy of using advanced machine learning algorithms for predicting the future values of gold and silver futures. Applying Gaussian Processes, Quantile Regression Forests, Extreme Learning Machines, and Support Vector Regression with an RBF kernel has shown significant improvements in prediction accuracy compared to traditional models. These advancements offer valuable tools for traders and investors, enabling them to navigate the complexities of the commodities market with greater confidence and precision. Further research and refinement of these models will likely yield even more accurate predictions, contributing to the ongoing evolution of financial forecasting methodologies.

REFERENCES

[1]. M. Ahmadi, N. Bashiri Behmiri, and M. Manera, "How is volatility in commodity markets linked to oil price shocks?," Energy Econ., vol. 59, pp. 11–23, 2016, doi: 10.1016/j.eneco.2016.07.006.

[2]. B. Kumar and A. Pandey, "Market efficiency in Indian commodity futures markets," J. Indian Bus. Res., vol. 5, no. 2, pp. 101–121, 2013, doi: 10.1108/17554191311320773.

[3]. S. Aboura, J. Chevallier, R. Jammazi, and A. K. Tiwari, "The place of gold in the cross-market dependencies," Stud. Nonlinear Dyn. Econom., vol. 20, no. 5, pp. 567–586, 2016, doi: 10.1515/snde-2015-0017.

[4]. I. Sengupta, C. Koner, N. K. Bhattacherjee, and S. Gupta, "Automated Student Merit Prediction using Machine Learning," Proc. - 2022 IEEE World Conf. Appl. Intell. Comput. AIC 2022, pp. 556–560, 2022, doi: 10.1109/AIC55036.2022.9848976.

[5]. A. K. Pradhan, I. Mittal, and A. K. Tiwari, "Optimizing the market-risk of major cryptocurrencies using CVaR measure and copula simulation," Macroecon. Financ. Emerg. Mark. Econ., vol. 14, no. 3, pp. 291–307, 2021, doi: 10.1080/17520843.2021.1909828.

[6]. B. Mondal and S. Gupta, "Execution Survey and State of the Art of Different ML-Based Ensemble Classifiers Approach Contextual Analysis of Spam Remark Location," in Proceedings of Third International Conference on Computing, Communications, and Cyber-Security. Lecture Notes in Networks and Systems, vol 421., Vol 421., M. (eds) Singh, P.K., Wierzchoń, S.T., Tanwar, S., Rodrigues, J.J.P.C., Ganzha, Ed. Springer, Singapore, 2023, pp. 311–323. doi: 10.1007/978-981-19-1142-2_24.

[7]. G. Buyukkara, C. C. Kucukozmen, and E. T. Uysal, "Optimal hedge ratios and hedging effectiveness: An analysis of the Turkish futures market," Borsa Istanbul Rev., vol. 22, no. 1, pp. 92–102, 2022, doi: 10.1016/j.bir.2021.02.002.

[8]. Z. H. Munim, C. S. Fiskin, B. Nepal, and M. M. H. Chowdhury, "Forecasting container throughput of major Asian ports using the Prophet and hybrid time series models," Asian J. Shipp. Logist., vol. 39, no. 2, pp. 67–77, 2023, doi: 10.1016/j.ajsl.2023.02.004.

[9]. J. Mann and J. N. Kutz, "Dynamic mode decomposition for financial trading strategies," Quant. Financ., vol. 16, no. 11, pp. 1643–1655, 2016, doi: 10.1080/14697688.2016.1170194.

[10]. F. Cabitza et al., "Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests," Clin. Chem. Lab. Med., vol. 59, no. 2, pp. 421–431, 2021, doi: 10.1515/cclm-2020-1294.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24JUL292

[11]. D. Shah, H. Isah, and F. Zulkernine, “Predicting the Effects of News Sentiments on the Stock Market,” Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018, pp. 4705–4708, 2019, doi: 10.1109/BigData.2018.8621884.
[12]. A. K. Singh, J. Patra, M. Chakraborty, and S. Gupta, “Prediction of Indian government stakeholder oil stock prices using hyper parameterized LSTM models,” 2022 Int. Conf. Intell. Controll. Comput. Smart Power, ICICCSP 2022, pp. 1–6, 2022, doi: 10.1109/ICICCSP53532.2022.9862425.
[13]. T. Mukherjee, S. Gupta, and A. Mitra, “Impact of share market based on global happening and future prediction,” in 6th Smart Cities Symposium (SCS 2022), 2022, pp. 384–389. doi: 10.1049/icp.2023.0601.
[14]. J. Patra, “Optimization of Intraday Trading in F & O on the NSE Utilizing BOLLINGER BANDS,” vol. 2, pp. 33–42, 2023, doi: 10.1007/978-981-99-3478-2_4.
[15]. I. Sengupta, S. Samanta, J. Patra, and S. Gupta, “Impact of Macroeconomic Indicators on the Indian Stock Market: A Study on NSE Nifty,” 2023 Int. Conf. Comput. Intell. Commun. Technol. Networking, CICTN 2023, vol. 16, pp. 275–282, 2023, doi: 10.1109/CICTN57981.2023.10140919.
[16]. P. Mukherjee et al., “Monitoring, Recognition and Attendance Automation in Online Class: Combination of Image Processing, Cryptography in IoT Security,” in Proceedings of International Conference on Network Security and Blockchain Technology., Singapore: Springer, 2022, pp. 18–27. doi: 10.1007/978-981-19-3182-6_2.
[17]. H. E. Kiziloz, “Classifier ensemble methods in feature selection,” Neurocomputing, vol. 419. pp. 97–107, 2021. doi: 10.1016/j.neucom.2020.07.113.
[18]. V. K. Joshi, “Impact of Fluctuation : Stock / Forex / Crude Oil on Gold,” SCMS J. Indian Manag., no. October-December, pp. 96–115, 2009.
[19]. Vashishtha and S. Kumar, “Development of financial derivatives market in India - A case study,” Int. Res. J. Financ. Econ., vol. 37, no. September, pp. 15–29, 2010.