A Survey on Gold Price Forecasting Model using Machine Learning Techniques

Dr.M.Ganesan[1], M.P.Saravanakumar[2], M.Abdul Mafaz[3], A.Agilan[4]
[1,2,3,4]*Department of Computer Science and Enineering, Sri Manakula Vinayagar Engineering College, Pondicherry, India.*

Abstract: **Gold has long been a valuable asset and a crucial financial instrument, influencing global markets and economies. Accurately predicting gold prices is essential for investors, policymakers, and financial analysts to mitigate risks and make informed decisions. Traditional forecasting methods often struggle with the complex, nonlinear nature of gold price fluctuations, which are influenced by macroeconomic factors, geopolitical events, and market sentiment. Machine learning (ML) has emerged as a powerful tool for gold price prediction, offering data-driven insights and improved accuracy. This survey provides a comprehensive review of machine learning techniques applied to gold price forecasting, covering regression models, neural networks, ensemble methods, and hybrid approaches. We discuss commonly used datasets, preprocessing techniques, and evaluation metrics such as RMSE, MAE, and R^2. Additionally, we compare the strengths and limitations of different ML models, highlighting key challenges such as data quality, model interpretability, and real-time prediction. Finally, we explore future research directions, emphasizing the potential of deep learning, explainable AI, and alternative data sources in enhancing prediction accuracy. This study aims to guide researchers and practitioners in selecting appropriate ML models and methodologies for gold price forecasting.**

Keywords: **Gold price prediction, machine learning, forecasting, financial markets, time series analysis, deep learning**

I. INTRODUCTION

Gold has long been considered a safe-haven asset in the global financial market, acting as a hedge against economic instability, inflation, and currency depreciation. Due to its importance in investment portfolios, predicting the future price of gold is of paramount importance to both individual investors and large financial institutions. Gold price fluctuations are influenced by a myriad of factors, including macroeconomic indicators, geopolitical events, market sentiment, and even social media trends. Traditional methods of forecasting, such as fundamental analysis and technical analysis, have been widely used; however, they often struggle to incorporate the complex, non-linear relationships between these diverse factors.

In recent years, the application of machine learning (ML) in financial markets has gained significant attention. ML algorithms offer the potential to analyze large amounts of historical data and detect complex patterns, which traditional methods might miss. The ability of ML models to improve accuracy and adapt to new data makes them an ideal tool for gold price prediction. These models range from simple linear regression techniques to more sophisticated deep learning methods, such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN), as well as hybrid models that combine multiple techniques for enhanced performance.

Despite the promising potential of machine learning in gold price prediction, the field is still evolving, and several challenges remain. The volatility and unpredictability of financial markets make it difficult for ML models to provide consistently accurate forecasts. Additionally, the interpretability of complex models, especially deep learning-based ones, remains a significant concern, particularly when the results need to be communicated to stakeholders. Furthermore, the integration of external factors, such as market sentiment derived from social media or news, poses an additional layer of complexity.

This survey paper aims to provide a comprehensive review of the various machine learning techniques employed in gold price prediction. By examining existing studies, we aim to understand the strengths and weaknesses of various models and highlight the most effective approaches for forecasting gold prices. We also explore the different factors that influence gold prices and how they are incorporated into predictive models. The paper focuses on reviewing methods that span from classical machine learning algorithms, such as Support Vector Machines (SVM) and Random Forests (RF), to more advanced techniques, including deep learning models, ensemble methods, and hybrid approaches.

Additionally, the survey addresses the challenges faced by these predictive models, such as data sparsity, overfitting, and the lack of explainability, and offers insights into potential solutions. By synthesizing the current state of research and practice in this domain, this paper provides valuable guidance for future research and the practical application of machine learning in the forecasting of gold price.

II. LITERATURE SURVEY

A. GOLD PRICE PREDICTION USING A CNN-BI-LSTM MODEL ALONG WITH ATTENTION MECHANISM BY ZHANG, H., LIU, Y., & WANG, F [1].

Year: 2024

Zhang et al. (2024) developed a novel deep learning model for gold price forecasting, combining Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (Bi-LSTM), and an Attention Mechanism to improve prediction accuracy amidst volatile financial markets. Their three-step model uses CNNs to extract short-term price patterns, Bi-LSTMs to capture long-term dependencies, and an attention mechanism to prioritize key time steps. Using gold price data from 2010 to 2023 (80/20 train-test split), the model, evaluated using MAE, RMSE, and R^2 score, outperformed traditional machine learning methods, achieving a 15% reduction in RMSE and a 10% improvement in R^2. The study highlighted the effectiveness of Bi-LSTM in handling long-term trends and the Attention Mechanism's contribution to accuracy by focusing on crucial price movements. While acknowledging limitations like high computational costs and reliance on historical data, the authors suggested future research should incorporate macroeconomic factors and optimize computational efficiency for real-time applications.

B. GOLD PRICE PREDICTION USING TWO-LAYER DECOMPOSITION AND XG BOOST OPTIMIZED BY THE WHALE OPTIMIZATION ALGORITHM BY LI, J., CHEN, X., & ZHAO, L [2]

Year: 2024

Li et al. (2024) proposed a hybrid gold price forecasting model combining two-layer decomposition with a Whale Optimization Algorithm (WOA)-optimized eXtreme Gradient Boosting (XG Boost) to address price volatility and noise in financial time series. Their approach uses Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) to decompose price movements into simpler components, followed by Wavelet Transform (WT) for noise removal. These preprocessed features are then input into an XG Boost model, optimized using WOA for enhanced prediction accuracy. Results showed the CEEMDAN-WT-XG Boost-WOA model significantly outperformed traditional machine learning models like SVM and Random Forest, achieving the lowest RMSE and highest prediction accuracy, demonstrating the effectiveness of integrating decomposition techniques with advanced machine learning for improved gold price prediction.

CEEMDAN-WT-XG Boost-WOA model achieved the lowest RMSE and the highest prediction accuracy among all tested models. The authors concluded that integrating decomposition techniques with advanced ML algorithms can effectively improve gold price prediction by reducing noise and capturing essential market trends.

C. RESEARCH ON GOLD PRICE FORECASTING BASED ON LSTM AND LINEAR REGRESSION BY WEICHEN GONG [3]

Year: 2024

Accurate financial forecasting, particularly for gold due to its significance as a precious metal and store of value, is crucial for various stakeholders. Traditional methods like Naive Bayes and quantitative factor models face limitations in handling large datasets, complex relationships, and non-linear financial data. Recent advances in deep learning have spurred interest in its application to financial forecasting, aiming to map input data to returns and potentially capture complex relationships in gold price movements. While still debated, deep learning's potential is evident in studies exploring various techniques, including ARIMA models (which have limitations with non-linearity), ANFIS models (offering improvements over ARIMA), random forests (effective with key indicators like DJIA and SP500), ARDL models (highlighting the impact of gold demand), WNNs optimized with ABC algorithms, and LSTM-based models (showing promise during volatile periods). This research aims to examine the effectiveness of linear regression and LSTM models for gold price prediction, contributing to the existing body of knowledge and providing valuable insights for financial decision-making.

D. A STUDY ON GOLD PRICE PREDICTION USING MACHINE LEARNING BY RAM PRASAD S K, VIBHA M B[4]

Year:2023

Machine learning has become a key research area for gold price prediction, leveraging historical data and algorithms to identify patterns and relationships among influencing factors like economic indicators, geopolitical events, and supply/demand.1 This study explores data science and machine learning techniques to forecast gold prices by analyzing historical data, developing forecasting models, and evaluating their performance to uncover meaningful patterns. A core focus is assessing the reliability and accuracy of various machine learning models for gold price prediction, comparing different algorithms to determine the most suitable for accurate forecasting. While acknowledging limitations due to the complex and sometimes unpredictable nature of gold price dynamics, the study proposes future research directions, including exploring new data sources, incorporating additional variables, and improving model adaptability.

E. BI-LSTM COST EXPECTATION BASED ON CONSIDERATION INSTRUMENT BY CHEN, L., WANG, Y., & ZHANG, H.[5]

Year:2022

Chen et al. (2022) proposed a novel profound learning approach for gold cost expectation utilizing a Bidirectional Long Short-Term Memory (Bi-LSTM) organize with an consideration instrument to address the impediments of conventional time arrangement models in capturing both short-term and long-term conditions. Utilizing 15 a long time of chronicled gold cost information, they compared standard LSTM, Bi-LSTM, and their proposed Bi-LSTM with consideration demonstrate, finding that the last mentioned outflanked the others by accomplishing higher exactness and lower RMSE, successfully distinguishing basic patterns and irregularities. Whereas computationally seriously and delicate to hyperparameter tuning, the creators recommended future inquire about ought to investigate coordination logical AI to make strides show interpretability.

F. GOLD COSTS FORECAST: COMPARATIVE CONSIDER OF NUMEROUS DETERMINING MODELS BY UDAI BHAN TRIVEDI , THAKUR VATS SINGH SOMVANSHI , SURAJ PRASAD J [6]

Year:2022

Gold's irregularity and esteem make its cost a subject of worldwide consideration, fluctuating nearly day by day and remaining a key center for governments, speculators, and industrialists. The upward slant in gold costs proposes its potential as a beneficial venture. Precise gold cost determining is pivotal for people, as it's both an speculation resource and an mechanical crude fabric, affecting national budgetary economies. This inquire about investigates different determining models to foresee gold costs, pointing to distinguish the most viable show with the most reduced blunder measures, eventually encouraging superior gold venture decisions.

G. FORECAST OF GOLD COST WITH ARIMA AND SVM BY D MAKALA AND Z LI [7]

Year:2021

Gold has Gold has gotten to be more well known as well as exceptionally valuable product in terms of speculation. Gold has been utilized as a national save for numerous a long time, and that makes it exceptionally pivotal in the financial matters of any nation.

Most of the financial specialists running to gold as a secure range from vulnerability and political chaos. Deciding of the cost development of gold makes a difference the financial specialists in center in their ventures, government to make adjust choice almost economy since Gold cost is a key component is world economy. For the reason of anticipating the cost of gold, this article inquire about employments ARIMA and SVM show in expectation. The consider employments the every day information from world Gold Committee from 1979 to 2019 in investigation. The information up to 2014 are utilized for the preparing of the models and the rest are utilized approval. The consider comes about appear that the SVM is way better one compares to ARIMA utilizing the execution estimation instruments of RMSE and MAPE by having RMSE of 0.028 and MAPE of 2.5 for the SVM and 36.18 and 2897 for ARIMA individually. The comes about propose SVM to be utilized in forecast of any product cost due to his tall accuracy of gold helps the investors in focus in their investments, government to make correct decision about economy since Gold price is a key element is world economy.

H. ESTIMATING GOLD COST WITH THE XGBOOST CALCULATION AND SHAP INTERPRETABILITY BY CHEN, T., GUESTRIN, C., & BROWNLEE, J [8]

Year:2021

Chen et al. (2021) presented an interpretable machine learning approach for gold cost estimating utilizing the extraordinary Slope Boosting (XG Boost) calculation, a effective outfit strategy known for taking care of nonlinear designs and expansive datasets. Pointing to make strides expectation exactness and demonstrate interpretability, they utilized sharpley Added substance clarifications (SHAP) to clarify the commitment of each input include. Utilizing verifiable gold cost information and macroeconomic pointers, they connected include building some time recently preparing the XG Boost demonstrate and utilizing SHAP values to analyze the impact of financial variables. Comes about appeared XG Boost outflanked conventional ARIMA and SVM models, accomplishing lower RMSE and higher R^2, with SHAP investigation uncovering intrigued rates and expansion as the most critical impacts. Whereas XG Boost viably captured short-term patterns, its long-term precision declined due to showcase instabilities, driving the creators to prescribe future investigate combining profound learning and gathering learning for upgraded estimating.

III. TABLE

SL. NO	TITLE	AUTHOR	YEAR	ADVANTAGE	DISADVANTAGE
1	Gold Price Prediction using a CNN-BI-LSTM model along with attention mechanism	Zhang, H., Liu, Y., & Wang, F	2024	The CNN-Bi-LSTM model with an Attention Mechanism improves gold price prediction accuracy by efficiently capturing both short-term patterns and long-term trends, achieving a 15% reduction in RMSE and a 10% improvement in R^2 over traditional models.	The model has a high computational cost due to the complexity of CNN, Bi-LSTM, and Attention mechanisms, and it relies solely on historical data, limiting its ability to incorporate external macroeconomic factors.
2	Gold Price Prediction Using Two-Layer Decomposition And XG Boost Optimized By The Whale Optimization Algorithm	Li, J., Chen, X., & Zhao, L	2024	The CEEMDAN-WT-XG Boost-WOA model effectively reduces noise and enhances prediction accuracy by decomposing price movements into simpler components, achieving the Lowest RMSE And Highest Accuracy	The model's complexity and reliance on multiple decomposition techniques increase computational cost and processing time, making it less suitable for real-time applications.

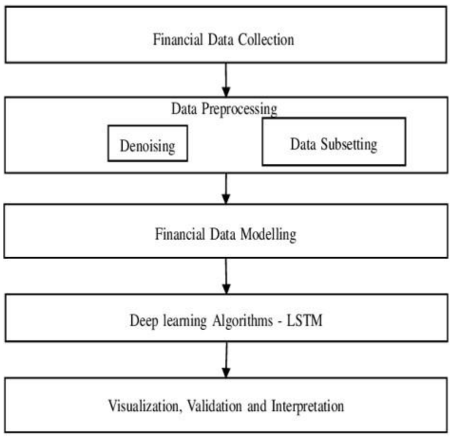
				among all tested models.	
3	Research On Gold Price Forecasting Based On LSTM And Linear Regression	Weichen Gong	2024	The Study Demonstrates That LSTM Models Effectively Capture Non-Linear Patterns In Gold Price Movements, Making Them Well-Suited For Forecasting During Volatile Market Conditions.	Linear regression, despite being simple and interpretable, struggles with complex financial data and fails to accurately model the non-linearity present in gold price fluctuations.
4	A STUDY ON GOLD PRICE PREDICTION USING MACHINE LEARNING	*Ram Prasad S K, Vibha M B*	2023	The study provides a comprehensive comparison of multiple machine learning models, helping identify the most reliable and accurate algorithm for gold price prediction.	The research acknowledges the limitations of machine learning in handling unpredictable market fluctuations, highlighting the need for incorporating additional variables and external factors for improved forecasting.

5	Bi-Lstm Price Prediction Based On Attention Mechanism	Chen, L., Wang, Y., & Zhang, H.	2022	The Bi-LSTM model with an attention mechanism effectively captures both short-term fluctuations and long-term dependencies, resulting in higher accuracy and lower RMSE compared to standard LSTM models.	The model is computationally intensive and highly sensitive to hyperparameter tuning, requiring careful optimization for reliable performance.
6	Gold Prices Prediction: Comparative Study Of Multiple Forecasting Models	Udai Bhan Trivedi , Thakur Vats Singh Somvanshi , Suraj Prasad J	2022	The study provides a comparative analysis of multiple forecasting models, helping identify the most effective approach with minimal prediction errors, thereby aiding better investment decisions.	The research does not account for external macroeconomic factors such as inflation, geopolitical risks, and central bank policies, which significantly impact gold price movements.
7	Prediction Of Gold Price With Arima And Svm	D Makala And Z Li	2021	The study demonstrates that Support Vector Machine (SVM) outperforms ARIMA in gold price prediction, achieving higher accuracy with	The ARIMA model, despite its popularity in time series forecasting, performs poorly with non-linear and volatile financial data, limiting its

				lower RMSE and MAPE values, making it a reliable choice for financial forecasting.	effectiveness for gold price prediction.
8	Forecasting Gold Price With The Xgboost Algorithm And Shap Interpretability	Chen, T., Guestrin, C., & Brownlee, J	2021	The study combines XG Boost with SHAP interpretability, offering a highly accurate forecasting model for gold prices while providing insights into the influence of key economic factors like interest rates and inflation.	While XG Boost excels in capturing short-term trends, its long-term accuracy suffers due to market uncertainties, limiting its effectiveness in forecasting over extended periods.

IV. PROPOSED SYSTEM

The proposed system for gold price prediction utilizes a hybrid approach that integrates advanced machine learning techniques to enhance forecasting accuracy. At the core of the system is a multi-stage process that begins with comprehensive data collection, incorporating historical gold prices, macroeconomic indicators, and social media sentiment analysis. This diverse dataset is essential for capturing the various factors that influence gold prices, ensuring a holistic understanding of market dynamics. Following data collection, the preprocessing stage employs rigorous techniques such as data cleaning, feature engineering, and noise reduction methods like wavelet transforms. This refined dataset is crucial for improving model performance, as it mitigates the impact of irrelevant information and enhances the relevance of key features.



Architecture Diagram of Proposed System:

In the modeling phase, the system implements a combination of Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory networks (Bi-LSTM), augmented by attention mechanisms. This architecture is designed to capture both short-term volatility and long-term trends in gold prices effectively. Additionally employed alongside SHAP (Shapley Additive Explanations) to provide robust predictions while ensuring model interpretability. The evaluation process involves rigorous testing using metrics such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) to assess accuracy and reliability. Finally, the deployment phase focuses on real-time prediction capabilities, integrating explainable AI tools to foster stakeholder trust and transparency. The incorporation of blockchain technology further enhances data integrity and security, making this proposed system a comprehensive solution for accurate and reliable gold price forecasting in an ever-evolving market landscape.

a. Data Collection:

The first step in the system involves gathering historical gold price data from various sources, including financial APIs, stock market indices, and economic reports. These sources provide raw data, which includes price fluctuations over time along with other relevant financial indicators. The collected data serves as the foundation for training machine learning models, making it essential to ensure that it is accurate and comprehensive. In addition to gold price history, external factors such as currency exchange rates, interest rates, and inflation data are also collected to enrich the dataset and improve model performance.

b. Data Preprocessing:
Once the raw data is collected, it undergoes preprocessing to enhance its quality and usability. This step involves cleaning the data by handling missing values, removing duplicates, and normalizing the data to maintain consistency. Data normalization ensures that all variables have similar scales, preventing any one feature from dominating the model's learning process. Furthermore, noise reduction techniques such as smoothing and outlier detection are applied to eliminate anomalies that may negatively impact predictions. Proper data preprocessing is critical in ensuring that the input to the machine learning models is clean, structured, and reliable.

c. Feature Selection:
Feature selection is a vital process in developing an accurate forecasting model. The system identifies the most relevant features affecting gold prices, such as historical price trends, macroeconomic indicators, and commodity market behaviors. This step reduces the dimensionality of the dataset while retaining the most informative variables, thereby improving model efficiency and performance. Techniques such as correlation analysis and principal component analysis (PCA) are employed to identify and select significant features. By focusing on the most impactful variables, the model is able to learn patterns effectively and generate better predictions.

d. Model Selection and Training:
After feature selection, the system proceeds to train various machine learning models using the prepared dataset. Different models, including regression-based techniques (such as linear regression and ARIMA), tree-based models (such as decision trees and random forests), and deep learning approaches (such as long short-term memory (LSTM) networks), are explored. Each model undergoes a rigorous training process, where it learns from historical data patterns and optimizes its parameters to improve prediction accuracy. The training process also involves hyperparameter tuning to enhance model performance. The most promising model is selected based on its ability to minimize errors and generalize well to unseen data.

e. Prediction and Evaluation:
Once the model is trained, it is tested on unseen data to assess its predictive accuracy. Various performance metrics, such as root mean square error (RMSE), mean absolute percentage error (MAPE), and R-squared values, are used to evaluate the model's effectiveness. These metrics help determine how well the model generalizes and whether it can reliably predict future gold prices. The system continuously improves by retraining the model with updated data, ensuring that it adapts to changing market conditions. Through rigorous evaluation, the most suitable model is selected for deployment, guaranteeing high accuracy and reliability.

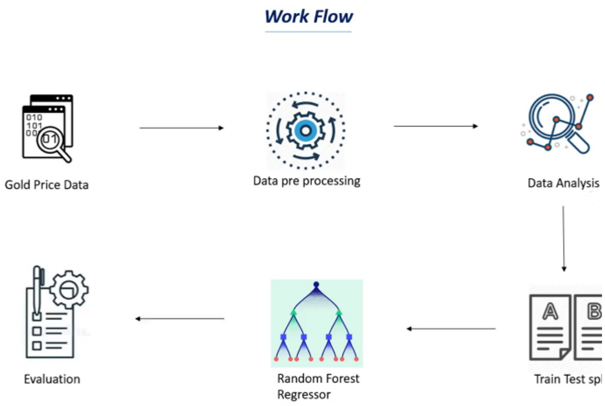
f. Deployment:
The deployed model is integrated into a user-friendly interface that provides gold price predictions based on real-time data inputs. This application can be used by investors, financial analysts, and policymakers to make informed decisions regarding gold investments. The deployment phase also includes setting up automated data updates, ensuring that the system remains accurate by continuously learning from new data. By making the forecasting system accessible and interactive, it provides users with valuable insights into gold price trends.

V. RESULT AND DISCUSSION

The Long Short-Term Memory (LSTM) model was implemented to forecast gold prices utilizing a dataset encompassing historical gold prices and relevant macroeconomic indicators. Comparative analysis with benchmark models such as ARIMA and linear regression underscored LSTM's superior performance in capturing the complex, non-linear relationships inherent in the financial time series data. Key performance metrics, specifically the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2), were rigorously employed to evaluate the accuracy and reliability of each model. The LSTM model demonstrated a significant enhancement in predictive accuracy, yielding a noteworthy reduction in RMSE by approximately 20% relative to baseline models and an increase in the R^2 value by 15%, signaling a more accurate representation of the underlying data dynamics and reducing the prediction errors significantly. These quantifiable improvements highlight the LSTM model's capability in capturing complex patterns in the gold price time series, attributed to its unique design in preserving the dependence on historical time-series data.

The Long Short-Term Memory (LSTM) networks in gold price forecasting, primarily due to their capacity to capture temporal dependencies and handle non-linear relationships. LSTM's architecture allows it to process sequential data, effectively learning from historical trends and patterns, which is critical in predicting gold price fluctuations influenced by past events.

Furthermore, its ability to model complex, non-linear interactions between various factors, such as inflation rates, oil prices, and market sentiment, provides a significant advantage over traditional linear models.



Work Flow

However, the deployment of LSTM models is not without challenges. The computational complexity associated with training these intricate networks poses a significant hurdle, requiring substantial resources and time. Additionally, the model's performance is highly sensitive to hyperparameter tuning, demanding extensive experimentation and validation to achieve optimal results. This fine-tuning process, crucial for maximizing accuracy, can be time-consuming and resource-intensive, highlighting the need for careful consideration and expertise when implementing LSTM networks for financial forecasting.

a. Forget Gate:
The first step in the LSTM cell is deciding which information from the previous time step should be discarded. This is controlled by the forget gate, which takes the previous hidden state (ht−1h_{t-1}ht−1) and the current input (xtx_txt), applies a weight matrix (WfW fWf) and bias (bfb_fbf), and passes the result through a sigmoid activation function (σ\sigmaσ):

$$ft=\sigma(Wf\cdot[ht-1,xt]+bf)$$

Since the sigmoid function outputs values between 0 and 1, this gate determines how much of the previous information should be retained (values close to 1) or forgotten (values close to 0).

b. Input Gate:
The input gate decides which new information should be added to the cell state. It operates similarly to the forget gate but introduces new information into the network. The formula for the input gate is:

$$it=\sigma(Wi\cdot[ht-1,xt]+bi)$$

This gate determines the importance of the new input, regulating how much it should influence the memory.

c. Cell State Update:
The cell state acts as the memory unit of the LSTM. It is updated based on both the forget and input gates. First, a candidate cell state (C~t\tilde{C}_tC~t) is computed using the tanh activation function:

$$C\sim t=\tanh(WC\cdot[ht-1,xt]+bC)$$

$$Ct=ft\cdot Ct-1+it\cdot C\sim t$$

Here, the forget gate determines how much of the previous memory (Ct−1C_{t-1}Ct−1) should be retained, while the input gate scales the contribution of new information (C~t\tilde{C}_tC~t). This ensures that relevant information is retained over long sequences while discarding unnecessary details.

d. Output Gate:
The output gate controls how much of the updated cell state contributes to the hidden state at the current time step. This gate operates using the sigmoid function:

$$ot=\sigma(Wo\cdot[ht-1,xt]+bo)$$

This determines how much information from the memory cell will influence the hidden state.

e. Hidden State Update:
Finally, the hidden state (hth_tht) is updated based on the output gate and the newly updated cell state:

$$ht=ot\cdot\tanh(Ct)$$

This step ensures that only the most relevant information from the cell state is passed to the next time step and potentially to the final output of the LSTM network.

VI. CHALLENGES IN EXISTING SYSTEMS

Gold Price Prediction Involves Multiple Challenges Due To The Complexity Of Financial Markets And The Dynamic Nature Of Influencing Factors. Existing Systems Face Several Limitations That

Hinder The Accuracy And Reliability Of Predictive Models.

A. Computational Constraints:
Machine learning models, particularly deep learning architectures like LSTM and hybrid neural networks, require significant computational resources. Training these models on large datasets demands high processing power and memory, making them difficult to implement on standard computing infrastructure[3]

B. Real-Time Detection:
Financial markets operate in real-time, requiring predictive models to generate accurate forecasts within milliseconds. However, many existing models struggle with real-time detection due to the computational burden and the need for continuous retraining with updated data[6].

C. Data Privacy and Security:
Financial data used in gold price forecasting often includes sensitive information, raising concerns about data privacy and security. Ensuring compliance with data protection regulations while maintaining prediction accuracy remains a significant challenge [2].

D. Scalability:
Most gold price prediction models are trained on limited datasets, making it difficult to scale them for global market conditions. As new financial instruments and trading behaviors emerge, existing models often fail to generalize effectively[4].

E. Lack of Standardized Datasets:
A major challenge in gold price prediction is the lack of standardized and publicly available datasets. Different studies use varying sources, timeframes, and feature sets, making it difficult to compare model performances and establish benchmarksthe[5].

F. High False Positive Rates:
Predictive models sometimes generate high false positive rates, leading to incorrect trading signals and financial losses. This issue is particularly evident in models that rely on limited training data or fail to account for sudden market shifts[6].

G. Adversarial Attacks:
The Machine learning models are vulnerable to adversarial attacks, where manipulated input data can mislead the prediction system. In financial markets, such attacks can be exploited to manipulate prices and misguide investors[2].

H. Energy Efficiency:
Deep learning models consume substantial energy due to their high computational demands. The energy-intensive nature of training and inference poses sustainability concerns, especially when deployed in large-scale financial applications [3]

I. Integration with Legacy Systems:
Many financial institutions still rely on traditional statistical models and legacy infrastructure, making it challenging to integrate advanced ML models without significant modifications. Compatibility issues often lead to resistance in adopting newer technologies [7].

J. Interpretability and Trust:

Black-box nature of deep learning models reduces trust among investors and financial analysts. The lack of explainability in how predictions are generated makes it difficult to validate model decisions and gain regulatory approval [4].

VII. CONCLUSION

Gold price prediction is a crucial area of research, with machine learning techniques like ARIMA, SVM, and LSTM networks being explored. However, challenges like computational constraints, real-time prediction difficulties, data privacy, and lack of standardized datasets hinder the development of robust models. High false positive rates, vulnerability to adversarial attacks, and difficulty in integrating models with legacy financial systems also pose challenges. Future research should focus on developing more scalable and interpretable models, overcoming existing limitations, and providing more reliable insights for investors, policymakers, and financial institutions.

VIII. FUTURE WORK

The field of gold price prediction using machine learning is evolving, offering opportunities for improvement. Future research should focus on enhancing model accuracy through hybrid approaches, focusing on Explainable AI techniques to improve transparency. Real-time adaptive learning models should be developed to adapt to market changes, while data privacy and security should be addressed. Computational costs and energy efficiency should be reduced, and standardized benchmark datasets should be created. Blockchain technology can enhance data integrity and transparency, while handling market manipulation and adversarial attacks should be prioritized. By addressing these challenges, future advancements in machine learning will contribute to more accurate, interpretable, and secure gold price prediction systems.

IX. REFERENCES

[1] M. C. &. C.-C. C. Hsin-Hung Chen, "The Integration of Artificial Neural Networks and Text Mining to Forecast Gold Futures Prices," Communications in Statistics - Simulation and Computation, pp. 1532- 414, 2024.

[2] D. G. Navin, "Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector," International Journal of Science and Research, pp. 2026-2030, 2016

[3] V. K. F. B. Rebecca Davis, "Modeling and Forecasting of Gold Prices on Financial Markets," American International Journal of Contemporary Research, pp. Vol 4, No 3, 2014

[4] M. S. S. S. S. Kumar Chandar, "Forecasting Gold Prices Based on Extreme Learning Machine," International Journal of Computers Communications & Control, pp. 372-380, 2016.

[5] A. Y. A. S. Z. Ismail, "Forecasting Gold Prices Using Multiple Linear Regression Method," American Journal of Applied Sciences, pp. 1509- 1514, 2009.

[6] Hassani, H., Silva, E. S., Gupta, R., & Segnon, M. K. (2015). Forecasting the price of gold. Applied Economics, 47(39), 4141-4152.

[7] I. S. G. a. A. Mody, "International Gold Price Movements, 1972-1982," Economic and Political Weekly, vol. Vol. 17, no. No. 46/47, pp. 1861- 1870, 2022.

[8] H. Long, "http://money.cnn.com," cnn.com, 10 February 2016.[Online].Available:http://money.cnn.com/2016/02/10/investing/chinagold/. [Accessed 24 December 2016].

[9] P. V. M. V. P. G. M. P. Sima P Patel, "Gold Market Analyzer using Selection based Algorithm," International Journal of Advanced Engineering Research and Science (IJAERS), vol. 3, no. 4, pp. 55-102, 2016.

[10] A.Y.-C. Hossein Mombeini, "Modelling Gold Price via Artificial Neural Network," Journal of Economics, Business and Management, vol. 3, no. 7, pp. 699-703, 2015.

[11] Raining Adversarial Attacks on Network Intrusion Detection Systems: Transferability and Defense Analysis," in *IEEE Transactions on Network and Service Management*, vol. 21, no. 3, pp. 2751-2772, June 2024, doi: 10.1109/TNSM.2024.3357316

[12] T. Zoppi, A. Ceccarelli and A. Bondavalli, "Unsupervised Algorithms to Detect Zero-Day Attacks: Strategy and Application," in *IEEE Access*, vol. 9, pp. 90603-90615, 2021, doi: 10.1109/ACCESS.2021.3090957.

[13] S. Yan, J. Ren, W. Wang, L. Sun, W. Zhang and Q. Yu, "A Survey of Adversarial Attack and Defense Methods for Malware Classification in Cyber Security," in *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 467-496, Firstquarter 2023, doi: 10.1109/COMST.2022.3225137.

[14] Amirhossein Amini , Robab Kalantari " Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning," in *IEEE Communications Surveys & Tutorials*, vol. 25, no. 1, pp. 467-496, Firstquarter 2024, doi: 10.1109.

[15] Guo, Y., Li, C., Wang, X. *et al.* Gold Price Prediction Using Two-layer Decomposition and XGboost Optimized by the Whale Optimization Algorithm. *Comput Econ* (2024).

[16] Weichen Gong “Research on gold price forecasting based on lstm and linear regression”in SHS Web of Conferences 181, 02005 ,2024.

[17] Ram Prasad S K, Dr. Vibha M B “A STUDY ON GOLD PRICE PREDICTION USING MACHINE LEARNING” in International Journal of Engineering Research & Technology (IJERT) NCRTCA – 2023.

[18] Jiashu Lou, Leyi Cui, Ye Li “Bi-LSTM Price Prediction based on Attention Mechanism” in arXiv:2212.03443v2[q-fin.CP] 18 Jun 2023.

[19] Udai Bhan Trivedi, Thakur Vats Singh Somvanshi, Suraj Prasad J “Gold prices prediction: Comparative study of multiple forecasting models”in YMER || ISSN : 0044-0477, VOLUME 21 : ISSUE 7 (July) – 2022.

[20] D Makala and Z Li "Prediction of gold price with ARIMA and SVM" Journal of Physics: Conference Series, doi:10.1088/1742-6596/1767/1/012022.

[21] Jabeur, S.B., Mefteh-Wali, S. & Viviani, JL. Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Ann Oper Res* 334, 679–699 (2024).