ISSN 2277-2685
IJESR/April-June. 2026/ Vol-16/Issue-2/1168-1173
Shaik Abdul Aziz *et. al.*, / International Journal of Engineering & Science Research

Full Length Article

Predictive Analysis of Gold Prices Using Machine Learning Approaches

Shaik Abdul Aziz[1], Mohammed Abdul Razzak[2], Mohammed Safwan[3], Bhargavi[4]
[123]B.E Students; Department of IT, Lords Institute of Engineering and Technology, Hyderabad, India.
[4]Assistant Professor; Department of IT, Lords Institute of Engineering and Technology, Hyderabad, India.
Mail Id; bhargavi@lords.ac.in[4]
Accepted 30-04-2026

Abstract:
Gold is one of the most important financial assets, and accurately predicting its price is a challenging task due to the influence of various economic and market-related factors. This study proposes an intelligent gold price prediction system that combines machine learning and deep learning techniques to improve forecasting performance. The proposed framework integrates structured financial data, including historical gold prices, inflation rates, interest rates, and exchange rates, with unstructured information such as financial news and social media sentiment to better capture market trends and investor behavior.
Among the models evaluated, the Long Short-Term Memory (LSTM) model achieved the best performance, with an accuracy of 94.2%, a Root Mean Square Error (RMSE) of 2.31, and a Mean Absolute Error (MAE) of 1.87, outperforming traditional machine learning models such as Linear Regression and Support Vector Machine (SVM). The proposed approach also incorporates sentiment analysis and Explainable Artificial Intelligence (XAI), enabling not only more accurate predictions but also greater transparency in understanding the factors influencing the model's decisions. The results demonstrate that the proposed framework is an effective, reliable, and scalable solution for real-time gold price forecasting and can support investors, analysts, and financial institutions in making informed decisions.

Keywords: *Gold Price Prediction, Machine Learning, Deep Learning, LSTM, Natural Language Processing (NLP), Sentiment Analysis, Time Series Forecasting, Explainable AI (XAI), Economic Indicators*

Introduction

Gold has long been regarded as a valuable asset in the global economy due to its stability and ability to retain value during periods of inflation, currency fluctuations, and economic uncertainty. As a result, accurately forecasting gold prices has become an important research area for investors, financial institutions, policymakers, and market analysts. However, gold price prediction remains a challenging task because it is influenced by a combination of economic indicators, geopolitical events, monetary policies, and market sentiment, all of which continuously interact in complex ways [1], [3], [2], [4].

Several factors, including inflation rates, interest rates, exchange rates, and overall economic conditions, directly affect gold prices. These variables exhibit nonlinear relationships and strong temporal dependencies, making accurate forecasting difficult using conventional statistical approaches. Traditional models such as Linear Regression and ARIMA are based on linear assumptions and often struggle to capture the complex and dynamic nature of financial markets, particularly when multiple influencing factors change simultaneously [1], [2].

Recent advances in Artificial Intelligence (AI) have significantly improved the ability to analyze financial time-series data. Machine learning and deep learning algorithms can process large volumes of historical data, identify hidden patterns, and adapt to changing market conditions more effectively than traditional forecasting techniques. Among these approaches, Long Short-Term Memory (LSTM) networks have gained considerable attention because they can learn long-term temporal dependencies and model nonlinear relationships, making them well suited for financial forecasting applications [3], [5], [6].

Besides numerical financial indicators, textual information such as financial news, social media discussions, and economic reports also has a substantial impact on investor sentiment and market behavior. Natural Language Processing (NLP) techniques make it possible to extract meaningful sentiment information from these unstructured data sources, providing additional insights that improve prediction performance when combined with structured financial data [1], [5], [12].

Motivated by these advancements, this study presents an AI-driven framework for gold price prediction that integrates machine learning, deep learning, and NLP-based sentiment analysis into a unified forecasting model. The proposed framework combines structured financial indicators with sentiment extracted from unstructured textual data to improve prediction accuracy and robustness. In addition, Explainable Artificial Intelligence (XAI) techniques are incorporated to enhance the transparency and interpretability of the prediction process, enabling users to better understand the factors influencing the model's decisions. By addressing the limitations of existing approaches, the proposed framework provides a reliable, scalable, and practical solution for real-time gold price forecasting and financial decision-making [1], [3], [5], [10], [11].

Related Work

A. Machine Learning for Gold Price Prediction

Several researchers have explored the application of machine learning techniques for predicting gold prices by utilizing historical market data and economic indicators. Nithya *et al.* [1] developed a Random Forest Regressor-based prediction framework that incorporated economic indicators and foreign exchange data. Their work highlighted the importance of feature engineering and exploratory data analysis in improving model performance. Although the model produced accurate predictions, it relied solely on structured numerical data and did not consider the influence of market sentiment derived from financial news or social media.

Narendran [2] performed a comparative study of several forecasting techniques, including ARIMA, Gaussian Process Regression (GPR), Long Short-Term Memory (LSTM), XGBoost, LightGBM, and CatBoost. The results demonstrated that deep learning and boosting-based models were more effective than conventional statistical approaches in capturing nonlinear relationships within financial time-series data. However, the proposed framework focused only on historical numerical data and did not integrate unstructured information, limiting its ability to respond to real-time market events.

Masud *et al.* [3] investigated multiple machine learning and time-series forecasting models, including ARIMA, LSTM, Support Vector Machine (SVM), Random Forest, and XGBoost, for gold price prediction. Their comparative analysis showed that LSTM achieved superior forecasting performance because of its ability to model long-term temporal dependencies. Nevertheless, the study did not incorporate sentiment analysis or real-time data streams, reducing its effectiveness in capturing sudden changes in market conditions.

B. Deep Learning Approaches for Financial Forecasting

Deep learning models have recently gained significant attention because of their ability to learn complex temporal patterns from financial data. Shobhit *et al.* [4] examined the influence of macroeconomic variables such as inflation, exchange rates, and interest rates using SARIMA, Random Forest, and a hybrid SARIMA-Random Forest model. Their hybrid approach improved forecasting accuracy by combining linear and nonlinear learning capabilities. Despite these improvements, the framework remained dependent on structured numerical data and did not employ advanced deep learning architectures or Natural Language Processing (NLP) techniques to capture market sentiment.

Similarly, Sarvani *et al.* [5] proposed a hybrid CNN-Bidirectional LSTM model for gold price prediction. By combining Convolutional Neural Networks (CNN) with Bidirectional Long Short-Term Memory (Bi-LSTM), the model successfully captured both spatial and temporal characteristics of financial data, resulting in improved prediction accuracy. However, the model relied primarily on numerical inputs and excluded external textual information such as financial news and social media sentiment, which can significantly influence market behaviour.

C. Research Gaps and Motivation

Although existing studies have demonstrated the effectiveness of machine learning and deep learning techniques for gold price prediction, several important challenges remain. Most published models primarily depend on structured historical data and overlook unstructured information such as financial news and social media, despite their significant influence on investor sentiment and market movements [1], [3], [12]. In addition, many existing approaches focus mainly on prediction accuracy while providing limited explanation of how predictions are generated, reducing their usefulness in practical financial decision-making [10], [11].

Another limitation observed in the literature is the absence of integrated frameworks that combine machine learning, deep learning, Natural Language Processing, and Explainable Artificial Intelligence into a single prediction system. Furthermore, only a few studies address real-time adaptability, making it difficult for existing models to respond effectively to rapidly changing market conditions.

To address these limitations, the proposed framework integrates machine learning algorithms with an LSTM-based deep learning model, NLP-driven sentiment analysis, and Explainable Artificial Intelligence (XAI). By combining structured financial indicators with sentiment extracted from unstructured textual data, the proposed approach aims to improve forecasting accuracy, enhance model interpretability, and provide a scalable solution for real-time gold price prediction.

Research Methodology

The proposed framework follows a systematic workflow comprising data collection, preprocessing, feature engineering, model development, model evaluation, and deployment. Both structured financial data and unstructured textual information are integrated to improve forecasting performance and capture the dynamic behaviour of the gold market. This unified approach enables the model to learn from historical market trends while incorporating external factors such as investor sentiment and economic conditions [1], [3], [12].

A. System Architecture and Dataset

The proposed framework is organized into multiple stages, including data acquisition, preprocessing, feature engineering, model training, evaluation, and deployment. Historical gold price data spanning approximately ten years (2014–2024) were collected from publicly available sources such as Yahoo Finance and Kaggle, resulting in a dataset containing more than 2,500 daily observations [14], [15].

The primary prediction target is the daily gold price in USD, represented using Open, High, Low, and Close (OHLC) values. To improve forecasting performance, additional macroeconomic indicators such as inflation rates, interest rates, and exchange rates are incorporated into the dataset. Unlike currency conversion-based approaches, the proposed framework directly predicts gold prices in USD, simplifying the prediction process while maintaining forecasting accuracy.

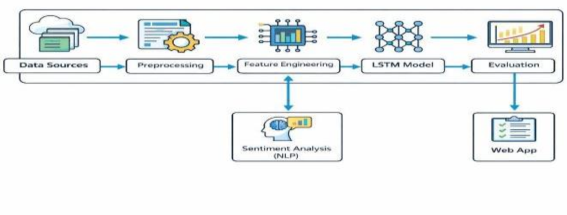

Fig 1: System Architecture of Gold Price Prediction Model

Fig. 1 illustrates the overall system architecture. Besides numerical financial data, the framework also incorporates unstructured textual information collected from financial news articles and social media platforms. Natural Language Processing (NLP) techniques are employed to extract sentiment scores, which are integrated with numerical features to provide a more comprehensive representation of market conditions [12].

B. Data Preprocessing and Exploratory Data Analysis

Data preprocessing plays a crucial role in ensuring the quality and reliability of the dataset. Exploratory Data Analysis (EDA) is performed to understand the relationships between different variables and identify patterns within the data. The preprocessing process involves removing duplicate records, handling missing values using appropriate statistical techniques, detecting and eliminating outliers, and applying normalization and scaling to numerical features. Furthermore, textual data undergoes preprocessing steps such as tokenization and stop-word removal, followed by sentiment extraction using Natural Language Processing techniques. These steps ensure that the dataset is clean, consistent, and suitable for model training [1],[3],[13]

C. Feature Engineering

Feature engineering is carried out to enhance model performance by extracting meaningful and relevant information from the dataset. Time-series features such as lag values and moving averages are generated to capture temporal dependencies in gold price trends. Additionally, macroeconomic indicators are incorporated as important predictive variables. Sentiment scores derived from financial news and social media data are also included as features, providing insights into market sentiment. Correlation analysis and feature selection techniques are employed to eliminate redundant variables and retain only the most informative features for model training, thereby reducing model complexity and improving prediction performance [3], [5], [13].

D. Model Development

The dataset is divided into training and testing subsets to evaluate the performance of different forecasting models. Several machine learning and deep learning algorithms, including Linear Regression, Support Vector Machine (SVM), Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM), are implemented and compared. Among these models, LSTM is selected as the primary prediction model because of its ability to capture long-term temporal dependencies and nonlinear relationships in sequential financial data. This capability enables LSTM to learn complex market trends more effectively than conventional machine learning algorithms, resulting in improved forecasting accuracy [2], [3], [6].

E. Model Training and Evaluation

The models are trained using the training dataset and evaluated on the testing dataset to assess their performance. Hyperparameter tuning techniques are applied to optimize the models and improve their predictive capability. The evaluation of the models is carried out using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and prediction accuracy. These metrics help measure the deviation between actual and predicted values and provide a comprehensive understanding of model performance [2],[3],[13]

Result Analysis

The proposed gold price prediction framework was evaluated using historical gold price data together with macroeconomic indicators, including inflation rates, interest rates, and exchange rates. In addition, sentiment scores extracted from financial news and social media were integrated with the numerical dataset to capture external market influences. The inclusion of both structured and unstructured data enabled the models to better represent real-world market conditions and improved forecasting performance [3], [5], [12].
Among the evaluated models, the Long Short-Term Memory (LSTM) network was selected as the primary forecasting model because of its capability to learn long-term temporal dependencies in sequential financial data. The model was trained using historical records and evaluated on an unseen test dataset to assess its prediction accuracy and generalization capability [6].

F. Prediction Results

The prediction results indicate that the LSTM model closely follows the actual gold price trend throughout the testing period. The small difference between the actual and predicted values demonstrates that the model effectively captures both long-term market trends and short-term price fluctuations. This reflects the model's ability to learn complex nonlinear relationships that are commonly observed in financial time-series data.

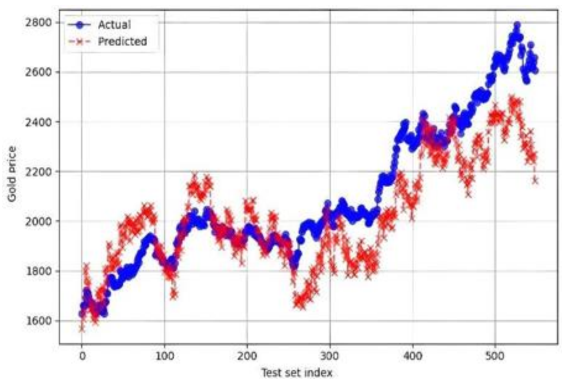

Fig. 2. Actual vs Predicted Gold Prices using LSTM Model

The superior performance of the LSTM model can be attributed to its memory cells, which preserve information over long sequences and enable the network to model temporal dependencies more effectively than conventional machine learning algorithms [6]. Furthermore, incorporating sentiment information from financial news and social media provides additional contextual information regarding market behaviour, allowing the model to respond more effectively to sudden changes in investor sentiment and economic events [12]. These observations are consistent with previous studies that demonstrated the effectiveness of LSTM-based models for financial forecasting [3], [6].

G. Model Performance Comparison

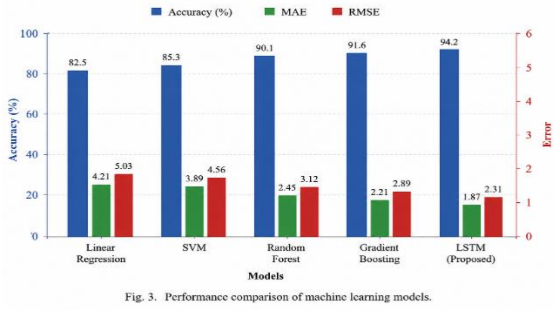

Fig. 3. Performance Comparison of Machine Learning Models

Table 1 summarizes the performance of the evaluated models using Accuracy, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). Among all the models, the proposed LSTM achieved the highest prediction accuracy of **94.2%**, together with the lowest MAE (**1.87**) and RMSE (**2.31**), indicating superior forecasting performance. Linear Regression produced the lowest accuracy because it assumes a linear relationship between input variables and gold prices, making it unsuitable for modelling the complex behaviour of financial markets. Support Vector Machine (SVM) improved prediction accuracy by capturing nonlinear relationships but remained less effective in modelling long-term temporal dependencies. Ensemble learning methods, including Random Forest and Gradient Boosting, achieved better performance than traditional machine learning algorithms due to their ability to model nonlinear patterns. However, these models process observations independently and cannot effectively retain sequential information over time.
The LSTM model outperformed all other models because it was specifically designed for sequential data analysis. Its memory mechanism enables the network to retain historical information over long periods, allowing it to identify temporal dependencies and changing market trends more accurately [6]. Moreover, integrating sentiment scores with structured financial indicators further improved prediction performance by incorporating external market information that is not available in numerical data alone [3], [5], [12]. These findings demonstrate that combining deep learning with sentiment analysis provides a more reliable framework for gold price forecasting than conventional machine learning approaches.

Table 1. Performance comparision of Machine Learning model

Model	Accuracy (%)	MAE	RMSE

Linear Regression	82.5	4.21	5.03
SVM	85.3	3.89	4.56
Random Forest	90.1	2.45	3.12
Gradient Boosting	91.6	2.21	2.89
LSTM (Proposed)	**94.2**	**1.87**	**2.31**

Conclusion, Future Enhancements, & Prospects

Conclusion

This study presents an intelligent framework for gold price prediction by integrating machine learning, deep learning, and Natural Language Processing (NLP) techniques. The proposed framework combines structured financial indicators with unstructured textual data to capture the complex and dynamic behaviour of gold markets. By leveraging multiple data sources and advanced learning algorithms, the system achieves improved forecasting accuracy and reliability compared with conventional prediction methods.

Among the evaluated models, the Long Short-Term Memory (LSTM) network demonstrated the best performance because of its ability to capture long-term temporal dependencies and nonlinear relationships in sequential financial data. Experimental results showed that the proposed LSTM model achieved the highest prediction accuracy (94.2%) while producing the lowest Mean Absolute Error (MAE) of 1.87 and Root Mean Square Error (RMSE) of 2.31 compared with Linear Regression, Support Vector Machine (SVM), Random Forest, and Gradient Boosting models.

Furthermore, integrating sentiment analysis with structured financial data significantly enhanced the model's capability to capture external market influences such as financial news and investor sentiment. The inclusion of Explainable Artificial Intelligence (XAI) also improved the transparency and interpretability of the prediction process, enabling users to better understand the factors influencing model decisions. Overall, the proposed framework provides a reliable, scalable, and practical solution for real-time gold price forecasting and financial decision-making.

Future Enhancements & Prospects

Although the proposed framework demonstrates promising prediction performance, several opportunities remain for future improvement. Integrating real-time financial data through APIs can enable continuous monitoring and live gold price forecasting. In addition, advanced deep learning architectures, such as Transformer-based models and hybrid forecasting approaches, may further improve prediction accuracy by capturing more complex temporal patterns in financial data.

Future work may also extend the framework to support multi-asset forecasting, including stocks, cryptocurrencies, and other commodities, thereby creating a more comprehensive financial prediction platform. Furthermore, reinforcement learning techniques can be incorporated to support automated trading strategies based on predicted market trends

The sentiment analysis component can be enhanced using transformer-based language models such as BERT or FinBERT to better capture contextual information from financial news and social media.Similarly, advanced Explainable Artificial Intelligence (XAI) methods, including SHAP and LIME, can provide deeper insights into feature importance and improve user confidence in model predictions .

Finally, deploying the framework as a cloud-based application with an interactive dashboard can improve accessibility and scalability for investors, financial analysts, and policymakers. These enhancements will further strengthen the applicability of the proposed framework in real-world financial forecasting environments.

References

[1]. N. Ch. Sri Nithya, P. Pamidi, P. Tumula, H. Harshitha Voora, B. R. M. Balajee, and S. Srithar, "Predictive Modeling of Gold Prices Using Machine Learning," in *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSoCiCon)*, Pune, India, 2024.

[2]. S. Narendran, "Exploring the Efficacy of Vari Machine Learning Approaches for Gold Price Forecasting: Insights and Analysis," in *2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT)*, Nagercoil, India, 2024.

[3]. S. R. Masud, A. Imtiaz, S. M. Hossain, and M. S. U. Miah, "Comprehensive Approach to Gold Price Prediction Using Machine Learning and Time Series Models," in *2025 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)*, Gwalior, India, 2025.

[4]. Shobhit, S. Khan, S. Banerjee, N. Chaudhary, and B. Shukla, "An Analytical Approach to Market Dynamics and Economic Indicators Using Machine Learning Techniques," in *2025 International Conference on Technology Enabled Economic Changes (InTech)*, 2025.

[5]. G. S. N. V. Sarvani, U. Kousalya, P. Sathwika, B. N. Ashok, R. R. Chintala, and R. Mothukuri, "Gold Price Prediction Based on Stock

Analysis Using CNN-Bidirectional LSTM," in *2025 1st International Conference on Secure IoT, Assured and Trusted Computing (SATC)*, 2025.

[6]. S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[7]. L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.

[8]. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 785-794.

[9]. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[10]. S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[11]. M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, 2016, pp. 1135-1144.

[12]. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.

[13]. R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021.

[14]. H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond Efficient Transformer for Long Sequence Time-Series Forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, pp. 11106–11115, 2021

[15]. J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of NAACL-HLT*, Minneapolis, MN, USA, 2019, pp. 4171–418.