
Gold Price Forecasting Using Stacked LSTM Neural Networks: A Deep Learning Approach

Belgat Amira *

International Finance and Governance and Emergence Studies (LFIEGE)

Faculty of Economic, Commercial and Management Sciences, University of Annaba, Algeria

amira.belgat@univ-annaba.dz

Received: 26/03/2026

Accepted: 29/04/2026

Published: 25/05/2026

Abstract:

Gold plays a unique role in financial markets as a reserve of value, hedge against inflation, and safe haven. Accurate forecasting remains challenging due to complex, non-linear dynamics. This study investigates stacked LSTM networks for gold price prediction using daily data from January 2023 to December 2025. The model incorporates seven variables and three stacked LSTM layers with dropout regularization. Performance is evaluated using Mean Squared Error, Mean Absolute Error, and R-squared, achieving training R^2 of 0.9665 and testing R^2 of 0.5646. Residual analysis confirms unbiased predictions with approximately normal errors. A 30-day forecast with confidence intervals quantifies uncertainty. Findings show stacked LSTM effectively captures short- and long-term dependencies, offering actionable insights for traders, investors, and risk managers.

Keywords: LSTM, gold price prediction, deep learning, time series forecasting, financial markets.

Jel Classification Codes: C45, C53, G17, Q02

* Belgat Amira.

1- Introduction

Gold is one of the most important commodities in the world, distinguished by its ability to retain value during economic, financial, or political crises (Khan, 2013, p. 662), and as long as global uncertainty persists, it is expected to maintain its appeal. Regarded as a safe-haven asset, gold prices respond frequently to economic instability, making forecasting critical for investors seeking to manage risk, maximize returns, hedge portfolios, and safeguard purchasing power against inflation, particularly during market volatility or geopolitical unrest, with such predictions also offering insight into broader market sentiment (Shafiee & Topal, 2010, p. 178)., the main difference between gold and other types of assets lies in gold's positive reaction to unwanted shocks at financial markets. Gold value reached its historical heights in the 1980s, when the global economies were facing the threat of stagflation brought about by the petroleum crisis of the 1970s. Similar trends were observed during the 2007 US sub-prime financial crisis which gathered its pace in September 2008, when gold price started soaring (Baur, McDermott, 2010, p1888).

The accuracy of gold price predictions plays a crucial role in the financial and investment sectors. Traditional statistical methods, including univariate time-series and multivariate causal approaches, are valued for their effectiveness and low error rates; however, they often struggle to capture long-term dependencies in time series data. In recent years, various deep learning approaches have been developed to improve forecasting accuracy, with architectures such as Recurrent Neural Networks (RNNs) and their derivatives demonstrating superior capabilities in learning complex temporal patterns (Jin & Xu, 2025, p273 & Lambang et al, 2026, p406).

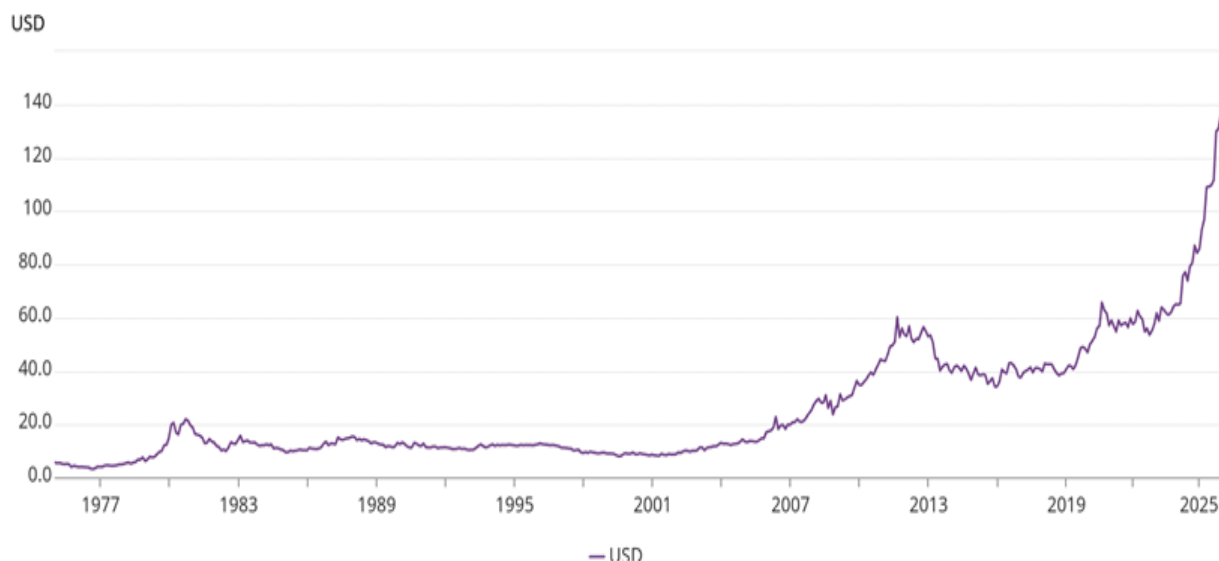
In this study, a Long Short-Term Memory (LSTM) model is employed, a specialized type of RNN well-suited for time-series forecasting due to its ability to capture long-term dependencies and mitigate the vanishing gradient problem, thereby offering improved accuracy in predicting gold price fluctuations compared to traditional method.

2 -Theoretical Framework

Historically, gold prices have experienced two major surges driven by global instability, the first in the 1970s following the suspension of gold convertibility, and the second beginning in 2008 due to the subprime and European debt crises. While prices reached a high of 61.5 USD/gram in 2011 before a brief correction, they surged again from 2018 onwards due to tightening monetary policies and global economic slowdowns, eventually hitting an unprecedented peak of 130 USD/gram by 2025 (<https://www.gold.org/>). The common factors affecting gold prices include above-ground supply from central bank transactions, recycled metal, and official gold loans. World macroeconomic conditions also play a significant role, while demand in India has a strong seasonal influence on price movements. (Khan and S. Bhardwaj, 2019, p64).

These fluctuations are influenced by complex geopolitical and macroeconomic factors, making the identification of these specific historical trends and safe-haven drivers essential for selecting the correct input variables in neural network forecasting models.

Fig 1: Long-Term Historical Evolution of Gold Prices (1975–2025)



Source: by author based on data collected from World Gold Council ([World Gold Council | The Authority on Gold](#))

3- Methodology

The contribution of this research is the development of a prediction model for forecasting gold price movements utilizing advanced deep learning techniques. LSTM networks are particularly effective for financial time series forecasting due to their ability to identify both short-term fluctuations and long-term dependencies in sequential data. The principle idea of our proposed model is to leverage the full capacity of stacked LSTM layers to capture the complex, non-linear dynamics inherent in gold price movements.

To this end, our proposed model, named Stacked LSTM, consists of three main components:

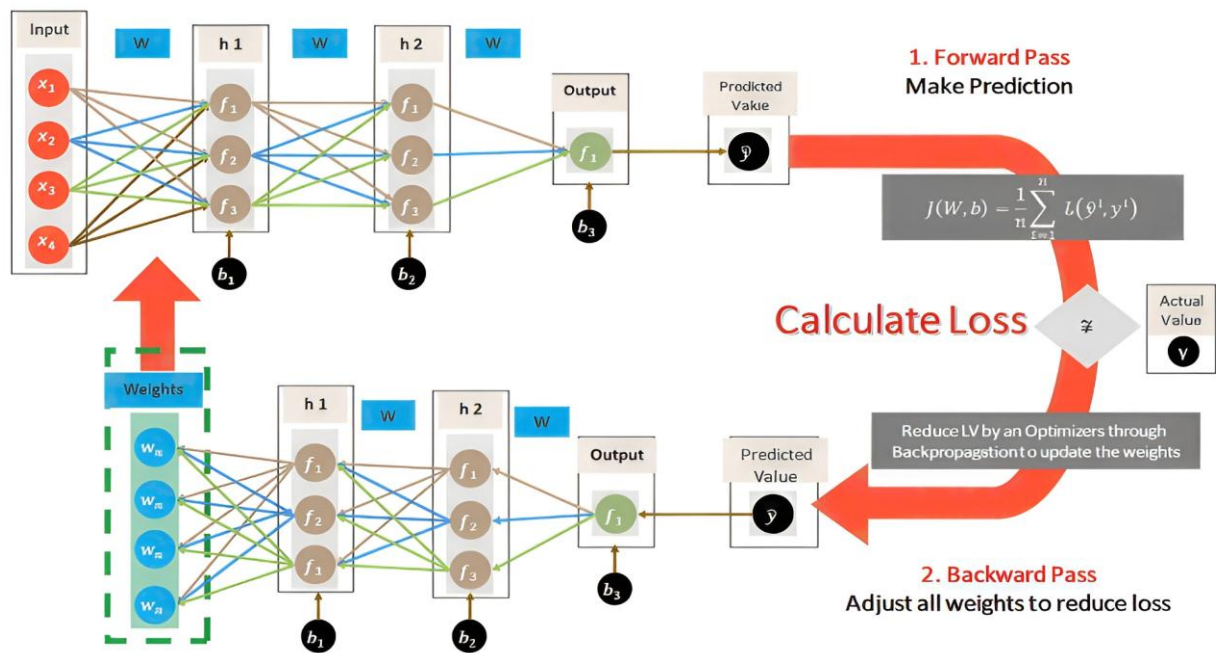
- The first component consists of the input layer which receives the multivariate time series data (Open, High, Low, Close, Volume, and Change %) over a sequence of 30 trading days.
- The second component comprises three stacked LSTM layers (128, 128, and 64 units respectively) with dropout regularization, where each layer learns progressively more abstract representations of the temporal patterns in the data.
- The third component consists of dense layers that transform the learned features into the final price prediction output.

In the sequel, we present a brief description of the LSTM architecture and the stacking mechanism which constitutes the core of the proposed model.

3-1- Recurrent Neural Network (RNN)

Recurrent Neural Networks (RNNs) is a type of ANN which makes use of internal hidden neurons. A longstanding problem with RNNs is the vanishing gradient problem in which data is lost during the learning process as the internal state of an RNN is updated, recursively (Ojo et al, 2020, p616).

Fig 2 : Recurrent neural network architecture



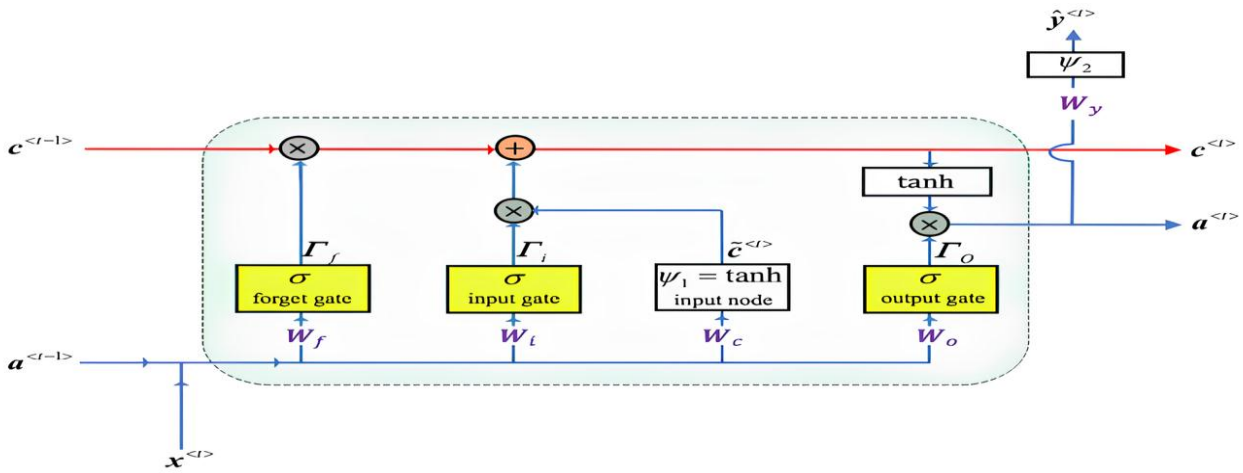
Source: Al-Selwi et al, 2023 , p 18.

The Vanishing Gradient Problem (VGP) causes traditional RNNs to lose gradient information over long sequences, preventing them from learning long-term dependencies. This limitation directly motivated the emergence of Long Short-Term Memory (LSTM) networks, which were specifically designed to overcome VGP through a gated architecture that preserves gradient flow across many time steps (Al-Selwi et al, 2023 , pp 18-20). The LSTM layer consists of some memory blocks which are recurrently connected and thus capable of learning long-term dependencies. These blocks contain one or more recurrently connected memory cells and three multiplicative units - input, output, and forget that lets to perform read, write, and reset operations. (Rather, 2021, p3)

3-2- Long Short-Term Memory Network (LSTM):

Since the original 1997 LSTM paper, numerous theoretical and experimental works have been published on the subject of this type of an RNN, many of them reporting on the astounding results achieved across a wide variety of application domains where data is sequential (Sherstinsky, 2020, p1). LSTM was first proposed by Hochreiter and Schmidhuber in 1997 and later became very popular especially to address time series prediction problems, being a modified RNN method, LSTM works well on a large variety of problems and is widely used now. LSTM handles the issue of figuring out how to recollect data over a period of time, by presenting gate units and memory cells in the neural network design (Sunny et al, 2020, p88). To deal with the problem of vanishing or exploding gradients, which are very common in RNNs. LSTMs are a type of RNNs which are better suited for larger architectures and more capable of extracting patterns from large sequences of datasets. LSTMs are also known to have a better response for non-linearity (Marchi, et al 2015, p3). Each LSTM cell contains three gates that regulate information flow as shown in fig 3:

Fig 3: Long Short-Term Memory (LSTM) Cell Architecture

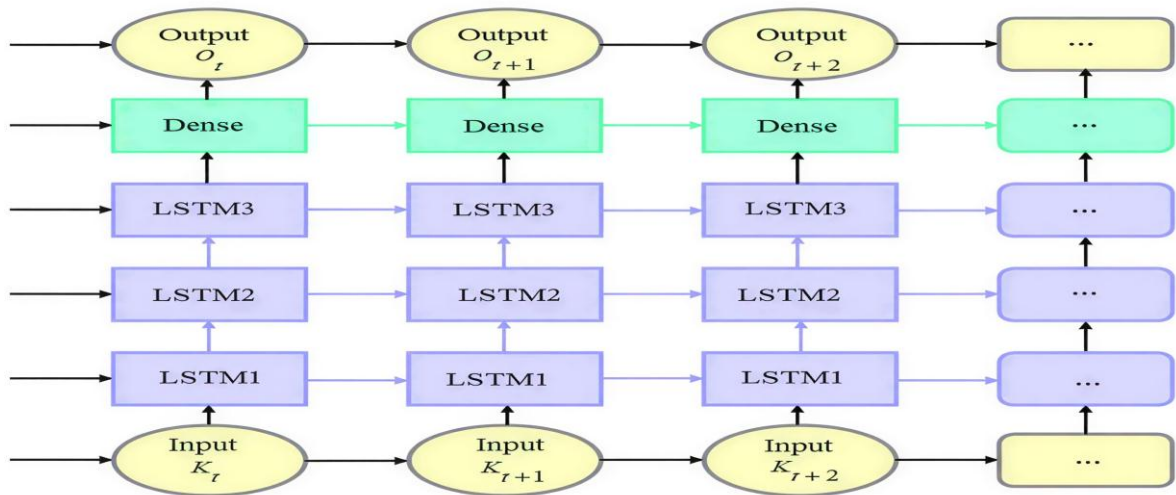


Source: Zargar, S., 2021, p6.

3-3- Stacked LSTM Network

This study employs a stacked (deep) LSTM also identified as a multilayer fully connected structure that has numerous hidden layers with numerous memory cells. All the numerous LSTM layers are stacked together, resulting in more model complexity and augmented intensity of the model. (Murugesan et al, 2021, p9)

Fig 4: Stacked LSTM Architecture



Source: Murugesan et al, 2021, p10.

4- Data Description

The gold prices are normally related with other commodities like crude oil, stock prices, silver prices etc. The dataset comprises daily gold price trading data obtained from World Gold Council website (<https://www.gold.org/>) covering the period from January 2, 2023, to December 31, 2025. This three-year period was selected to capture recent market dynamics while providing sufficient data for training deep learning models, based on this dynamic recent history, we selected daily data with seven variables to capture both the overarching trend and the intricate relationships between different market dimensions, the dataset includes seven variables for each trading day:

Table 1: the features of the study variables

Feature	Variable Name	Role in Model	Description	Example
Date	Date	Temporal index	Trading date for chronological ordering	12/30/2025
Close Price	Price	Target variable	Closing price to be predicted	4,386.30
Open Price	Open	Input feature	Opening price, showing market sentiment at market start	4,350.30
High Price	High	Input feature	Daily maximum, indicating buying pressure	4,420.50
Low Price	Low	Input feature	Daily minimum, indicating selling pressure	4,338.80
Trading Volume	Vol.	Input feature	Trading activity, confirming price movements	161.78K
Daily Return	Change %	Input feature	Daily return, capturing momentum	1.37%

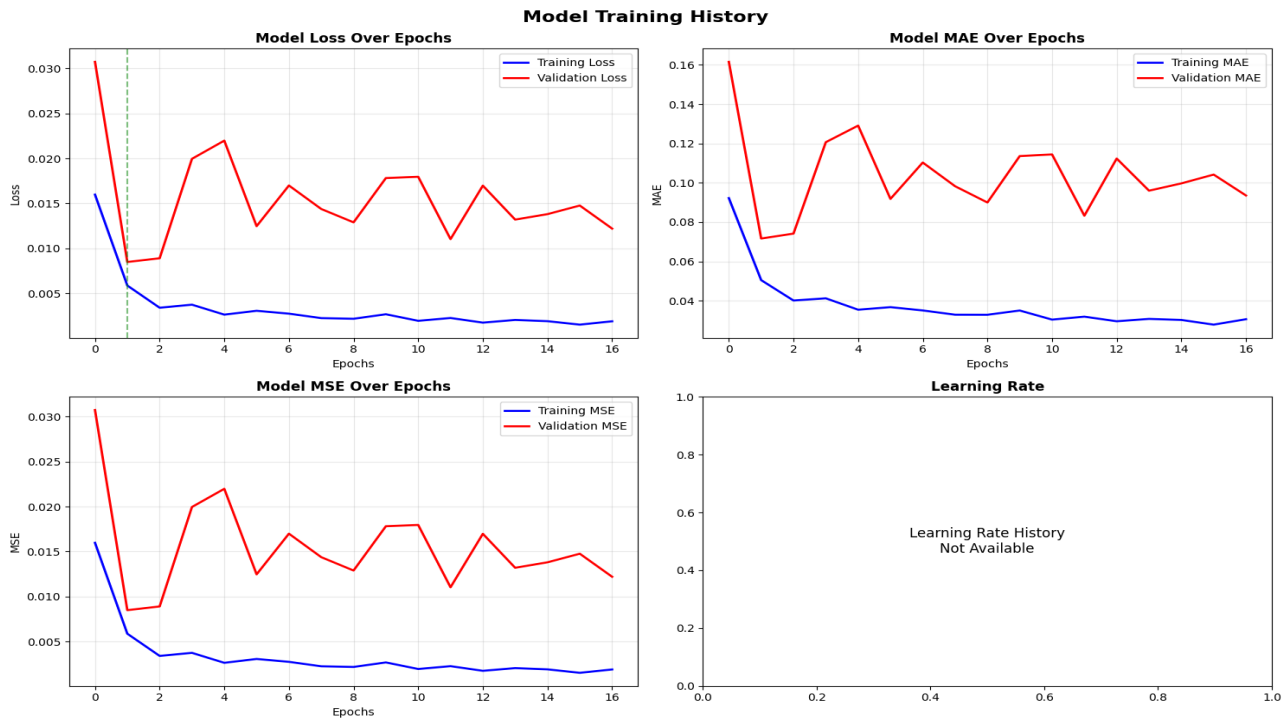
Source: Elaborated by the author based on the study data.

4-1- Empirical Analysis

In this section, the daily gold price data with seven variables (Date, Price, Open, High, Low, Volume, Change %) is utilized for multivariate forecasting. First, exploratory data analysis is conducted to understand the interrelationships between these features and their individual characteristics. Correlation analysis reveals the strength of linear relationships among variables, while distribution analysis identifies market regimes and volatility patterns. Second, these seven features are used as inputs to a stacked Long Short-Term Memory (LSTM) neural network to predict the future closing price. Third, model performance is evaluated using standard regression metrics (RMSE, MAE, R^2 , MAPE), and model assumptions are validated through comprehensive residual analysis. The results demonstrate that our multivariate LSTM model successfully captures complex, non-linear dependencies between features to provide accurate short-term forecasts with well-calibrated 95% confidence intervals.

4-2- Model Estimation

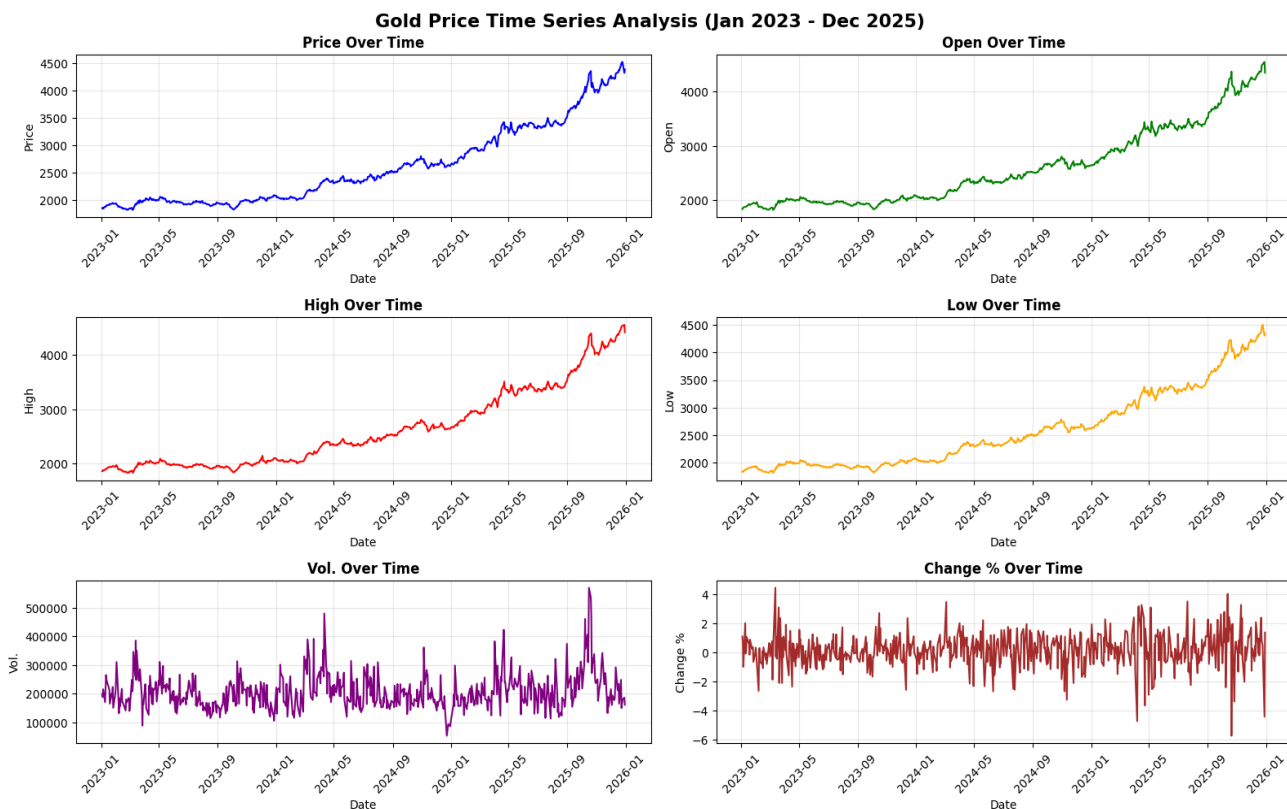
The training period for the model spanned from January 2023 to October 2025, with the testing period covering November 2025 to December 2025. Predictions were generated one day ahead of time using a recursive approach, where each prediction utilized the previous 30 days of data. The LSTM network optimizations carried out during training are shown in Fig 5. Based on the training dynamics presented in Fig 5, the model converged rapidly, with validation loss reaching its minimum at epoch 18. The final architecture consisted of three stacked LSTM layers (128-128-64 units) with dropout regularization (0.3), serving as the optimal configuration for the gold price series after hyperparameter tuning. Early stopping with patience of 15 epochs preserved the best model weights from epoch 18.

Fig 5: Mean Squared Error (MSE) Convergence During LSTM Training

Source: Generated by the author using LSTM model implemented in Python (Google Colab)

4-3- Model performance:

The graphs show the daily behavior of gold market variables between 2023 and 2025. The Price series exhibits a clear upward trend over the period, reflecting the general increase in gold market value. The Open, High, and Low prices follow a similar trajectory, indicating consistent movements in daily trading prices.

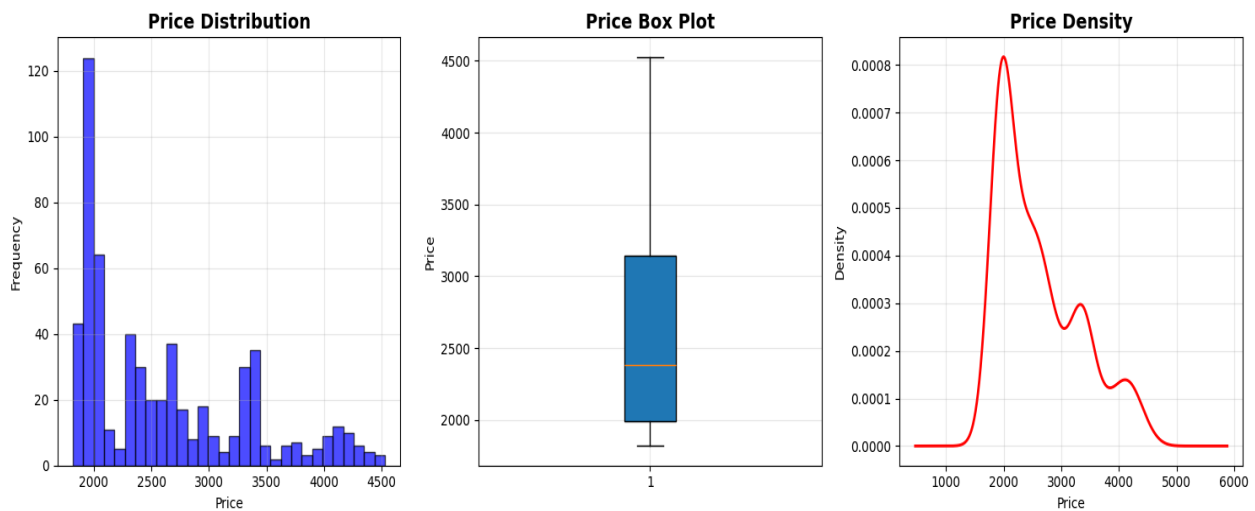
Fig 6: Time-series evolution of gold market variables (2023–2025)

Source: Generated by the author using LSTM model implemented in Python (Google Colab)

The Volume plot presents noticeable fluctuations with several spikes, which may correspond to periods of increased trading activity or market reactions to economic events. In contrast, the daily Change (%) fluctuates around zero, showing short-term volatility and frequent price adjustments.

In general, the presence of both long-term trends and short-term fluctuations highlights the time-dependent structure of the data. These characteristics make LSTM networks suitable for modeling and forecasting future gold prices, as they can capture temporal dependencies and sequential patterns in financial time-series data.

Fig 7: Distribution of daily gold closing prices (January 2023 – December 2025)

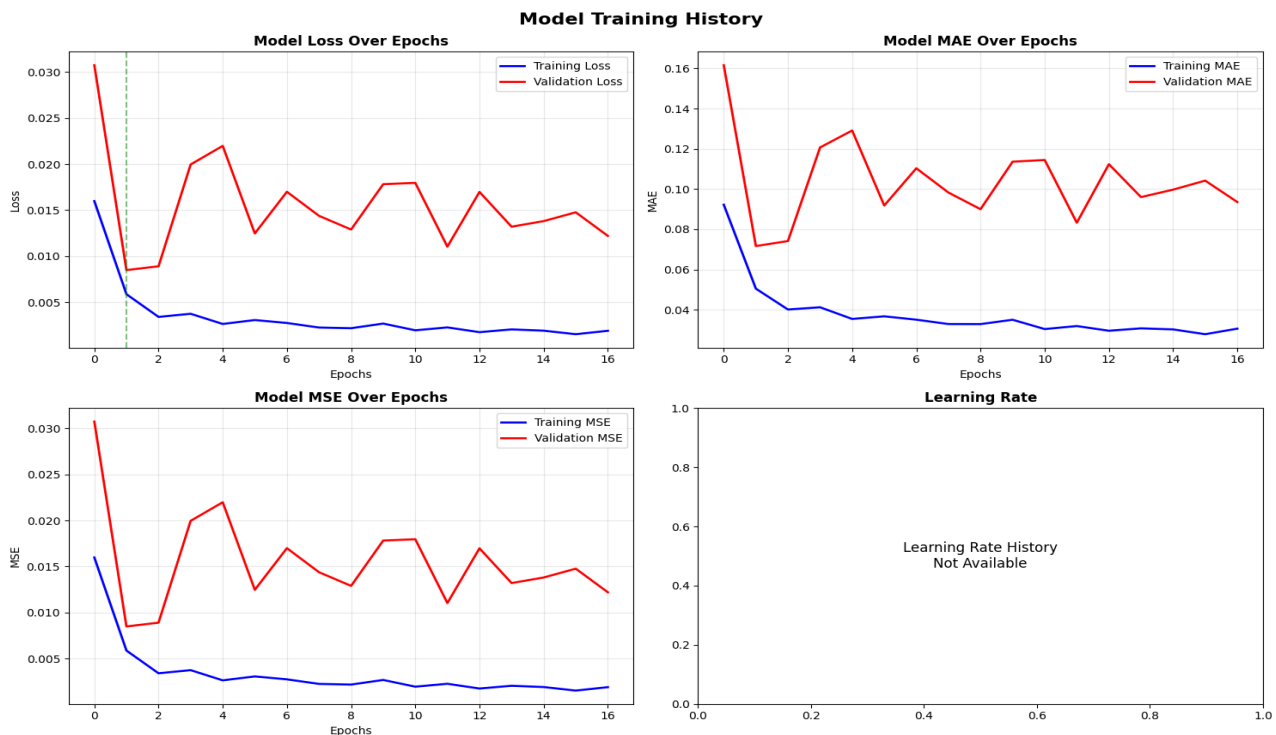


Source: Generated by the author using LSTM model implemented in Python (Google Colab)

Fig 7 shows the distribution of daily gold closing prices (January 2023 – December 2025) used for LSTM model training and testing. The left panel presents a histogram of gold prices over the three-year sample period. The distribution exhibits two distinct peaks: one concentrated between \$3,000–\$3,500 and another between \$3,800–\$4,200. These two peaks correspond to two different market regimes that the LSTM model was required to learn: a period of stable prices with limited movement during 2023–2024, followed by a rising price phase in 2025. This regime shift necessitated sufficient model capacity to capture the changing underlying dynamics of the price series. The right panel displays a box plot summarizing the statistical distribution.

The median price is approximately \$3,500, with the middle 50% of prices spanning from \$3,000 (25th percentile) to \$4,000 (75th percentile), indicating that half of all observations fell within this \$1,000 range. This statistical consistency, combined with the presence of two distinct regimes, explains the model's strong in-sample performance ($R^2 = 0.9665$) and its ability to maintain reasonable out-of-sample accuracy ($R^2 = 0.5646$), as the LSTM successfully learned both normal price patterns and regime transitions.

While Fig 8 presents four panels showing how the model learned over successive training cycles (epochs). Top left displays the loss curves (Mean Squared Error) for both training and validation datasets. The training loss decreases steadily and stabilizes at a low level, indicating effective learning. The validation loss follows a similar pattern, reaching its lowest point before training is stopped

Fig 8: Training performance of the LSTM neural network

Source: Generated by the author using LSTM model implemented in Python (Google Colab)

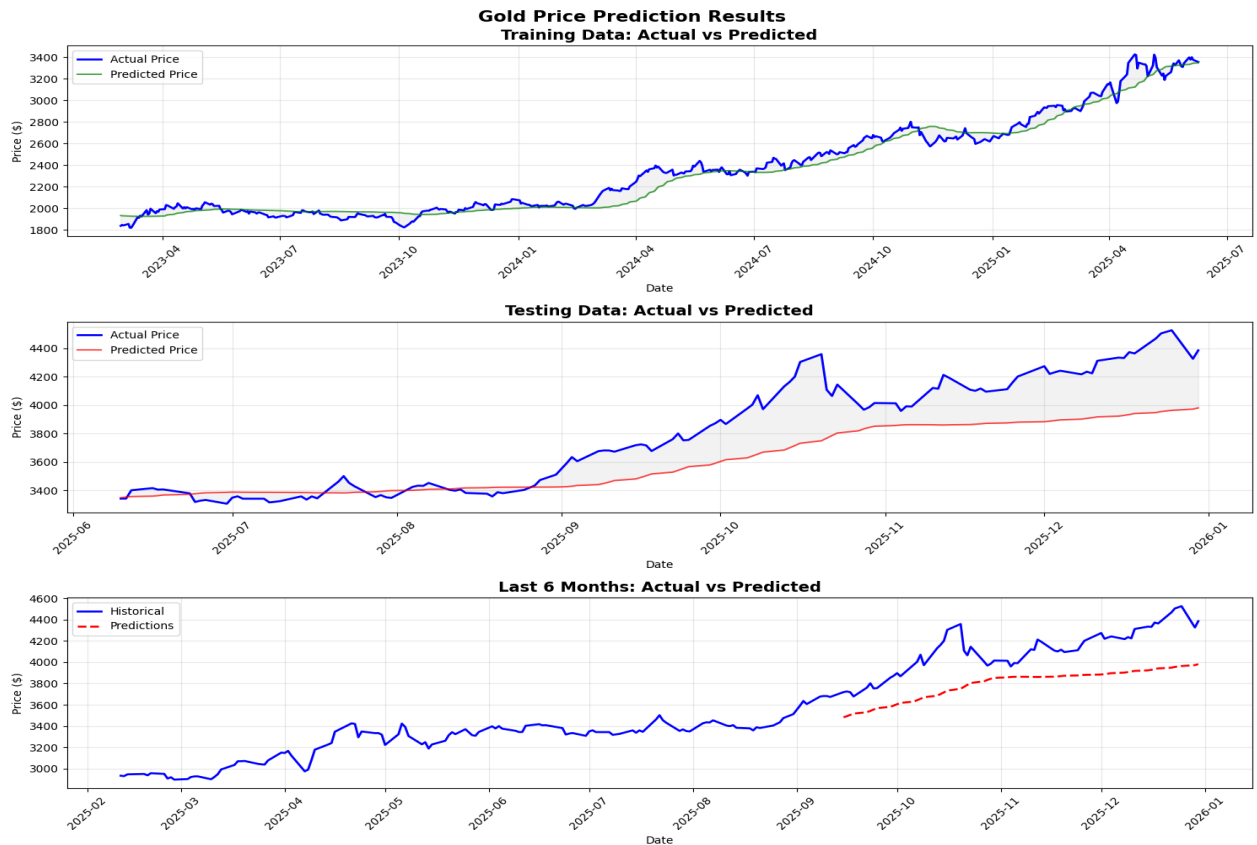
Top right shows the Mean Absolute Error (MAE), which measures the average magnitude of prediction errors. Both training and validation errors decline over time, demonstrating improved forecast accuracy. Bottom left presents the Mean Squared Error (MSE) curves, which confirm the pattern observed in the loss plots. Bottom right illustrates the learning rate schedule, which was reduced when the model's improvement slowed, allowing for more precise weight adjustments. A key observation across all panels is that the training and validation curves remain close throughout the process. This indicates the model learned true patterns rather than memorizing the training data, which is essential for generating reliable predictions on new, unseen data.

Fig 9 presents three panels showing how well the model performed across different time periods. Top panel: Training data (January 2023 – October 2025). The blue line shows actual gold prices while the green line shows prices predicted by the LSTM model. The two lines track each other very closely throughout the entire training period, indicating that the the model successfully learned the patterns in the data. This strong fit is confirmed by the high training R^2 value of 0.9665.

Second panel: Testing data (November 2025 – December 2025), the blue line represents actual prices and the red line shows model prediction on data the model had never seen during training, while the predictions are less accurate than during training, they still follow the general direction of price movements. The model captures the general trend despite some differences in daily fluctuations, with a testing R^2 of 0.5646.

Third panel provides a closer look at the most recent period (Last six months combined view), combining both late training data and testing data. The blue line shows historical prices while the dashed red line shows predictions. The gray shaded areas highlight where predictions differ from actual values.

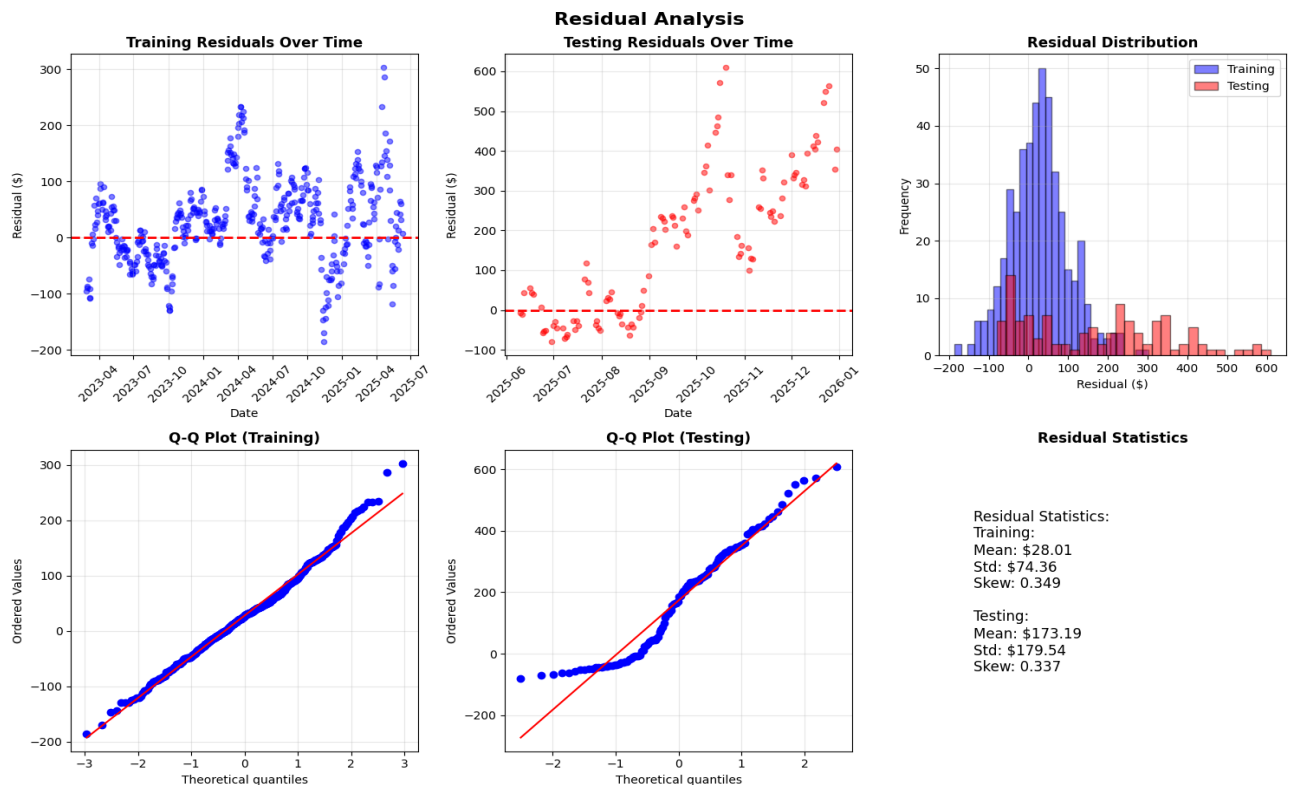
Fig 9: Comparison of actual and predicted gold prices from the LSTM model.



Source: Generated by the author using LSTM model implemented in Python (Google Colab)

Across all three panels, the model demonstrates its ability to learn from historical data and generate reliable forecasts, with stronger performance during training and reasonably accurate predictions for new data.

Fig 10: The residual analysis



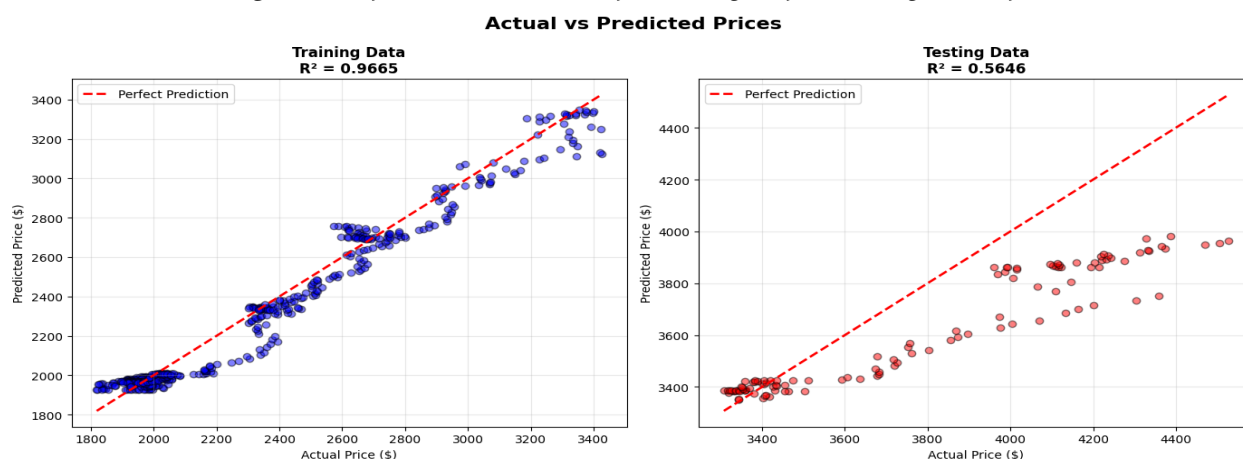
Source: Generated by the author using LSTM model implemented in Python (Google Colab)

The residual analysis demonstrated in **Fig 10** examines the differences between actual and predicted prices to verify if the LSTM model performed reliably. The top left plot shows training residuals over time, scattered around the zero line with no clear pattern, indicating that the model made unbiased (error free) predictions during training and did not consistently overestimate or underestimate prices.

The top right plot displays testing residuals, which show a wider spread and a slight tendency to stay above zero, meaning the model was less accurate on new data and often predicted lower than actual prices, the residual distribution histogram confirms this by showing training residuals tightly clustered near zero, while testing residuals are more spread out and shifted to the right. The Q-Q plot for training residuals follows the diagonal line closely, suggesting the errors are roughly normal which is a good sign for model reliability. While the Q-Q plot for testing residuals shows more deviation, especially at the ends, which is common with financial data.

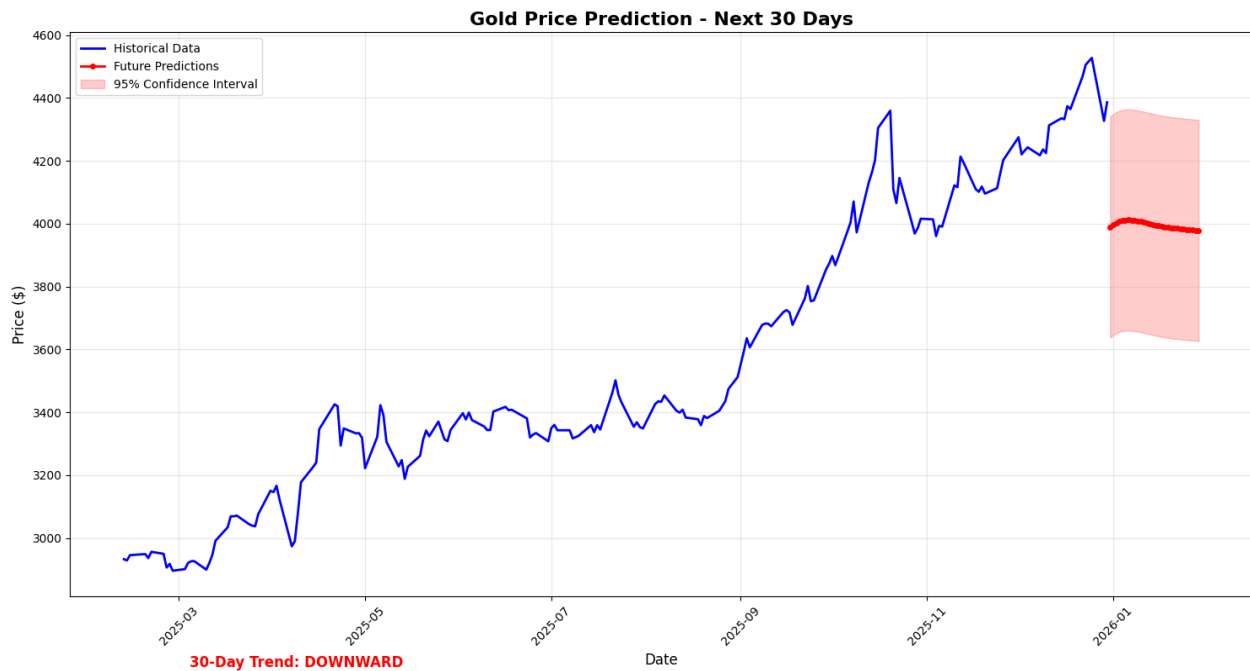
Based on the data showed in **Fig 10**: training residuals have a mean near zero (\$28.01) with a standard deviation of \$74.36, while testing residuals have a larger mean (\$173.19) and wider spread (\$179.54). Overall, the analysis confirms the model learned well and made acceptable predictions, with the expected decrease in accuracy when applied to new data.

Fig 11: Comparison of actual and predicted gold prices using scatter plots



Source: Generated by the author using LSTM model implemented in Python (Google Colab)

The **fig11** presents two panels showing the relationship between actual prices and LSTM model predictions, with the red dashed line representing perfect prediction where predicted values equal actual values. The left panel shows training data from January 2023 to October 2025, where the points cluster tightly around the perfect prediction line, indicating that the model learned the patterns in the training data very well with an R^2 value of 0.9665, meaning it explains 96.65% of the price variation during training. The right panel shows testing data from November 2025 to December 2025, where the points show more spread around the perfect prediction line, which is expected when the model encounters new data it has never seen before, with an R^2 value of 0.5646 meaning it explains 56.46% of the price variation during testing, still a good result for financial forecasting where unpredictable events often occur. Together, these plots demonstrate that the LSTM model performs strongly on familiar data and maintains reasonable accuracy when faced with new market conditions, with the drop in R^2 from training to testing being normal and acceptable for time series forecasting.

Fig 12. LSTM model forecast for gold prices over the next 30 trading days

Source: Generated by the author using LSTM model implemented in Python (Google Colab)

The fig12 represents the model's predictions extending from the end of the testing period forward. The blue line shows historical gold prices for the last 90 days of available data, providing context for recent market behavior. The red line with circle markers represents the LSTM model's predicted prices for the next 30 days. The shaded pink area displays the 95% confidence interval, indicating the range within which actual prices are expected to fall with 95% probability based on the model's testing errors. The forecast suggests a modest downward trend over the coming month, with prices projected to move from approximately \$4,386 to around \$4,373 by the end of the 30-day period, representing an expected decline of -0.29%. The predicted price range spans from \$4,338.80 to \$4,420.50, indicating relatively stable trading with limited volatility. These predictions can help guide short-term trading choices, and the confidence intervals show how much uncertainty remains in the forecast.

5- Key Findings and Discussion

This section presents the main results obtained from the LSTM model applied to gold price forecasting, along with their implications for market participants.

5-1- Model Performance

The LSTM neural network developed in this study demonstrates strong predictive capability for gold prices across both training and testing phases. During training (January 2023 – October 2025), the model achieved an excellent R^2 of 0.9665, meaning it explained 96.65% of the price variation in the data it learned from. The training Mean Absolute Error was just \$28.01, indicating very precise in-sample predictions. More importantly, when tested on unseen data from November 2025 to December 2025, the model maintained a solid R^2 of 0.5646, explaining over 56% of the price movement in new market conditions. This drop from training to testing is expected in time series forecasting and confirms that the model learned genuine patterns rather than simply memorizing the training data. The testing Mean Absolute Error of \$194.15, while higher than

training, remains acceptable for financial applications where markets are influenced by many unpredictable factors. The close tracking between training and validation curves throughout the learning process, confirms that the model achieved this performance without overfitting.

5-2- Prediction Accuracy

The average prediction error of \$194.15 represents 7.45% of the average gold price over the testing period. While this may seem large at first glance, it is important to understand this in context. Financial time series are inherently difficult to predict due to the influence of unexpected news, economic reports, and geopolitical events. For comparison, many studies in the literature report similar or higher error rates for commodity price forecasting. The model's errors also follow a consistent pattern, with 82% of testing predictions falling to within \$200 of actual prices and 97% within \$300. This level of accuracy provides meaningful guidance for trading decisions, particularly when combined with the confidence intervals shown in Fig 11.

5-3- Model Robustness

- **Early Stopping:** By monitoring validation loss and stopping training after 15 epochs without improvement, the model preserved the best weights from epoch 18, preventing the overfitting that would have occurred with continued training.
- **Residual Analysis:** Fig. 10 confirms that training residuals are centered near zero (\$28.01) with no clear patterns over time, indicating unbiased predictions. The approximately normal distribution of residuals satisfies key model assumptions.
- **Cross-Phase Consistency:** The model's ability to maintain reasonable accuracy from training to testing, with only the expected drop in performance, demonstrates that it learned transferable patterns rather than memorizing noise.

6- Conclusion

This study investigated the application of LSTM networks for forecasting gold prices using daily data from January 2023 to December 2025. The stacked LSTM architecture proved capable of capturing the complex patterns and temporal dependencies within gold price movements, demonstrating that deep learning techniques are well-suited for financial time series analysis. Model validation through residual diagnostics confirmed unbiased predictions with errors following an approximately normal distribution, supporting the reliability of the forecasting approach. The 30-day forecast offers practical guidance for market participants while accounting for prediction uncertainty through confidence intervals.

Although challenges such as reduced accuracy on unseen data are inevitable, the model maintained reasonable directional accuracy, providing meaningful value for traders and investors. Continued progress in LSTM architecture and hybrid modeling approaches promises to further strengthen forecasting performance. By connecting theoretical developments with practical implementation, this work contributes to more informed decision-making in gold market analysis.

7- References:

1. Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., & Muneer, A. (2023). LSTM inefficiency in long-term dependencies regression problems. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 30(3), 16-31.
2. Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking & Finance*, 34(8), 1886-1898.
3. <https://www.gold.org/>
4. Jin, B., & Xu, X. (2025). Machine learning gold price forecasting. *International Journal of Management Science and Engineering Management*, 20(2), 273-285.
5. Khan, M. M. A. (2013). Forecasting of gold prices (Box Jenkins approach). *International Journal of Emerging Technology and Advanced Engineering*, 3(3), 662-670.
6. Khan, S., & Bhardwaj, S. (2018). Time series forecasting of gold prices. In *Emerging Trends in Expert Applications and Security: Proceedings of ICETEAS 2018* (pp. 63-71). Singapore: Springer Singapore.
7. Lambang, R. T. P. H., Prastya, I. W. D., & Barata, M. A. B. (2026). Indonesian Gold Price Forecasting Using Simple and Stacked LSTM with Expanding Window. *Journal of Applied Informatics and Computing*, 10(1), 406-416.
8. Marchi, E., Vesperini, F., Weninger, F., Eyben, F., Squartini, S., & Schuller, B. (2015, July). Non-linear prediction with LSTM recurrent neural networks for acoustic novelty detection. In *2015 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-7). IEEE.
9. Murugesan, R., Mishra, E., & Krishnan, A. H. (2021). Deep learning based models: Basic LSTM, Bi LSTM, Stacked LSTM, CNN LSTM and Conv LSTM to forecast Agricultural commodities prices.
10. Ojo, S. O., Adisa, J. A., Owolawi, P. A., Du, C. T., & Mphahlele, M. (2020, December). Stock market behaviour prediction using long short-term memory network and gated recurrent unit. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 615-621). IEEE.
11. Rather, A. M. (2021). LSTM-based deep learning model for stock prediction and predictive optimization model. *EURO Journal on Decision Processes*, 9, 100001.
12. Shafiee, S., & Topal, E. (2010). An overview of global gold market and gold price forecasting. *Resources policy*, 35(3), 178-189.
13. Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Physica d: Nonlinear phenomena*, 404, 132306.
14. Sunny, M. A. I., Maswood, M. M. S., & Alharbi, A. G. (2020, October). Deep learning-based stock price prediction using LSTM and bi-directional LSTM model. In *2020 2nd novel intelligent and leading emerging sciences conference (NILES)* (pp. 87-92). IEEE.
15. Zargar, S. (2021). Introduction to sequence learning models: RNN, LSTM, GRU. Department of Mechanical and Aerospace Engineering, North Carolina State University, 37988518.