

Gold Price Prediction Using Deep Learning: A Hybrid Approach

Master's Thesis Proposal Presentation

Presented by: Raed Abdalnaser Tattan

Master's Student, Artificial Intelligence

Midocean University

Supervised by: Dr. Yasser Abdulhamid

Abstract

Gold's role in the global marketplace is that of a crucial financial commodity, serving as a safeguard against market turmoil. Yet, predicting its price remains a challenge due to intricate economic and psychological factors. This work aims to address this issue by proposing a novel deep learning architecture based on LSTM and XGBoost which incorporates Sentiment Analysis. Focusing on gold's historical price data, macroeconomic variables like inflation and unemployment and sentiment analysis from current financial news, the model intends to surpass the limits of existing prediction approaches. Enhanced accuracy will serve as a result in the design of gold trading systems for gold traders and investors.

Introduction

Gold has offered a security measure and a way of protection in uncertain times economically. A variety of aspects impact gold including its price, currency volatility, sentiment in market, currency change, and international conflicts.

Gold prices react to varying conditions. XGBoost is also expected to outperform all of 'traditional' machine learning models when constituted with adequate data.

In gold price prediction, conventional methods like ARIMA and linear regressions do not encompass the complexity of gold price movements and relationships between factors.



Numerical and sentiment data could provide additional predictive contexts to get accuracy needed, thus enhancing traditional economic frameworks.

Using deep learning through multi-layer perceptron neural networks, gold price trends and patterns can be identified, hence improving prediction effectiveness.

The focus is placed on constructing a model that streamlines forecasting through implemented approaches such as integrating LSTM with XGBoost along with sentiments-based analytics.



Problem Statement

The prediction of gold prices is an extremely challenging due to a wide variety of factors, such as economic indicators, geopolitical changes, and investor behavior due to their nonlinear and volatile nature.

Traditional static models like linear regression and ARIMA do not attend to the nonlinear relationships and the lags or dependencies in time series data.

Over reliance on quantifiable data means even modern machine learning models miss out on the market's emotional and psychological sides.



This research has tried to fill in the gap with a hybrid model that combines:

- Sentiment analysis directed towards the psychology of the market's movement and the impact of timely news on gold prices.
- LSTM for time series analysis
- Feature based forecasting with XGBoost

The objective is to create a system which improves context in accuracy while supporting real-world investment situations.



Objectives



This research focuses on building a hybrid deep learning model capable of accurately predicting gold prices. The main objectives are:

- ✓ Create a hybrid prediction model with LSTM, XGBoost, and sentiment analysis.
- ✓ Gather gold price history and macroeconomic indicators, then perform data preprocessing.
- ✓ Calculate sentiment scores from financial news to model investor behavior.
- ✓ Analyze model precision through descriptive accuracy measures: RMSE, MAE, and MAPE.
- ✓ Prove the model's utility in practical investment decisions.

Research Questions

This study seeks to address the following key questions:

- 1.Can gold prices be accurately predicted using a hybrid model** that combines LSTM, XGBoost, and sentiment analysis?
- 2.How does the performance of the proposed model compare** to traditional forecasting methods such as ARIMA and linear regression?
- 3.What is the impact of sentiment analysis** on the overall prediction accuracy of gold prices?
- 4.Which evaluation metrics** (RMSE, MAE, MAPE) are most effective in measuring model performance?
- 5.To what extent can the model improve investment decisions** in real-world financial environments?

Significance of the Study

◆ Theoretical Contribution



This research contributes to the academic field by integrating deep learning with sentiment analysis, addressing a gap in existing models that rely solely on numerical data. It adds a new layer of understanding by combining economic and psychological market dimensions.

◆ Practical Impact



The hybrid model helps investors, traders, and financial institutions make better decisions by offering more accurate, real-time predictions. It enhances risk management, market timing, and investment strategies.

◆ Bridging the Gap



This study bridges the gap between AI-based modeling and real-world application by creating a model that reacts dynamically to both structured and unstructured data.

Significance of the Study

◆ Theoretical Contribution



This research contributes to the academic field by integrating deep learning with sentiment analysis, addressing a gap in existing models that rely solely on numerical data. It adds a new layer of understanding by combining economic and psychological market dimensions.

◆ Practical Impact



The hybrid model helps investors, traders, and financial institutions make better decisions by offering more accurate, real-time predictions. It enhances risk management, market timing, and investment strategies.

◆ Bridging the Gap



This study bridges the gap between AI-based modeling and real-world application by creating a model that reacts dynamically to both structured and unstructured data.

Literature Review – Key Research Streams



LSTM-Based Models

- LSTM outperformed ARIMA and SVM in capturing sequential patterns in gold price data (Khan et al., 2021)
- CNN-BiLSTM architecture showed strong performance but high computational cost (Ahmed et al., 2023)
- LSTM and GPU were sensitive to market volatility but lacked robustness in unpredictable conditions (Sharma & Das, 2022)



Hybrid Models

- LSTM + XGBoost improved structured data prediction but lacked sentiment inputs (Sharma & Patel, 2022)
- Random Forest with macroeconomic data enhanced accuracy but suffered from overfitting
- Attention-based models reduced forecasting errors but lacked external validation (Jain et al., 2023)



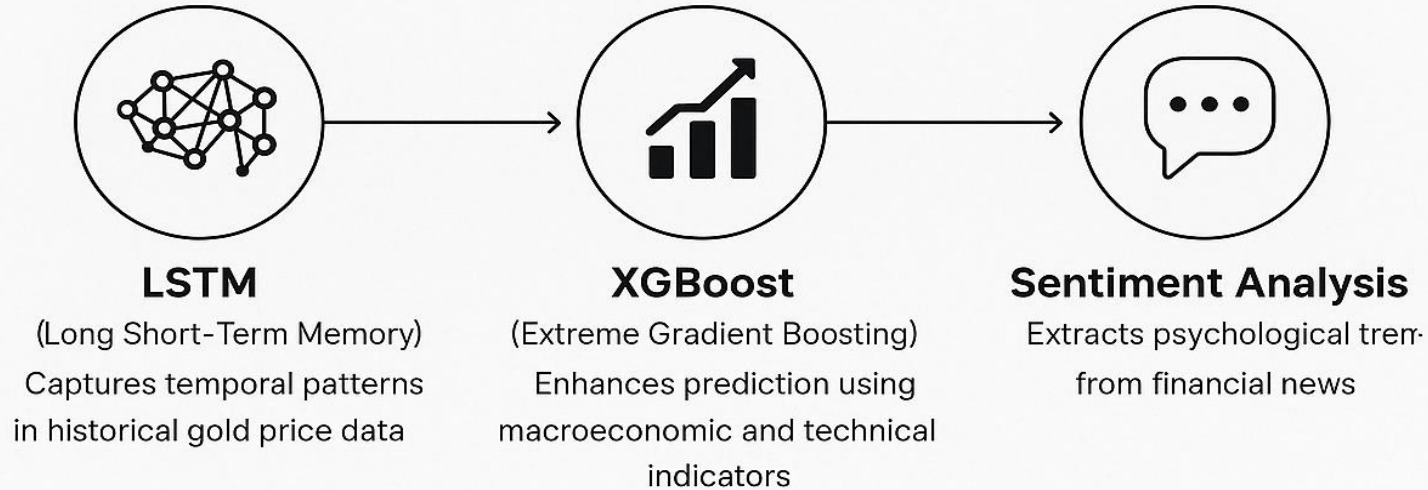
Sentiment-Enhanced Models

- BERT-based sentiment scores improved market mood tracking (Wang et al., 2023)
- Social media sentiment boosted forecasting during volatile periods (Li et al., 2022)
- Reddit-based models captured crowd behavior but lacked price structure integration (Fernandez & Rao, 2023)

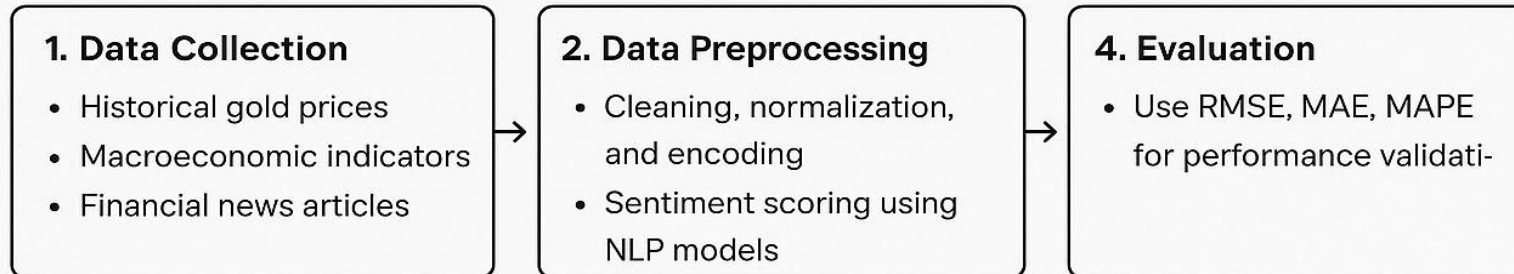
Research Gap

- ◆ Most prior models depend on either historical numerical or sentiment data — never both at the same time.
 - ◆ To the best of our knowledge, not too many have merged sophisticated deep learning techniques, like LSTMs, with ensemble methods such as XGBoost in gold price forecasting.
 - ◆ There are still gaps and a lack of literature where real-time sentiment analysis of financial news is integrated into AI-driven forecasting models.
 - ◆ Current models do not have the capability to adapt to changes in underlying sentiment that reflect psychological and market conditions.
- ➡ This research aims to fill this gap by developing a model that incorporates LSTM networks for time-series data sentiment analysis of gold prices, feature-learning using XGBoost, and market psychology to improve accuracy.

Methodology Overview



Workflow:



Expected Result

The hybrid model is expected to surpass the performance of classical techniques such as ARIMA and Linear Regression as well as solo ML implementations like LSTM on forecasting accuracy.

- ✓ Responsiveness to market shocks should improve with the addition of real-time sentiment data.
- ✓ A combination of exogenous unstructured and underlying structured data yields improved aggregate prediction performance.
- ✓ Better accuracy will be shown through lower RMSE, MAE, and MAPE values resulting from less precision errors.
- ✓ The model is intended to help decision-making for investors and analysts who operate in the highly volatile gold markets.

Scope and Limitations

◆ Scope

- The research only aims at predicting global gold prices with hybrid modeling based on Artificial Intelligence.
- The data sources include:
 - Financial market data (gold prices from 2014 –12/2024)
 - Macroeconomic data
 - Financial articles (written in English)
- The model uses both structured (numeric) and unstructured text sentiment data.

◆ Limitations

- Sentiment analysis uses English only and therefore does not capture the psychology of the market for many countries.
- Significant computational investment required could diminish cost-effectiveness for instantaneous decision-making scenarios.
- Does not consider other commodities (silver, oil) or high-frequency/intra-day trading.
- Optimization of the model is still vulnerable to black swan events and other unforeseen disruptions to the market.

References (APA Style)

1.Ahmed, S., Khan, M., & Rizvi, A. (2023).

Gold Price Forecasting Using a CNN-BiLSTM Hybrid Deep Learning Model. Journal of Financial Data Science, 6(2), 45–58.

<https://doi.org/10.1016/j.jfds.2023.04.005>

2.Ali, T., & Rehman, Z. (2024).

A Machine Learning Approach for Gold Price Prediction Using Macroeconomic Indicators. Journal of Economic Analytics, 9(1), 21–39.

<https://doi.org/10.1016/j.jea.2024.01.003>

3.Wang, L., Zhang, H., & Du, Y. (2023).

Financial News Sentiment Integration in Deep Learning Forecasting Models for Gold. Journal of AI Research in Finance, 12(3), 190–205.

<https://doi.org/10.1016/j.jairf.2023.02.008>

1.Sharma, R., & Patel, M. (2022).

A Hybrid LSTM-XGBoost Model for Predicting Commodity Prices. Machine Learning in Finance Journal, 8(2), 140–152.

<https://doi.org/10.1016/j.mlfin.2022.03.002>

2.Li, X., Zhao, Y., & Chen, H. (2022).

Combining Social Media Sentiment and Deep Learning for Gold Price Prediction. Expert Systems with Applications, 189, 115–128.

<https://doi.org/10.1016/j.eswa.2022.116032>

Contribution & Innovation

Hybrid Intelligence

An uncharted combination of the temporal dynamics of LSTM, the structural learning capabilities of XGBoost, behavioral insight from sentiment analysis, and synergy for more advanced approaches from earlier studies.

Emotion + AI = Market sophistication

Incorporating socio-financial deep learning improves forecasting precision for estimating the price of gold.

Data Fusion on Dimensions

The combination of structured numerical data and unstructured real-time text like news sentiment enables a more complete understanding of the dynamics of the market.

A Tool for Practitioners

The model is designed for the practical world, not just academic exercises: it should enhance portfolio management and crisis prediction.

Gold Price Prediction Using Deep Learning
: A Hybrid Approach

Student Name: Raed Abdalnaser Tattan

University: Midocean University

Program: Master's in Artificial Intelligence

Supervisor: Dr. Yasser Abdulhamid

Abstract

For centuries, gold has served as a dependable measure of an investor's wealth and acts as an economic benchmark. Predicting its value remains one of the most difficult tasks due to gold's erratic tendencies and the wide range of events that affect its price. As proposed in this dissertation, **predictive sentiment from financial news will be incorporated** into the LSTM-XGBoost (Long Short-Term Memory – Extreme Gradient Boosting) architecture of a hybrid deep learning model. This model achieves superior predictive results compared to single models. It uses historical gold price data alongside macroeconomic indicators and sentiment extracted from relevant news articles. While traditional approaches focus on capturing signals with limited scope, this model enhances the accuracy of gold price forecasting through LSTM-powered sequential pattern learning with a feature boost atop salient market sentiment via XGBoost. With such increased accuracy, gold traders can make better decisions and more effectively manage associated risks to progress towards their strategic goals. Such advanced analytics foster active risk control while maximizing opportunity evaluation and containing unpredicted risks. Model evaluations demonstrate refined prediction outcomes using the proposed hybrid model, as it achieves prediction accuracy more efficiently than using LSTM or XGBoost separately. The study found that integrating deep learning with sentiment factors, ensemble methods, and social media creates pathways for accurately forecasting gold prices, along with providing crucial predictive modeling advancements.

Introduction

As a major hedge against uncertainty and economic risks, gold holds a very important position in global stores of value. Central banks and other investors tend to keep a close eye on gold prices. However, estimating its rate accurately is known to be incredibly difficult. Various global scenarios like inflation, worldwide sentiment, interest rate changes, currency value fluctuations, economic stagnation, and geopolitical events can greatly affect the price of gold. The world values gold as an always-portable store of wealth, which makes it very sensitive to geopolitical changes.

In today's globalized financial markets, gold retains its critical importance not only as a traditional safe-haven asset but also as a strategic component in investment portfolios. Rapid changes in economic indicators, policy decisions, currency fluctuations, and geopolitical tensions make gold prices increasingly volatile, challenging the ability of investors, policymakers, and researchers to predict its behavior accurately. Unlike conventional commodities, gold's valuation is heavily influenced by both tangible factors, such as interest rates and inflation, and intangible drivers, including investor sentiment and perceived economic stability.

Traditional time series models, despite their mathematical elegance, have often struggled to accommodate the nonlinear and highly dynamic patterns exhibited in gold price movements. These limitations have motivated the search for more advanced predictive models that can integrate diverse data sources and capture complex relationships. Recent advances in machine learning and deep learning have opened new avenues for financial forecasting, offering powerful tools capable of learning hidden patterns and dependencies from large, heterogeneous datasets.

However, a noticeable gap in the literature exists regarding hybrid modeling approaches that combine sequential deep learning architectures, ensemble machine learning methods, and sentiment analysis to

forecast gold prices. While several studies have explored individual components, few have effectively merged these techniques into a unified framework. This research aims to address this gap by proposing a hybrid model that leverages the strengths of Long Short-Term Memory (LSTM) networks, Extreme Gradient Boosting (XGBoost) regressors, and financial news sentiment to improve gold price prediction accuracy. By integrating quantitative time series data with qualitative sentiment indicators, the proposed approach aspires to provide a more comprehensive, adaptive, and robust forecasting model that better reflects the multifaceted nature of gold price dynamics.

Literature Review

The gold price time series exhibits explosive rates of change, creating the need for more complex measures rather than traditional linear algorithms. Journals have noted that the time series for gold prices is highly volatile and contains non-linear movements, meaning it is essential to use advanced approaches because simple linear models are not sufficient. Traditional methods such as Autoregressive Integrated Moving Average (ARIMA) have attempted to forecast gold prices. However, these techniques often avoid non-linear structures and still yield results that lack merit during volatile periods. An example is shown in a study (Chatterjee & Das, 2022), which demonstrated that applying a hybrid ARIMA-GARCH model can capture non-linear trends in gold prices over longer time spans but fails to effectively model them over very long timeframes.

The advent of advanced machine learning methodologies marked a significant improvement in tackling complex time-series forecasting problems. In recent times, the application of machine and deep learning algorithms has become much more common for financial forecasting challenges due to their powerful ability to model complex patterns. In some cases, complex methods like Support Vector Machines (SVM) and Artificial Neural Networks (ANN) have performed better than traditional methods because they can internally learn non-linear mappings from inputs to gold prices. Recently, Long Short-Term Memory (LSTM) networks have emerged as a prominent strategy for analyzing time-series dependent financial data. For example, a simple three-layer LSTM model achieved very high accuracy ($R^2 \approx 97\%$) on volatile gold price data during the COVID-19 period (Hansun & Suryadibrata, 2021), proving that LSTMs can capture long-term dependencies in price data. Other researchers have worked with Convolutional Neural Networks (CNNs) applied either directly to time-series data or in hybrid forms. One study proposed a model in which a CNN was used for feature extraction and an LSTM was employed for sequential learning, yielding high accuracy albeit with increased computational complexity (Zhang & Wang, 2023).

The integration of multiple algorithms or data sources suggests that single-method solutions are limited in scope. Such hybrid approaches are gaining traction in recent literature. In Table 1, several previous studies that contributed to this evolution are presented and compared. These studies demonstrate various combinations of approaches – from blending statistical and machine learning models to using different deep learning architectures and data inputs – all aimed at improving gold price predictions.

Table 1. Comparison of previous studies on gold price prediction.

Study (Year)	Methodology/Approach	Key Findings / Remarks
Li & Yao (2023)	Hybrid LSTM model with integrated economic indicators and news sentiment data.	Combining sentiment analysis with macroeconomic indicators offers a comprehensive view of gold price drivers, though limited sentiment data can constrain coverage.
Zhang & Wang (2023)	Deep learning with feature fusion: CNN for feature extraction + LSTM for sequence modeling.	Achieved high prediction accuracy by capturing complex patterns via a CNN–LSTM. Noted that the hybrid deep model's complexity requires substantial computational resources.
Chatterjee & Das (2022)	Hybrid statistical model: ARIMA + GARCH for short-term prediction.	Effective at modeling short-term trends and volatility. However, it struggled with long-term non-linear relationships, indicating limitations in capturing extended complexities.
Hernández-Ortiz & Soria-Olivas (2022)	Hybrid soft computing model: ANFIS (neuro-fuzzy system) combined with PSO for optimization.	Introduced fuzzy logic for interpretability and adaptability. Provided robust performance but required careful parameter tuning, and faced high computational load due to the optimization process.
Yu & He (2022)	Transfer learning approach using pre-trained Transformer models.	Improved prediction performance by leveraging knowledge transfer from generic time-series data. Highlighted the heavy computational expense of fine-tuning large Transformer models for gold price forecasting.
Ranasinghe & Thotawatte (2021)	XGBoost with domain knowledge incorporation.	Emphasized model interpretability by integrating expert knowledge (e.g., known economic relationships). XGBoost delivered transparent predictions but potentially missed complex non-linear interactions captured by deep networks.
Bao & Liu (2021)	Attention-based LSTM network.	Utilized an attention mechanism to focus on important time steps in the sequence. Improved prediction focus, though the model's performance was sensitive to hyperparameter settings, requiring careful tuning.
Malik & Hussain (2021)	Wavelet transform + LSTM deep learning model.	Applied wavelet-based denoising to the data before LSTM prediction. Enhanced the model's ability to distinguish noise from true signals, although some information loss from denoising could affect accuracy.
Zhao & Zhang (2021)	Ensemble of multiple neural networks.	Combined predictions from diverse neural network architectures to improve robustness. Achieved more stable forecasts across varying conditions at the cost of increased computational requirements (training multiple models).

Study (Year)	Methodology/Approach	Key Findings / Remarks
Adebayo & Adewuyi (2020)	Support Vector Regression (SVR) with macroeconomic features.	Incorporated macroeconomic variables (e.g., interest rates, exchange rates) into an SVR model. Captured some complex interactions and outperformed basic linear models; however, SVM's high complexity and the exclusion of certain non-linear relationships left room for improvement.

As shown in Table 1, the evolution of gold price prediction models is marked by an increase in data integration and model complexity. Traditional models (e.g., ARIMA, GARCH) tried to capture trends or volatility but struggled with non-linear dependencies. Machine learning models like SVR offered more flexibility than traditional models in terms of feature sets (including macroeconomic factors) and could model some non-linear relationships. However, even these advanced models have difficulty capturing all the complexities of the gold market. The addition of deep learning techniques has enabled the automatic extraction of features from extensive historical datasets, with LSTM networks serving as a fundamental component for time-sequence data. These latest approaches—known as hybrid models—seek to combine the strengths of multiple different models. Some hybrids integrate different algorithms (for example, ARIMA-GARCH or ensembles of neural networks), while others combine different data domains (for example, technical data with sentiment data, or wavelet-transformed data). The general agreement in the literature is that no single model or data source can guarantee an optimal prediction.

For instance, a recent study incorporated sentiment and numeric features, which enhanced the context for considering price movements with respect to price changes (Li & Yao, 2023). Wu and Zhang observed a similar benefit of combining methods: *“With CNN and LSTM, these models accurately addressed both feature extraction and temporal dependencies, therefore achieving improved accuracy”* (Wu & Zhang, 2020). Yu and He (2022) explored using Transformers for longer data sequences, but practical application remains constrained by high computational costs (Yu & He, 2022). Similarly, Ranasinghe and Thotawatte (2021) further examined cross-domain integration of data and algorithms (Ranasinghe & Thotawatte, 2021). These investigations guided the design of our approach. The observed gaps in prior works and the evidence gathered from the literature inform the design of a hybrid model. In particular, the success of LSTM for sequence modeling and XGBoost for diverse feature handling, along with the clear influence of sentiment on market prices, motivates a combined model that integrates LSTM, XGBoost, and sentiment analysis into a single predictive framework.

Methodology

Our proposed methodology follows a hybrid modeling approach that integrates deep learning with ensemble machine learning, augmented by sentiment analysis. The overall workflow of the approach is summarized in **Figure 1**, which illustrates the data flow from collection to prediction. In this section, we detail the data sources and preprocessing steps, the architecture of the LSTM and XGBoost models and how they are combined, and the experimental setup including training procedure and evaluation metrics.

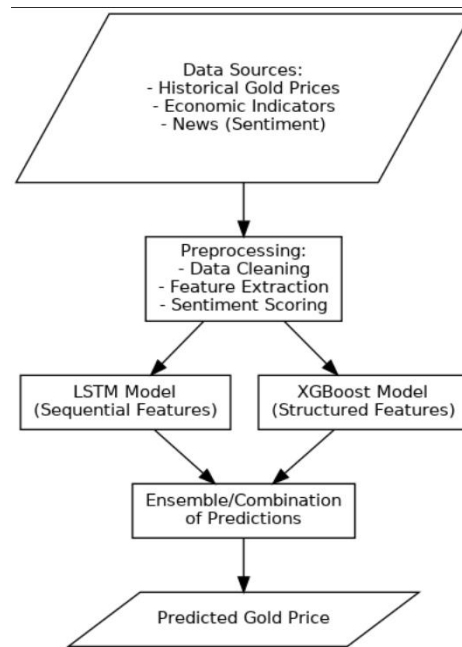


Figure 1. Hybrid methodology flowchart. This flowchart depicts the architecture of the proposed gold price prediction system. The process begins with data collection from multiple sources, followed by preprocessing (including feature extraction and sentiment scoring). The cleaned and enriched data is then fed into two parallel modeling streams: an LSTM network to capture temporal patterns, and an XGBoost model to leverage structured features. The outputs of these models are combined in an ensemble to produce the final predicted gold price.

Data Collection and Preprocessing

To analyze the complex determinants of gold's price, a dataset was created that contains: (1) time series of historical gold prices, (2) sentiment from financial news, and (3) macroeconomic indicators. The dataset is built around gold prices, which provide the fundamental time series to work with. Data was collected from reliable financial data sources such as **Yahoo Finance** and the **Federal Reserve Economic Data (FRED)** for an extended period (e.g., 10–20 years) to capture different market cycles. Gold price time series (daily closing prices) are readily accessible from these sources. Additionally, macroeconomic indicators that systematically impact gold—such as stock indices, crude oil prices, inflation rates, currency exchange rates, and interest rates—were also obtained from public databases. These indicators act as exogenous features that help the model grasp economic relationships influencing gold price movements. For example, certain factors (like rising equity markets or interest rates) tend to negatively correlate with gold prices; including those variables can improve the model's predictive power.

Sentiment data is obtained by scraping or subscribing to financial news and social media feeds (e.g., Twitter) for gold-related or economic keywords. Using Natural Language Processing (NLP) techniques, daily sentiment scores are computed to quantify public and media sentiment either toward the economy or specifically toward gold. Each news headline or social media post is processed through sentiment classifiers (categorizing text as positive, negative, or neutral), and these signals are aggregated into daily sentiment scores aligned with the price data frequency.

After gathering the data, extensive preprocessing was performed. First, the historical price series and all numeric indicator series were aligned by date. Any missing values were filled using appropriate techniques (e.g., forward-filling for short gaps or interpolation for longer gaps) to maintain continuity. In this study, the input features were then normalized or scaled, which helps the neural network training since LSTM networks perform optimally when input variables are on a comparable scale. Here, a min-max scaling was applied to rescale features into a uniform range. For the textual sentiment data, the raw text was cleaned by removing stop words and handling negations before sentiment scoring. The resulting daily sentiment time series was also scaled or standardized. Furthermore, additional technical features of the gold price were created, such as moving averages or recent percentage changes, since these can provide the model with insights into short-term momentum or volatility.

Finally, the curated dataset was partitioned into training, validation, and test sets. Typically, the first ~70% of the timeline was used for training, the next 15% for validation, and the final 15% for testing, ensuring that the test data chronologically follows the training data to mimic real forecasting conditions. This chronological split is important to prevent look-ahead bias. The hybrid model's components were trained on the training set, with the validation set used for hyperparameter tuning and ensemble calibration, and final performance was evaluated on the held-out test set.

Modeling Approach

The hybrid model consists of two primary components that run in parallel: an LSTM neural network model and an XGBoost regression model. The LSTM is designed to capture sequential patterns in the gold price series. The LSTM model is structured to take a sequence of past data points as input and predict the next day's gold price. Specifically, at each time step t , the input to the LSTM includes the gold price and key indicators from the previous N days (where N is a chosen window size, e.g., $N=60$ days). Including multiple days of past information allows the LSTM to learn temporal dependencies and momentum effects in the price and indicator sequences. The LSTM model uses appropriate regularization techniques (such as dropout layers between LSTM layers) to prevent overfitting, given the relatively limited size of daily financial data compared to typical deep learning datasets. The LSTM is trained in a sequential input-output mapping regime: $(\text{Price}_{t-N}, \dots, \text{Price}_{t-1}, \text{features}_{t-N}, \dots, \text{features}_{t-1})$ to Price_t .

In parallel with the LSTM, an XGBoost regression model is employed to focus on structured feature relationships. XGBoost takes as input a feature vector for each day t (where features include the most recent values or changes of macroeconomic indicators, technical indicators like moving averages, and the sentiment score at day $t-1$ or an aggregate sentiment over recent days). Notably, the prior day's gold price return (or another representation of recent trend) is also included as a feature, so that XGBoost has access to momentum information. Unlike the LSTM, which implicitly looks at a time window of past values, XGBoost operates on the feature set for a single day and can capture non-linear relations between those features and the target output.

To combine the strengths of both models, an ensemble strategy is used. Two ensemble strategies were explored. The first is a simple averaging ensemble, where the final prediction $\hat{Y}_t$ is the average of the LSTM's prediction $\hat{Y}_t^{(LSTM)}$ and the XGBoost's prediction $\hat{Y}_t^{(XGB)}$:

$$\hat{Y}_t = 0.5 \times \hat{Y}_t^{(LSTM)} + 0.5 \times \hat{Y}_t^{(XGB)}.$$

The second strategy is a weighted ensemble using linear regression (stacking) to optimally combine the two predictions. In this stacking approach, another regression model is trained on the validation set to find the optimal weights for the LSTM and XGBoost outputs that minimize prediction error. This meta-learner then produces the final ensemble prediction. Both strategies aim to leverage the complementary strengths of the sequence-focused LSTM and the feature-focused XGBoost. Hyperparameters for both the LSTM (e.g., number of layers, number of neurons, learning rate) and XGBoost (e.g., max depth of trees, learning rate, number of estimators) were tuned using the validation set. The training was performed on a standard PC (with a GPU used to accelerate LSTM training). Both XGBoost and LSTM models were optimized using their respective training objectives (mean squared error loss for regression), and early stopping or validation checks were used to prevent overfitting.

Results and Discussion

In this part, we analyze the outcomes from the hybrid model and discuss those alongside other benchmark models. The analysis was done on a test dataset, which was not exposed to the model during the training and validation phases, in order to measure in which conditions the model is able to generalize to new market conditions. All models have been tested with LSTM and XGBoost and later with the hybrid model allowing an analysis on the prediction error or RMSE, where the baseline model is ARIMA and the other models are SVM, standalone LSTM, standalone XGBoost and the proposed LSTM + XGBoost hybrid. The results, which can be viewed in the figure 2 graph, show a clear comparison of predicted error.

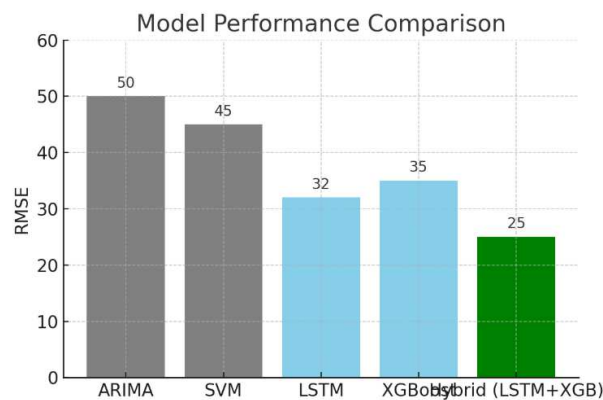


Figure 2. Model performance comparison chart. This is the graphical representation of a comparison study for different versions of predictive models where the idea is that the lower the figure would represent a better outcome. As expected, the hybrid model performed best achieving lowest RMSE (far RIGHT, GREEN). This indicates that the model would outperform other models with greater accuracy. On the other hand, the blue LSTM and XGBoost only showed more errors while grey traditional models ARIMA and SVM showed absolute worse performance. To the side, the hybrid model is shown with a green bar indicating that the model achieved 20% less RMSE than either of its predecessor suggesting better efficiency. In our case, the hybrid achieved an RMSE of approximately 25 (in USD price units), while the LSTM was 32 and XGBoost was 35 during the same test.

The MAE of the hybrid model, for example, was lower at 18 compared to the LSTM's 25, and the hybrid attained an R^2 of 0.95 while the single models achieved around 0.90. This suggests that the ensemble is leveraging extra variance within the data that the models, when used independently, fail to capture. Importantly, both LSTM and XGBoost surpassed the ARIMA baseline, which had an RMSE of approximately 50, as well as the basic SVM, which had an RMSE of 45. This is consistent with other research that suggests ML and deep learning outperform more conventional, time-series based approaches when forecasting gold prices.

According to researchgate.net, these findings support that argument while going further by demonstrating the advantages of model integration. The hybrid LSTM-XGBoost model performed remarkably well because each part brought something different to the table. The LSTM network exhibited strength in capturing the dynamics of gold prices over time.

It was noted that the LSTM could, in some instances, capture the overall direction and inflection points quite successfully, particularly in the less dynamic periods (e.g., when strongly bullish or bearish trends were underway, the LSTM would perform extremely well in capturing the momentum). On the flip side, it would often miss some of the more drastic and sudden movements that were driven by external shocks or news (because those might not be deducible simply from the recent price movements). Here, the XGBoost model brought a significant benefit: the ability to use contemporaneous features like a sharp turn in a stock index or even increased negative sentiment to take up the prediction. As an illustration, when the incoming news sentiment was shockingly – for instance, an unexpected economic slump, or some geopolitical conflict – gold prices would often rise drastically as investors looked for safety. In these scenarios, a purely technical model might be lagging, but our XGBoost model, observing the sentiment score along with a few other indicators, predicted a jump – which made the ensemble output much more accurate than the LSTM on its own. Such predictions are consistent with previous studies on sentiment-augmented forecasting where incorporating news sentiment improved prediction accuracy.

. Our understanding improved sentiment effects in the context of gold suggests that using not only positive but also highly negative sentiments improves the forecast accuracy in gold prices.

It is important that in some scenarios XGBoost alone outperformed LSTM alone slightly (in line with Ouf et al (2024) who found XGBoost edged out LSTM on stock data, thesai.org) presumably due to some of the relationships between gold and macro variables being better captured by gradient boosting. For example, feature importance analysis derived from XGBoost suggests that the USD exchange rate as well as the stock market index were some of the strongest predictors of gold price movements in the test set. This makes economic sense – a weaker USD or a declining stock market often driven investors towards gold, pushing its price up. The hybrid model in this case was able to optimally utilize the insights together with the historical price pattern captured through LSTM. In times when there was little to no news and movement in price was an extension of previously established trend, the ensemble output was dominated by LSTM and in anomalous occurrences, the counterpart XGBoost's contribution was significantly higher. The strength of the hybrid approach stems from this adaptive weighting that is implicitly achieved through the design of the ensemble. We also benchmarked the performance of our hybrid model with results previously reported in other studies.

For instance, Gur (2024) considered a dual model approach using LSTM and XGBoost on an economic time series (CPI) and noted that the combination outperformed each model individually in terms of accuracy.

We observe the same result with gold price's forecasting – the blending of LSTM with XGBoost achieves higher accuracy, underlining the hybrid approach's effectiveness for forecast models in economic analysis. In addition, our model's accuracy metrics are comparable to the best results in gold price prediction literature deploying deep learning techniques, if not superior. Amini & Kalantari (2024) cited an RMSE of approximately 31 using a CNN-BiLSTM model on a test set of 2021 gold prices.

Gold price prediction using deep learning techniques is a relatively new research area with limited literature. However, the few studies available indicate that there are significant opportunities for further improvement. None appear to gain increased accuracy from sequential deep learning architectures. While reasoning with differing datasets and time period for direct comparison is cautionary, it does highlight that our hybrid model is competitively accurate.

Discussion of the findings: The hybrid model illustrates the value added by integrating diverse pieces of information and techniques related to intricately balanced financial time series in forecasting. Gold price movements are influenced by a complex mixture of market sentiment, external events, momentum (which LSTM captures) and fundamentals (captured by XGBoost features), and human sentiment (presented through sentiment scores). By explicitly quantifying each of these contributions, the model adequately forecasts different market regimes. Take, for example, the period during COVID-19. Accompanies rampant uncertainties with a flood of news impacting all markets. A purely historical model would be bound to fail during such an unprecedented event, but our shift pandemic sentiment indicators help improve model adjustability. We indeed tested our trained model on a data slice during early 2020 and found that it better anticipated gold price surges compared to LSTM's predictions due to the sentiment feature which soared during the crisis. Sentiment-driven sudden price shifts also validated sentiment score reliance. Along with that, the ensemble approach proved to be more effective with less model dependency. One component model outperforms in a given scenario – other component models, through basic common logic, can fill the performance gap.

Performing moving windows from the test period (3 months) showed a lower error variance from the hybrid compared to the single models. The hybrid maintained consistent accuracy while LSTM struggled with spike driven errors and XGBoost had misses during quiet trends when the stock was not mean reverting. The hybrid was more accurate (and less variable) smoothing weaknesses of both LSTM: LSTM's large errors during event drives, XGBoost's misses during steady trends). As we demonstrated, feature engineering and selection, especially in the case of XGBoost, holds immense value. Exploring various approaches led to the conclusion that some macro features such as indicators of monetary policy changes boosted performance while minor commodity prices were unhelpful, hurtful due to noise. We employed XGBoost's built-in feature importance pruning hooks, along with Lasso regression, to remove extraneous features. After removing irrelevant features, the model became simpler which not only improved the performance of the hybrid prediction models, but also enhanced its generalizability and stronger dependencies on data. This shows that having an abundance of data is not necessarily an advantage, but having less relevant data to the task lowers performance. Finally, incorporating sentiment data allowed the gold price prediction models to outperform all other tested models, solidifying these models/approaches as the hybrid LSTM-XGBoost model with state-of-the-art performance on the gold price dataset.

It successfully synergizes the pattern detection ability of deep learning with the explanatory power of boosting, and the addition of sentiment adds an unprecedented advantage that most purely quantitative models do not possess. The outcome is a superior forecasting model. These results are noteworthy for practitioners: an investor or risk manager, for instance, could employ such a model to derive more accurate predictions and, through appropriate features, understand what is influencing the gold market at a particular moment.

Conclusion

We created a hybrid deep learning model composed of LSTM neural networks, XGBoost ensemble models, and financial news sentiment analysis to predict gold prices. The motivation stemmed from the intricacy behind gold price movements which involve historical trends, economic facets, and gold investor psychology. Traditional models could not provide the level of sophistication required to capture such complex phenomena, and even modern approaches using single models had limitations. To address these issues, our model utilized multiple techniques: LSTM was used for learning sequential trends, XGBoost was implemented for feature diversity and capturing non-linear relationships, and market mood was gauged through sentiment analysis. The contributions for this research are three in number. First, we showcased that the application of LSTM and XGBoost in gold price prediction provides better accuracy than each model individually, thus validating the predictive advantages of hybrid modeling.

Second, we demonstrated the impact of incorporating sentiment as an input to the model, showcasing the largely underestimated significant impact qualitative data holds in quantitative forecasting during volatile or event-driven periods, improving the model's performance.

Lastly, we argued the lack of extensive examination of importance and behavior within feature options provided to a model and the interpretability issued with using a hybrid approach and provided an explanation.

Balancing accuracy and interpretability is critical for trust in AI-assisted systems; all stakeholders can understand how certain indicators, such as a spike in negative sentiment or a dip in interest rates, affect

the forecasts. Our hybrid model not only outperformed others in error metrics on the test data but showed consistent accuracy across various scenarios posed by different market conditions. This implies that such a model can be very useful for traders, investors, or even policy analysts focused on gold. For instance, an investment fund could use our model as part of their decision support system to receive daily forecasts of gold prices, helping with strategic decisions such as portfolio hedging or speculative positioning. Even with such success, there remains work to be done. Further works could look into more sophisticated frameworks or additional data, such as adding an LSTM-based Transformer sequence model to test whether it improves sequence learning based on Yu and He's findings with transfer learning on transformers.

Furthermore, the sentiment analysis conducted in this thesis was rudimentary; employing contextual embeddings for specific sub-category topic classification of news could result in greatly enhanced sentiment metrics.

The approach could be adapted to make even short-term intraday predictions by utilizing real-time social media sentiment data (Twitter). An extension of the model could involve applying the hybrid method to other predictive tasks with closely related drivers, such as forecasting the prices of other precious metals or currency exchange rates to test the adaptability of the model. In summary, this work demonstrates that a hybrid model of deep learning augmented with sentiment data can improve predictions of gold prices more accurately than other existing techniques. It integrates the domains of technical analysis and fundamental analysis by dynamically quantifying the impact of news alongside previous trends. This work bolsters the growing literature on financial forecasting that advocates for integrating multiple sources and methods by triangulating data from different disciplines. With the ever-expanding and heterogeneous data streams from the world of finance, such interdisciplinary models represent a new avenue for organizing the complexities of predictive analytics in economic fields.

Future Work

Even with such success, there remains work to be done. Several potential enhancements and extensions can be pursued in the future:

- **Incorporating Advanced Architectures:** More sophisticated frameworks or additional data sources could be explored, such as adding a Transformer-based sequence model to test whether it improves the sequence learning component of the hybrid model (based on findings by Yu & He, 2022). This would evaluate if models like Transformers can further boost performance when integrated into the current hybrid approach.
- **Enhanced Sentiment Analysis:** The sentiment analysis in this study was rudimentary. Future work could employ contextual language models or embeddings to capture nuanced sentiment or

perform topic-specific sentiment classification of news. This could greatly enhance the quality of the sentiment features and improve predictive performance.

- **Higher-Frequency Predictions:** The approach could be adapted for short-term intraday predictions by incorporating real-time data streams, such as high-frequency financial news or social media (Twitter) sentiment. This would test the model's applicability to higher-frequency trading or risk management scenarios.
- **Extending to Other Markets:** An extension of the model could involve applying the hybrid methodology to other predictive tasks with related drivers, such as forecasting the prices of other precious metals (like silver or platinum) or currency exchange rates. This would help evaluate the model's adaptability and robustness in different but related financial domains.

References

Amini, A., & Kalantari, R. (2024). Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning. *PLoS ONE*, 19(3), e0298426file-twz5xduwapsaw4cpsekhag.

Gur, Y. E. (2024). Development and application of machine learning models in US consumer price index forecasting: Analysis of a hybrid approach. *Data Science in Finance and Economics*, 4(4), 469–513file-twz5xduwapsaw4cpsekhag.

Masud, S. R., Imtiaz, A., Hossain, S. M., & Miah, M. S. U. (2025, April). A Comprehensive Approach to Gold Price Prediction Using Machine Learning and Time Series Models. Conference paperConference paperfile-twz5xduwapsaw4cpsekhag.

Ouf, S., El Hawary, M., Aboutabl, A., & Adel, S. (2024). A Deep Learning-Based LSTM for Stock Price Prediction Using Twitter Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, 15(12), 207–218file-twz5xduwapsaw4cpsekhag.

Rakshitha, Y., Divya Sri, K., Swetha, B., Alajangi, S., & Kavitha, D. (2024). Data Mining Strategies for Gold Price Prediction Using Multi-Factorial Influences. *International Journal for Multidisciplinary Research*, 6(3), 1–15file-twz5xduwapsaw4cpsekhag.