Gold Price Forecasting with a Hybrid Queuing Search Algorithm and Extreme Gradient Boosting Regressor Model

Yingying Dai
Finance Office, ShanDong JiaoTong University, Jinan 250300, Shandong, China
Institute of International Education, New Era University College, Kajang, Selangor, Malaysia, China
E-mail: daiyingying666@163.com

Keywords: gold price fluctuations, forecasting, technical indicators, extreme gradient boosting regressor, queuing search algorithm.

Received: August 7, 2025

Gold is a conventional haven metal in wealth management and risk diversification tactics since its price trends mostly display signs of future political and economic trends. The forecasting of future prices of gold in the process of regression analysis is extremely challenging, considering their unpredictability and volatility. A hybrid model for a machine learning model is formed in this study by combining the Extreme Gradient Boosting Regressor (XGBR) with optimization techniques Mayfly algorithm (MA) and Queuing Search Algorithm (QSA). The hybrid model is developed in this study by using daily data points from 2011 to 2023, based on the variables Exponential Moving Average (EMA), Simple Moving Average (SMA), On-Bal Volume (OBV), Average True Range (ATR), along with Open, High, Low, volume, and Close patterns. 5-fold Cross-Validation is applied to test the performance of the model on 80% of the data used for training and 20% data used for testing. The QSA-XGBR model gives a better R^2 statistic with a value of 0.998 on the training data and 0.997 on the test data compared to other models, including Decision Tree (DT), Transformer, iTransformer, LSTM, XGBR, MA-XGBR, and Bi-LSTM with regard to accuracy. The SHAP value of each variable explains the importance of the variable. The variables with the highest contribution to the model are the Low variable, the High variable, and the Open variable. Additionally, this research assesses the generalization of the QSA-XGBR model by applying it to silver and crude oil, demonstrating its robust performance across multiple asset classes. Finally, this study demonstrated the significance of the improvement in forecast accuracy gained through the proposed model through the Diebold-Mariano test. For analysts and investors, the QSA-XGBR model provides insightful information that could be improved by macroeconomic variables and testing under volatile circumstances.

Povzetek: Študija predstavlja hibridni model QSA-XGBR za zelo natančno napovedovanje cen zlata, ki presega obstoječe metode in se dobro obnese tudi pri drugih sredstvih.

1 Introduction

Gold is a distinctive and prized substance that possesses traits of a tradable item, a valuable metal, and a form of money. Gold is the predominant option for investment among all the precious metals [1]. Gold serves not only as a means of adornment and embellishment in jewelry but also has significant value as a fundamental resource for industrial manufacturing. As a result, the gold market attracts considerable interest from governments, institutional investors, and individuals both domestically and internationally [2].

The expansion of the financial industry has elevated the gold market to a notable investment arena, putting it on par with stock, futures, and bond markets [3]. The price and trade volume of the gold market fluctuate dramatically and unpredictably due to various variables that interact within the complex and interlinked system of the gold market [4]. Creating models and predicting volatility have become very important considerations in areas like risk management, asset allocation, commodity valuation, and creating policy creation. Furthermore, it has been an important undertaking in financial marketplaces, attracting significant interest from academics and practitioners for quite some time [5]. Forecasting the value of gold is a critical economic concern due to government involvement [6].

Various Gold Price (GP) studies and forecasts rely on the application of econometric models and ML technologies, respectively [7], [8]. The other efficient econometric technique to be applied for nonstationary TS data prediction is the Autoregressive Integrated Moving Average (ARIMA) model, developed by Box and Jenkins and sometimes known as the Box-Jenkins model [9]. Guha and Bandyopadhyay employed an ARIMA in their analysis to forecast future GPs in India [10]. Generally, the ARIMA model is regarded as among the most well-regarded, largely because of its reliable statistical characteristics. However, with an inherent requirement for homoscedasticity, one major drawback is its limiting

representative TS features strictly to a linear form structure, thus resulting in the models deriving incorrect results with instances where TS features have a variable variance. The solution to this problem comes in the form of three models, namely: Exponential Generalized Autoregressive Conditional Heteroscedasticity (EGARCH) [11], Generalized Autoregressive Conditional Heteroscedasticity (GARCH) [11], and Autoregressive Conditional Heteroscedasticity (ARCH) [11].

Li et al. [12] put forth a model in their research that employs the GARCH approach to evaluate and predict the price of gold. Their research results indicate that GARCH-based methods considerably boost the correctness of predicting gold's volatility. Though the forecasting strategies that rely on econometric models have inherent limitations. Specifically, the models struggle to effectively learn from data as the sample size increases. The ML-based forecasting approaches may successfully overcome the previously described difficulty. For example, an ensemble learning method.

The gradient boosting technique has been completely improved with the introduction of XGBR by Chen and Guestrin [13]. The total amount of leaf nodes (LNs) and the adjustment of the LN score for the L2 regularization term define the tree's complexity in the XGBR technique. To mitigate overfitting, an L2 regularization term is incorporated into each Layer Normalization's score. A regularization component is incorporated into the XGBR cost function to regulate the model's complexity. The regularization term may be defined as the total of the squared L2 norms of the scores generated for every LN in the tree, plus the number of LNs. Given the bias-variance trade-off, this added regularization factor will decrease the variance of the learned model; hence, it will be easier to interpret and with less overfitting risk. Hence, XGBR works out to be better than its previous approach for a Gradient Boost Decision Tree (GBDT) [14], [15]. To achieve better results with real-world applications, studies must be made about hybrid algorithms, modifying current algorithms, and building new ones [16], [17]. Many optimizers were developed for use in many fields because optimizers are important; they have helped improve the performance of models when applied together with them [18], [19]. The salp swarm algorithm [20], the slime mold algorithm [21], and the African vulture's optimization algorithm are a few instances of metaheuristic algorithms [22]. In this research, a new metaheuristic algorithm called QSA is presented. It draws inspiration from the actions exhibited by humans in queues throughout various activities, such as shopping, waiting for buses, and completing transactions. It may be compared to the previously published human activity-based algorithm.

In light of the defects identified in prior studies and the dire need to improve models for more reliable and credible gold price forecasts, the research questions guiding this study are drawn from the following:

- Can an ensemble-based model, such as XGBR, capture nonlinear and volatile dynamics that exist in gold price movements, whereas traditional ML models, such as Decision Trees, cannot?
- How much can the metaheuristic optimizer QSA further improve the predictive performance of XGBR?
- Compared to other models such as DT, XGBR, and MA-XGBR, how does the proposed QSA-XGBR model perform in terms of accuracy and error reduction for different types of evaluation metrics?
- For the shorter term, will a model trained on a set of various technical indicators for trend, momentum, volatility, and volume-based features have better performance?

These research questions serve to provide a conceptual framework for the subsequent methodology and analysis. The purpose of this paper is to investigate whether the integration of an XGBR with the QSA can significantly enhance the performance of daily gold price forecasting by using 12 years of data. Thus, this paper proposes a hybrid QSA-XGBR model based on a range of trend, momentum, volatility, and volume-based technical indicators and evaluates the performance against a set of benchmark models comprising DT, stand-alone XGBR, and MA-XGBR.

The contribution of this study is divided into the following categories:

- This research compares various single and combined models, such as DT, XGBR, and XGBR with MA optimizer, and the suggested QSA-XGBR hybrid model, systematically. The findings revealed that the QSA-XGBR model demonstrates improved predictive performance.
- By incorporating technical indicators like Exponential Moving Average (EMA), SMA, OBV, ATR, and historical data such as Open, High, and Low, the feature set is enhanced, allowing for a more detailed comprehension of market dynamics.
- This study marks the initial effort to employ a QSA-XGBR model for forecasting shifts in GPs.

The findings provide a helpful instrument for financial institutions, policy, and shareholders in the improvement of investment risk mitigation, enhancement of decision-making, and identification of profitable opportunities in the gold market.

This work is structured as outlined: Section 2 discusses the literature on GP predicting models. Section 3 presents a comprehensive elucidation of the data and the interrelationships among variables. Section 4 details the methodology utilized in the suggested framework. Section 5 provides the empirical findings and examination. Section 6 offers the final thoughts.

2 Literature review

Liang et al. [23] put forth a novel decomposition-ensemble approach to boost the precision of GP predictions. They utilized the improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN) enhanced approach to break down the initial gold prices into separate layers with different frequencies. Then, they

utilize a hybrid forecasting approach of Convolutional Neural Networks (CNN), Convolutional Block Attention Module (CBAM), and Long Short-Term Memory (LSTM) to make predictions for every sublayer. The experimental outcomes demonstrated that combining CBAM, LSTM, and CNN enhanced the modelling ability and boosted the prediction precision. Besides, with the use of the decomposition algorithm of ICEEMDAN, the ACC has been significantly improved compared with the existing decomposition methods.

Cohen et al. [24] conducted a work to explore the capability of current ML methods for predicting fluctuations in GPs. For this research, the data used consisted of previous values of significant global stock market indexes, the S&P500 VIX volatility index, leading product futures contracts, and bond yields from numerous countries for up to 10 days prior. Three ML models, RF, GBRT, and XGBoost, were used to predict GP variations. The outcomes indicated that the inclusion of diversified financial variables and moving averages could be used to enhance the ACC of forecasting the future GP. The GBRT and XGBoost models could be helpful for making informed decisions about investing in gold.

Jabeur et al. [25] introduced a new method to forecast the variations in GPs with high ACC. Their research work presented a comparison among a set of ML models that included a newly designed approach called CatBoost and XGBoost. The data has demonstrated that XGBoost is better than other high-performing models. The article further suggested that the Shapley additive explanations (SHAP) technique would be useful for policymakers to understand the predictions of complex ML models and give an understanding of the importance of many elements affecting the GP. In summary, Jabeur et al. showed that incorporating XGBoost with the SHAP technique significantly improved GPP's ACC. Table 1 provides a summary of similar works.

Table 1: Similar studies concerning forecasting gold prices.

Authors / Year	Models	Dataset	Performance Metrics	Key Limitations	References
Cohen et al. (2023)	RF, GBRT, XGBoost + stepwise variable selection	Daily data (2011–2023), engineered features: global indices, VIX, commodities, bond yields	MSE = 0.0000186	High-dimensional dataset; no deep-learning comparison; lacks emerging-market data	[24]
Cohen & Aiche (2023)	CNN–Bi-LSTM with grid search optimization	44 years daily gold closing prices (1978–2021)	R^2 = 0.95	High computational cost; no macroeconomic input; no multi-step forecast	[26]
Ahmad et al. (2025)	Linear Regression	Daily gold futures (2021–2025)	R^2 = 0.96	ML models struggled with volatility; no deep learning; univariate	[27]
Zangana & Obeyd (2024)	LSTM and BiLSTM + SHAP explainability	Gold prices + macro factors (oil, FX rates, inflation, interest rates)	R^2 = 0.95	Limited historical window; only LSTM-based models; stationarity preprocessing reduces interpretability	[28]
Liang et al. (2022)	Hybrid ICEEMDAN decomposition + LSTM, CNN, CBAM (Convolutional Block Attention Module)	Daily COMEX gold futures prices and spot price (2010–2020)	Future price R^2 = 0.9939	Limited investigation of multivariable inputs or external shock factors; high computational complexity because of decomposition and deep learning layers.	[29]

- **Comparative analysis with related works**

A review of similar works in the study area of gold price forecasting reveals a number of gaps that spurred the creation of the suggested QSA-XGBR model. There are several gaps in these works. First, most models cannot learn complex nonlinearities in changes in the price of gold because they are based on fixed architectures with little hyperparameter tuning. Second, despite the benefits demonstrated in numerous financial applications, metaheuristic algorithms are still infrequently employed in gold forecasting, either by themselves or in conjunction with potent ensemble learners like XGBR. Third, rich sets of technical indicators, such as momentum and volatility features, are frequently underutilized in current research, which makes them crucial for capturing abrupt price changes. Lastly, there are not enough studies that systematically compare the outcomes of optimized and

unoptimized models to determine the true value added by sophisticated search algorithms.

Further research will be needed to fill the significant gaps in the current research on GPPs. A substantial amount of investigation has been carried out on the prediction of short- to medium-term outcomes, but very little has been done on the creation of long-term predictions, which are essential for investment decisions and strategic planning. Metrics for research evaluation could make model comparisons simpler to understand and improve the efficacy of each model's relative potency. Finally, for models to be genuinely effective in the real world, it is imperative to make sure they are robust and applicable to a range of market conditions. It is possible to develop more thorough and reliable methods for GPP by addressing these flaws. Many players in the commodities and financial markets would benefit from this. It should be noted that GP time-series forecasting is an exciting domain due to its importance in global financial markets. However, chaotic, noisy, and non-stationary gold futures time-series data create tough challenges for prediction. Classical ML models, like DTs, are not sophisticated enough to grasp the rich nonlinear dynamics of GP fluctuations and thus provide suboptimal forecasting performance. Even ensemble methods that show outstanding performance, like the XGBR, usually must be further optimized for high predictive ACC to be realized. Although some optimization algorithms, like the MA, have been researched in similar areas, their applications remain minimal when it comes to forecasting gold futures. Apart from that, new types of optimization techniques, such as the QSA, have not been well integrated with ML models so far to assess their potential capability in enhancing prediction ACC. The current state of research mainly addresses individual or simple combination models, without any systematic exploration of the advantages of hybrids by integrating advanced optimization with robust ensemble methods. An outstanding fraction of this existing research also overlooks diversified feature engineering that boosts model performance. Thus, full utilization of the trend, momentum, and volatility indicators EMA, SMA, OBV, and ATR is not considered in previous works, but gives a wider scope that thus restricts the prevailing models from capturing the market dynamics. The following study will bridge these gaps by proposing a hybrid QSA-XGBR model, using advanced optimization algorithms with a diverse set of technical indicators to substantially raise the forecasting ACC of gold futures prices. This paper presents a new and robust approach that overcomes the limitations identified in the literature through comparative analysis against single and combination models.

3 Dataset

3.1 Data description

The aim of this work is to predict the everyday GP using examination. The data included in this study were gathered for training purposes from January 3, 2011, to June 4, 2021, and for testing purposes from June 7, 2021, to December 29, 2023. The data is partitioned into several sets for training and testing purposes. The train-test split method is a procedure for gauging the performance of ML algorithms by evaluating their forecasts on information not utilized to prepare the model. Prices in this article are thus based on the global data, which describes aggregated international market statistics and indicators for a thorough and reliable analysis of GP trends. To guarantee ACC and trustworthiness, the information utilized in this research was obtained from dependable financial sources like Investing.com and Yahoo Finance. Both sources, Investing.com and Yahoo Finance, publish their data publicly, thus increasing the transparency and reproducibility of this research. However, this procedure is not suitable for small datasets or when dealing with imbalanced classes. As shown in Fig. 1, it is important to mention that eighty percent of the data was utilized for training, while twenty percent was utilized for assessing the pattern.

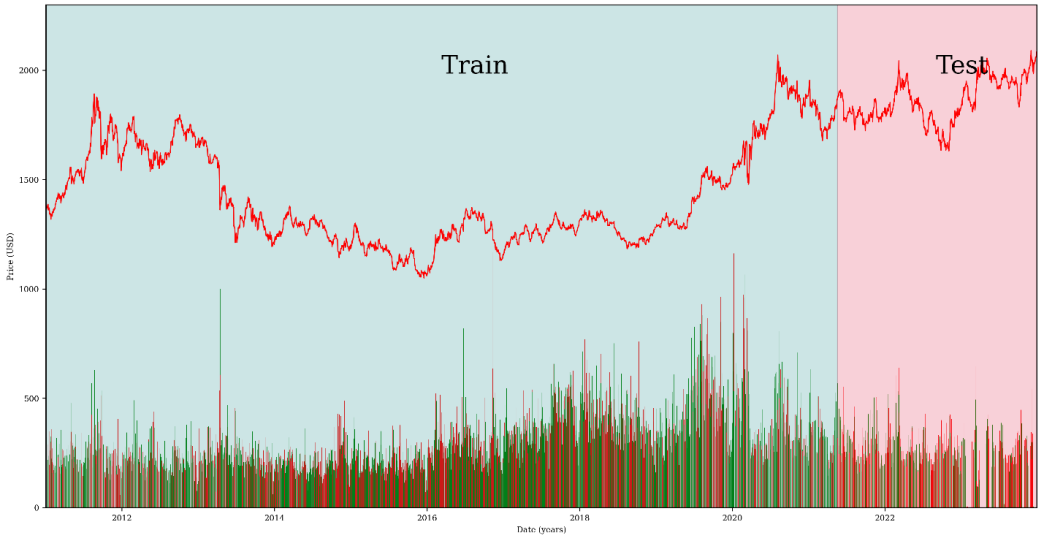


Figure 1: Dividing the data set into train and test.

This study has examined various categories of variables. The first category consists of technical indicators such as ATR, Money Flow Index (MFI), Relative Strength Index (RSI), SMA, Moving Average Convergence Divergence (MACD), EMA, Momentum, Williams, Stochastic, and OBV. The second category is based on historical data, including Open, High, Low, Close (OHLC), and volume. Examining and investigating factual data and its dispersion is a fundamental duty in descriptive statistics across several study fields, and it frequently acts as a basis for thorough analyses. Comparing data distributions involves various analysis tasks that encompass a wide range of aspects. These tasks range from analyzing global characteristics, such as the type, shape, and skewness of distribution, to analyzing local characteristics, which include comparing the frequencies of values or estimating differences at an individual level. Fig. 2 displays data distribution, showing dispersion in the data points over different values in a data set.

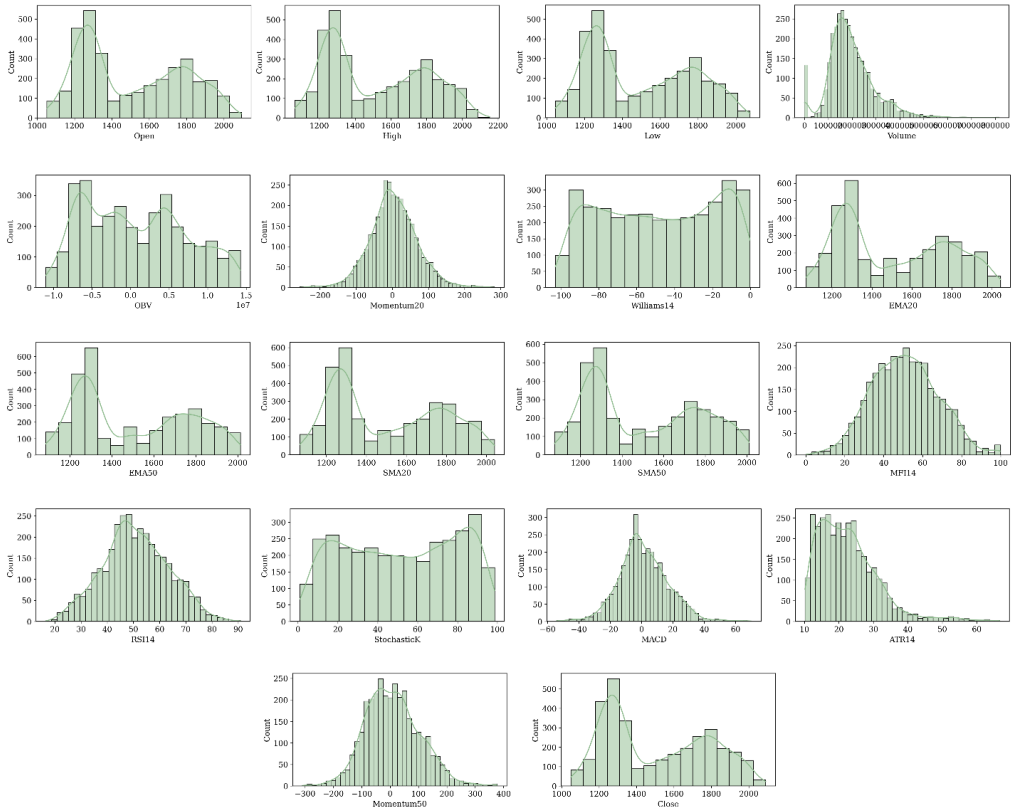


Figure 2: The distribution of historical variables and technical indicators

3.2 Correlation and statistical analysis

This study focuses on the correlation between predictor factors and GPs. However, correlation analysis is considered one of the best means through which one can decipher the strength of this association, as noticed from various studies, one of which is the work of Haque et al. [30]. The central thrust of the present study will be to establish a correlation that explores the relationship between 2 variables that have some logical link. It also tries to find the direction and strength of such a relationship so that it can describe it precisely. This kind of connection reveals how changes in one element would impact the other element, either in a positive or negative manner.

The Pearson cross-correlation values show the observed qualitative relationships between gold and the chosen variables. This coefficient gauges the degree of linear relationship and the extent of linear correlation between a pair of variables. It gives a value between $+1$ and -1, initially presented by Pearson in 1895. Pearson (1895) determined that a coefficient of $+1$ denotes a perfect positive association, 0 signifies no association, and -1 signifies a perfect negative association. The Pearson correlation coefficient is calculated by taking the covariance of a pair of variables and dividing it by multiplying their Standard Deviations (SDs). $\rho_{c,s}$ represents the correlation coefficient between two variables: P_c (the gold closing price) and P_s (one of the predictor variables). cov represents the covariance between P_c and P_s, which measures how the two variables change together. σ_{P_c} and σ_{P_s} represent the standard deviations of the gold closing price (P_c) and the predictor variable (P_s), respectively.

$$\rho_{c,s} = \frac{\text{cov}\,(P_c, P_s)}{\sigma_{P_c} \times \sigma_{P_s}} \tag{1}$$

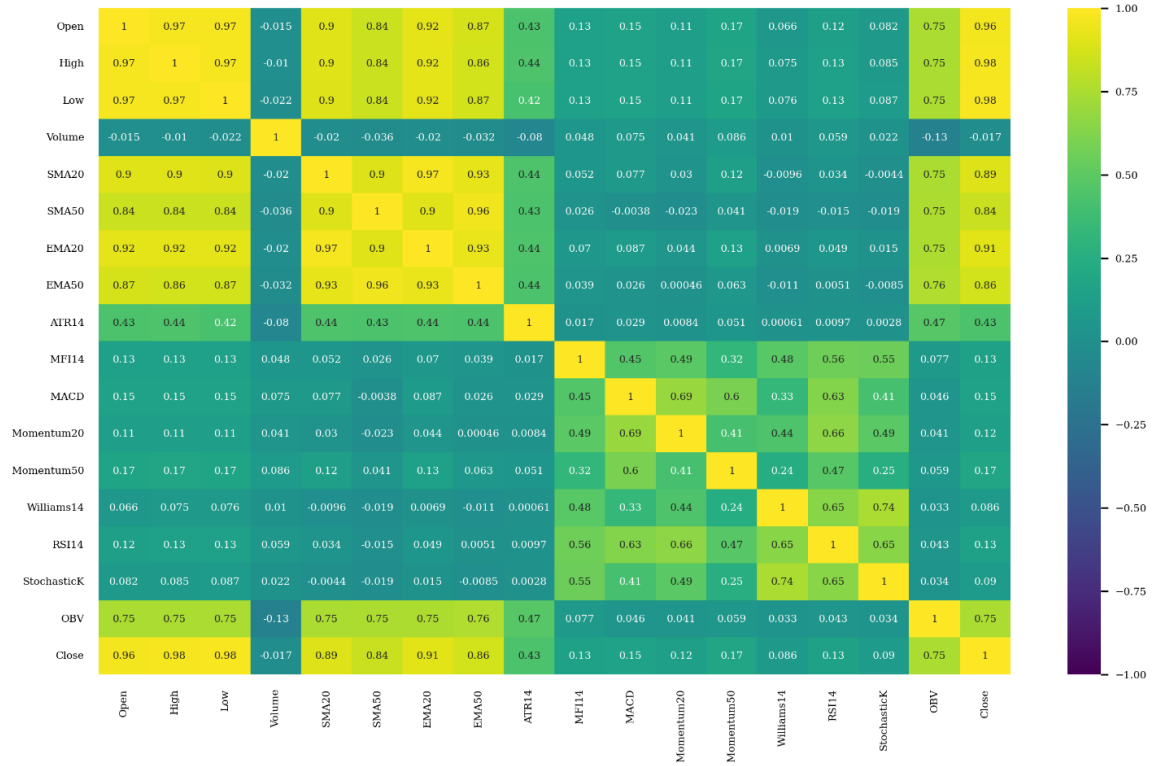


Figure 3: The Pearson correlation coefficients between the key features in the dataset

It is important to mention that if the variable falls between the range of -0.3 and +0.3, it signifies a lack of correlation with gold. The analysis of Fig. 3 indicates a positive link between gold and variables exceeding +0.3 or falling below -0.3. It is noteworthy that gold and the following variables have a positive correlation: open, high, low, SMA20, SMA50, EMA20, EMA50, ATR14, and OBV. Consequently, to predict the closing GP, these variables are employed as input variables. The report displayed in Table 2 thoroughly examines the dataset. Table 2 gives a detailed statistical outline of the input and output property information. This process guarantees full comprehension of the facts. The table contains a large number of statistical measures, including the mean, variance, count, maximum (max), SD, coefficient of variation (CV), and minimum (min). The data may be accurately and thoroughly analyzed using these measures.

Table 2: Descriptive statistics of the dataset features for gold price forecasting

	count	mean	Std.	min	max	variance	CV
Open	3296	1508.662	273.9128	1051.5	2094.4	75028.24	18.15601
High	3296	1519.222	276.2637	1062.7	2152.3	76321.61	18.18455
Low	3296	1497.213	271.2242	1045.4	2074.6	73562.57	18.11527
SMA20	3296	1506.519	270.9458	1067.1	2040.8	73411.62	17.98489
SMA50	3296	1503.591	267.0761	1075.13	2011.208	71329.62	17.76255
EMA20	3296	1506.501	270.3935	1069.573	2042.812	73112.63	17.94844
EMA50	3296	1503.622	265.9901	1084.339	2012.246	70750.75	17.68996
ATR14	3296	22.63149	8.843433	10.16093	66.64821	78.2063	39.07578
OBV	3296	698078.8	6495604	-1.1E+07	14255360	4.22E+13	930.4973
Close	3296	1508.327	273.8426	1049.6	2089.7	74989.75	18.15538

3.3 Indicators used in technical analysis

Financial forecasts rely on technical indicators due to their ability to effectively track and identify trends, filter out irrelevant fluctuations in TS data, promptly adapt to market volatility, and reveal underlying elements that drive alter. This work employs indicators like momentum, trend, and volatility to examine nonlinear dynamic elements, including asymmetry, tendency, nonlinearity, and long memory properties. Different attributes are taken out to efficiently record nonlinear dynamics in time series and ensure the strength of the gold forecasting model. These features include volatility indicators like ATR, trend indicators like SMA and EMA, and momentum oscillators like OBV.

3.3.1 Volatility indicators

J. Welles Wilder, Jr. originally developed the volatility indicator for commodities, and it is employed in technical

analysis [31] Eliminating the price enables the determination of the actual range and the ATR. In this process, adjusting the past price does not impact volatility when a consistent value is either included or taken away from each price. When short-term and long-term contracts are aggregated, reverse adjustments are frequently implemented. Nevertheless, the conventional method employed to compute stock price volatility is not constant. Analysts and futures traders frequently employ the ATR to assess volatility. Conversely, stock market participants and experts often depend on the SD of logarithmic price ratios. The computation of ATR can be ascertained in the following manner:

$$T = \max[(\text{high} - - \text{low}), |\text{high} - - \text{close}_{\text{prev}}|, |\text{low} - - \text{close}_{\text{prev}}|] \quad (2)$$

$$A_t = \frac{A_{t-1} \times (n-1) + T_t}{n} \quad (3)$$

$$OBV = \begin{cases} OBV(\text{prev}) + \text{Volume}; & \text{Close} - \text{Close}(\text{prev}) > 0 \\ OBV(\text{prev}) - \text{Volume}; & \text{Close} - \text{Close}(\text{prev}) < 0 \\ OBV(\text{prev}); & \text{otherwise} \end{cases} \quad (5)$$

If the ultimate price exceeds the prior closing value, then: Prior OBV + Current Volume = Current OBV. Suppose the CP falls from the previous CP, Prior OBV - Current Volume = Current OBV. In the event that the CPs are the same as the earlier CP, then: Prior OBV = Current OBV (no alteration).

3.3.3 Trend indicators

Moving averages (MAs) are frequently employed to examine data points by calculating several TS subsets series averages. In addition to measuring momentum and identifying regions of support and resistance, MAs validate trends. MAs can trend horizontally, upwardly, or downwardly, although signals can take some time to appear. However, the indicator may give erroneous signs if the period is too long. The indicator's sensitivity to price fluctuations increases with the specified decreasing time. By smoothing the price TS, these indicators track the price. When used in isolation, they identify the trend's direction rather than forecasting the price's movement. Reduced duration averages result in a stronger correlation with fluctuations in the underlying asset values, whereas longer-term averages only detect major patterns [34].

A SMA is a statistical computation employed to examine data elements by creating a string of means derived from various segments of the complete data set. The SMA indicator, often employed as a tool, signifies the basic mean of GPs across a specific timeframe.

$$SMA_t = \frac{1}{T_{SMA}} \sum_{i=k-T_{SMA}+1}^{k} C_i \quad (6)$$

T_{SMA} indicates how many periods were utilized to compute the average. C_i refers to the CP of the gold, and k expresses the current observed period's relative Location (Loc).

The exponentially decreasing weight elements of the EMA are used as a first-order infinite impulse response

$$A = \frac{1}{n}\sum_{i=1}^{n} T_i \quad (4)$$

In this formula, the true range (T) accounts for yesterday's Closing Price (CP) if it falls outside today's high-low range, with A denoting the ATR.

3.3.2 Momentum oscillators

Volume indicators provide useful information when applied to technical analysis. Additionally, there is a favorable correlation between them and the direction of prices, and traders who utilize this knowledge outperform those who do not [32]. Created by Joe Granville in 1963, the OBV indicator is a cumulative measure that increases volume when prices close higher and decreases volume when prices close lower. It serves to measure the force behind buying and selling activity. Chartists can confirm trends and identify divergences with the use of this indicator [33].

filter. The EMA, which is superior to the SMA in identifying market patterns, can be defined as

$$EMA_t = C_{i-1} \times p - EMA_{t-1} \times (p-1) \quad (7)$$

Here, p signifies the extent of weighting reduction, representing a consistent smoothing element ranging from 0 to 1. C_{i-1}refers to the prior period's CP. $EMA_{t-1}$and EMA_t reflect the EMA's values from the earlier and current computations, respectively. The smoothing factor, or p, is expressed inside the specified range [0,1] as $p = \frac{2}{T_{\text{SMA}}+1}$. Previous observations are discounted at a faster rate with a higher p-value [35].

Moving averages of previous CPs are used to construct the MACD [36]. A first-order momentum oscillator is created by the difference between the slow and fast EMAs.

3.4 Historical data

Opening Price (OP): The OP refers to the initial price of each share when trading begins. The initial price offers a dependable prediction of the daily fluctuation in the financial market. The OP does not necessarily have to be the same as the closing price of the previous day because the financial market operates like an auction when buyers and sellers come together to negotiate with the highest bidder.

Day's highest/lowest price: The preceding day's highest and lowest prices are documented, offering insights into the typical fluctuation of the market during a trading day and its impact on the final CP.

CP: An adjusted CP signifies the CP of an asset that has been revised to account for any payouts or corporate activities that happened before the commencement of the following trading session.

When conducting a thorough analysis of historical returns or assessing recent returns, the adjusted CP is commonly utilized.

To guarantee the methodological soundness and reproducibility of the suggested forecasting model, precise data preprocessing procedures were carried out. The portals Investing.com and Yahoo Finance, which offer continuous and carefully chosen financial time-series data, provided the daily OHLC values and the technical indicators that go along with them. Technical indicators like SMA, EMA, ATR, Momentum, and OBV were methodically extracted from the raw price series in strict compliance with their canonical mathematical formulation. The 14-day ATR and the 20- and 50-day moving averages were among the window lengths that were explicitly stated.

Min-max normalization was applied to all derived indicators and continuous input variables to enhance numerical stability during training and to support the convergence behavior of QSA-driven hyperparameter optimization. Financial time series outliers are typically an indication of actual market behavior rather than errors in the context of price spikes, as has long been known. In particular, the outlier observation was replaced with the average of the observations that came right before and after it. The interpolation algorithm adds none of its own and preserves the time series' trend and smoothness.

Furthermore, because of technical tools like the SMA, EMA, and ATR, which smooth out transient series irregularities with a window, additional series smoothing is accomplished without removing important financial data.

Lastly, the data is divided temporally in an 80/20 ratio to preserve causality in time and prevent any kind of information leakage. All of these data preprocessing procedures work together to create a transparent and repeatable academic foundation.

4 Methods and materials

4.1 Research outline

A detailed description of the processes involved in the development and evaluation of the QSA-XGBR forecasting model for gold prices is discussed in this section. Data acquisition and feature extraction are the initial processes involved in conducting the study. These have been arranged systematically for the purposes of this investigation. Fig. 4 highlights the entire procedure undertaken in this investigation.

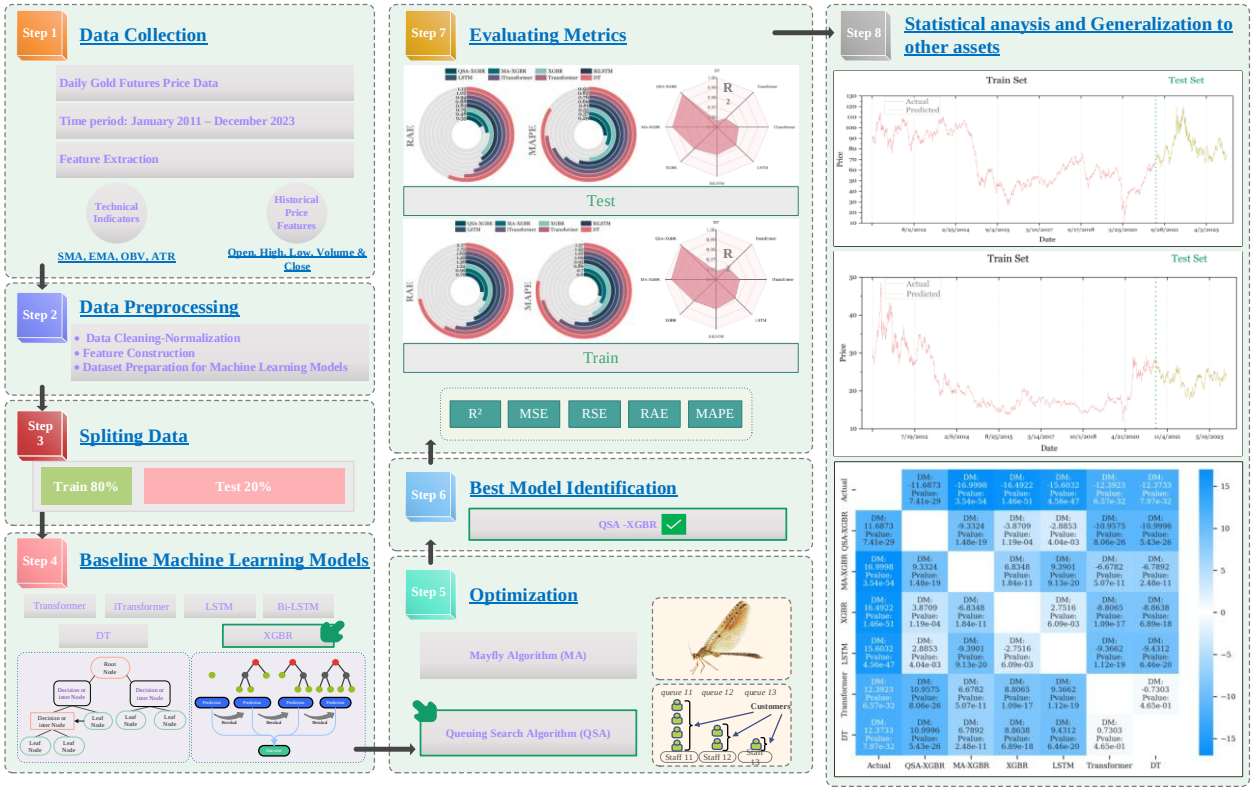


Figure 4: The outline of the research and proposed model to predict the gold price.

4.2 Models

4.2.1 Decision tree algorithm

DT regression is a common non-parametric ML method that establishes a relationship between input variables (X) and a constant target variable (Y) through dividing the input area into separate regions using a repetitive splitting process. The crucial aim of this splitting approach is to reduce the heterogeneity inside each node [37]. The commonly used metrics for quantifying impurity are MSE, MAE, and R^2. The purity of a node can be calculated using Eq. (8) to determine the MSE.

$$\text{MSE} = \frac{\sum (y_i - \hat{y})^2}{N} \tag{8}$$

Here, N signifies the overall count of samples within the node. The variable y_i indicates the real target value for the ith sample, while $\hat{y}$ signifies the mean target value of the samples located in the node.

4.2.2 Extreme gradient boosting regressor

Regression using a DT approach requires the minimization of OFs that apply to the training data sets. Stated differently, the training data set utilized in decision-based approaches is optimized through the employment of classifiers and tree regression parameters.[38] The three main sections contain descriptions of tree-type regression and categorization. 1) LN 2) internal node 3) As depicted in Fig. 5, the primary or Root Node (RN) stands out as the most prominent element of the classification and regression tree [39].

A binary decision-making method can be used to carry out the procedure of splitting the RN or internal node. With the help of the provided technique, data sets are categorized at the root and arranged in the interior nodes. The final class is obtained by resuming the previous operation, resulting in the LNs. A robust network can be created by assigning weights to each classification and regression tree during the training process, using the gradient-boosting approach [40]. It is possible to train a dataset with n examples, m characteristics or features, and y as the model target or output using the equation developed for the tree-based method [41].

$$y_i = \sum_{k=1}^{N} f_k(X_i) \tag{9}$$

Wherein the function's necessary condition is as follows [41].

$$f = \{f(X) = \omega_{qx}\}, (q: m \rightarrow (T, \omega) \varepsilon T) \tag{10}$$

During the classification stage, the OF will diminish errors and enhance the prediction model's resilience. By following the steps outlined in the XGBR diagram, the OF can be developed as outlined below [42]:

$$L = \sum_{i=1}^{N} l(\hat{y}_i, y_i) + \sum_{k=1}^{N} \xi(f_k) \tag{11}$$
$$\xi(f_k) = \gamma T + 0.5\lambda\omega^2$$

In the XGBR approach, the symbols $l, \xi, \gamma,$ and λ represent the loss function, minimal loss, regulation coefficient, and regulation functions, respectively. The function of the above elements exhibits a convex relationship, subject to differentiability conditions, restrictions on division and the intricacy of the tree-structured model. These elements play a role in the facilitation of variance expansion in the model. By computing the given equation and reducing the OF, it is possible to modify the final results [43].

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \tag{12}$$

The XGBR model offers a significant advantage in its ability to utilize a shrinkage approach. Implementing this methodology will mitigate the issue of overfitting in DT-based models. To provide further clarification on this method, it is possible to define the learning factor and apply regulation to the learning rate to control the distribution of weights across the categories [42].

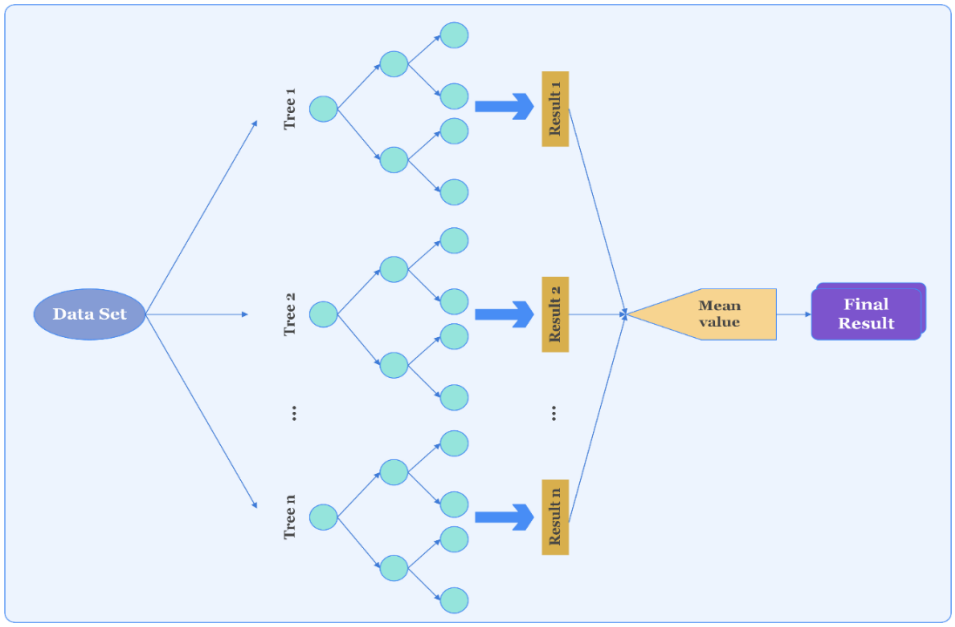


Figure 5: The structure of an extreme gradient boosting regressor model.

4.2.3 Long short-term memory

The LSTM model is a sort of RNN meant for detecting long-term dependence in a series of data. In time series forecasting problems, when current observations are affected by past observations, RNN networks can be helpful, such as LSTMs. Since LSTMs use a special type of memory, called "LSTM cells," to remember information, they will be able to find complicated patterns. Therefore, finances can be predicted using RNN networks.

4.2.4 Bidirectional LSTM

The Bi-LSTM model is an extended version of the traditional LSTM model, in which it allows the model to process the input sequence of forward and backward

directions. The ability of the Bi-LSTM model to analyze both past and future directions of the input sequences gives a proper understanding of the sequence. This improves the performance of the model on time series forecasting. Since the prediction needs past and future data for accurate prediction, the Bi-LSTM model could be appropriate for financial time series.

4.2.5 Transformer

Instead of processing the input data step-by-step, the Transformer model's self-attention mechanism enables simultaneous processing. For this reason, the model will be very effective in picking up the dependencies and patterns in time-series data, especially the long-range ones. For this reason, the Transformer model is extremely useful when it becomes relevant to understand the dependency between two far-apart points, such as in financial forecasting.

4.2.6 iTransformer

iTransformer refers to a transformer architecture with additional enhancements over the model performance. The main advantage of the iTransformer over the vanilla transformer architecture in earlier models is that it improves upon some of the shortcomings found in the latter. This supports its application in financial time series forecasting tasks, such as forecasting changes in the price of gold.

4.3 Optimizer

4.3.1 Mayfly algorithm

MA is a novel nature-inspired hybrid structured swarm intelligence algorithm that integrates the benefits of GA, FA, and PSO. MA draws inspiration from the courtship, mating, and reproduction behaviors of mayflies. Mayfly (MF) behavior influences certain operations in PSO, such as crossover, hence enhancing the algorithm's performance. Fig. 6 shows that the fundamental operational premise of the MA algorithm involves the stochastic generation of populations, including male and female mayflies. Every MF serves as a potential solution to the problem, positioned randomly throughout the search space, and subsequently assessed for its performance. The flying path of each MF is influenced by its unique experiences and interactions with other mayflies. Each MF modifies its flight path based on its best Loc (pbest) and the group's current best Loc (gbest), according to the outcomes [42].[44].

Multiparameter complex models are not easy to tune. One class of automatic optimizers that may serve this purpose is the metaheuristic optimizer. These optimizers will result in an optimum Hyperparameter (HP) set within a very short frame of time with a good degree of ACC. As shown in Table 3, in the current paper, the MA optimizer has been utilized to boost the HPs of the XGBR model. It was done by setting up the lower and upper boundaries of the HPs; from there, it allowed the optimizer to have a best-possibly-efficient search to come up with optimum solutions. This approach proved generally very successful in efforts towards the optimization of complicated multi-parameter models.

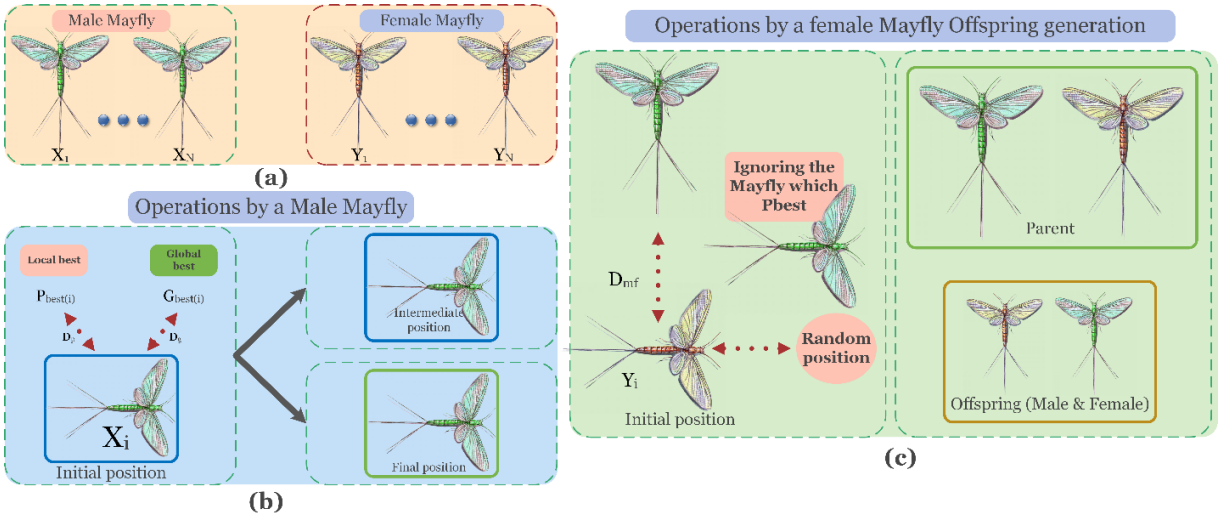


Figure 6: The operations of the Mayfly algorithm during the optimization process.

4.3.2 Queuing search algorithm

The QSA, devised by Zhang et al.[45], emulates human queuing behaviors. It considers various overarching phenomena, including the effect of consumers adhering to the queue on the pace of service, the actions of individual customers, and the impact of other customers on their service experience when the queue order is not upheld. As shown in Fig. 7, the QSA framework achieves equilibrium between exploitation and exploration through 3 distinct stages referred to as business1, business2, and business3. Fig. 8's flowchart aids in the explanation of QSA. The precise particulars of these measures are outlined in reference [46].

The HPs of the model are illustrated in Table 3, which the QSA optimizer computes for the best results by adjusting these HPs.

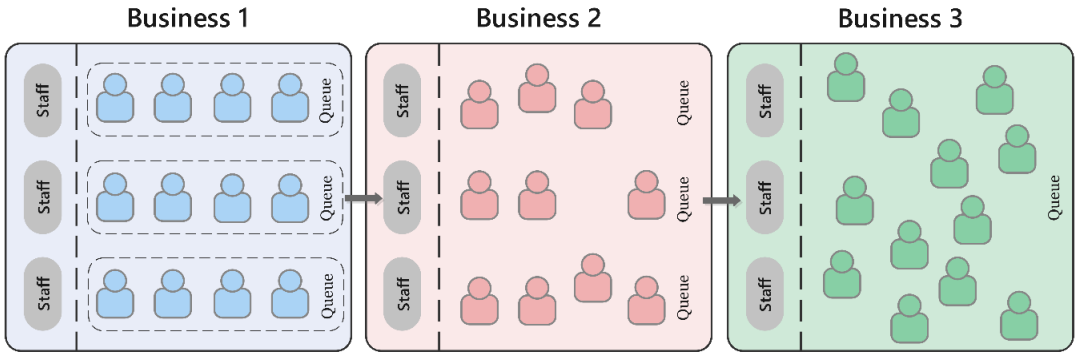


Figure 7: The operational flow of the queuing search algorithm.

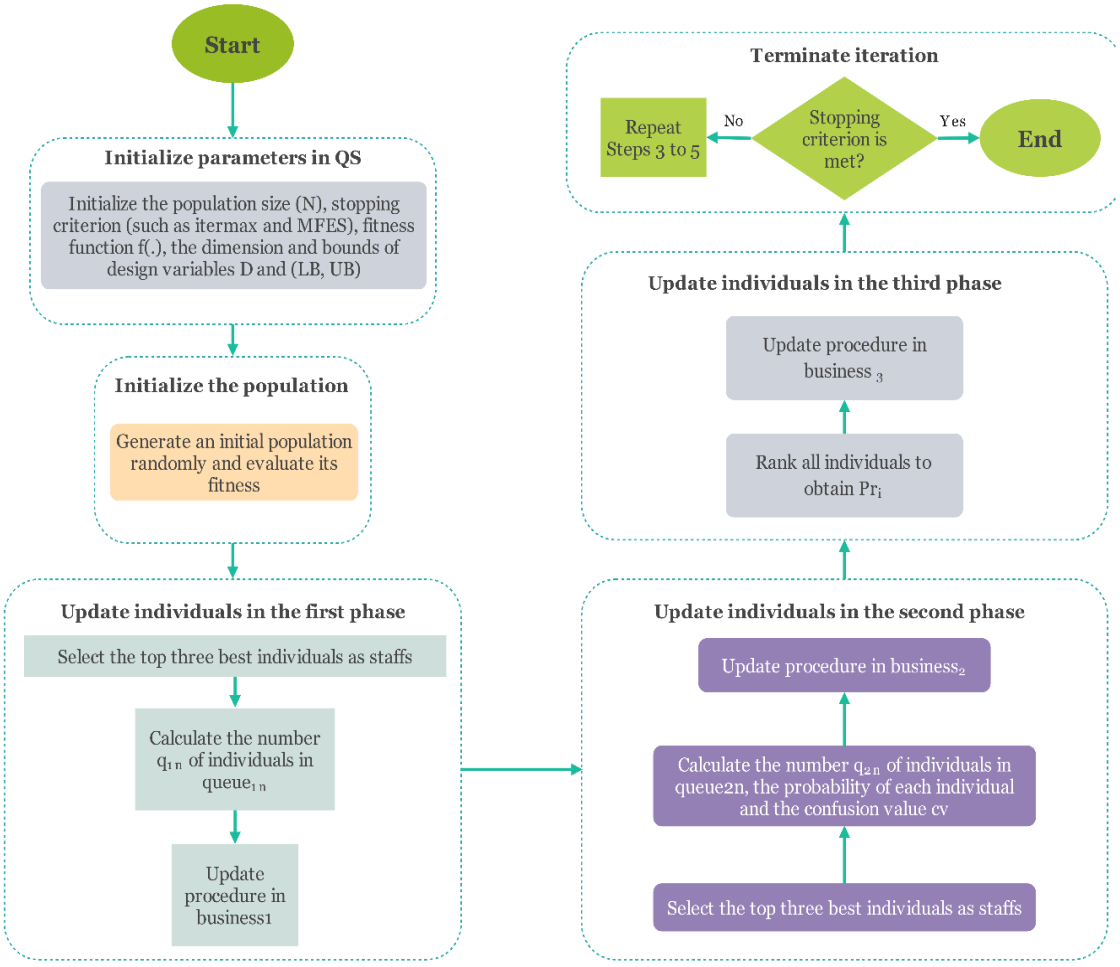


Figure 8: Flowchart of the QSA algorithm.

Table 3: XGBR hyperparameter search space during the optimization process and the optimal hyperparameter values obtained using the MA and QSA algorithms.

	n-estimators	Gamma	Reg-lambda	Reg-alpha	Learning-rate	training complexity (FLOPs)	training time(s) per core	testing time(s) per core
Lower bound	100	0	0	0	0.001	-	-	-

Upper bound	2000	1	1	5	1	-	-	-
QSA	500	0.78	0.67	0.540	0.010	3.58E+08	51.15	0.35
MA	500	0.76	0.79	0.455	0.015	4.44E+08	65.32	0.37

4.3.3 Performance measures

This study verifies the efficiency and correctness of the suggested model in predicting variations in GPs, and it contrasts its outcomes with those from alternative predictive models. To improve the thoroughness of the assessment, the typical metric is utilized to judge how well the prediction performs. Moreover, these measurements are extensively employed to evaluate the disparities between predictions generated by forecasting models and real values. This article evaluates the forecasting performance of the models using 5 popular assessment metrics: MAPE, RAE, RSE, MSE, and R^2. The following definitions are used for these measures:

MAPE is a statistical measure for calculating the ACC of a model. It is presented using percentages. The computation involves determining the absolute variation between the forecasted and observed values. This difference is then divided by the observed value. Subsequently, the mean of these percentages is computed. MAPE is a statistical measure utilized to examine a model for its correctness. A lower MAPE signifies that a model performs well.

$$\text{MAPE} = \left(\frac{1}{\text{n}} \sum_{\text{i}=1}^{\text{n}} \left| \frac{\text{y}_\text{i} - \hat{\text{y}}_\text{i}}{\text{y}_\text{i}} \right| \right) \times 100 \tag{30}$$

The RSE serves as a statistical metric in regression analysis to examine the accuracy of a model's forecasts. It measures how far apart the observed values are from the projected values. A lower RSE score signifies greater precision. The RSE is determined using the subsequent calculation:

$$\text{RSE} = \frac{\sum_{\text{i}=1}^{\text{n}} (\text{y}_\text{i} - \hat{\text{y}}_\text{i})}{\sum_{\text{i}=1}^{\text{n}} (\bar{\text{y}} - \hat{\text{y}}_\text{i})} \tag{31}$$

The MSE measures the deviation between the actual and expected values. It is computed by dividing the difference between the approximated and real values by the squared differences and then taking an average of all these squared differences; hence, the less the value, the better the model's performance.

$$\text{MSE} = \frac{1}{\text{N}} \sum_{\text{k}=0}^{\text{n}} \binom{\text{n}}{\text{k}} (\text{Fi} - \text{Yi})\text{b}^2 \tag{32}$$

The RAE is found by separating the total of absolute errors by the total of actual values. The resulting value is then multiplied by one hundred so it can be shown as a percentage.

$$\text{RAE} = \frac{\sum_{r=1}^{p} (r_i - r_i^T)^2}{\sum_{r=1}^{p} (r_i - \bar{r}_i)^2} \times 100 \tag{33}$$

The R^2 score is a numerical metric that evaluates the ACC of a model in accurately describing the data. The model's ACC is assessed using this metric; as it approaches 1, the model's quality increases.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{34}$$

5 Results and discussion

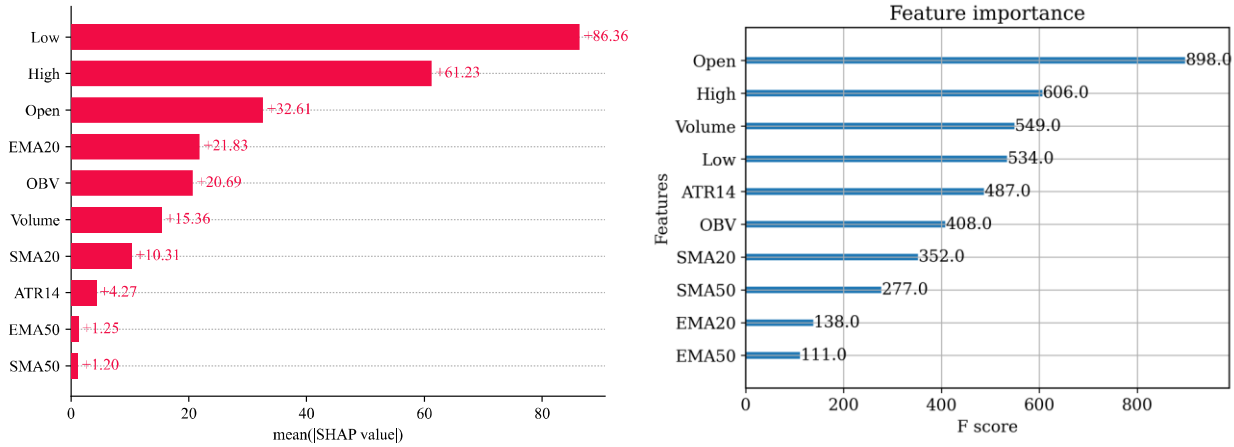
5.1 SHAP values and feature importance

SHAP values were used in this feature importance calculation to show how each feature contributed to the output. By illustrating the extent of the influence of each feature on the output, SHAP values are useful for illustrating how features influence a given output.

The values of SHAP and the graphic representation in Fig. 9 (a) reveal that the factors that influence the gold prices the most are the Low and High factors, with SHAP mean values of 86.36 and 61.23, respectively. The Open and EMA20 factors have SHAP values of 32.61 and 21.83, respectively. It is also apparent that the model is slightly influenced by factors like OBV, Volume, SMA20, ATR14, and the SMA50.

Fig. 9 (b) presents the values of the F-scores, where the values denote the level of significance that the features possess in the development of the model. The features Open and High possess the highest values of the F-score with 898.0 and 606.0, respectively.

These feature importance analyses, which are carried out using SHAP and F-score, guarantee the interpretability of the model and offer a deeper understanding of the underlying factors influencing its predictions.



(a) (b)

Figure 9: Feature importance analysis using SHAP values and F-score.

5.2 Analysis and comparisons

The assessment metrics for both training and testing sets of these various models DT, Transformer, iTransformer, LSTM, Bi-LSTM, XGBR, MA-XGBR, and QSA-XGBR reflect remarkable performance differences as seen in Table 4 and Figs 10 and 11. The DT model is considered to have the lowest predictive accuracy among single models: It has an R^2 of 0.958 for the training set and 0.956 for the testing set, meaning that it is not efficient at predicting the nonlinear dynamics of gold price fluctuations.

Compared to the QSA-XGBR model, the R^2 value of the LSTM and Bi-LSTM models shows a reasonable improvement over the DT model. The LSTM model's values on the test and train datasets are 0.976 and 0.977, respectively. The Bi-LSTM model's values for the train and test datasets are 0.980 and 0.978, respectively, indicating a very slight improvement. The purpose of using recurrent neural networks like LSTM and BiLSTM in the model is to detect the long-term dependencies in the data. If the gold prices show both the forward and backward connections, then the concept of the Bi-LSTM model proves to be very helpful in detecting the temporal connections in the data.

The Transformer model's test and train R^2 statistics are 0.962 and 0.966, while the iTransformer model, a variation of the Transformer model, has values of 0.971 and 0.973, suggesting that both models function fairly well. Due to their sequential structure, both models avoid the drawbacks of RNN networks like LSTM because they use self-attention techniques that enable them to process data in parallel.

On the other hand, with an R^2 of 0.981 on both training and testing sets, the single model XGBR performs considerably better. The main reason for this boost is due to the ensemble-based boosting mechanism of XGBR. XGBR builds several weak decision trees one after another in an attempt to correct errors of previous iterations. By combining several trees through gradient boosting, XGBR can model more complex nonlinear relationships more precisely, thereby improving its generalization on new data. In addition to the foregoing points, regularization parameters in XGBR further stabilize the model and avoid overfitting problems.

When optimization algorithms are implemented, the performance highly improves. The R^2 of 0.994 for the train and 0.993 for the test in the case of the MA-XGBR hybrid model depicts how the learning stability and error reduction ability of XGBR are enhanced through the metaheuristic tuning of hyperparameters. However, the QSA-XGBR outperforms both the baseline and MA-optimized counterparts with the highest accuracy, having R2 values of 0.998 for the train and 0.997 for the test. This superior performance further underlines the strong ability of the Queuing Search Algorithm to conduct efficient exploration and fine-tuning in the hyperparameter space.

The proposed QSA-XGBR model will work very efficiently for three main reasons. First, QSA effectively balances exploration and exploitation without getting stuck in a suboptimal local minimum. Second, hyperparameter choices of QSA are more stable because its convergence behavior is smoother. Third, XGBR benefits a lot from optimized tuning since it is inherently strong as a gradient-boosted ensemble learner. If these characteristics are merged into one hybrid model, the QSA-XGBR model will be able to grasp complex nonlinear patterns and minute changes in gold price time-series data with much greater precision.

Table 4: Evaluation criterion results for various models during the training and testing sets.

Models/Metrics		DT	Transformer	iTransformer	LSTM	Bi-LSTM	XGBR	MA-XGBR	QSA-XGBR
Train	MSE	970.19	877.13	729.71	547.33	494.61	316.44	201.13	128.92
	RSE	44.07	40.34	33.14	31.50	27.10	25.17	20.06	16.06
	RAE	2.17	1.75	1.63	1.49	1.36	1.24	0.99	0.79

	MAPE	1.57	1.43	1.22	1.09	0.95	0.89	0.70	0.60
	R^2	0.958	0.966	0.973	0.977	0.980	0.981	0.994	0.998
Test	MSE	443.90	395.80	363.59	309.07	255.82	219.66	80.66	43.57
	RSE	29.84	27.94	26.08	24.61	22.92	20.99	12.72	9.35
	RAE	1.13	1.02	0.94	0.88	0.83	0.79	0.48	0.35
	MAPE	0.93	0.82	0.76	0.69	0.61	0.55	0.37	0.29
	R^2	0.956	0.962	0.971	0.976	0.978	0.981	0.993	0.997

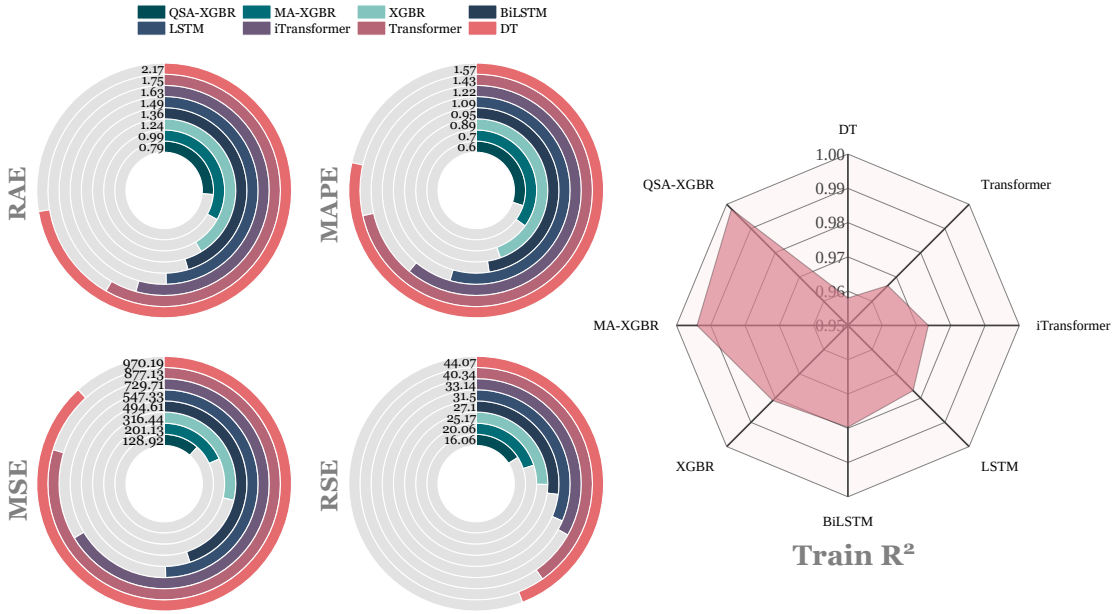


Figure 10: The proposed model outcomes throughout the training phase

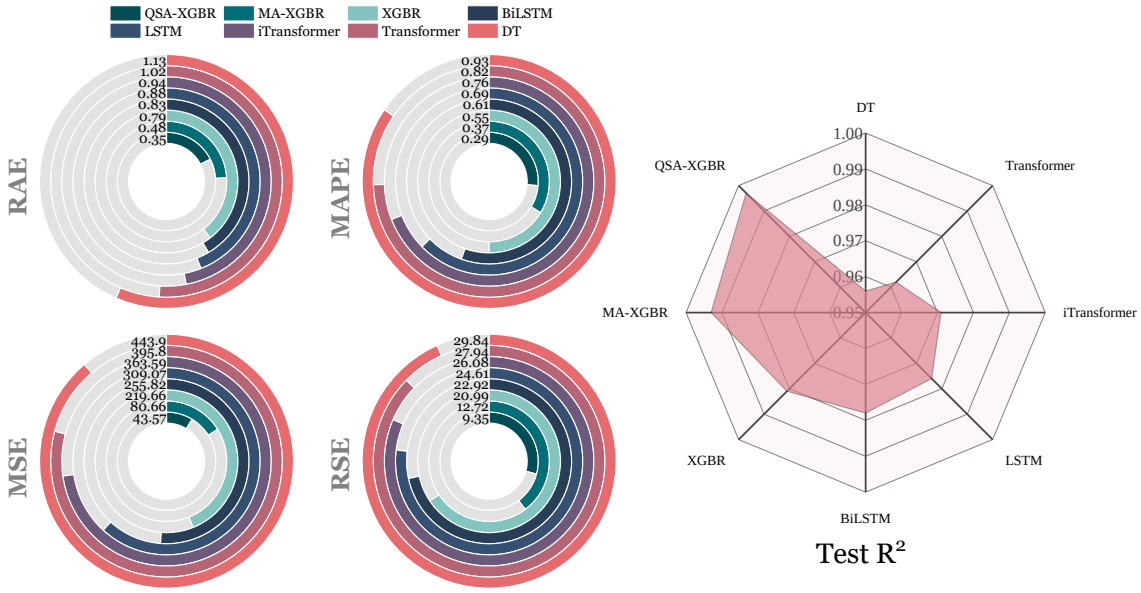


Figure 11: The proposed model outcomes throughout the testing phase

The QSA-XGBR model can become one avenue to reduce investment risks while seeking profitable opportunities, especially for practitioners who are investors or financial analysts. As shown, the capability of the model to closely align predicted values with the actual trend of GPs enhances confidence in applying it for real-

world decision-making. By decreasing the error in prediction, this model will also support stakeholders in managing market volatility, making more correct decisions regarding trading, and lessening the impact of sudden market fluctuations. The presented model may be adopted for different commodities or other financial instruments, although the focus of this research has fallen on gold futures prices.

Fig. 12 Demonstrates the observed and forecasted values for the training and testing datasets. It is important to note that the suggested model has superior performance on return points and its predicted curve is very similar to the actual curve.

Figure 12: Prediction curve generated using the QSA-XGBR approach over the training and testing sets

This research might be informative for future researchers. Furthermore, future research could consider integrating macroeconomic elements, namely, interest rates and inflation, or even social media sentiment analysis, to potentially enhance the model's predictive capabilities. Besides, the QSA-XGBR model could be tested with other financial instruments or very volatile market periods for further insight into its robustness and scalability. The comparison with other emerging optimization algorithms could finally show further paths toward the improvement of hybrid models.

The QSA-XGBR model's robustness was thoroughly examined by means of the explicit high-volatility periods that are associated with the COVID-19 pandemic in 2020. The dataset used for this purpose obviously included several major periods of disruption, for example, the crash of 2020, the instability of 2018, the correction of gold's price in 2013, and the inflation-driven shocks of 2022-2023. During such periods, the learning algorithms face abrupt changes in the data structure, heavy-tailed fluctuations, and sudden increases in volatility. ATR and OBV are, respectively, technical indicators that are very efficient in capturing volatility clustering and unusual trading activity; these certainly help the model to adapt faster to the unstable regimes. The model will be able to pinpoint the parameter settings that continue to work even when the distribution of data is far from the optimal configurations because of the metaheuristic hyperparameter optimization by QSA. In such a scenario, the hybrid QSA-XGBR model can assert its generalizability under shock-induced dynamics since it empirically kept lower prediction errors compared to the average benchmark models during the turbulent times. Its robustness is validated through a strong performance in a variety of historical stress events, but the separate crisis-only subsample analysis is not part of the current investigation. Future research should concentrate on explicit regime-specific validation to provide a more thorough check of the robustness of the method used.

5.3 Robustness analysis

Table 5 illustrates how the QSA-XGBR model performs using 5-fold cross-validation with varying levels of accuracy and performance for all of the folds included in the dataset. As a way to assess the effectiveness of QSA-XGBR, five different metrics were used: R^2, MSE, MAPE, RAE, and RSE. Through cross-validation, QSA-XGBR can be tested on many different subsets of data instead of relying on a single data subset for testing, thereby demonstrating that it has potentially overfitted to this dataset and was able to generalize effectively between multiple datasets. Based on the information provided in Table 5, the performance of the QSA-XGBR model is consistently efficient, capturing the majority of the variance (approximately 99.2% to 99.8% for all five folds). Therefore, R^2 adequately describes the proportion of variance from the predictions of QSA-XGBR that is attributable to gold prices, and better reflects the model's accuracy in predicting gold prices compared to other models.

The MSE value represents the average square difference between the predicted and actual values. With MSE values on the lower end (folds 4 and 5), it appears that as cross-validation continues, predictions become more accurate. This indicates that the model can effectively minimize errors as it progresses through later folds. MAPE represents how much each model prediction differs from the actual value, with lower MAPE indicating more accurate predictions. Furthermore, as MAPE continues to decrease consistently across folds, it indicates that the model's ability to provide accurate predictions is improving with additional folds in cross-validation.

The two performance metrics evaluated, RAE and RSE, show improvement throughout the folds, indicating a decrease in error and greater consistency of errors as the model learns from the training data. Based on the results of 5-Fold Cross Validation, the QSA-XGBR model has proven to be an efficient and robust prediction algorithm across numerous data subsets. Furthermore, and most importantly, the model consistently performs at a high level of predictive ability, demonstrating the capability to generalize across varying market conditions, including times of extreme market volatility and/or market instability.

Table 5: QSA-XGBR model performance metrics for 5-folds.

K-Fold	validation set				
	R^2	MSE	MAPE	RAE	RSE
fold1	0.992	211.86	0.76	1.01	20.82
fold2	0.994	181.18	0.71	0.97	18.73
fold3	0.995	155.43	0.68	0.85	17.45
fold4	0.997	135.54	0.61	0.81	16.17
fold5	0.998	125.92	0.59	0.77	15.86

5.4 Statistical analysis

This research formally compared the QSA-XGBR model's performance against the rest using the DM Test. The DM Test confirms the differences in forecast accuracy. The fig .13 shows the Diebold-Mariano test statistics, or DM values for short, and their respective p-values for each pair. The DM test shows that the model with lower values of its p-values, all of which are below 0.05, outperforms others, including the QSA-XGBR model, in terms of forecast accuracy.

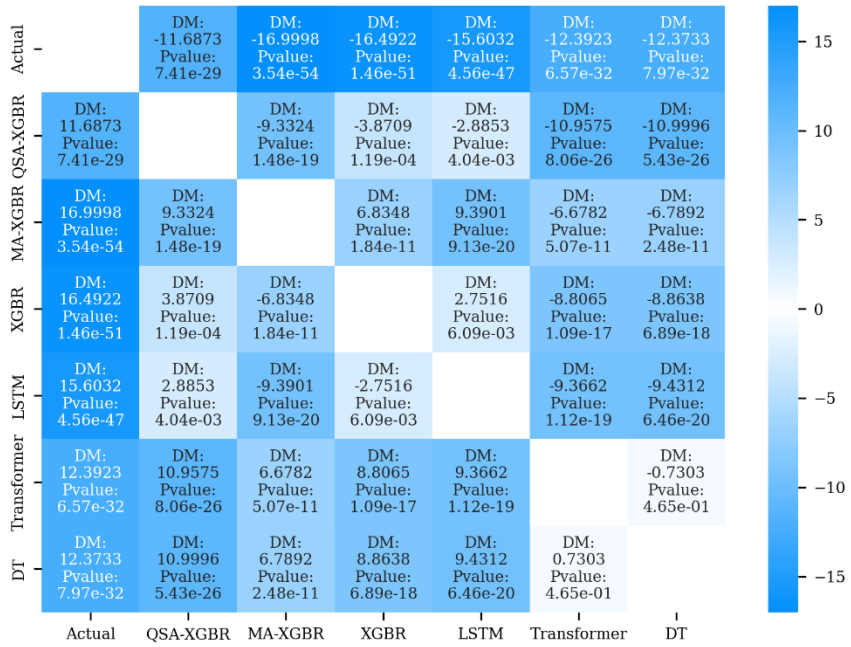


Figure 13: Pairwise comparisons between the QSA-XGBR model and other forecasting models and their respective Diebold-Mariano test statistics, and their associated p-values.

5.5 Generalization to other assets

A forecasting model's robustness and usefulness in real-world markets depend on its capacity to generalize across various financial assets. Even though the central focus and aim of this study was the forecasting of gold prices, one of the most essential subsequent steps will involve using the model on other assets like silver and crude oil prices. This will enable the evaluation of the model for retaining its ability for different market dynamics and characteristics by checking the feasibility of executing the pricing forecasts for other commodities. The suggested QSA-XGBR model is a useful and effective way to model time-series data and capture market trends, volatility, and other significant market behaviors. To make sure that this works

well for gold as well as other commodities with comparable market dynamics, such as silver and crude oil, testing on other assets is crucial. This raises its potential for widespread application in the financial industry, encompassing commodities trading, risk management, and investment strategies. Fig. 14 and 15 below show the expected and current prices for silver and crude oil, respectively. The QSA-XGBR model remains strongly fitted on a variety of assets, from high market volatility to providing accurate forecasts, according to the testing results using the same methodology as for gold on these assets. The model's capacity to manage market volatility and generalize to other assets is demonstrated by the test set predictions, which show a high degree of agreement with actual values. Fig. 14 illustrates how accurately the model predicts silver prices. The estimated values in this instance are remarkably close to the real price of silver. This implies that this model can accurately represent the price dynamics of an extremely volatile asset like silver. The model's accuracy in predicting the crude oil prices is depicted in Fig. 15. It is clear how accurate the model is in predicting the crude oil prices, despite the fluctuating prices and the market crises.

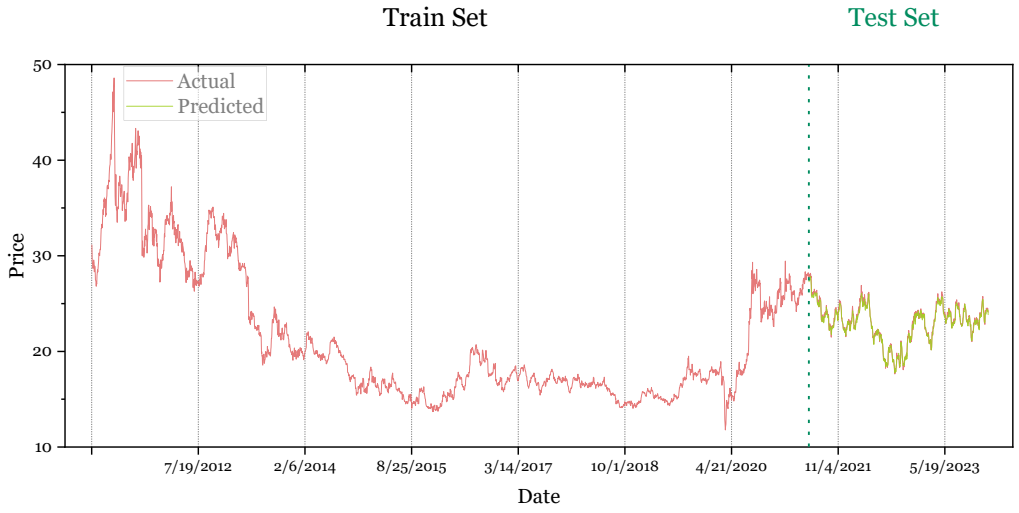


Figure 14: Prediction curve produced over the silver using the QSA-XGBR method.

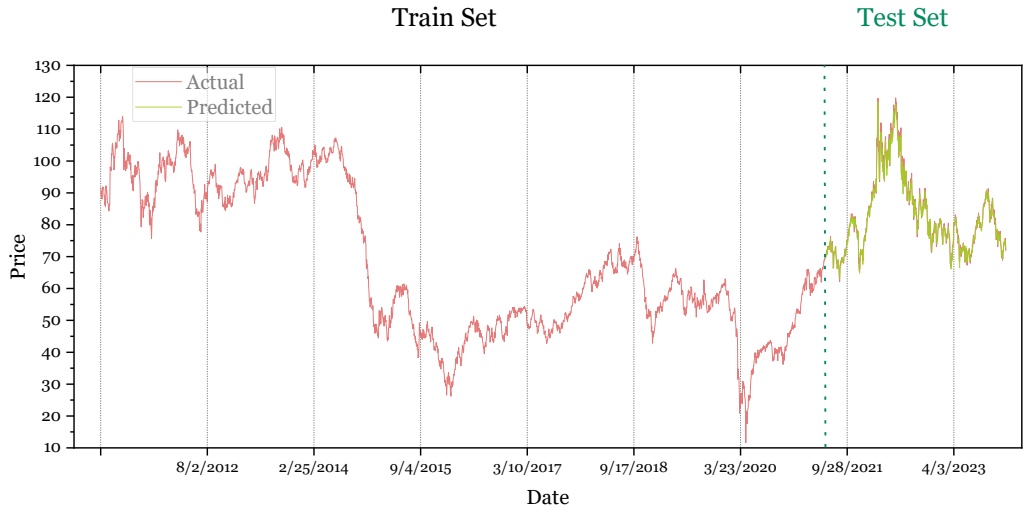


Figure 15: Prediction curve produced over the silver using the QSA-XGBR method.

Despite market volatility and the ensuing shock effects, the model's predictive power makes it an incredibly useful analytical tool for traders, investors, and financial analysts. The increasing trend in the number of research activities focusing on the usage of the hybrid optimization method in financial forecasting stems from the efficiency of the QSA-XGBR in multiple markets. It presents an empirical manifestation that will serve as the basis for further development in other markets. Thus, the generalization ability of the QSA-XGBR model on other commodities, such as silver and crude oil, proves the efficiency and robustness of the model on different financial platforms. The ability of the model to perform self-adaptation is of utmost significance for financial models and provides a bright future for the study of commodity prices.

5.6 Consequences and application

By implementing the QSA-XGBR model to forecast gold futures prices, one potential result is the ability to reduce investment risks. Precise forecasting is of great use to investors when making prudent decisions. Increased ACC with this model, therefore, hugely helps minimize losses and maximize returns. The better the prediction with QSA-XGBR, the better the investor will be able to create a strategy regarding their investments. With better insight into future price changes, investors can make prudent adjustments in their portfolios of investment through value addition to their portfolios of investment in anticipation of an upturn in the market. Because of the enormous influence the commodity has on the economy, governments and organizations indeed do monitor the prices of gold closely. Accurate forecasting would help policymakers in reaching decisions on monetary policies, trade rules, and international reserve management, and, as a result, the proposed model of QSA-XGBR will also support proper risk management more effectively in financial markets for economic stability. GP is of wide importance; hence, extensive use of the advanced forecasting model is necessary. This might have a key impact on economics. Precise price forecasting will help different industries' stakeholders make more informed decisions, hence fostering economic stability and growth both nationally and internationally. Therefore, gold producers and investors can use the QSA-XGBR model to enhance the ACC of GP predictions.

A closer look at the predictive behavior of the model shows the QSA-XGBR framework to be more reliable. The result is a very high out-of-sample accuracy (R^2 = 0.997), which can be attributed to the large range of trend, momentum, price-based, and volatility indicators, along with QSA-driven hyperparameter optimization that improves generalization and reduces variance. The dataset comprises more than 3,200 daily observations over 12 years and shows constantly low values of MSE, MAPE, RAE, and RSE on test data not seen so far. Besides this, XGBR's built-in mechanisms of regularization protect the model against fitting noise.

Because the dataset covers many real-world crisis periods, including the 2013 gold price correction, 2018 market instability, 2020 COVID-19 crash, and the inflationary and geopolitical shocks of 2022–2023, it exposes the model to a broad range of unstable regimes. By employing indicators such as ATR and OBV, the model can capture stressful market dynamics; meanwhile, QSA helps the parameters remain stable under the influence of such extreme volatility. Even so, subsample testing focused on crisis periods is still very beneficial in research, such as focusing on the 2020 crash.

These considerations have, therefore, guided the current study to consciously focus its effort on next-day, short-term forecasting, due to its most substantial implications for daily risk management and even real-time trading. This is clearly demonstrated in the strong performance of the QSA-XGBR architecture within this naturally volatile and noise-dominated horizon. Further extensions can, nevertheless, introduce long-horizon forecasting by combining the use of macroeconomic variables with decomposition frameworks and multi-step prediction techniques.

The high predictive performance of the QSA-XGBR model will strongly impact the area of financial forecasting in real-world applications. This implies that, in highly susceptible markets, precise one-day forecasts may significantly enhance decision-making processes, such as those involved in trade timing, improving risk-adjusted portfolio strategies, and developing responsive algorithmic trading systems. Thus, benefits gained from such enhancements in predictive precision have substantial implications for financial practitioners, since even minimal forecast error reductions can yield quite sizeable financial gains in cases of volatile market conditions. While it works considerably for short-term forecasting horizons, more care is needed when trying to use this model for longer forecasts. Multi-step forecasting introduces cumulative uncertainty and very high sensitivity to a number of exogenous economic and geopolitical factors not specified within this framework. Longer-horizon versions of the model would likely need methodological extensions, such as decomposition techniques, macroeconomic indicators, or regime-switching mechanisms to model structural changes in market behavior. Custom feature engineering and re-optimization procedures may be necessary when generalizing across asset classes such as stocks, commodities, and cryptocurrencies, since doing so requires careful consideration of potentially different volatility profiles, fluctuating liquidity, and market microstructure effects. Its performance on a wider range of assets and forecasting horizons should thus be examined in future research that focuses on model resilience and scalability in a wide variety of financial settings.

6 Conclusion

In the gold market, many variables interact in a nonlinear and complex fashion. Consequently, the GP is keenly observed by backers, administrations, and groups across the globe. By gaining an understanding of the pattern of variations in the GP, an important positive economic influence can be made. Nevertheless, because of the various elements that affect this market and the fluctuating nature of GPs, it remains remarkably challenging to predict the GP with precision. This work assesses the predicting ACC of the QSA-XGBR model for GPs. The QSA-XGBR model accurately forecasts GPs, which might lead to increased profitability for producers and policymakers.

- Moreover, this study utilized a dataset consisting of daily data for exploring the link between predictor variables and the prices of gold.
- The ensemble learning methods take into account the features that are picked at the input using trend indicators (EMA, and SMA), momentum oscillators (OBV), volatility indicators (ATR), and historical data (open, high, low, and close).

- Comparative analyses were carried out to evaluate the effectiveness of the suggested innovative model by using the DT, XGBR, and MA-XGBR models.
- Regarding empirical findings, the QSA-XGBR model outperforms other forecasting models for the GP. It exhibits the lowest values for MSE, MAPE, RSE, and RAE, and the highest value for R^2. The results unequivocally demonstrate that the QSA-XGBR model surpasses all others in performance.

In conclusion, employing QSA-XGBR to predict the closing price of gold futures for the subsequent trading day can assist investors in mitigating investment risks and identifying potential investment prospects.

Adding a metaheuristic optimizer does increase the computational cost, given that the algorithm requires multiple evaluations of potential hyperparameter configurations. This may render the model of limited utility in cases requiring extremely rapid model retraining or real-time model updating, but the gained performance more than compensates for the added expense. Secondly, although technical indicators and historical price data have proved their worth in financial forecasting, the framework could be insensitive to macroeconomic, geopolitical, or sentiment-driven factors that could impact the dynamics of changes in the price of gold. In this respect, the predictive power of the model will also be limited during periods when these types of external shocks become dominant and cannot be captured by historical trends.

Although the model performed exceptionally well on the dataset in question, any generalization to other financial assets, regimes of the market, or longer horizons would still have to be validated. Therefore, a future line of studies should be directed at the extension of this model to multi-asset settings, the incorporation of alternative data sources such as news sentiment, macroeconomic indicators, or regime-switching features, and exploring model compression or surrogate optimization strategies that can reduce computational overhead. These improvements will help increase the applicability and operational effectiveness of the proposed framework.

Competing of interests

The authors declare no competing of interests.

Authorship contribution statement

Yingying Dai: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Author statement

The manuscript has been read and approved by all the authors, the requirements for authorship, as stated earlier in this document, have been met, and each author believes that the manuscript represents honest work.

Ethical approval

All authors have been personally and actively involved in substantial work leading to the paper, and will take public responsibility for its content.

References

[1] P. Zhang and B. Ci (2020). Deep belief network for gold price forecasting, *Resources Policy*, 69, 101806. https://doi.org/10.1016/j.resourpol.2020.101806

[2] K. Gangopadhyay, A. Jangir, and R. Sensarma (2016). Forecasting the price of gold: An error correction approach, *IIMB management review*, 28(1): 6–12. https://doi.org/10.1016/j.iimb.2015.11.001

[3] E. Bouri, A. Jain, P. C. Biswal, and D. Roubaud (2017). Cointegration and nonlinear causality amongst gold, oil, and the Indian stock market: Evidence from implied volatility indices, *Resources Policy*, 52, 201–206. https://doi.org/10.1016/j.resourpol.2017.03.003

[4] F. Wen, X. Yang, X. Gong, and K. K. Lai (2017). Multi-scale volatility feature analysis and prediction of gold price, *Int J Inf Technol Decis Mak*, 16(01):205–223. https://doi.org/10.1142/S0219622016500504

[5] Z. Alameer, M. Abd Elaziz, A. A. Ewees, H. Ye, and Z. Jianhua (2019). Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm, *Resources Policy*, 61, 250–260. https://doi.org/10.1016/j.resourpol.2019.02.014

[6] L. Fang, B. Chen, H. Yu, and Y. Qian, (2018). The importance of global economic policy uncertainty in predicting gold futures market volatility: A GARCH-MIDAS approach, *Journal of Futures Markets*, 38(3): 413–422. https://doi.org/10.1002/fut.21897

[7] H. Abderazek, A. R. Yildiz, and S. M. Sait (2019). Mechanical engineering design optimisation using novel adaptive differential evolution algorithm, *International Journal of vehicle design*, 80(2–4): 285–329. https://doi.org/10.1504/IJVD.2019.109873

[8] B. S. Yıldız *et al.* (2023). A novel hybrid arithmetic optimization algorithm for solving constrained optimization problems, *Knowl Based Syst*, 271, 110554. https://doi.org/10.1016/j.knosys.2023.110554

[9] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time series analysis: forecasting and control*. John Wiley & Sons. DOI: 10.1111/jtsa.12194

[10] B. Guha and G. Bandyopadhyay (2016). Gold price forecasting using ARIMA model, *Journal of Advanced Management Science*, 4(2). DOI: 10.1109/ViTECoN58111.2023.10157017

[11] E. Jianwei, K. He, H. Liu, and Q. Ji (2023). A novel separation-ensemble analyzing and

forecasting method for the gold price forecasting based on RLS-type independent component analysis, *Expert Syst Appl*, 120852. https://doi.org/10.1016/j.eswa.2023.120852

[12] X. Li, D. Li, X. Zhang, G. Wei, L. Bai, and Y. Wei (2021). Forecasting regular and extreme gold price volatility: The roles of asymmetry, extreme event, and jump, *J Forecast*, 40(8): 1501–1523. https://doi.org/10.1002/for.2781

[13] T. Chen and C. Guestrin (2016). Xgboost: A scalable tree boosting system, in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794. https://doi.org/10.1145/2939672.2939785

[14] J. Zhang, M. Xiao, L. Gao, and Q. Pan (2018). Queuing search algorithm: A novel metaheuristic algorithm for solving engineering optimization problems, *Appl Math Model*, 63, 464–490. https://doi.org/10.1016/j.apm.2018.06.036

[15] S. Kumar *et al.* (2023). Chaotic marine predators' algorithm for global optimization of real-world engineering problems, *Knowl Based Syst*, 261, 110192. https://doi.org/10.1016/j.knosys.2022.110192

[16] T. Kunakote *et al.* (2022). Comparative performance of twelve metaheuristics for wind farm layout optimisation, *Archives of Computational Methods in Engineering*, 1–14. https://doi.org/10.1007/s11831-021-09586-7

[17] P. Mehta, B. S. Yildiz, S. M. Sait, and A. R. Yildiz (2022). Hunger games search algorithm for global optimization of engineering design problems, *Materials Testing*, 64(4): 524–532. https://doi.org/10.1515/mt-2022-0013

[18] H. Abderazek, S. M. Sait, and A. R. Yildiz (2019). Optimal design of planetary gear train for automotive transmissions using advanced meta-heuristics, *International Journal of Vehicle Design*, 80(2–4): 121–136. https://doi.org/10.1504/IJVD.2019.109862

[19] B. S. Yıldız, P. Mehta, N. Panagant, S. Mirjalili, and A. R. Yildiz (2022). A novel chaotic Runge Kutta optimization algorithm for solving constrained engineering problems, *J Comput Des Eng*, 9(6): 2452–2465. , https://doi.org/10.1093/jcde/qwac113

[20] Z. Meng, B. S. Yıldız, G. Li, C. Zhong, S. Mirjalili, and A. R. Yildiz (2023). Application of state-of-the-art multiobjective metaheuristic algorithms in reliability-based design optimization: a comparative study, *Structural and Multidisciplinary Optimization*, 66(8): 191. https://doi.org/10.1007/s00158-023-03639-0

[21] D. Gürses, S. Bureerat, S. M. Sait, and A. R. Yıldız (2021). Comparison of the arithmetic optimization algorithm, the slime mold optimization algorithm, the marine predator's algorithm, the salp swarm algorithm for real-world engineering applications, *Materials Testing*, 63(5): 448–452. https://doi.org/10.1515/mt-2020-0076

[22] D. Gürses, P. Mehta, S. M. Sait, and A. R. Yildiz (2022). African vultures optimization algorithm for optimization of shell and tube heat exchangers, *Materials Testing*, 64(8): 1234–1241. https://doi.org/10.1515/mt-2022-0050

[23] M. M. Manoj and L. S. Kurian (2022). Gold Price Prediction," in *National Conference on Emerging Computer Applications*. https://doi.org/10.5281/zenodo.6183215

[24] N. Rana, M. S. A. Latiff, S. M. Abdulhamid, and H. Chiroma, (2020). Whale optimization algorithm: a systematic review of contemporary applications, modifications and developments, *Neural Comput Appl*, 32(20): 16245–16277. https://doi.org/10.1007/s00521-020-04849-z

[25] Y. Liang, Y. Lin, and Q. Lu (2022). Forecasting gold price using a novel hybrid model with ICEEMDAN and LSTM-CNN-CBAM, *Expert Syst Appl*, vol. 206, p. 117847. https://doi.org/10.1016/j.eswa.2022.117847

[26] G. Cohen and A. Aiche (2023). Forecasting gold price using machine learning methodologies, *Chaos Solitons Fractals*, 175, 114079. https://doi.org/10.1016/j.chaos.2023.114079

[27] S. Ben Jabeur, S. Mefteh-Wali, and J.-L. Viviani (2024). Forecasting gold price with the XGBoost algorithm and SHAP interaction values," *Ann Oper Res*, 334(1):679–699. https://doi.org/10.1007/s10479-021-04187-w

[28] M. A. Haque, E. Topal, and E. Lilford (2015). Relationship between the gold price and the Australian dollar-US dollar exchange rate, *Mineral Economics*, 28, 65–78. https://doi.org/10.1007/s13563-015-0067-y

[29] V. Chang, T. Li, and Z. Zeng (2019). Towards an improved Adaboost algorithmic method for computational financial analysis, *J Parallel Distrib Comput*, 134, 219–232. https://doi.org/10.1016/j.jpdc.2019.07.014

[30] F. Gallegos-Erazo (2022). Technical Indicator for a Better Intraday Understanding of Uptrends or Downtrends in the Financial Markets using Volume Transactions as a Trigger., in *ICAI Workshops*, , 182–194. https://doi.org/10.1016/j.joitmc.2024.100398

[31] J. E. Granville (2018). *Granville's new key to stock market profits*. Pickle Partners Publishing. https://doi.org/10.1145/3716895.3717007

[32] Y. Zhu and G. Zhou (2009). Technical analysis: An asset allocation perspective on the use of moving averages, *J financ econ*, 92(3):519–544. https://doi.org/10.1016/j.jfineco.2008.07.002

[33] S. Karasu and A. Altan (2022). Crude oil time series prediction model based on LSTM network with chaotic Henry gas solubility optimization, *Energy*, 242, 122964. https://doi.org/10.1016/j.energy.2021.122964

[34] N. H. L. Kan, Q. Cao, and C. Quek (2024). Learning and processing framework using Fuzzy Deep Neural Network for trading and portfolio

rebalancing, *Appl Soft Comput*, 152, 111233. https://doi.org/10.1016/j.asoc.2024.111233

[35] S. Christa, V. Suma, and U. Mohan (2022). Regression and decision tree approaches in predicting the effort in resolving incidents, *Int J Bus Inf Syst*, 39(3): 379–399. https://doi.org/10.1504/IJBIS.2022.122342

[36] B. Dou *et al.* (2023). Machine learning methods for small data challenges in molecular science, *Chem Rev*, 123(13): 8736–8780. https://doi.org/10.1021/acs.chemrev.3c00189

[37] O. Sagi and L. Rokach (2021). Approximating XGBoost with an interpretable decision tree, *Inf Sci (N Y)*, 572, 522–542. https://doi.org/10.1016/j.ins.2021.05.055

[38] Z. Zhang, Y. Huang, R. Qin, W. Ren, and G. Wen (2021). XGBoost-based on-line prediction of seam tensile strength for Al-Li alloy in laser welding: Experiment study and modelling, *J Manuf Process*, 64, 30–44. https://doi.org/10.1016/j.jmapro.2020.12.004

[39] J. Abdi, F. Hadavimoghaddam, M. Hadipoor, and A. Hemmati-Sarapardeh (2021). Modeling of CO2 adsorption capacity by porous metal organic frameworks using advanced decision tree-based models, *Sci Rep*, 11(1): 24468. https://doi.org/10.1038/s41598-021-04168-w

[40] S. A. Madani, M.-R. Mohammadi, S. Atashrouz, A. Abedi, A. Hemmati-Sarapardeh, and A. Mohaddespour (2021). Modeling of nitrogen solubility in normal alkanes using machine learning methods compared with cubic and PC-SAFT equations of state, *Sci Rep*, 11(1): 24403. https://doi.org/10.1038/s41598-021-03643-8

[41] M.-R. Mohammadi *et al.* (2021). Modeling hydrogen solubility in hydrocarbons using extreme gradient boosting and equations of state, *Sci Rep*, 11(1): 17911. https://doi.org/10.1038/s41598-021-97131-8

[42] K. Zervoudakis and S. Tsafarakis (2020). A mayfly optimization algorithm, *Comput Ind Eng*, 145, 106559. https://doi.org/10.1016/j.cie.2020.106559

[43] K. Rajwar, K. Deep, and S. Das (2023). An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges, *Artif Intell Rev*, 56(11): 13187–13257. https://doi.org/10.1007/s10462-023-10470-y

[44] H. Yan (2024). Research on Gold Price Prediction Based on LSTM Modeling, *Advances in Economics, Management and Political Sciences*, 94, 202–210. DOI:10.1051/shsconf/202418102005

[45] I. E. Livieris, E. Pintelas, and P. Pintelas (2020). A CNN–LSTM model for gold price time-series forecasting, *Neural Comput Appl*, 32(23): 17351–17360. https://doi.org/10.1007/s00521-020-04867-x

[46] N. Tripurana, B. Kar, S. Chakravarty, B. K. Paikaray, and S. Satpathy, Gold Price Prediction Using Machine Learning Techniques. DOI: 10.18280/ijcmem.130120

[47] A. Amini and R. Kalantari (2024). Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning, *PLoS One*, 19(3): March, doi: 10.1371/journal.pone.0298426.