



Forecasting gold price with the XGBoost algorithm and SHAP interaction values

Sami Ben Jabeur¹ · Salma Mefteh-Wali² · Jean-Laurent Viviani³

Accepted: 24 June 2021 / Published online: 23 July 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Financial institutions, investors, mining companies and related firms need an effective accurate forecasting model to examine gold price fluctuations in order to make correct decisions. This paper proposes an innovative approach to accurately forecast gold price movements and to interpret predictions. First, it compares six machine learning models. These models include two very recent methods: the eXtreme Gradient Boosting (XGBoost) and CatBoost. The empirical findings indicate the superiority of XGBoost over other advanced machine learning models. Second, it proposes Shapley additive explanations (SHAP) in order to help policy makers to interpret the predictions of complex machine learning models and to examine the importance of various features that affect gold prices. Our results illustrate that the utilization of XGBoost along with SHAP approach could provide a significant boost in increasing the gold price forecasting performance.

Keywords Gold price · XGBoost · CatBoost · Shapley additive explanations

JEL Classification C22 · C45 · C53 · G12

1 Introduction

Gold has been an important precious metal for centuries. It is a major financial asset for countries and a key component of the global monetary reserves for trading and currency hedging (Capie et al., 2005; Wen et al., 2017). It also plays a prominent role in investments,

✉ Salma Mefteh-Wali
salma.mefteh@essca.fr

Sami Ben Jabeur
sbenjabeur@gmail.com

Jean-Laurent Viviani
jean-laurent.viviani@univ-rennes1.fr

¹ Institute of Sustainable Business and Organizations, Confluence: Sciences Et Humanités - UCLY, ESDES, 10 place des archives, 69002 Lyon, France

² ESSCA School of Management, 1 rue Lakanal, 49003 Angers, France

³ University of Rennes 1, CNRS, CREM-UMR 6211, 35000 Rennes, France

especially to hedge against adverse financial events. Indeed, in times of financial turmoil leading stock indices to decrease, the prices of precious metals tend to move in the opposite direction. Furthermore, forecasting gold price fluctuations is a crucial issue for investors, for mining projects and related companies and in general for any agent who sees gold as an indicator of the future performance of the world economy.

Up to now, several studies have been performed to predict the commodity price as gold price. There are three main categories of prediction methods: (1) classical methods (2) artificial intelligence methods (3) hybrid approaches. First, traditional mathematical models such as Autoregressive Integrated Moving Average (Parisi et al., 2008), jump and dip diffusion (Shafiee & Topal, 2010a, 2010b) and the multi-linear regression (Escribano & Granger, 1998; Kearney & Lombra, 2009) have been used for gold price forecasting. Those classical methods mostly describe the linear relationship between variables through specific *ex ante* analytical formulation. Second, with the recognition of the nonlinear and complex characteristics of the gold price system (Alameer et al., 2019), intelligent models, especially artificial neural networks have been useful to predict volatile financial variables, which are quite difficult to guess with classical statistics and econometrics. They are one of the most important types of machine learning models, which have been introduced and examined for forecasting commodity prices (Khashei & Bijari, 2010; Lineesh et al., 2010; Parisi et al., 2008). The characteristics of artificial intelligence methods that make them appropriate for prediction are their nonlinear structure, flexibility and data-driven learning process. Third, the hybrid approaches are a combination of artificial neural networks and several models to forecast the fluctuations of a variable such commodity prices, market returns, etc. For example, Kristjanpoller and Hernandez (2017) proposed a hybrid model (hybrid ANN-GARCH) to predict the volatility of the spot prices of gold, silver, and copper. To forecast long term gold price fluctuations, Alameer et al. (2019) used a whale optimization algorithm as a trainer to learn the multilayer perceptron neural network.

The main novelty of this study is twofold. It assesses the performance of machine learning models to show the success of XGBoost in forecasting the gold price. Then it is the first, to the best of our knowledge, to analyze the importance of individual features of gold price fluctuation using SHAP (SHapley Additive exPlanation). Gold price has significant nonlinear, time-varying, many influence factors in consequence it is especially important to detect the most influential factors first and then to combine them in order to improve prediction accuracy. Other studies aim to accomplish this task by using hybrid models: one kind of method is used to detect factors and machine learning to combine them. For instance, Chen and Zhang (2019) use a Projection Pursuit (PP) algorithm for the factors and Neural Network for prediction.

This paper makes two major contributions to the literature. The first is beginning the forecast with six machine learning models and to compare their performance in the gold price prediction. Those models are linear regression, neural networks, random forest, and three gradient methods based on decision trees that are Light Gradient Boosting Machine (LightGBM), CatBoost algorithm and eXtreme Gradient Boosting (XGBoost). To the best of our knowledge, this research represents the first attempt to use CatBoost, LightGBM and XGBoost models for forecasting gold price fluctuations. The best fit model is identified according to the performance criteria including coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE). We show that XGBoost performs better than other machine learning techniques to predict the gold price. We provide evidence of the benefits of using artificial intelligence to improve forecasting and more specifically show that XGBoost is a successful forecasting procedure. Second, in terms of model interpretation—which is especially important when using machine learning models

that are often difficult to interpret—several studies have started to take advantage of SHAP (Ribeiro et al., 2016; Štrumbelj and Kononenko, 2014). We have applied SHAP for the first time. To the best of our knowledge, SHAP interaction values have not yet been applied to analyze financial data set. Several studies have shown that gold prices may be affected by many predictors, such as inflation (Alameer et al., 2019; Beckmann & Czudaj, 2013), currencies (Beckmann & Czudaj, 2013; Kristjanpoller & Minutolo, 2016), metals (Bhatia et al., 2018; Schweikert, 2018), crude oil (Elie et al., 2019; Kanjilal & Ghosh, 2017; Sephton & Mann, 2018) and exchange rates (Akbar et al., 2019; Singhal et al., 2019). The SHAP method allows assigning each factor an importance value for gold price prediction. Applying methods based on explanations for a complex model to interpret gold price forecasts is of great interest as it allows understanding how the model behaves.

The rest of this paper is structured as follows. Section 2 provides a literature review of the time series forecasting models and factors that influence gold prices. Section 3 provides a description of the data. Section 4 presents the methodology and summarizes six machine learning models to forecast gold price and the method to interpret the predictions generated by these complex models. The results obtained are discussed in Sect. 5. Finally, the conclusion is put forward in Sect. 6.

2 Related literature

In order to improve the quality of gold price predictions, our study examines two important steps of a prediction process: first, the choice of the most relevant input variables and second, the choice of the best statistical model. Even if no consensus has emerged on which macroeconomic variables should be taken into account as primary drivers of gold-price fluctuations, empirical studies often use a common set of variables such as exchange rates, precious metals and mineral commodities prices, oil prices, and inflation. In their survey, O'Connor et al. (2015) show that gold prices have a broad range of predictors, covering the fields of commodities, financial variables, macroeconomic data, and interest rates.

The literature has demonstrated that exchange rates have a strong power for forecasting commodity prices (Chen et al., 2010) and especially gold-price fluctuations (Pukthuanthong & Roll, 2011; Reboredo, 2013). Bodart et al. (2015) provided evidence of the relationship between exchange rates and commodity prices for developing countries that export these commodities. Ciner (2017) confirmed that the exchange rate of South African rand has a significant predictive power to forecast palladium and platinum prices, and to a lesser degree silver prices. Sari et al. (2010) argued that precious metals respond to any shock in the exchange rate or shock in the prices of other precious metals.

In effect, the relationship between gold and other precious metals is complex. Studies report a co-integration relationship between gold and silver and the role of financial crises, but empirical studies do not reach a consensus concerning the direction of causality. Bhatia et al. (2018), in contrast to Sensoy (2013), show the existence of two-way causality among precious metals. Studies also found that there are cointegrating relationships between prices of mineral commodities (Kucher & McCoskey, 2017; Liu et al., 2019; Roberts, 2009; Rossen, 2015; Wu & Hu, 2016; Yue et al., 2015). Batten et al. (2015) report evidence of time-varying spillover effects between precious metal prices, which can be interpreted as evidence of time-varying market integration.

Several studies have confirmed that, because oil still being one of the most popular sources of energy used, crude oil prices are the leading cause of commodity price volatility

(Abd Elaziz et al., 2019; He et al., 2010; Lardic & Mignon, 2008; Shafiee & Topal, 2010a, 2010b). Behmiri and Manera (2015) examined the influence of oil price shocks on metal (including gold and silver) price volatility. They found that the price volatility of these metals has been influenced by oil price shocks. The oil price is considered as one (with inflation) of the main macroeconomic variables that influence the gold price (Batten et al., 2010; Tully & Lucey, 2007). It is generally recognized that there is a positive correlation between gold and crude oil prices (Teetranont et al., 2018). That is also the case in a recent study of Mo et al., (2018) that explored the dynamic linkages between the USD and the gold and crude oil markets. Bedoui et al. (2019) also found that gold, oil and USD exchange rates have strong connections during a period of crisis. Furthermore, Cologni and Manera (2008) observed that the rising oil prices increase metal prices by the inflation effects.

In effect, inflation is another major macroeconomic variable that influences gold prices (Batten et al., 2010; Fortune, 1987; Mahdavi & Zhou, 1997). Following Shafiee and Topal (2010a, 2010b), the two most important variables that explain gold price behavior are the oil price and the inflation rate. Correlation between gold price and inflation rate is negative (Shafiee & Topal, 2010a, 2010b). In a recent study, Alameer et al. (2019) indicate that all the previously presented variables (crude oil, iron, silver, and copper prices, exchange rates and inflation rates of US and China) have a high forecasting power on gold prices. Finally, it is also well known that precious metals play a prominent role especially to hedge against adverse financial events, the so-called “safe haven” hypothesis in the literature (Baur & Lucey, 2010; Baur & McDermott, 2010). Kang et al. (2017) demonstrate that gold and silver could apparently also benefit from a flight-to-quality phenomenon during financial crises. It is the reason why authors introduced market indices to explain gold price fluctuations (Liu et al., 2017; Pierdzioch et al. 2016; Kristjanpoller & Minutolo, 2015). Akbar et al. (2019) in Pakistan and Singhal et al. (2019) in Mexico demonstrate the same effect by studying the dynamic relationships among gold price, stock price index and exchange rate. In very recent studies (Risse, 2019; Zhang & Ci, 2020) combine CPI of U.S., the federal funds rate, crude oil future price, the nominal effective exchange rate and the Dow Jones index, as inputs for gold price forecasting.

Numerous statistical approaches have already been used to predict gold price fluctuations. Artificial neural networks (NN) are one of the most important types of machine learning methods, which have been examined for forecasting commodity prices (Khashei & Bijari, 2010; Lineesh et al., 2010; Parisi et al., 2008). Recent studies have combined Artificial Neural Networks with other machine learning approaches in order to improve prediction efficiency. Ramyar and Kianfar (2017) demonstrated the superiority of the MLP neural network compared to the vector autoregressive model (VAR) to forecast crude oil prices. Alameer et al. (2019) compared to a recent meta-heuristic method called whale optimization algorithm (WOA) as a trainer to learn the multilayer perceptron neural network to other models, including the classic NN, particle swarm optimization for NN (PSO–NN), genetic algorithm for NN (GA–NN), and grey wolf optimization for NN (GWO–NN). Deep learning is another approach to improve the predicting ability of traditional ANN. Deep learning algorithms has three main advantages: they improve the speed of network training, they avoid being trapped in local minima and they solve the multi-layer network training problems (Sezer et al., 2020). Recently, these methods have been used in time series forecasting.

Zheng et al. (2019) proposed an improved Deep Belief Networks (DBN) for forecasting exchange rates and determined that it worked better than traditional methods. Zhang and Ci (2020) show the superiority of a DBN model (compared to ARIMA or classical NN),

composed of restricted Boltzmann machines (RBM) for pre-training and a layer of supervised back-propagation (BP) for fine-tuning, in gold price forecasting. Chen et al. (2020) went a step further in this direction by using extreme learning machine (ELM) for time series forecasting and specifically gold price. Compared with most of the other machine learning algorithms such as support vector machine and deep learning method, ELM boast a faster learning speed.

Besides linear regression and NN models, we have chosen to implement trees boosting methods. Those hybrid models are able to overcome the limitations of the individual models and increase the forecasting accuracy by combining the advantages of both linear and nonlinear models (Khashei & Bijari, 2011). Hybrid models have already been used to predict gold price: Wen et al. (2017) used ensemble empirical mode decomposition (EEMD), SVM and ANN to analyze and predict the gold price series, Herawati et al. (2017) utilized a traditional recurrent neural network (RNN) and EEMD, Zhu and Zhang (2018) developed a hybrid model using an ANN, principal component analysis (PCA) and genetic algorithm (GA) Hybridization of times series analysis methods and machine learning has also been used to predict gold prices. Several papers (see Kumar, 2018 for a recent example and the references) propose to combine of ARIMA (for the linear part of the time series) and ANN (for the nonlinear component), more original is the combination of discrete wavelet transform with support vector machine made by Risse (2019). Very recently, Du et al. (2020) combine ELM and hybrid approaches to analyze the traits of metal prices. On the whole, these aforementioned scholars have confirmed the higher prediction performance of hybrid models than that of the individual methods.

The main idea of trees boosting methods is to combine decision tree methods represented in our paper by random forests and gradient boosted methods. The interest for random forests remain important, for instance in a very recent paper, Pierdzioch and Risse (2020) use multivariate random to forecasts of a vector of returns of four precious metal prices (gold, silver, platinum, and palladium). They show that multivariate forecasts are more accurate than univariate forecasts. Gradient represents the slope of the loss function, so if gradients are large in some points, it means that these points are important for finding the optimal split point. Recent research (Pierdzioch et al., 2015a, b) already uses the boosting approach to study the determinants of returns of the price of gold. The combination of decision trees and gradient boosting methods has the advantages of good training effect and not easily over-fitting. Specifically, in this paper, we will compare XGBoost, with two very recent boosting methods: LightGBM and CatBoost. These algorithms use different splitting methods in order to increase learning speed, prevent overfitting and improve performance.¹ They have already been used to analyze financial data sets (Basak et al., 2019; Ma et al., 2018; Huck, 2019) but never to predict commodity price fluctuations.

In practical financial decision situations, decision makers not only need to make accurate predictions, they also have to justify how predictions are obtained and why a given decision is taken (for instance to refuse a loan based on the financial status of the borrower; or to buy/sell a financial asset, in our example gold). In the academic world, it is also important to understand the causal relationships between variables and the hierarchy of causes. To answer these practical and theoretical concerns, we use SHAP to explain the output of the machine learning model. The idea of SHAP is to show the contribution of each feature to run the model output from the base value of explanatory variables to

¹ More detailed features of these algorithms will be presented in the methodology section.

Table 1 Data and variables

Variables	Signification	Period	Source (Freely available)
Gold	Gold price	Monthly	https://www.gold.org
Silver	Silver price	Monthly	https://www.investing.com
Oil	Crude oil price	Monthly	https://www.eia.gov
USD_EUR	Exchange rate (USA/EUR)	Monthly	https://fxtop.com/
USD_CNY	Exchange rate (USA/CNY)	Monthly	https://fxtop.com/
Inflation	Inflation Rate (US)	Monthly	https://tradingeconomics.com
SP_500	S&P 500	Monthly	https://www.investing.com
Iron_Ore	Iron Ore	Monthly	https://www.investing.com

Table 2 Descriptive statistics

	Gold	Inflation	Iron_Ore	Oil	Silver	SP_500	USD_CNY	USD_EUR
Mean	701.96	184.46	1244.92	43.68	10.94	1154.11	6.77	0.84
Median	415.95	182.20	18.80	31	6.47	1128.20	6.83	0.83
Maximum	1813.50	258.501	15,360.02	133	48.58	3230.78	8.72	1.17
Minimum	254.80	108.70	0.05	11	3.56	211.78	3.20	0.63
Std. Dev	454.012	43.41	2503.114	29.12	8.21	710.66	1.52	0.11
Skewness	0.79	0.04	2.73	0.86	1.52	0.76	0.72	0.85
Kurtosis	2.10	1.77	11.57	2.61	5.21	3.05	2.55	3.56
Observations	408	408	408	408	408	408	408	408

The table reports descriptive statistics of full sample

the model output value. In short, SHAP values represent a feature's responsibility for a change in the model output. Features pushing the prediction higher are distinguished from those pushing the prediction lower. To our knowledge, SHAP interaction values have not yet been applied to analyze financial data set.

3 Data and variables

In this paper, we investigate the effect of several explanatory variables on gold price, which is given in US dollars. The data covers the period from January 1986 to December 2019, including 408 monthly observations. This study has been divided into training (80%) and test (20%) samples in order to compare the performances of different machine learning models. We randomly partition the dataset by selecting 80% of the data as the training data set and the remaining 20% as the testing set. This method is commonly used in many of the previous research (e.g. Abellán & Mantas, 2014; Antunes et al., 2017; Ben Jabeur et al., 2020). Also, Gholamy et al., (2018) show that that the best results are obtained if we use 20–30% of the data for testing, and the remaining 70–80% of the data for training. Table 1 gives more information about data and variables used in this study. Table 2 provides descriptive statistics of time series data and Table 3 presents the correlation matrix between variables. It is clear that gold price is significantly correlated with all the predictor variables.

Table 3 Correlation matrix

Variables	Gold	Inflation	USD_CNY	USD_EUR	Silver	Oil	SP_500	Iron_Ore
<i>Gold</i>	1							
Inflation	0.828*** (0.000)	1						
USD/CNY	-0.121** (0.015)	0.338*** (0.0001)	1					
USD/EUR	-0.382*** (0.001)	-0.236*** (0.000)	0.053 (0.282)	1				
Silver	0.932*** (0.000)	0.729*** (0.000)	-0.078 (0.116)	-0.422*** (0.000)	1			
Oil	0.798*** (0.000)	0.770*** (0.000)	0.057 (0.249)	-0.528*** (0.000)	0.837*** (0.000)	1		
SP_500	0.688*** (0.000)	0.899*** (0.000)	0.295*** (0.000)	0.021 (0.666)	0.532*** (0.000)	0.545*** (0.000)	1	
Iron_Ore	-0.340*** (0.000)	-0.639*** (0.000)	-0.478*** (0.000)	0.120** (0.015)	-0.328*** (0.000)	-0.445*** (0.000)	-0.546*** (0.000)	1

**At the 5% level

***Significant coefficient at the 1% level

p-values are in parentheses

4 Methodology

In this section, we present six machine learning models to forecast the gold price. We describe the metrics that can be used to evaluate their performance. We also present the SHAP approach to interpret the results provided by machine learning models.

4.1 Machine learning models

4.1.1 Linear regression

Linear regression is a statistical analysis that analyzes the effect of selected independent variables² on an explanatory dependent variable. Linear regression uses the ordinary least squares method to estimate the linear relationship between variables. The function of forecasting models are expressed as follows:

$$Y_t = \beta_0 + \beta_1 X_1 + \dots + \beta_n X_n + \varepsilon_t$$

where Y_t is the expected value at time t , X_t is a vector of k predictors variables at time t , β_j is the estimated coefficients, and ε_t is a random error term at time t .

Several previous studies have shown that linear regression is less accurate in forecasting compared to advanced methods (Risse, 2019) and suffers from various statistical restrictions, such as endogeneity and multicollinearity (Baker et al., 2020; Pesaran & Smith, 2019).

4.1.2 Neural networks

Artificial neural networks are widely used in forecasting commodity prices (Ewees et al., 2020; Kristjanpoller & Minutolo, 2015, 2016). Neural networks (NN) are a set of formal neurons associated with layers and operating in parallel. In a network, each subgroup processes independently from the others and transmits the result of its analysis to the next subgroup. The first layer is called the input layer. It will receive the source data that we want to use for the analysis. Its size is therefore directly determined by the number of input variables. The second layer is a hidden layer, in the sense that it has only an intrinsic utility for the network and has no direct contact with the outside. The third layer is called the output layer. It gives the result obtained after compilation by the network of the data entered in the first layer. Each neuron collects the information provided by the neurons of the previous layer and then calculates its activation potential. This potential is then transformed by a function to determine the pulse sent to the neurons of the next layer. The output of the hidden layer is calculated as follows:

$$Y_i = \frac{1}{\left[1 + e^{-\left(\sum_{j=1}^N w_{ij} x_j - b_j\right)} \right]}$$

² Semantically, we note that the terms variable and feature are identical. The former tends to be used in statistics and the latter tends to be used in computer science.

where w_{ij} is the weight of the i th hidden neuron and b_j is the base of the second hidden layer. Multi-layer perceptrons are the most commonly used models. Several processing layers allow them to realize non-linear relationships between input and output.

4.1.3 Random forest

Recently, several studies have shown the effectiveness of random forest regression (RF) in economics and finance (Krauss et al., 2017; Loureiro et al., 2018; Mercadier & Lardy, 2019). RF is a tree-based regression approach. According to Babar et al. (2020), RF has been extensively used in recent years due to its robust performance compared to other traditional models. It is an ensemble learning framework proposed by Breiman (2001), which is built on the association of a multitude of regression trees. RF can be constructed by randomly sampling a feature subgroup for each decision tree. Based on bootstrap sampling, the generation of a random subset of each base tree model and the pruning of linear nodes is created on the same sample. After the training process, the predicted values of gold price are expressed as follows:

$$Y_i = \frac{1}{T} \sum_{k=1}^T l_k(x)$$

where $l_k(x)$ is a set of k th learner random tree learners and X is the vector of T input variables. The trees are constructed using the binary recursive partitioning.

4.1.4 Light gradient boosting machine (LightGBM)

LightGBM is a novel gradient boosting framework proposed by Ke et al. (2017). It employs gradient-based one-side sampling to fix the split point via computing variance gain. The LightGBM algorithm built two novel approaches, which are the gradient-based one-side sampling and the exclusive feature bundling (Sun et al., 2019). The estimated function of LightGBM integrates a number of T regression trees and defined as follows:

$$Y_i = \sum_{h=1}^T f_h(x)$$

where $f_i(x)$ denotes the regression trees. In LightGBM, Newton's method was used to estimate the objective function.

Several studies showed that the LightGBM provides more efficient and accurate performance than advanced machine learning algorithms. According to Sun et al. (2019), the advantages of LightGBM can be reflected in fast training speed, low memory consumption and good model accuracy.

4.1.5 CatBoost algorithm

CatBoost is also a new gradient descent algorithm proposed recently by Prokhorenkova et al. (2018). This supervised machine learning algorithm consists in classifying categorical data using the gradient boosting on decision trees. The decision tree is created by dividing the training data set into similar instances. According to Prokhorenkova et al. (2018), CatBoost uses ordered boosting and an innovative algorithm for processing categorical features. CatBoost

outperforming other boosting techniques in terms of performances. The function of decision tree h can be written as:

$$h^t = \arg \min \frac{1}{N} \sum (-f^t(X_k, Y_k) - h(X_k))^2$$

where X_k is the random vector of N input variables, Y_k is the outcome, and f function is a least squares approximation by the Newton method. Moreover, CatBoost uses oblivious decisions in order to improve the efficiency, to enhance execution speed and to solve the problem of over-fit.

4.1.6 XGBoost algorithm

Recently, XGBoost has been utilized in various disciplines, such as, energy (De Clercq et al., 2020; Ma et al., 2020), health care (Guo et al., 2019; Singh et al., 2019), and credit scoring (Jiang et al., 2019; Xia et al., 2017). XGBoost, developed by Chen and Guestrin (2016), is an algorithm that incorporates the boosting model proposed by Friedman (2001). Normalization is used in the objective function to reduce model complexity, to prevent overfitting and to make the learning process faster. Importantly, XGBoost is an ensemble model which consists of an efficient implementation of decision trees, in order to produce a combined model whose predictive performance is better than individual techniques used alone. According to Mo et al. (2019), the output function is calculated as follows:

$$\hat{Y}_i^T = \sum_{k=1}^T f_k(x_i) = \hat{y}_i^{T-1} + f_T(x_i)$$

where \hat{Y}_i^{T-1} is the generated tree, $f_T(x_i)$ is the newly created tree model, and T is the total number of tree models.

In XGBoost, several parameters need to be regulated to maximize the power of model performance and to prevent overfitting problems. Cross-validation technique has been used to find the optimal combination of parameters. In our study, for parameter tuning we use tenfold cross-validation. The optimal hyper-parameters values selected after cross-validation are col sample by tree: 0.7; learning rate: 0.05; number of iterations: 500, max depth: 5, subsample: 0.7; minimum sum for instance weight needed in a child: 4; number of parallel threads: 4, and silent: 1.

4.2 Performance metrics of models

The forecasting performance of the six machine learning models is evaluated by five common evaluation metrics: the root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and the coefficient of determination (R^2). The specific definition of each of these metrics can be calculated as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{h=1}^N (\hat{Y}_h - Y_h)^2}$$

$$MSE = \frac{1}{N} \sum_{h=1}^N (\hat{Y}_h - Y_h)^2$$

$$MAE = \frac{1}{N} \sum_{h=1}^N |\hat{Y}_h - Y_h|$$

$$R^2 = \frac{\sum_{h=1}^N (\hat{Y}_h - \bar{Y}_h)^2}{\sum_{h=1}^N (Y_h - \bar{Y}_h)^2}$$

4.3 SHAP (SHapley Additive exPlanation) approach for results interpretation

Machine learning has great potential in forecasting times series data. But researchers do not usually explain their predictions, which is a barrier to the adoption of machine learning. To overcome this problem, Lundberg and Lee (2017) proposed a SHAP approach for interpreting predictions for different techniques including LightGBM, NGBoost, CatBoost, XGBoost, and Scikit-learn tree models. SHAP helps users to interpret the predictions of complex models. SHAP was initially proposed by Shapley in 1953 and it is based on game theory (Shapley, 1953). It allows us to explain the prediction of a specific input (X) by calculating the impact of each feature to the prediction. The estimated Shapley value is calculated as follows:

$$\hat{\phi}_j = \frac{1}{K} \sum_{k=1}^K ((\hat{g}(x_{+j}^m) - \hat{g}(x_{-j}^m)))$$

where $\hat{g}(x_{+j}^m)$ is the prediction for x, but with a random number of feature values.

Lundberg et al., (2018) proposed TreeSHAP, for gradient boosting models, among them XGBoost. TreeSHAP offers a rich visualization of each feature attribution that improves over classic feature importance and partial dependence plots. According to Lundberg et al., (2018) the TreeSHAP interaction values can be estimated as follows:

$$\phi_{i,j} = \sum_{S \subseteq N \setminus \{i,j\}} \frac{|S|!(M - |S| - 2)!}{2(M - 1)!} \delta_{ij}(S)$$

when $i \neq j$, $\delta_{ij}(S) = f_x(S \cup \{i,j\}) - f_x(S \cup \{i\}) - f_x(S \cup \{j\}) + f_x(S)$, M is the number of features, and S all feature subsets. SHAP values advance our understanding of tree models by including feature importance, feature dependence plots, local explanations and summary plots.

Table 4 Machine learning models performance on testing dataset

Models	RMSE	MSE	MAE	R ²
XGBoost	34.921	1219.500	21.968	0.994
CatBoost	36.285	1316.636	23.488	0.993
Random Forest	42.113	1773.566	26.787	0.991
LightGBM	45.592	2078.698	28.793	0.990
Neural Network	195.961	38,400.748	155.425	0.807
Linear Regression	71.325	5087.391	53.693	0.974

This table presents forecasting performance of the six machine learning models used: the root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and the coefficient of determination (R²)

5 Results analysis

5.1 Comparison of models' performance

We have estimated model performance using the root mean square error (RMSE), mean square error (MSE), mean absolute error (MAE), and the coefficient of determination (R²). To compare the six machine learning models, a validation test sample (20%) has been used. The findings confirm the high capacity of explanatory variables to predict gold price, which R² ranging from 0.807 for linear regression to 0.994 for XGBoost shown in Table 4.

Table 4 presents a comparison of the predictive capacity of the six models. The model with the lowest values of RMSE, MSE and MAE, and the highest value of R² is considered the best forecasting model. As depicted in Table 4, XGBoost provides the highest R² value coupled with the lowest RMSE, MSE and MAE among all the models used in this study. XGBoost is followed by CatBoost and RF, whereas linear regression and neural networks lead to the worst results. This indicates the advantage of the XGBoost over traditional forecasting techniques used in time series data. To assess the predictive power of our analysis, we have also illustrated the performance of different models in Fig. 1. This figure shows that the gold price forecasted using XGBoost is extremely near to reality for the test data. These findings are in line with the results of Xia et al. (2017) and Climent et al., (2019) who reported the performance of XGBoost compared to traditional models in bankruptcy prediction and credit scoring. Nevertheless, the worst results of neural networks could be explained by the small size of the sample. According to Lago et al. (2018), deep learning models require large amounts of data to be correctly computed.

5.2 Feature analysis

SHAP allows interpreting the effect of the influence of the input variables in the output. SHAP helps policymakers to interpret machine learning models; it calculates the variables' importance.

Figure 2 displays the SHAP summary plot that orders variables based on their importance to affect the gold price. We can see that the silver price is the most important feature in the model. This result supports the findings in Schweikert (2018) who reported a strong dependence on the long-run relationship between gold and silver prices. Additionally,

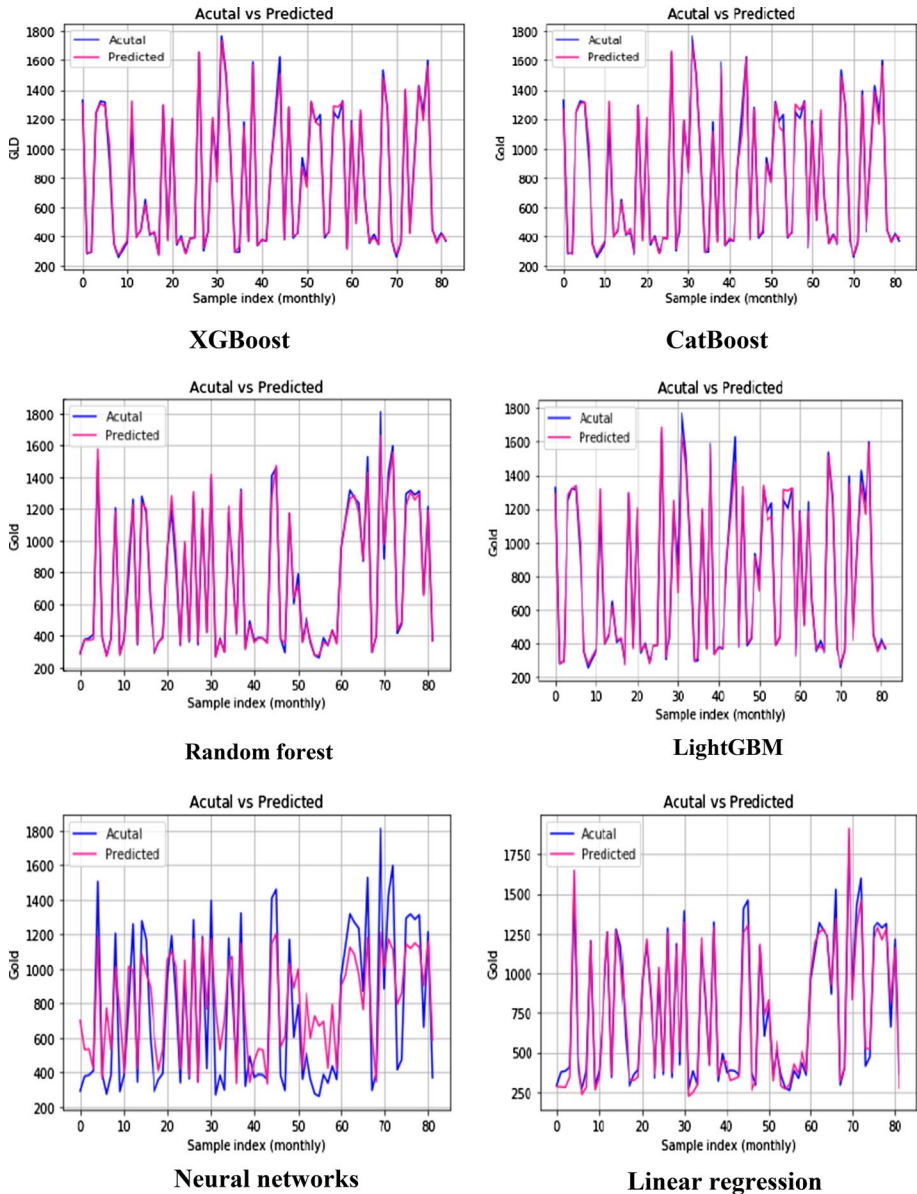


Fig. 1 Performances of six machine learning models over test sample

higher values of silver price result in higher SHAP values, which relate to a higher probability that gold price increases. Inflation is the next most important feature, in that order, and higher values of this variable correspond to higher chance that gold price increases. This is in line with the results of Kristjanpoller and Minutolo (2015) who documented that inflation is correlated with gold price fluctuations. In contrast, lower values of iron ore correspond to a higher chance of gold price increases. This finding is in contrast with Alameer

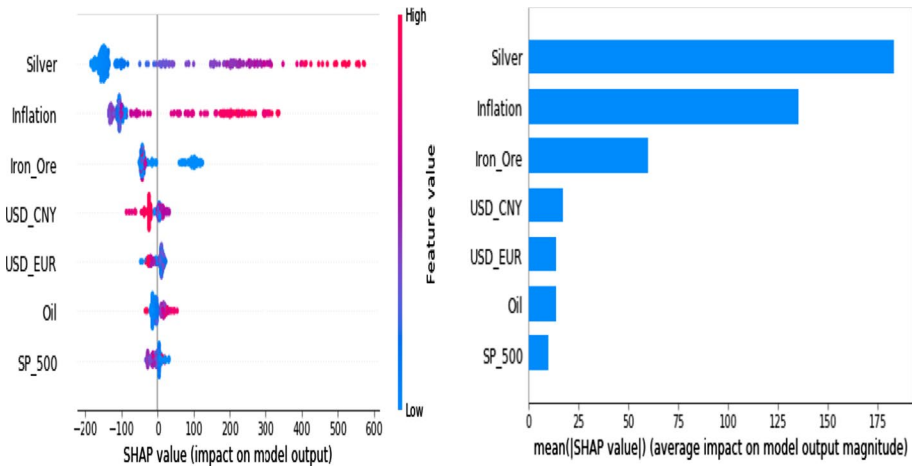


Fig. 2 On the left, SHAP summary plot of the XGBoost model. The higher SHAP value of a feature, the higher gold price levels. On the right, the relative importance for each feature, obtained by taking the average absolute value of the SHAP values

et al., (2019), who found a positive relationship between iron ore and gold price. The relationship between gold price and crude oil price is particularly evident when higher gold price results in increased crude oil price. Again, this finding is consistent with the results of Kanjilal and Ghosh (2017) and Singhal et al. (2019), who documented that crude oil price is a major macroeconomic determinant to guide gold price movement. Finally, SHAP importance variables in Fig. 2 indicate that high values of gold price results in low values of S&P 500. This result is consistent with the findings reported by Piñeiro-Chousa et al. (2018), who found that gold returns and the S&P 500 index is negatively related.

To further examine the relationship between features and the outcome, SHAP dependence plots show how a variable's value impacts the prediction (y axis) of every observation in the dataset. We display SHAP dependence plots in Fig. 3. SHAP dependency plots may depict both the major impact of individual predictor variables as well as the interactions between them. By global interpretability, we can see on the whole sample, the positive or negative contribution of each feature to the prediction score. For example, in Fig. 3-d, we can examine the impact of China's exchange rate when the price of crude oil increases from 20 to 80. The red points display higher values of crude oil price, and the blue points represent lower ones, which reveal that increasing China's exchange rate increases the volatility of gold price. When China's exchange rate is low, SHAP values for high crude oil prices are above zero, which suggests that increasing China's exchange rate increases the gold price. In contrast, SHAP values for low crude oil price are under zero, which indicates that increasing China's exchange rate while crude oil price is low reduces the chance of increasing gold price.

Figure 3-g shows the effect of S&P 500 and inflation on the movements of gold price. The SHAP values indicate that the impact of S&P 500 starts positively; that is, increasing S&P 500 when it is below 30, results in higher gold price. Between 0 and 30, the linkage becomes negative, and an increase in S&P 500 causes a high gold price. In addition, high values of S&P 500 and inflation show that the gold price tends to be higher.

Let's see how SHAP can help us to obtain local knowledge. By local interpretability, we can measure how the feature values contribute to the prediction score of each observation

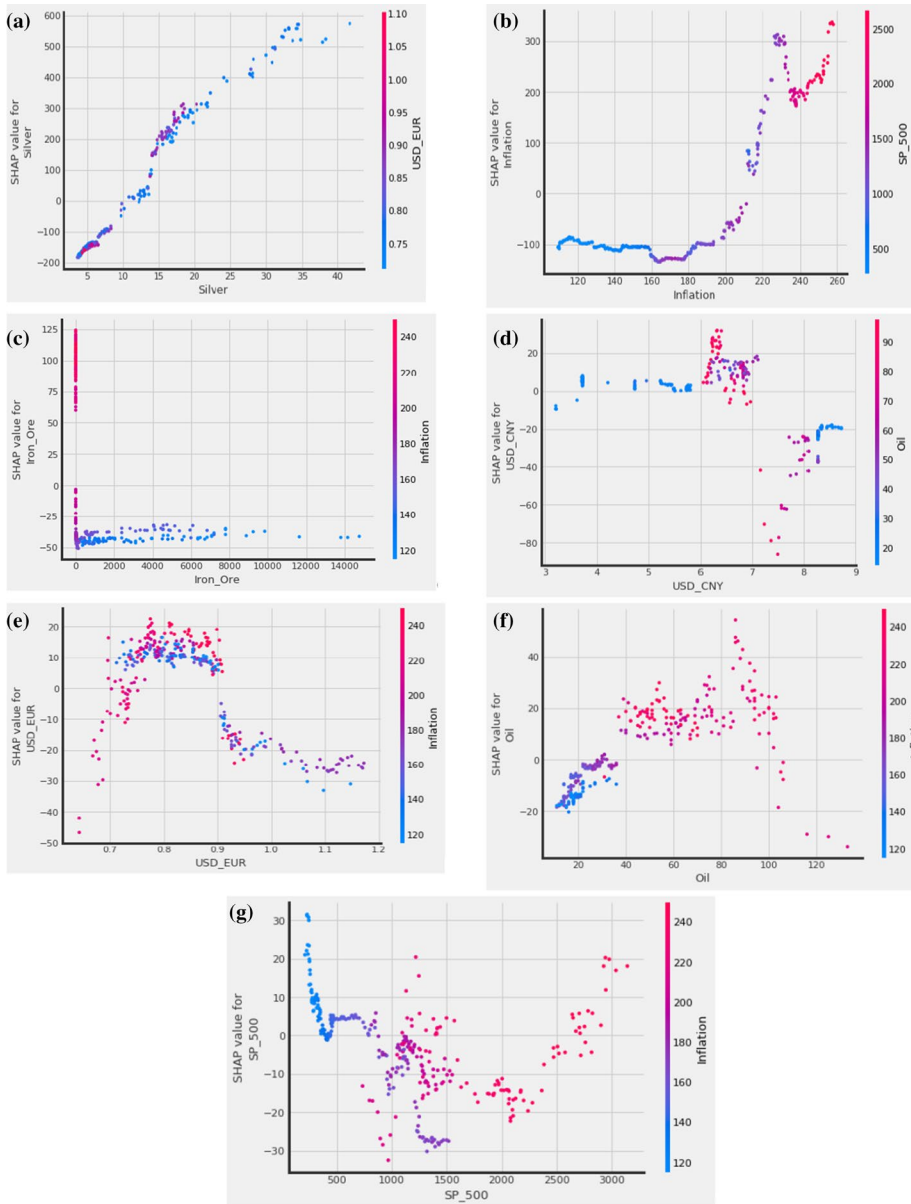


Fig. 3 SHAP dependence plots. The x-axis is the value of the feature pressure and the y-axis is the SHAP value. The red values represent the high values of the variable, whereas the blue signifies the low values

in the sample separately. The traditional attribute importance algorithm, like linear models, only gives the importance value of a global attribute over the whole dataset, while SHAP value will provide importance value for each observation separately. Figure 4 shows the marginal effect of seven features on the predicted gold price level of the XGBoost model in the training dataset. In Fig. 4, the red color represents that the feature increased the gold

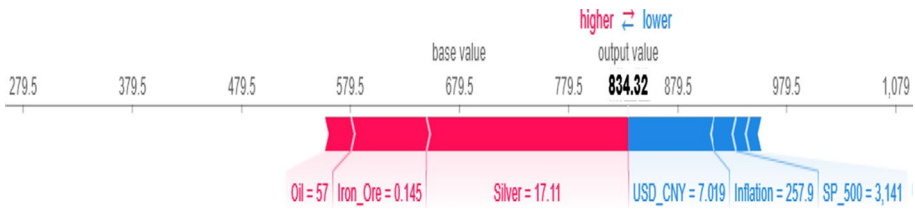


Fig. 4 Explanation of the first prediction generated by the XGBoost model using tree SHAP

price, while the blue color shows that the feature decreased the gold price. We have chosen the first observation prediction by XGBoost for illustration. The gold price prediction is equal to 834.32 significantly higher as the features have the following characteristics: China's exchange rate (7.019), inflation (257.9) and S&P 500 (3.141) located in the blue zone, which indicates those variables driving the gold price towards lower values. Features located in the red zone; Silver (17.11), Iron Ore (0.145) and crude oil (57), which indicate those features, push the gold price towards higher values.

6 Conclusion

In this paper, we compare six different machine learning models to determine which one is more suitable to predict the gold price. The findings show that XGBoost, which provides the best result over the other competitive techniques and outperforms all well-known benchmark models. Furthermore, the results provide significant correlations among gold prices and all the predictor variables considered, i.e. crude oil, iron and silver price, USD_EUR and USD_CNY exchange rates, S&P 500 and inflation rate of US. This result demonstrates that those variables display a high capacity to forecast future gold price volatility.

Moreover, this study presents the SHAP method to unify the field of interpretable machine learning. Indeed, the proposed technique provides a rich visualization of individualized feature attributions that improves the interpretability of gold price fluctuations. Moreover, Tree SHAP's advance our understanding of tree models. It offers an insightful means to interpret the findings from a complex framework such as XGBoost and also extract nonlinear relationships of features on the output of a model. We show how Shapley additive explanations can be used to interpret the outputs of XGBoost designed to predict the gold price.

As practical implications, our results offer some meaningful implications to the investors and policy makers. First, the choice of an accurate technique should represent an effective forecasting tool for central banks and investors. Central banks need to know the gold price movements in order to secure certain transactions or to make strategic reserves. For investors, gold helps to diversify the portfolio and serves as insurance in the case of risk and volatility, in other words, gold can be used as investors' safe haven. If the direction of the gold market is successfully predicted, the investors might be better guided and earn a safe return. Second, our findings suggest that traders may benefit from the XGBoost algorithm and SHAP interaction values in their decisions. Successful forecasting procedure empowers traders to make decisions and plan for the future in order to enhance favorable scenarios. Third, our study will profit to policy makers as it gives a list of factors including oil as indicators for gold price. The SHAP method offers a powerful and insightful measure of the importance of every input variable in the prediction of future gold price fluctuations.

Our study presents a limitation common to all similar studies. Although the model proposed can achieve very accurate predictions, it should also be acknowledged that markets rely on a number of variables, like geopolitical decisions that can result in unpredictable movements.

Finally, future research might extend our work by considering additional variables, such as political or commercial factors as well as phases of economic instability, which are generally determining factors of the price of gold. Also, as another direction for future research is the application of the proposed model in forecasting other commodities prices. Moreover, it would be interesting to include one or more computational cost factors in the comparison of different forecasting models. A mathematical formulation based upon some operational research process would allow a more objective comparison. It would also be promising to equip some of the approaches presented above with a preprocessing stage to provide some hybrid adaptive approach (Saâdaoui, 2012). Moreover, developing multimodal extensions by first proceeding with an unsupervised clustering could also lead to sufficiently robust approaches to better capture the outliers present in the data (Saâdaoui, 2012). Metaheuristic optimization methods could therefore be useful to identify and estimate this type of method (Rabbouch et al., 2020).

References

- Abd Elaziz, M., Ewees, A. A., & Alameer, Z. (2019). Improving adaptive neuro-fuzzy inference system based on a modified salp swarm algorithm using genetic algorithms to forecast crude oil price. *Natural Resources Research*. <https://doi.org/10.1007/s11053-019-09587-1>
- Abellán, J., & Mantas, C. J. (2014). Improving experimental studies about ensembles of classifiers for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 41, 3825–3830. <https://doi.org/10.1016/j.eswa.2013.12.003>
- Akbar, M., Iqbal, F., & Noor, F. (2019). Bayesian analysis of dynamic linkages among gold price, stock prices, exchange rate and interest rate in Pakistan. *Resources Policy*, 62(April), 154–164. <https://doi.org/10.1016/j.resourpol.2019.03.003>
- Alameer, Z., Elaziz, M. A., Ewees, A. A., Ye, H., & Jianhua, Z. (2019). Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. *Resources Policy*, 61(February), 250–260. <https://doi.org/10.1016/j.resourpol.2019.02.014>
- Antunes, F., Ribeiro, B., & Pereira, F. (2017). Probabilistic modeling and visualization for bankruptcy prediction. *Applied Soft Computing*, 60, 831–843. <https://doi.org/10.1016/j.asoc.2017.06.043>
- Babar, B., Luppino, L. T., Boström, T., & Anfinsen, S. N. (2020). Random forest regression for improved mapping of solar irradiance at high latitudes. *Solar Energy*, 198, 81–92. <https://doi.org/10.1016/j.solener.2020.01.034>
- Baker, R., Forrest, D., & Pérez, L. (2020). Modelling demand for lotto using a novel method of correcting for endogeneity. *Economic Modelling*, 84, 302–308. <https://doi.org/10.1016/j.econmod.2019.04.021>
- Basak, S., Karb, S., Sahaa, S., Luckyson, K., & Sudeepa, R. D. (2019). Predicting the direction of stock market prices using tree-based classifiers. *North American Journal of Economics and Finance*, 47, 552–567.
- Batten, J. A., Ciner, C., & Lucey, B. M. (2010). The macroeconomic determinants of volatility in precious metals markets. *Resources Policy*, 35(2), 65–71.
- Baur, D. G., & Lucey, B. M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Finance Review*, 45, 217–229. <https://doi.org/10.1111/j.1540-6288.2010.00244.x>
- Baur, D. G., & McDermott, T. K. (2010). Is gold a safe haven? International evidence. *Journal of Banking and Finance*, 34, 1886–1898. <https://doi.org/10.1016/j.jbankfin.2009.12.008>
- Beckmann, J., & Czudaj, R. (2013). Gold as an inflation hedge in a time-varying coefficient framework. *North American Journal of Economics and Finance*, 24(1), 208–222. <https://doi.org/10.1016/j.najef.2012.10.007>
- Bedoui, R., Braïek, S., Guesmi, K., & Chevallier, J. (2019). On the conditional dependence structure between oil, gold and USD exchange rates: Nested copula based GJR GARCH model. *Energy Economy*, 80, 876–889.

- Behmiri, N. B., & Manera, M. (2015). The role of outliers and oil price shocks on volatility of metal prices. *Resources Policy*, 46, 139–150. <https://doi.org/10.1016/j.resourpol.2015.09.004>
- Ben Jabeur, S., Sadaoui, A., Sghaier, A., & Aloui, R. (2020). Machine learning models and cost-sensitive decision trees for bond rating prediction. *Journal of the Operational Research Society*, 71(8), 1161–1179. <https://doi.org/10.1080/01605682.2019.1581405>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Bhatia, V., Das, D., Tiwari, A. K., Shahbaz, M., & Hasim, H. M. (2018). Do precious metal spot prices influence each other? Evidence from a nonparametric causality-in-quantiles approach. *Resources Policy*, 55, 244–252. <https://doi.org/10.1016/j.resourpol.2017.12.008>
- Bodart, V., Candelon, B., & Carpentier, J.-F. (2015). Real exchanges rates, commodity prices and structural factors in developing countries. *Journal of International Money and Finance*, 51, 264–284. <https://doi.org/10.1016/j.jimonfin.2014.11.021>
- Capie, F., Mills, T. C., & Wood, G. (2005). Gold as a hedge against the dollar. *Journal of International Financial Markets, Institutions and Money*, 15, 343–352.
- Chen, L. & Zhang, X. (2019). Gold price forecasting based on projection pursuit and neural network. *IOP Conf. Series: Journal of Physics: Conf. Series* 1168 06 2009. IOP Publishing Doi:<https://doi.org/10.1088/1742-6596/1168/6/062009>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- Chen, Y.-C., Rogoff, K. S., & Rossi, B. (2010). Can exchange rates forecast commodity prices? *Quarterly Journal of Economics*, 125, 1145–1194. <http://www.jstor.org/stable/27867508>
- Chen, Y., Xie, X., Zhang, T., Bai, J., & Hou, M. (2020). A deep residual compensation extreme learning machine and applications. *Journal of Forecasting*, 39, 986–999. <https://doi.org/10.1002/for.2663>
- Ciner, C. (2017). Predicting white metal prices by a commodity sensitive exchange rate. *International Review of Financial Analysis*, 52, 309–315. <https://doi.org/10.1016/j.irfa.2017.04.002>
- Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. *Journal of Business Research*, 101, 885–896. <https://doi.org/10.1016/j.jbusres.2018.11.015>
- Cologni, A., & Manera, M. (2008). Oil prices, inflation and interest rates in a structural cointegrated VAR model for the G-7 countries. *Energy Economics*, 30, 856–888. <https://doi.org/10.2139/ssrn.843505>
- De Clercq, D., Wen, Z., Fei, F., Caicedo, L., Yuan, K., & Shang, R. (2020). Interpretable machine learning for predicting biomethane production in industrial-scale anaerobic co-digestion. *Science of the Total Environment*. <https://doi.org/10.1016/j.scitotenv.2019.134574>
- Du, P., Wang, J., Yang, W., & Tong, N. (2020). Point and interval forecasting for metal prices based on variational mode decomposition and an optimized outlier-robust extreme learning machine. *Resources Policy*, 69, 101881.
- Elie, B., Naji, J., Dutta, A., & Uddin, G. S. (2019). Gold and crude oil as safe-haven assets for clean energy stock indices: Blended copulas approach. *Energy*, 178, 544–553. <https://doi.org/10.1016/j.energy.2019.04.155>
- Escribano, A., & Granger, C. W. J. (1998). Investigating the relationship between gold and silver prices. *Journal of Forecasting*, 17, 81–107.
- Ewees, A. A., Elaziz, M. A., Alameer, Z., Ye, H., & Jianhua, Z. (2020). Improving multilayer perceptron neural network using chaotic grasshopper optimization algorithm to forecast iron ore price volatility. *Resources Policy*. <https://doi.org/10.1016/j.resourpol.2019.101555>
- Fortune, J. N. (1987). The inflation rate of the price of gold, expected prices and interest rates. *Journal of Macroeconomics*, 9, 71–82.
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29, 1189–1232.
- Gholamy, A., Kreinovich, V., Kosheleva, O., (2018). Why 70/30 or 80/20 relation between training and testing sets : A pedagogical explanation. Departmental Technical Reports (CS) 1–6.
- Guo, J., Yang, L., Bie, R., Yu, J., Gao, Y., Shen, Y., & Kos, A. (2019). An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring. *Computer Networks*, 151, 166–180. <https://doi.org/10.1016/j.comnet.2019.01.026>
- He, Y., Wang, S., & Lai, K. K. (2010). Global economic activity and crude oil prices: A cointegration analysis. *Energy Economics*, 32, 868–876. <https://doi.org/10.1016/j.eneco.2009.12.005>

- Herawati, S., Firmansyah, A., Latif, M., & Aeri, R. (2017). Implementing method of ensemble empirical mode decomposition and recurrent neural network for gold price forecasting. *Journal of Engineering Research and Applications*, 7(11), 39–43.
- Huck, N. (2019). Large data sets and machine learning: Applications to statistical arbitrage. *European Journal of Operational Research*, 278, 330–342.
- Jiang, C., Wang, Z., & Zhao, H. (2019). A prediction-driven mixture cure model and its application in credit scoring. *European Journal of Operational Research*, 277(1), 20–31. <https://doi.org/10.1016/j.ejor.2019.01.072>
- Kang, S. H., McIver, R., & Yoon, S.-M. (2017). Dynamic spillover effects among crude oil, precious metal, and agricultural commodity futures markets. *Energy Economics*, 62, 19–32.
- Kanjilal, K., & Ghosh, S. (2017). Dynamics of crude oil and gold price post 2008 global financial crisis—New evidence from threshold vector error-correction model. *Resources Policy*, 52(March), 358–365. <https://doi.org/10.1016/j.resourpol.2017.04.001>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 2017-Decem(Nips)*, 3147–3155.
- Kearney, A., & Lombra, R. E. (2009). Gold and platinum: Toward solving the price puzzle. *The Quarterly Review of Economics and Finance*, 49, 884–892.
- Khashei, M., & Bijari, M. (2010). An artificial neural network model for time series forecasting. *Expert System and Applications*, 37, 479–489. <https://doi.org/10.1016/j.eswa.2009.05.044>
- Khashei, M., & Bijari, M. (2011). A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Application Soft Computation*, 11, 2664–2675. <https://doi.org/10.1016/j.asoc.2010.10.015>
- Krauss, C., Anh, X., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500 R. *European Journal of Operational Research*, 259(2), 689–702. <https://doi.org/10.1016/j.ejor.2016.10.031>
- Kristjanpoller, W., & Hernandez, E. (2017). Volatility of main metals forecasted by a hybrid ANN-GARCH model with regressors. *Expert Systems with Applications*, 84, 290–300. <https://doi.org/10.1016/j.eswa.2017.05.024>
- Kristjanpoller, W., & Minutolo, M. C. (2015). Gold price volatility: A forecasting approach using the artificial neural network-GARCH model. *Expert Systems with Applications*, 42(20), 7245–7251. <https://doi.org/10.1016/j.eswa.2015.04.058>
- Kristjanpoller, W., & Minutolo, M. C. (2016). Forecasting volatility of oil price using an artificial neural network-GARCH model. *Expert Systems with Applications*, 65, 233–241. <https://doi.org/10.1016/j.eswa.2016.08.045>
- Kucher, O., & McCoskey, S. (2017). The long-run relationship between precious metal prices and the business cycle. *Quarterly Review of Economics and Finance*, 65, 263–275. <https://doi.org/10.1016/j.qref.2016.09.005>
- Kumar, S. (2018). Prediction of gold and silver prices in an emerging economy: Comparative analysis of linear, nonlinear, hybrid, and ensemble models. *The Journal of Prediction Markets*, 12(3), 63–78.
- Lago, J., De Ridder, F., & De Schutter, B. (2018). Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy*, 221(April), 386–405. <https://doi.org/10.1016/j.apenergy.2018.02.069>
- Lardic, S., & Mignon, V. (2008). Oil prices and economic activity: An asymmetric cointegration approach. *Energy Economics*, 30, 847–855. <https://doi.org/10.1016/j.eneco.2006.10.010>
- Lineesh, M., Minu, K., & John, C. J. (2010). Analysis of nonstationary nonlinear economic time series of gold price: A comparative study. *International Mathematical Forum*, 5, 1673–1683.
- Liu, C., Hu, Z., Li, Y., & Liu, S. (2017). Forecasting copper prices by decision tree learning. *Resources Policy*, 52, 427–434. <https://doi.org/10.1016/j.resourpol.2017.05.007>
- Liu, Y., Li, H., Guan, J., Liu, X., Guan, Q., & Sun, Q. (2019). Influence of different factors on prices of upstream, middle and downstream products in China's whole steel industry chain: Based on Adaptive Neural Fuzzy Inference. *System Resources Policy*, 60, 134–142. <https://doi.org/10.1016/j.resourpol.2018.12.009>
- Loureiro, A. L. D., Miguéis, V. L., & da Silva, L. F. M. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, 114(January), 81–93. <https://doi.org/10.1016/j.dss.2018.08.010>
- Lundberg, S. M., Erion, G. G., & Lee, S. -I. (2018). *Consistent Individualized Feature Attribution for Tree Ensembles. 2*.
- Lundberg, S. M., and Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems, 2017-Decem(Section 2)*, 4766–4775.

- Ma, J., Ding, Y., Cheng, J. C. P., Jiang, F., Tan, Y., Gan, V. J. L., & Wan, Z. (2020). Identification of high impact factors of air quality on a national scale using big data and machine learning techniques. *Journal of Cleaner Production*. <https://doi.org/10.1016/j.jclepro.2019.118955>
- Mahdavi, S., & Zhou, S. (1997). Gold and commodity prices as leading indicators of inflation: Tests of long-run relationship and predictive performance. *Journal of Economics and Business*, 49(5), 475–489.
- Mercadier, M., & Lardy, J. P. (2019). Credit spread approximation and improvement using random forest regression. *European Journal of Operational Research*, 277(1), 351–365. <https://doi.org/10.1016/j.ejor.2019.02.005>
- Mo, B., Nie, H., & Jiang, Y. (2018). Dynamic linkages among the gold market, US dollar and crude oil market. *Physica A*, 491, 984–994. <https://doi.org/10.1016/j.physa.2017.09.091>
- Mo, H., Sun, H., Liu, J., & Wei, S. (2019). Developing window behavior models for residential buildings using XGBoost algorithm. *Energy and Buildings*. <https://doi.org/10.1016/j.enbuild.2019.109564>
- O'connor, F. A., Lucey, B. M., Battend, J. A., & Baure, D. G. (2015). The financial economics of gold—A survey. *International Review of Financial Analysis*, 41, 186–205.
- Parisi, A., Parisi, F., & Diaz, D. (2008). Forecasting gold price changes: Rolling and recursive neural network models. *Journal of Multinational Financial Management*, 18, 477–487. <https://doi.org/10.1016/j.mulfin.2007.12.002>
- Pesaran, M. H., & Smith, R. P. (2019). A Bayesian analysis of linear regression models with highly collinear regressors. *Econometrics and Statistics*, 11, 1–21. <https://doi.org/10.1016/j.ecosta.2018.10.001>
- Pierdzioch, C., Risse, M., & Rohloff, S. (2015a). Forecasting gold-price fluctuations: A real-time boosting approach. *Applied Economics Letter*, 22, 46–50.
- Pierdzioch, C., Risse, M., & Rohloff, S. (2015b). A boosting approach to forecasting gold and silver returns: Economic and statistical forecast evaluation. *Applied Economics Letter*, 22, 46–50.
- Pierdzioch, C., Risse, M., & Rohloff, S. (2016). Fluctuations of the real exchange rate, real interest rates, and the dynamics of the price of gold in a small open economy. *Empirical Economics*, 51(4), 1481–1499. <https://doi.org/10.1007/s00181-015-1053-5>
- Pierdziochu, P., & Risse, M. (2020). Forecasting precious metal returns with multivariate random forests. *Empirical Economics*, 58, 1167–1184. <https://doi.org/10.1007/s00181-018-1558-9>
- Piñeiro-Chousa, J., López-Cabarcos, M. Á., Pérez-Pico, A. M., & Ribeiro-Navarrete, B. (2018). Does social network sentiment influence the relationship between the S&P 500 and gold returns? *International Review of Financial Analysis*, 57, 57–64. <https://doi.org/10.1016/j.irfa.2018.02.005>
- Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems, 2018-December* (Section 4), 6638–6648.
- Pukthuanthong, K., & Roll, R. (2011). Gold and the dollar (and the Euro, Pound, and Yen). *Journal of Banking and Finance*, 35, 2070–2083.
- Rabbouch, B., Saâdaoui, F., & Mraïhi, R. (2020). Empirical-type simulated annealing for solving the capacitated vehicle routing problem. *Journal of Experimental and Theoretical Artificial Intelligence*, 32(3), 437–452. <https://doi.org/10.1080/0952813X.2019.1652356>
- Ramyar, S., & Kianfar, F. (2017). Forecasting crude oil prices: A comparison between artificial neural networks and vector Autoregressive models. *Computational Economics*. <https://doi.org/10.1007/s10614-017-9764-7>
- Reboredo, J. C. (2013). Is gold a safe haven or a hedge for the U.S. dollar? Implications for risk management. *Journal of Banking and Finance*, 37, 2665–2676.
- Risse, M. (2019). Combining wavelet decomposition with machine learning to forecast gold returns. *International Journal of Forecasting*, 35(2), 601–615. <https://doi.org/10.1016/j.ijforecast.2018.11.008>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Modelagnostic interpretability of machine learning. arXiv preprint arXiv:1606.05386.
- Roberts, M. C. (2009). Duration and characteristics of metal price cycles. *Resources Policy*, 34, 87–102. <https://doi.org/10.1016/j.resourpol.2009.02.001>
- Rossen, A. (2015). What are metal prices like? Co-movement, price cycles and long-run trends. *Resources Policy*, 45, 255–276. <https://doi.org/10.1016/j.resourpol.2015.06.002>
- Saâdaoui, F. (2012). A probabilistic clustering method for US interest rate analysis. *Quantitative Finance*, 12(1), 135–148. <https://doi.org/10.1080/14697681003591712>
- Sari, R., Hammoudeh, S., & Soytaş, U. (2010). Dynamics of oil price, precious metal prices, and exchange rate. *Energy Economics*, 32, 351–362.
- Schweikert, K. (2018). Are gold and silver cointegrated? New evidence from quantile cointegrating regressions. *Journal of Banking and Finance*, 88, 44–51. <https://doi.org/10.1016/j.jbankfin.2017.11.010>
- Sensoy, A. (2013). Dynamic relationship between precious metals. *Resources Policy*, 38, 504–511.

- Sephton, P., & Mann, J. (2018). Gold and crude oil prices after the great moderation. *Energy Economics*, 71, 273–281. <https://doi.org/10.1016/j.eneco.2018.02.022>
- Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90, 106181.
- Shafiee, S., & Topal, E. (2010a). An overview of global gold market and gold price forecasting. *Resources Policy*, 35, 178–189. [4].
- Shafiee, S., & Topal, E. (2010b). An overview of global gold market and gold price forecasting. *Resources Policy*, 35, 178–189. <https://doi.org/10.1016/j.resourpol.2010.05.004>
- Shapley, L. S., (1953). A value for n-person games. *Contrib. to Theory Games*. pp. 307–317.
- Singh, N., Singh, P., & Bhagat, D. (2019). A rule extraction approach from support vector machines for diagnosing hypertension among diabetics. *Expert Systems with Applications*, 130, 188–205. <https://doi.org/10.1016/j.eswa.2019.04.029>
- Singhal, S., Choudhary, S., & Biswal, P. C. (2019). Return and volatility linkages among International crude oil price, gold price, exchange rate and stock markets: Evidence from Mexico. *Resources Policy*, 60(January), 255–261. <https://doi.org/10.1016/j.resourpol.2019.01.004>
- Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and information systems*, 41(3), 647–665.
- Sun, X., Liu, M., & Sima, Z. (2019). A novel cryptocurrency price trend forecasting model based on Light-GBM. *Finance Research Letters*, December. <https://doi.org/10.1016/j.frl.2018.12.032>
- Teetranont, T., Chanaim, S., Yamaka, W., & Sriboonchitta, S. (2018). Investigating relationship between gold price and crude oil price using interval data with copula based GARCH. In V. Kreinovich, S. Sriboonchitta, & N. Chakpitak (Eds.), *Predictive Econometrics and Big Data* (pp. 656–669). Springer International Publishing.
- Tully, E., & Lucey, B. M. (2007). A power GARCH examination of the gold market. *Research in International Business and Finance*, 21(2), 316–325.
- Wen, X., Yang, X., & Gong, K. K. L. (2017). Multi-scale volatility feature analysis and prediction of gold price. *International Journal of Information Technology and Decision*, 16, 205–223.
- Wu, D., & Hu, Z.-H. (2016). Structural changes and volatility correlation in nonferrous metal market. *Transactions Nonferrous Metals Society of China*, 26, 2784–2792. [https://doi.org/10.1016/S1003-6326\(16\)64395-9](https://doi.org/10.1016/S1003-6326(16)64395-9)
- Xia, Y., Liu, C., Li, Y. Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78, 225–241. <https://doi.org/10.1016/j.eswa.2017.02.017>
- Yue, Y., Liu, D., & Xu, S. (2015). Price linkage between Chinese and international nonferrous metals commodity markets based on VAR-DCC-GARCH models. *Transactions Nonferrous Metals Society of China*, 25, 1020–1026. [https://doi.org/10.1016/S1003-6326\(15\)63693-7](https://doi.org/10.1016/S1003-6326(15)63693-7)
- Zhang, P., & Ci, B. (2020). Deep belief network for gold price forecasting. *Resources Policy*, 69, 101806.
- Zheng, J., Fu, X., & Zhang, G. (2019). Research on exchange rate forecasting based on deep belief network. *Neural Computing and Applications*, 31, 573–582.
- Zhu, Y., & Zhang, C. (2018). Gold price prediction based on pca-ga-bp neural network. *Journal of Computer and Communications*, 6(7), 22–33.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.