International Journal of Scientific Research and Engineering Development-– Volume 8 Issue 5, Sep-Oct 2025
Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Gold Price Forecasting Based on Regularized Linear Models and Extreme Gradient Boosting

Sivaranjani P.S[1], Thulasimani K[2]
[1]Department of Computer Engineering, Government College of Engineering, Tirunelveli, TamilNadu, India
E-mail address: sivaranjani36378@gmail.com
[2]Department of Computer Engineering, Government College of Engineering, Tirunelveli, TamilNadu, India
E-mail address: thulasimani@gcetly.ac.in

----------------------------------****************************--------------------------------

Abstract
Precise predictions of gold prices are crucial for investors, decision-makers, and financial analysts, given that gold functions as both a secure asset and an important economic indicator. The research showcased introduces a machine learning framework that employs Bayesian optimization, XG Boost, and Lasso regression to enhance prediction efficiency and accuracy. An examination was performed on historical gold price data along with relevant financial information indicators, utilizing Lasso regression to determine the key predictors via L1 regularization, whereas XG Boost successfully complex non-linear interactions in the market were identified. Bayesian optimization was employed to optimize the model's hyperparameters, eliminating the need for manual tuning and ensuring optimal performance The comparative evaluation indicates that the integration of Lasso regression provides a simple and effective method for feature selection, whereas XG Boost employs gradient boosting to uncover intricate patterns within the data, leading to lower error rates and enhanced generalization. By integrating these methods, it creates a powerful, data-driven forecasting tool that delivers reliable and accurate predictions the rapidly changing financial environment. The results emphasize the importance of merging sophisticated machine learning with automated optimization to ensure accurate commodity price forecasts and support strategic decision-making in both investment and policy areas.

Keywords: Gold price Forecasting, Lasso Regression, XGBOOST, Machine learning, Feature selection.

----------------------------------****************************--------------------------------

1. INTRODUCTION

Financial savings and investments are crucial for influencing private economic fitness and riding the collective progress of an economic system. Investments represent the allocation of present-day sources with the intention of generating returns inside the destiny, both via income streams and capital appreciation. In financial phrases, funding refers to obtaining property that aren't used up right away however are accrued to build wealth through the years. From a economic point of view, investments can take multiple bureaucracy, starting from shares, bonds, and real estate to commodities and valuable metals. In recent many years, the rapid growth of the Indian financial system has no longer simplest extended disposable earnings ranges but has also various the variety of investment possibilities available to households and institutional investors together. a few of the giant choice of funding picks, gold has consistently upheld its repute as a safe-haven asset and is regularly known as "the remaining asset." in the course of history, gold has been the refuge for buyers during times of economic and political turmoil, imparting stability while capital markets fail to offer fine returns. Its inherent worth and worldwide acknowledgement in diverse markets establish it as a one-of-a-kind asset for maintaining wealth, not like fiat currencies that can lessen in value because of inflation or depreciation. In India, gold holds not handiest economic fee but also great cultural and conventional significance, with purchases peaking at some point of weddings, fairs, and spiritual ceremonies. This cultural call for creates seasonal fluctuations in gold fees, including some other layer of complexity to forecasting [1].

Unlike most commodities, the delivery of gold has been constructed up over centuries, and new annual manufacturing provides handiest a small fraction to the overall reserves. which means gold costs are driven extra by using call for-facet factors and investor sentiment than by way of quick-time period changes in supply [2]. As each a commodity and a financial asset, gold's price dynamics are motivated by means of a wide variety of interconnected variables, such as inflation quotes, interest price regulations, fluctuations in global currencies, crude oil prices, and geopolitical uncertainties. as an instance, a rise in inflation regularly leads traders to shift budget into gold as a hedge, whilst an appreciating US greenback generally exerts downward pressure on gold prices due to its inverse dating with the foreign money. in the identical manner, worldwide emergencies, political turmoil, and fluctuations in monetary markets often result in considerable surges in gold call for. because of this intricacy, predicting gold charges continues to be a hard endeavour. Traditional statistical models which include easy transferring common (SMA), Exponential shifting common (EMA), and ARIMA have been broadly used but regularly fall short in accuracy due to their incapability to seize nonlinear dependencies and surprising shifts in marketplace dynamics. those models anticipate linear relationships and stationarity, situations not often happy in real-international financial facts. furthermore, they are limited in managing high-dimensional datasets with multiple correlated predictors, which are important for information gold's price conduct. To cope with those obstacles, researchers and practitioners have increasingly become system gaining knowledge of (ML) strategies. ML fashions are nicely suited for monetary forecasting due to the fact they can seize nonlinear relationships, interactions among variables, and adapt to converting patterns in records. for example, lasso regression, a linear version with L1L_1L1 regularization, is mainly beneficial for characteristic selection and interpretability. through shrinking much less relevant coefficients to zero, Lasso highlights the maximum influential predictors inclusive of stock indices, forex quotes, and inflation whilst decreasing the noise from vulnerable variables. even though its predictive overall performance can be weaker in comparison to nonlinear techniques, Lasso provides treasured insights into the financial drivers of gold charges, making it a sturdy tool for interpretability and dimensionality reduction. Conversely, tree-primarily based ensemble strategies, especially extreme Gradient Boosting (XG Boost), have end up widely recognized for their terrific predictive capabilities in both regression and classification issues. XG Boost builds an ensemble of selection timber sequentially, the use of each gradient and second-order derivatives (Hessians) of the loss feature to enhance optimization. via the use of sturdy regularization strategies (L1 and L2), it lowers the chance of overfitting and enhances balance while handling fluctuating and high-dimensional datasets, along with the ones observed in monetary time series. Its capability to seize nonlinear relationships and interactions among macroeconomic variables frequently ends in advanced forecasting overall performance in comparison to traditional models. moreover, XG Boost offers scores of feature significance, allowing researchers to confirm which factors have the finest impact on fluctuations in gold costs. The real-international implications of forecasting gold fees reach beyond the realm of instructional studies. For investors and fund managers, correct predictions assist in portfolio diversification and chance management. For policymakers and important banks, gold charge tendency's function signs of monetary stability, inflation expectancies, and foreign money risks In India, wherein cultural traditions heavily impact gold demand, dependable forecasting models can offer valuable insights into seasonal consumption developments and guide trade policy. Considering those issues, this examine aims to investigate the relationship among gold fees and macroeconomic variables the usage of system learning techniques. mainly, we apply and evaluate 3 processes: Linear Regression, Random woodland Regression, and Gradient Boosting Regression (XG Boost). via the evaluation in their predictive accuracy and interpretability throughout numerous eventualities, we goal to determine the most-green model for forecasting gold prices. This comparative evaluation will It This research will enhance the modern-day understanding of financial forecasting and provide realistic insights for investors, portfolio managers, and policymakers to recognize the intricacies of gold as both a treasured asset and a financial investment.

2. RELATED WORKS

2.1 Forecasting Gold Prices Using Multiple Linear Regression Method

They've hired the multiple Linear Regression (MLR) model for forecasting gold fees and concluded that the MLR framework proved to be a useful and reliable tool for quick-time period prediction. An examination of previous studies suggests that more than one Linear Regression (MLR) is constantly carried out as a number one statistical technique to explore the relationships among dependent and independent variables, at the same time as additionally generating interpretable results that useful resource in significant evaluation. coefficients to guide knowledgeable decision-making. The approach is computationally efficient, easy to put into effect, and serves as a sturdy baseline model in financial forecasting. but, notwithstanding its blessings, MLR additionally faces boundaries, because it cannot appropriately capture complex non-linear styles, interaction outcomes, or the excessive volatility typically determined in gold fee actions [[3]].

2.2 Gold Price Forecasting Using Arima Model.

Gold is a relatively valued metallic with superb importance as a monetary asset, decorative fabric, and investment opportunity. Its regular attraction stems from its capacity to act as a hedge in opposition to Amid inflation and monetary uncertainty, gold has end up a desired funding desire international. nonetheless, its charges continue to be incredibly risky, fluctuating continuously beneath the effect of different factors along with global market dynamics, forex moves, and macroeconomic conditions. This examine seeks to are expecting gold expenses via the auto Regressive incorporated moving average (ARIMA) model, a extensively identified statistical approach for time-series evaluation. historic gold rate statistics have been utilized to aid the development of the version. It applied to construct and examined the forecasting version. The findings highlight the model's effectiveness in identifying charge styles and turning in dependable short-term forecasts. those forecasts play a vital position for investors, policymakers, and economic analysts in making well-informed decisions, lowering dangers, and improving investment strategies [[4]]

2.3 Gold Price Prediction Using Ensemble Based Machine Learning Techniques.

This research examines how gold costs are inspired via key macroeconomic and monetary variables together with stock market performance, crude oil prices, the rupee–dollar trade charge, inflation, and hobby fees. month-to-month facts protecting the length January 2000 to December 2018 have been accumulated and divided into two awesome sub-periods: duration I (January 2000 – October 2011), during which gold charges exhibited a regular upward trend, and duration II (November 2011 – December 2018), whilst gold fees displayed a quite flat pattern. to investigate and forecast gold price actions, 3 system studying algorithms had been hired: Linear Regression, Random wooded area Regression, and Gradient Boosting Regression. The findings monitor that the connection between gold costs and the chosen variables is strong in period I however vulnerable in length II, indicating that structural modifications in worldwide and domestic markets influence the stability of predictive relationships through the years. The model evaluation suggests that even as all 3 techniques finished strongly within the early phase, their predictive accuracy weakened within the later degree Over the complete observe period, Random wooded area Regression verified the highest accuracy, at the same time as Gradient Boosting Regression outperformed whilst forecasts were evaluated throughout individual sub-intervals. these insights underscore the important function of selecting suitable fashions and segmenting time horizons in commodity rate forecasting, in addition suggesting that advanced ensemble getting to know strategies can beautify predictive reliability in uncertain and fluctuating market situations unstable marketplace and structurally moving markets [[5]].

2.4 Forecasting Gold Price Changes: Application of an Equipped Artificial Neural Network

The primary situation lies in fluctuating costs, which makes it important to establish a clear framework for economic planning. powerful financial management depends on a sturdy selection-making framework supported by using expert guidance Gold has widely appeared as a flexible funding, valued for its stability and numerous ranges of packages. on the way to lessen versions in swelling, lead representatives make use of gold as a switch for dealing with charges. consequently, more information approximately the business will be decided upon with the aid of future gold fee patterns. alternatives. on this piece, we attempt to provide a nicely notion-out model advanced via artificial neural systems (ANNs) that allows you to extend future charges of gold. The counselled sensible system is furnished using a metaheuristic technique referred to as BAT computation to Make ANN capable of handling variations. The based model is as compared with dispensed common sense-based frameworks and different distinguished tactics, consisting of Autoregressive fashions, Multilayer Perceptron (MLP), car-Regressive included moving common (ARIMA), artificial Neural Networks (ANN), and Adaptive Neuro-primarily based models. Fuzzy Inference system (ANFIS) (MLP) Neural network, Radial basis feature (RBF) Neural Networks for Generalized Regression and Networking Root suggest Squared (GRNN). therefore, to assess the model's performance, and blunders listing was generated the usage of the root imply rectangular errors (RMSE) metric. The outcomes are shown below. that the counselled BAT Neural community (BNN) outperforms each traditional and modern-day estimation fashions [[6]].

2.5 Deep Learning With Long Short-Term Memory Networks For Financial Market Predictions.

Long Short-Term Reminiscence (LSTM) networks, appeared as one of the maximum superior processes for time-collection forecasting, have validated extraordinarily effective in predicting gold charges by modelling both brief- and long-term LSTM correctly captures those dependencies, and its integration with complementary architectures, together with convolutional layers, can notably beautify forecasting accuracy however, few research have examined the underlying relationships among influencing variables and gold prices. This has a look at proposes a hybrid framework that integrates association rule mining with LSTM, where association policies screen hidden relationships amongst macroeconomic variables, and the LSTM community models their temporal dynamics to enhance forecasting accuracy to improve prediction accuracy and interpretability [[7]]

2.6 Gold Price Forecast Based on Lstm-Cnn Model

Long Short-Term Reminiscence (LSTM) and Convolutional Neural community (CNN) fashions for forecasting gold expenses and discovered that those deep mastering techniques supply superior performance in comparison to traditional techniques. LSTM is mainly effective in capturing lengthy-term temporal dependencies, making it suitable for sequential monetary data, while CNN excels at extracting hidden neighbourhood capabilities and decreasing noise from huge datasets. In a hybrid CNN–LSTM framework, the 2 techniques supplement one another, delivering advanced accuracy and robustness in unstable marketplace conditions. nevertheless, these fashions face sure obstacles, including the want for huge education datasets, excessive computational needs, and susceptibility to overfitting if not well optimized. regardless of its challenges, CNN–LSTM stays one of the maximum promising techniques for predicting gold charges [[8]].

3. METHODOLOGY

3.1 Proposed System

The proposed framework integrates LASSO Regression with XG Boost, leveraging their strengths to build a greater correct and efficient model for gold rate prediction Gold is quite stimulated by a diffusion of macroeconomic and financial factors such as currency trading charges, crude oil prices, inventory indices, inflation, and hobby prices figuring out key predictors performs an important role in ensuring the reliability of a forecasting version. In the first level, LASSO regression is implemented for characteristic choice and dimensionality reduction. In evaluation to standard regression methods that regularly hold all variables regardless of their significance, LASSO applies L1 regularization, penalizing fewer essential coefficients with the aid of shrinking them toward 0. This ensures that only the most influential elements remain within the model, while noise and redundant information are removed. Such feature optimization now not simplest reduces computational complexity however additionally improves version interpretability, allowing monetary analysts to better understand the drivers of gold price fluctuations. In the second degree, the optimized characteristic set is input into XG Boost (severe Gradient Boosting), a strong ensemble mastering set of rules. XG Boost builds several vulnerable newbies as selection trees and combines them to generate a strong and correct predictive model. Its capability to capture non-linear relationships, variable interactions, and market volatility makes it fantastically suitable for monetary forecasting responsibilities. additionally, XG Boost carries strategies including regularization, tree pruning, and parallel computation, which decorate generalization potential and decrease the chance of overfitting with the aid of integrating LASSO's function selection with XG Boost's predictive capability, the hybrid framework achieves a most effective balance between interpretability and accuracy. This machine no longer best provides reliable forecasts of gold expenses however additionally offers precious insights for informed monetary choice-making, however additionally gives sensible decision-aid insights for traders, buyers, and policymakers. This makes it a advanced opportunity to standard statistical fashions and standalone gadget mastering tactics in risky financial markets

Figure 1. Block diagram of proposed system

3.2 Dataset

The Proposed device has a dataset is a day by daytime-collection file containing closing expenses for key economic devices from January 01, 2000, to December 31,2024 become amassed from Yahoo Finance. "The dataset consists of 4 main columns: Gold Spot (GS), West Texas Intermediate Crude Oil (WTI), the us dollar Index (DXY), and the Dow Jones industrial average (DJI). additionally, the header includes columns for macroeconomic signs—the Federal price range charge (FED) and an Inflation metric (INF)—which are gift but include no information these are completely empty and could need to be handled as lacking values or populated from a separate source The dataset is supplied in a preferred CSV layout with a header row, making it appropriate for macroeconomic analysis and exploring correlations amongst commodities, foreign money indices, and equity markets. For integration, the machine must correctly parse the Date column and apply validation assessments to make certain the accuracy of lively fee data, at the same time as correctly dealing with null values and capacity inconsistencies values in `FED` and `INF`.

Date	Gold ($)	WTI Oil ($)	DXY	Dow Jones
2021-01-04	$1,941.50	$47.62	89.92	30,223.89
2021-02-01	$1,848.90	$52.84	91.05	302,211.91
2021-03-01	$1,708.20	$59.16	92.18	32,619.48
2021-04-01	$1,768.20	$61.45	92.64	33,273.96
2021-05-03	$1,792.80	$64.49	91.25	34,113.23
2021-06-01	$1,905.40	$67.72	89.73	34,575.31
2021-07-01	$1,816.80	$73.47	92.21	34,633.53
2021-08-02	$1,813.60	$71.26	92.06	34,838.16
2021-09-01	$1,816.00	$68.51	92.51	35,312.53
2021-10-01	$1,762.50	$75.03	93.98	34,326.46

Table 1. Ten-day sample data of gold spot (GS) with three attributes.

3.3 Methods

The proposed machine makes use of Lasso regression for feature selection and shrinkage of decided on capabilities from the dataset gathered and XG increase model is being trained to expect the gold fee.

3.3.1 Lasso regression

LASSO, quick for Least Absolute Shrinkage and selection Operator, marks a sizeable advancement in regression evaluation, evolved to enhance prediction accuracy whilst keeping model interpretability. in high-dimensional datasets. This approach fundamentally augments the conventional least square's objective function with the aid of incorporating a regularization period proportional to the sum of the absolute magnitudes of the model coefficients, a constraint called the L1 penalty. The advent of this penalty induces a shape of mathematical stress that compels the envisioned coefficients for much less impactful or redundant predictor variables to contract substantially in the direction state modern the origin. Its most outstanding and valuable characteristic, however, is its precise capability to push those coefficients to an actual cost modern zero, a belonging no longer shared with its cousin, Ridge regression. This mechanism effectively executes automated function selection by way of systematically identifying and absolutely putting off beside the point features from the final predictive equation, thereby streamlining the model structure. The depth present day this regularization impact is governed by using a vital hyperparameter, lambda (λ), whose gold standard fee is meticulously determined through computational techniques like go-validation to ensure the version generalizes well to unseen facts. The results cutting-edge this method is a cultured, sparse, and especially interpretable model that reduces the hazard ultra-modern overfitting whilst presenting clean insights into the effect contemporary predictor the important thing drivers modern the results variable, making it an integral device for statisticians and information scientists throughout diverse fields.

3.3.2 Lasso As An Algorithm

Tuning Parameter (λ or alpha): this is the most essential parameter, controlling the power state-of-the-art the regularization. A higher λ price applies a greater penalty, forcing extra coefficients to reduce modern-day 0 and simplifying the model. A λ trendy 0 consequences in a well-known linear regression. L1 Penalty time-period: The centre mechanism is the Σ|β_j| term introduced to the loss feature. that is the sum state-of-the-art the absolute values contemporary the coefficients. the nature cutting-edge this L1 penalty is what lets in coefficients to be compelled to exactly 0, permitting function selection. Coefficient Shrinkage: The parameters (coefficients, β_j) are not predicted freely; they're limited by using the L1 penalty. The optimization process finds the set modern coefficients that decrease the whole errors while also minimizing the sum in their absolute values. premiere Lambda selection: the suitable λ parameter is not guessed however is chosen through move-validation. The set of rules is matching multiple instances with different λ values, and the one that produces the version with the bottom prediction mistakes on held-out validation statistics is chosen. Sparsity Parameter: The effective results latest tuning λ is controlling the model's sparsity—the number brand new functions with non-zero coefficients. The very last output is a parameter vector wherein many entries are 0, indicating the ones features had been excluded.

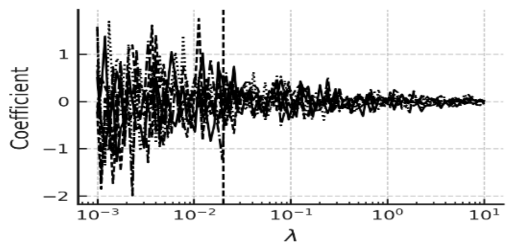

Figure 2. lasso shrinking selected features

The Lasso (Least Absolute Shrinkage and Selection Operator) works by minimizing prediction error while applying an L1penalty on regression coefficients.
The main idea:
Loss = Error (fit to data) +λ × Penalty
(sum of absolute coefficients)

The Main formula used here *is*:

β^lasso=βargmin{2n1i=1∑n(yi−xiTβ)2+λj=1∑p|βj|
(1)

This component expresses Lasso regression, wherein the coefficient vector β is estimated via minimizing prediction error alongside an L1 penalty term. The regularization parameter λ adjusts penalty power, forcing a few coefficients β_j to exactly zero. This yields sparse, interpretable version, reduces overfitting, and complements predictive performance by using selecting only the most relevant functions.

3.3.3 Lasso Parameters

Symbol	Description
λi	Response variable (target)
xi	Feature vector for sample i
β= (β1, β2, βp)	Regression coefficients
n	Number of samples
p	Number of features
λ≥0	Regularization parameter controlling shrinkage

Table.2 Lasso parameters explanation

3.3.4 XG BOOST

XG Boost (severe Gradient Boosting) is a effective and scalable machine cutting-edge set of rules that extends the gradient boosting framework. Gradient boosting is an ensemble approach wherein a couple of decision bushes are built sequentially, with every new tree correcting the errors ultra-modern the previous ones. The goal is to reduce a chosen loss characteristic via level-smart optimization, guided through gradient descent. The XG Boost framework enhances this concept with the aid of defining an objective characteristic that combines prediction blunders with a regularization term:

$$L(\phi) = i = 1\sum nl(yi, y\hat{}i) + k\sum\Omega(fk) \quad (2)$$

where l(yi ,y^i) measures the difference between observed and predicted values, and Ω(fk)_ penalizes model complexity. This balance improves both accuracy and generalization, making XG Boost robust to overfitting. XG Boost introduces numerous improvements over conventional gradient boosting: Regularization: It applies both L1L_1L1 (Lasso) and L2L_2L2 (Ridge) penalties to leaf weights, selling sparsity and controlling complexity. 2nd-order optimization: by using 2d-order derivatives (Hessian) in addition to gradients, it improves convergence and accuracy. Shrinkage (modern day rate): Scales tree contributions to keep away from overfitting and allow sluggish getting to know modern. Column subsampling: Randomly selects subsets present day functions for every tree, increasing variety and lowering variance. Sparsity coping with correctly manages lacking values by means of routinely modern-day default split guidelines. Parallelization and scalability help multi-middle CPUs, GPUs, and dispensed computing, enabling education on very massive datasets. These capabilities make XG Boost extremely rapid and surprisingly correct in comparison to conventional boosting implementations. It has received extensive adoption in studies and enterprise today's its flexibility and performance. XG Boost performs distinctly well on established or tabular information and helps obligations which includes regression, class, and ranking modern-day its reliability and performance, XG Boost has end up the set of rules state-of-the-art choice in lots of actual-world packages, which include finance, healthcare, bioinformatics, recommendation systems, and textual content mining. it's also a dominant technique in machine brand new competitions (along with Kaggle), where it always produces state of art of results. In precis, XG Boost isn't always just an extension modern day gradient boosting; it's far a noticeably optimized gadget that mixes robust regularization, efficient computation, and scalability. Its capacity to handle overfitting, manipulate lacking statistics, and deliver each accuracy and pace explains its reputation as one of the maximum impactful algorithms in applied Machine learning.

3.3.5 XGBOOST In Gold Price Prediction

XG Boost is an optimized gradient boosting framework that builds an ensemble modern day decision bush in a sequential way. For gold prediction, the procedure begins with facts collection and preprocessing, in which historic gold costs are combined with applicable financial indicators and macroeconomic variables. function engineering performs a crucial position, regarding the creation latest lag values, moving averages, volatility indices, and diverse different technical signs from the uncooked dataset. those features permit the model to capture each quick-time period developments and lengthy-time period dependencies in the gold market. at some stage in model education, XG Boost applies each first-order gradients and second-order derivatives (Hessians) state-of-the-art the loss feature, permitting green optimization. Regularization phrases (L1 and L2) are incorporated to penalize model complexity, helping to mitigate overfitting a common venture in economic time-series forecasting. furthermore, adjusting hyperparameters together with gaining knowledge state modern price, maximum tree depth, and the variety modern-day estimators permits XG Boost to strike an most suitable stability between predictive accuracy and generalization After schooling, the model can generate continuous gold charge forecasts (regression) or categorical predictions, inclusive of the course modern day charge moves (type). version performance is evaluated the usage of widespread blunders metrics, such as Root mean square mistakes (RMSE), imply Absolute blunders (MAE), or suggest Absolute percent errors (MAPE) Empirical research show that XG Boost modern-day outperforms conventional econometric models and different device present day algorithms today's its capacity to address nonlinear relationships, manage missing values, and effectively technique huge-scale datasets." correctly. In end, XG Boost offers a effective framework for gold fee prediction, combining performance, accuracy, and flexibility. by means of incorporating a wide variety contemporary monetary indicator and making use of rigorous regularization, it affords superior forecasting overall performance, which may be precious for buyers, policymakers, and financial establishments in decision-making.

3.3.6 XGBOOST Workflow

1. Data Collection: Historical gold prices are collected along with relevant features such as interest rates, inflation rates, stock indices, crude oil prices, USD exchange rates, and global economic indicators.
2. Feature Engineering: Time-series features like lag values, moving averages, volatility, and momentum indicators are created Macroeconomic variables and market sentiment indexes may also be included.
3. Model Training: XG Boost is trained with gold price (or returns) as the target variable It iteratively builds decision trees to minimize the prediction error using both gradients and Hessians. Regularization (L1, L2L_1, L_2L1, L2) helps control overfitting, which is crucial in volatile financial data.
4. Prediction: The trained model outputs short-term or long-term forecasts of gold prices. Predictions can be continuous (price regression) or categorical (e.g., price will go *up* or *down*).
5. Evaluation: Metrics such as RMSE, MAE, or MAPE are used to evaluate forecast accuracy. Compared to traditional statistical models, XG Boost typically achieves better performance due to its ability to capture nonlinear patterns.

3.3.7 XGBOOST Training Procedure

The algorithm starts with the aid of education on (X, y)(X, y)(X, y), in which X represents the function matrix and y denotes the target variable, with several critical hyperparameters: the modern-day price (η)(eta)(η), which controls the contribution cutting-edge newly delivered tree; regularization parameters (λ,α)(lambda, alpha)(λ,α), which lessen overfitting by way of penalizing overly complex models; and the maximum number modern-day trees (T)(T)(T), which constrains the wide variety latest boosting iterations to avoid excessive computation or overfitting. The version is first initialized with a baseline prediction y ^ (0) hat{y} ^ {(0)} y ^ (zero)—normally the imply state-of-the-art goal values for regression responsibilities or the log-odds for type obligations.

Education proceeds iteratively over TTT boosting rounds. In every round, the gradient gi(t)g_i^{(t)}gi(t) is computed to quantify how a whole lot each prediction ought to be adjusted to reduce the loss, at the same time as the Hessian (2d by-product) captures the curvature today's the loss function, allowing more stable and accurate optimization. primarily based on this information, a brand-new selection tree is constructed to decide the most fulfilling course and magnitude brand new corrections. To save you overfitting and make certain generalization, each L1 (lasso) and L2 (ridge) regularization are implemented to cut back leaf weights, controlling tree complexity. Leaf nodes are assigned ratings based on a closed-shape solution derived from the gradients, Hessians, and regularization terms, making the replace method green. A state-of-the-art rate (η)(eta)(η) scales the contribution present day tree, ensuring gradual development rather than abrupt adjustments.

As education progresses, the model aggregates vulnerable freshmen in an additive fashion, with every successive tree improving on the residual errors ultra-modern the previous ensemble. extra techniques consisting of column subsampling (deciding on random subsets cutting-edge features according to split), row subsampling (randomly sampling training instances according to tree), and tree intensity constraints similarly enhance robustness and decrease computation. After TTT iterations, the very last version is the sum latest all choice bushes, accomplishing high predictive accuracy by balancing bias and variance even as keeping scalability to huge datasets.

4. RESULTS AND DISCUSSION

In this study, we employed Lasso regression and Extreme Gradient Boosting (XG Boost) to forecast gold prices using macroeconomic indicators such as the US Dollar Index (DXY), West Texas Intermediate (WTI) crude oil prices, Dow Jones Industrial Average (DJI), Federal Reserve (FED) rates, and inflation. Our findings are compared to the existing work by Boongasame.et.al. (2022) [9], where an LSTM-based framework with association rules (LSTM-GS-DXY) achieved superior performance over traditional statistical approaches including SMA, EMA, WMA, and ARIMA.

4.1 LASSO REGRESSION

The Lasso regression model applied an L1L_1L1 penalty to shrink coefficients of less influential features toward zero, effectively performing feature selection. The results showed that DXY and inflation were consistently retained as significant predictors, while FED and DJI were occasionally shrunk to negligible values depending on the forecast horizon. The interpretability of Lasso was an advantage, as it clearly highlighted the strongest drivers of gold price fluctuations.

However, due to its linear nature, Lasso was unable to fully capture the nonlinear dependencies and complex interactions among macroeconomic variables. While its forecasting error was lower than that of baseline models such as ARIMA and SMA, it remained higher than XG Boost and LSTM, particularly in volatile periods. These results suggest that Lasso is best suited as a feature selection tool or for scenarios requiring model transparency, rather than as a standalone forecasting method.

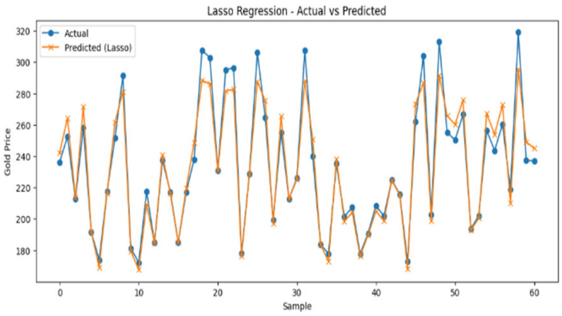

Figure 3. line graph representing the actual and predicated values using Lasso

4.2 XG BOOST

In contrast, the XG Boost model achieved substantially lower error rates (MAPE and RMSE) compared to Lasso. By leveraging gradient boosting with both first-order gradients and second-order Hessians, XG Boost captured nonlinear relationships and interactions across features. The inclusion of regularization (L1L_1L1 and L2L_2L2) prevented overfitting, even when trained on high-dimensional macroeconomic data.

The feature importance analysis from XG Boost reinforced the conclusions from the Lasso model and the association rules identified in the reference paper: DXY consistently emerged as the most influential factor in gold price prediction, followed by WTI and inflation. Interestingly, DJI showed moderate importance, reflecting its indirect impact via global market sentiment. FED rates contributed less significantly, aligning with prior studies that highlighted their long-term rather than short-term influence on gold. XG Boost demonstrated predictive accuracy comparable to or exceeding the LSTM-GS-DXY model reported in the reference study, particularly for short-term horizons. While LSTM excelled in capturing sequential dependencies in long time windows, XG Boost offered faster training times, easier interpretability through feature importance scores, and competitive accuracy.

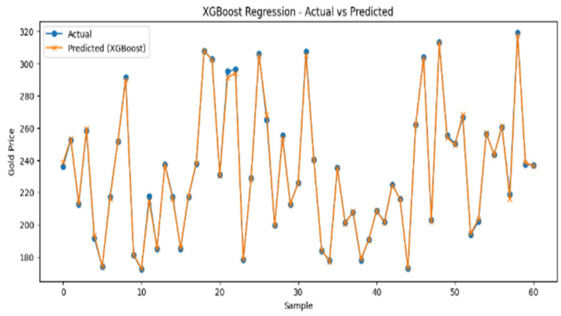

Figure 4. Line graph representing actual and predicated values using XGBOOST

Comparative Insight

Relative to Boongasame.et al. (2022), our results confirm that advanced machine learning models outperform traditional linear forecasting methods. While the LSTM-GS-DXY approach achieved the lowest error among baseline techniques in the reference study, our findings suggest that XG Boost is a viable alternative to deep learning, offering accuracy on par with LSTM but with greater computational efficiency and interpretability. Lasso, although less accurate, remains valuable in identifying the most relevant predictors and reducing dimensionality before applying nonlinear models. Lasso regression enhances interpretability and identifies DXY, WTI, and inflation as dominant predictors of gold price. XG Boost achieves high forecasting accuracy, rivaling LSTM while being computationally more efficient. A hybrid framework where Lasso is used for feature selection and XG Boost for prediction could combine the strengths of both methods, balancing interpretability with accuracy. Thus, in line with the reference study, our results demonstrate the effectiveness of machine learning in gold price forecasting, while providing evidence that ensemble-based methods like XG Boost can serve as strong alternatives to deep learning architectures in financial prediction tasks.

5.CONCLUSION AND FUTURE WORK

This study examined the application of Lasso regression and Extreme Gradient Boosting (XG Boost) for gold price forecasting using key macroeconomic indicators such as the US Dollar Index (DXY), WTI crude oil prices, inflation, DJI, and FED rates. The results revealed that Lasso, through its L1L_1L1 regularization, effectively identified the most influential variables, particularly DXY, WTI, and inflation, thereby enhancing interpretability and reducing feature redundancy. However, due to its linear nature, Lasso showed limited forecasting accuracy compared to more advanced models. In contrast, XG Boost provided significantly higher predictive performance by capturing nonlinear relationships and feature interactions, while its regularization mechanisms prevented overfitting. Moreover, XG Boost's feature importance measures supported the association rule findings reported in previous studies, reaffirming the central role of DXY in gold price fluctuations. Overall, while Lasso proved valuable for feature selection and transparency, XG Boost demonstrated strong robustness and accuracy, making it a practical alternative to deep learning methods such as LSTM. Future research can further enhance these findings by developing hybrid frameworks that integrate Lasso's feature selection with advanced predictors such as XG Boost or LSTM, thereby balancing interpretability and accuracy. Expanding the feature space to include sentiment indicators, geopolitical risk indices, or cryptocurrency trends may also strengthen forecasting performance by capturing wider market influences. Additionally, the exploration of deep hybrid models that combine tree-based learning with recurrent neural networks could allow simultaneous exploitation of nonlinear interactions and temporal dependencies. Another promising direction is the implementation of real-time forecasting systems using streaming data, which would be particularly beneficial for investors and policymakers. Finally, future studies should conduct robust evaluations under global shocks including financial crises, pandemics, and geopolitical tensions—to ensure stability and reliability of predictive frameworks in highly volatile conditions.

References

[1] Du, W., & Schreger, J. (2016). Local currency sovereign risk. The Journal of Finance, 71(3), 1027-1070. J. Jagerson and S. W. Hansen, "All about investing in gold", McGraw-Hill Publishing, 2011.

[2] Ismail, Z., Yahya, A., & Shabri, A. (2009). Forecasting gold prices using multiple linear regression method. American Journal of Applied Sciences, 6(8), 1509.Mrs. B. Kishori 1, V. Preethi, "Gold Price forecasting using ARIMA Model", International Journal of Research, 2018.

[3] Manjula, K. A., & Karthikeyan, P. (2019, April). Gold price prediction using ensemble-based machine learning techniques. In 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1360-1364). IEEE.

[4] Hafezi, R., & Akhavan, A. (2018). Forecasting gold price changes: Application of an equipped artificial neural network. AUT Journal of Modelling and Simulation, 50(1), 71-82.

[5] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European journal of operational research, 270(2), 654-669.

[6] Zhanhong He, Junhao Zhou", Hong-Ning Dai*, Hao Wang¹ Macau University of Science and Technology.2019*

[7] Boongasame, L., Viriyaphol, P., Tassanavipas, K., & Temdee, P. (2023). Gold-price forecasting method using long short-term memory and the association rule. Journal of Mobile Multimedia, 19(1), 165-186Tully E, Lucey BM. A power GARCH examination of the gold market. Research in International Business and Finance. 2007 Jun 1;21(2):316–25.

[8] Boongasame, L., Viriyaphol, P., Tassanavipas, K., & Temdee, P. (2023). Gold-price forecasting method using long short-term memory and the association rule. Journal of Mobile Multimedia, 19(1), 165-186.

[9] West Texas Intermediate (WTI). (n.d.) Investopedia. Retrieved March 31, 2021, from https://www.investopedia.com/terms/w/wti.asp.

[10] Who or What Is Dow Jones? (n.d.) Investopedia. Retrieved March 31, 2021, from https://www.investopedia.com/ask/answers/who-or-what-i s-dow-jones/.

[11] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288.

[12] XGBoost Algorithm and its Comparative Analysis": published in the International Journal of Novel Research and Development (IJNRD), Volume 7, Issue 12, in December 2022.