**GOLD PRICE FORECASTING MODELS:**

**A COMPARATIVE STUDY OF PREDICTIVE ALGORITHMS**



**UPES, DEHRADUN, UTTARAKHAND**

**MASTERS OF BUSINESS ADMINISTRATION**

**(BUSINESS ANALYTICS)**

**2022-24**

SUBMITTED BY:

**ADHERSH B**

**R116222001**

SAP: 500100074

UNDER THE SUPERVISION OF:

**Dr. INDER SINGH**

Associate Professor

SCHOOL OF BUSINESS, UPES, DEHRADUN, UTTARAKHAND

**Student Declaration**

I hereby declare that the dissertation work entitled “Gold Price Forecasting Models:

A Comparative Study of Predictive Algorithms” submitted to the School of Business,

University of petroleum and energystudies is a record of original work done by me under the guidance of Associate Professor Dr. Inder Singh.

The information and data given in the report are authentic to the best of my knowledge. This

Dissertation is not submitted to any institution for the award of any degree, diploma, or fellowship or published at any time before. This information is pure of academic interest

Adhersh B

SAP ID: 500100074

Enrollment Number: R116222001

MBA (Business Analytics) 2022-24

School of Business, UPES

**Acknowledgement**

Reflecting on the completion of my academic journey, I want to express my gratitude to those who significantly contributed to my dissertation. I wish to extend my sincere gratitude to

Dr. Inder Singh, my mentor, whose guidance and expertise have been invaluable throughout this academic journey. His insights, encouragement, and unwavering support have played a pivotal role in shaping the direction and quality of this research.

Furthermore, I would like to express my appreciation to Mr. Rahul Nainwal, Dean of the School of Business at UPES. Under his leadership, the academic environment at UPES has been conducive to intellectual growth and exploration. Also, I would like to thank those who directly or indirectly helped and guided me to complete our dissertation in the shed, including faculty of the School of Business, UPES.

Finally, I am deeply appreciative of all my colleagues and friends who provided encouragement, support, and constructive feedback during the course of this project.

Every individual mentioned above has played a crucial role in making this report possible, and I extend my heartfelt appreciation to each one of them for their contributions.

**Mentor Agreement Form for Dissertation**

I Adhersh, student of MBA (Business Analytics) Semester IV, with Enrolment number

R116222001and SAP ID 500100074 has undertaken dissertation in the topic “Gold Price Forecasting Models: Comparative Study of Predictive Algorithms” under the guidance of

Dr. Inder Singh and will submit the final Dissertation latest by 31st of March 2024.

Student Name: Adhersh B Mentor Name: Dr. Inder Singh

Signature: Signature:

**Table of Content**

|  |  |  |
| --- | --- | --- |
| Chapter No. | Particular | Page No. |
| 1 | **Part 1: Introduction and Background** |  |
| 1.1 Introduction to Gold Price Forecasting Models |  |
| 1.2 Importance of Gold Price Prediction in Financial Markets |  |
| 1.3 Overview of Predictive Algorithms in Forecasting |  |
| 1.4 Objectives of the Study |  |
| 1.5 Scope and Limitations |  |
| Part 2: Research Work | | |
|  | 2.1 Literature review |  |
| 2.2 Gaps in Existing Literature |  |
| 2.3 Research Objective |  |
| 3 | 3.1 Research Methodology |  |
| 3.3 Research Design |  |
| 3.4 Data Collection Sources |  |
| 3.5 Data Preprocessing |  |
| 3.6 Predictive Algorithms Under Study |  |
| 3.7 Evaluation Metrics |  |
| 4 | Data Analysis |  |
| 5 | Results and Discussion |  |
| 6 | Conclusion |  |
| 7 | References |  |

**Tables and Figures**

|  |  |
| --- | --- |
| Tables | Figures |
| Table 1 | Figure 1 |
| Table 2 | Figure 2 |
| Table 3 | Figure 3 |
| Table 4 | Figure 4 |
| Table 5 | Figure 5 |
| Table 6 | Figure 6 |
| Table 7 | Figure 7 |
| Table 8 | Figure 8 |
|  | Figure 9 |
|  | Figure 10 |
|  | Figure 11 |
|  | Figure 12 |
|  | Figure 13 |
|  | Figure 14 |
|  | Figure 15 |
|  | Figure 16 |
|  | Figure 17 |
|  | Figure 18 |
|  | Figure 19 |

**Abstract**

The forecasting of gold prices plays a pivotal role in the global financial landscape, influencing investment decisions, risk management strategies, and policy formulations. With the increasing volatility and complexity of financial markets, the demand for accurate and reliable gold price forecasting models has surged. This dissertation presents a comprehensive comparative study of predictive algorithms used in gold price forecasting, aiming to evaluate their effectiveness, accuracy, and applicability across various market conditions.

The study begins with an introduction to the significance of gold as a financial asset and the importance of accurate price predictions in today's dynamic economic environment. A thorough literature review explores existing gold price forecasting models, highlighting their methodologies, strengths, and limitations. The research methodology section outlines the data sources, preprocessing techniques, predictive algorithms under study, and evaluation metrics employed to assess the performance of each model.

Through rigorous data analysis and comparative evaluation, this dissertation identifies the strengths and weaknesses of different forecasting models, offering insights into their predictive capabilities and reliability. The findings contribute to the existing body of knowledge on gold price forecasting, providing valuable guidance for investors, financial analysts, and policymakers.

The comparative study aims to bridge the gap between theory and practice, offering actionable insights that can inform investment strategies, risk management approaches, and policy decisions related to gold investments. The research underscores the importance of selecting appropriate predictive algorithms based on market conditions, data quality, and forecasting objectives, thereby enhancing the accuracy and reliability of gold price predictions.

**1.1 Introduction to Gold Price Forecasting Models**

The volatility of gold prices has long captured the attention of investors and analysts alike. This precious metal, revered for its intrinsic value and aesthetic appeal, plays a dual role as both a financial asset and a cornerstone in the jewellery industry. The recent fluctuations in gold prices have sparked renewed interest and speculation, prompting a closer examination of the underlying economic factors driving these changes.

Gold's allure as a safe-haven asset during times of economic uncertainty and inflationary pressures is well-documented. Investors often flock to gold as a hedge against currency devaluation and market volatility, contributing to its price resilience and demand. Additionally, the intricate dynamics between global economic indicators, geopolitical tensions, and monetary policies further complicate the landscape, influencing gold's price trajectory.

While the fluctuation in gold prices may seem unpredictable to the untrained eye, there exists a rich body of literature exploring the economic rationale behind these movements. Researchers have delved into the complex interplay of supply and demand factors, macroeconomic indicators, and market sentiments shaping gold prices.

One prevalent approach to deciphering the economic logic behind gold price fluctuations involves constructing a causal model. In this framework, gold prices are treated as the dependent variable, influenced by a set of independent variables representing potential causal factors. These factors, grounded in existing research, can range from interest rates and inflation expectations to geopolitical stability and investor sentiment. Employing multiple regression analysis, researchers can quantify the relationship between these variables and gold prices, offering insights into their predictive power.

However, the use of regression techniques in forecasting gold prices is not without challenges. Issues such as multicollinearity, heteroscedasticity, and autocorrelation can distort results and compromise the accuracy of predictions.

The primary objective of this study is to explore existing literature to identify relevant factors that can serve as independent variables for forecasting gold prices. By employing multiple regression techniques and addressing statistical pitfalls, this research aims to develop a predictive model capable of estimating future gold prices accurately. Through rigorous analysis and comparison with actual market data, this study seeks to shed light on the underlying economic forces driving gold price fluctuations and provide valuable insights for investors and policymakers navigating the intricate world of precious metal markets.

**1.2 Importance of Gold Price Prediction in Financial Markets**

Gold, often referred to as the "ultimate store of value," holds a unique position in the global financial ecosystem. Its price movements are influenced by a myriad of factors, including macroeconomic indicators, geopolitical events, currency fluctuations, and investor sentiment. Given its significance, accurate prediction of gold prices is crucial for various stakeholders in the financial markets. Below are some key reasons highlighting the importance of gold price prediction:

Risk Management:

Accurate gold price forecasting enables investors, fund managers, and financial institutions to develop effective risk management strategies. By anticipating potential price movements, market participants can hedge their exposure to gold price volatility, thereby safeguarding their investments and portfolios against adverse market conditions.

Investment Decisions:

Gold serves as an essential asset class for diversifying investment portfolios and preserving wealth. Accurate price predictions empower investors to make informed decisions regarding asset allocation, timing of investments, and entry/exit points, optimizing returns and minimizing risks.

Policy Formulation:

Central banks and policymakers monitor gold prices closely as part of their monetary policy frameworks. Accurate forecasting of gold prices helps policymakers in formulating effective monetary policies, managing foreign exchange reserves, and stabilizing domestic currencies.

Economic Indicators:

Gold prices often serve as a barometer for economic health and market sentiment. Fluctuations in gold prices can reflect underlying economic trends, inflation expectations, and market confidence. Hence, reliable gold price predictions can provide valuable insights into broader economic conditions and potential market trends.

Trading and Speculation:

For traders and speculators, gold price forecasting is vital for identifying short-term trading opportunities and executing profitable trades. Real-time analysis and accurate predictions enable traders to capitalize on market trends, leverage trading strategies, and maximize profits.

Research and Analysis:

Gold price forecasting models contribute to academic research, financial analysis, and market studies. By evaluating the performance of various predictive algorithms and methodologies, researchers can advance the field of financial forecasting, develop innovative models, and enhance the predictive accuracy of gold price forecasts.

In conclusion, gold price prediction plays a pivotal role in the financial markets, influencing investment decisions, risk management strategies, policy formulations, and market analysis. The ability to forecast gold prices accurately provides stakeholders with a competitive edge, enabling them to navigate the complexities of the global financial landscape more effectively.

**1.3 Overview of Predictive Algorithms in Gold Price Forecasting**

The forecasting of gold prices is a complex task that requires sophisticated analytical tools capable of capturing the multifaceted nature of financial markets. Various predictive algorithms have been employed to model and forecast gold price movements, each with its unique strengths, limitations, and applicability. This section provides a comprehensive overview of some prominent predictive algorithms utilized in gold price forecasting:

1. **Multi-Variable Regression:**
   * Multi-variable regression analysis is a classical statistical method employed to establish a linear relationship between multiple independent variables and a dependent variable. In the context of gold price forecasting, this algorithm enables analysts to quantify the impact of various economic indicators, market sentiments, and geopolitical events on gold prices, providing a structured framework for predictive modelling.
2. **ARIMA (Autoregressive Integrated Moving Average):**
   * The Autoregressive Integrated Moving Average (ARIMA) model is a time series forecasting technique widely adopted in financial forecasting due to its ability to capture temporal dependencies and inherent volatility in data. ARIMA combines autoregressive (AR), differencing (I), and moving average (MA) components to model non-stationary time series data, making it particularly suitable for analysing and predicting gold price dynamics.
3. **SARIMA (Seasonal Autoregressive Integrated Moving Average):**
   * The Seasonal Autoregressive Integrated Moving Average (SARIMA) model extends the ARIMA framework by incorporating seasonal components to account for periodic fluctuations and seasonal variations in time series data. SARIMA is adept at capturing both short-term fluctuations and long-term trends influenced by seasonal factors, making it a valuable tool for modeling and forecasting gold prices.
4. **Random Forest Regressor:**
   * The Random Forest Regressor is an ensemble learning algorithm that leverages a collection of decision trees to make predictions. By aggregating the outputs of multiple decision trees, Random Forest Regressor can handle non-linear relationships and complex interactions between predictors, offering robust and accurate predictions for gold price forecasting.
5. **Long Short-Term Memory (LSTM):**
   * Long Short-Term Memory (LSTM) networks are a type of recurrent neural network (RNN) specifically designed for sequential data and time series forecasting. LSTMs excel at capturing long-term dependencies in time series data, making them highly effective for modeling and predicting gold prices, which exhibit complex temporal patterns and nonlinear dynamics.

In summary, the selection of an appropriate predictive algorithm for gold price forecasting necessitates a nuanced understanding of the data characteristics, underlying market dynamics, and forecasting objectives. Each algorithm offers a unique approach to modeling and prediction, with distinct advantages and challenges. This dissertation aims to rigorously evaluate and compare the performance of these algorithms, providing empirical insights into their effectiveness and applicability in gold price forecasting.

**1.4 Objectives of the Study**

The primary objective of this dissertation is to conduct a comparative study of predictive algorithms employed in gold price forecasting, with the aim of evaluating their effectiveness, accuracy, and reliability in predicting future gold price movements. To achieve this overarching goal, the study is structured around the following specific objectives:

1. To Review Existing Literature:
   * To provide a comprehensive review of the existing literature on gold price forecasting models, focusing on the methodologies, techniques, and algorithms utilized in previous studies.
2. To Identify Key Predictive Algorithms:
   * To identify and select a set of representative predictive algorithms commonly used in gold price forecasting, including Multi-Variable Regression, ARIMA, SARIMA, Random Forest Regression, and Long Short-Term Memory (LSTM).
3. To Collect and Preprocess Data:
   * To collect relevant historical data on gold prices, along with associated economic indicators, market trends, and influencing factors.
   * To preprocess the collected data to ensure consistency, accuracy, and suitability for analysis, including data cleaning, normalization, and feature engineering.
4. To Implement Predictive Models:
   * To implement the selected predictive algorithms using appropriate software tools and programming languages, ensuring proper configuration and parameter tuning for each model.
5. To Evaluate Model Performance:
   * To evaluate the performance of each predictive algorithm based on predefined evaluation metrics, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R-squared value.
   * To compare the predictive accuracy, reliability, and computational efficiency of the algorithms across different forecasting horizons and market conditions.
6. To Interpret and Analyse Results:
   * To interpret the findings from the comparative analysis, identifying the strengths, weaknesses, and limitations of each predictive algorithm.
   * To provide insights into the factors influencing the predictive performance of the algorithms and their implications for gold price forecasting.
7. To Provide Recommendations:
   * To provide actionable recommendations for stakeholders, including investors, financial analysts, and policymakers, on selecting appropriate predictive algorithms for gold price forecasting based on specific requirements and objectives.

By addressing these objectives, this dissertation aims to contribute to the existing body of knowledge on gold price forecasting, offering valuable insights into the comparative performance of predictive algorithms and their applicability in real-world financial forecasting scenarios.

**1.5 Scope and Limitations**

**Scope of the Study**

The scope of this dissertation encompasses a comprehensive comparative analysis of predictive algorithms utilized in gold price forecasting. The study focuses on evaluating the effectiveness, accuracy, and reliability of selected algorithms, namely Multi-Variable Regression, ARIMA, SARIMA, Random Forest Regressor, Gradient Boosting Machine (GBM), and Long Short-Term Memory (LSTM). The research aims to provide empirical insights into the performance of these algorithms across different forecasting horizons and market conditions, thereby aiding stakeholders in making informed decisions related to gold investments and financial planning.

The study will utilize historical data on gold prices, economic indicators, market trends, and influencing factors sourced from reliable financial databases and publications. The data collection and analysis will adhere to rigorous academic standards, ensuring the integrity, validity, and reliability of the research findings.

**Limitations of the Study**

While this dissertation aims to provide a comprehensive comparative analysis of predictive algorithms in gold price forecasting, it is important to acknowledge certain limitations that may impact the scope and generalizability of the findings:

1. **Data Availability and Quality**:
   * The availability and quality of historical data on gold prices and related variables may vary, potentially affecting the accuracy and reliability of the predictive models.
2. **Model Assumptions**:
   * Each predictive algorithm is based on certain assumptions and constraints that may not fully capture the complex dynamics and uncertainties inherent in financial markets.
3. **Parameter Sensitivity**:
   * The performance of machine learning algorithms such as Random Forest Regressor, GBM, and LSTM is sensitive to parameter tuning, which may influence the comparative analysis results.
4. **Market Volatility**:
   * Financial markets, including the gold market, are subject to inherent volatility and external shocks that may impact the predictive accuracy of the models.
5. **Model Interpretability**:
   * Some advanced machine learning algorithms, like LSTM and Random Forest, may lack interpretability, making it challenging to interpret and explain the underlying factors driving the predictions.
6. **Generalizability**:
   * While the study aims to provide valuable insights into the comparative performance of predictive algorithms, the findings may not be universally applicable across all markets and time periods due to varying market conditions and dynamics.

Despite these limitations, this dissertation seeks to contribute meaningfully to the field of gold price forecasting by offering a rigorous comparative analysis of predictive algorithms and highlighting their strengths, weaknesses, and applicability in real-world financial forecasting scenarios. The research findings are intended to serve as a valuable resource for investors, financial analysts, and policymakers in navigating the complexities of the gold market and making informed decisions.

**2.Research Work**

**2. 1 Literature Review**

Gold price forecasting has garnered considerable attention in financial research due to its profound implications for investment strategies and market dynamics. Various methodologies, including regression models and advanced statistical techniques, have been employed to capture the intricate patterns and influencing factors affecting gold prices. This literature review focuses on seminal studies that have utilized regression models and other predictive algorithms to forecast gold prices, laying the groundwork for the comparative analysis undertaken in this dissertation.

**Feng, Y., Liu, X., Zhang, Y. (2021):**

This study compares the performance of the Facebook Prophet algorithm, ARIMA, and Exponential Smoothing (ETS) in forecasting gold prices. Developed by Facebook, Prophet excels in handling time series data and capturing seasonality and trends, requiring minimal hyperparameter tuning but facing challenges with complex external factors .On the other hand, ARIMA and ETS, traditional forecasting models, are adept at capturing linear dependencies, trend, and seasonality but may falter with non-linear patterns and external influences .Comparative research indicates that Prophet outperforms ARIMA and ETS in accuracy and computational efficiency .However, to enhance predictive accuracy, incorporating external variables and exploring ensemble methods have been suggested Despite Prophet's promising performance, its limitations in handling external factors .necessitate caution, highlighting the need for further research to improve forecasting robustness.

[**Mustafa Yurtsever**](https://www.researchgate.net/profile/Mustafa-Yurtsever?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19)**(2021):**

Predicting gold prices is challenging due to its multifactorial and non-linear nature, influenced by various external factors such as market conditions, economic crises, oil prices, tax policies, and interest rates. Multivariate models have been shown to outperform univariate models in this complex environment. This study examined the impact of gold price, crude oil price, exchange rate index, stock market index, and interest indicators from 2001 to 2021. Using LSTM, Bi-LSTM, and GRU methods, models were evaluated based on RMSE, MAPE, and MAE metrics. The LSTM model demonstrated superior performance with MAPE of 3.48, RMSE of 61,728, and MAE of 48.85, indicating its effectiveness in predicting gold prices amidst various influencing factors.

**Abken (1980):**

Abken's study in 1980 stands as one of the pioneering works in gold price forecasting, employing a multiple regression approach to model gold prices based on interest rates and lagged gold prices. Despite the model's modest explanatory power, Abken emphasized the economic rationale behind gold price fluctuations, identifying factors such as political and economic uncertainty, supply-demand dynamics, inflation, and government auction policies as key determinants influencing gold prices.

**D. N, P. R. Prajapati, R. M, R. Ramesh, A. Kodipalli and R. Joy Martis (2021):**

This study employed linear regression, bidirectional LSTM, vanilla LSTM, and stacked LSTM algorithms for gold price forecasting, evaluating their performance using R2, RMSE, MAE, and MAPE metrics. The bidirectional LSTM exhibited the highest error rate, while vanilla and stacked LSTM models outperformed other methods, showcasing their efficacy. Compared to existing literature, the proposed models demonstrated superior accuracy. Despite advancements, accurately forecasting gold prices remains challenging due to unpredictable external factors.

**Koutsoyiannis (1983):**

Koutsoyiannis explored the relationship between gold prices and the U.S. economy, highlighting a negative correlation between the U.S. dollar and gold prices. Using regression analysis, Koutsoyiannis demonstrated that gold prices are more sensitive to the U.S. economic situation rather than global economic factors, providing valuable insights into the underlying dynamics driving gold price movements.

**Vural (2003):**

Vural conducted a comprehensive study in 2003, employing regression analysis to investigate the relationship between gold prices and various causal variables, including USD/Euro parity, Dow Jones industrial production index, oil prices, interest rates, and other commodity prices. The study revealed a negative relationship between gold prices and interest rates, USD/Euro parity, and Dow Jones index, while silver, oil, and copper prices exhibited a positive correlation with gold prices.

**Topçu (2010):**

Topçu examined the relationship between gold prices and several variables, including Dow Jones index, USD exchange rates, oil prices, U.S. inflation rates, and global money supply (M3), using a multivariate regression model. The study identified significant relationships between gold prices and these variables, highlighting the influential role of oil prices and U.S. inflation rates in shaping gold price movements.

**E A Selvanathan (1991):**

In the context of gold price forecasting, Selvanathan demonstrated that a simple ARIMA model was effective and cost-efficient for predicting London daily gold prices. This literature review underscores the ARIMA model's versatility and reliability across various forecasting scenarios, setting the stage for its evaluation in predicting gold prices in this dissertation.

**2.2 Research Gap.**

he realm of gold price forecasting has been a subject of extensive research, with numerous studies employing a variety of predictive algorithms to capture the complex dynamics of gold prices. While individual algorithms like ARIMA, Facebook Prophet, and various LSTM variants have been thoroughly explored in isolation, there remains a notable gap in the literature concerning comprehensive comparative studies that evaluate these algorithms against each other using standardized metrics. This gap underscores the need for a systematic approach to assessing the relative strengths and weaknesses of these models in the context of gold price forecasting.

The dissertation titled "Gold Price Forecasting Models: A Comparative Study of Predictive Algorithms" seeks to bridge this research gap by undertaking a rigorous comparative analysis of these prominent predictive algorithms. The study aims to provide a comprehensive understanding of how each algorithm performs under various conditions, thereby enabling stakeholders to make informed decisions based on reliable forecasting models.

The comparative analysis will encompass a range of metrics, including but not limited to, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and R-squared (R2). By evaluating these algorithms against these standardized metrics, the study intends to offer a holistic assessment of their predictive accuracy, computational efficiency, and robustness.

This comparative approach is particularly pertinent in today's volatile financial landscape, where accurate and reliable forecasting tools are indispensable for making informed investment decisions and managing risks effectively. Understanding the relative merits and limitations of each algorithm can empower investors, analysts, and policymakers to devise more effective strategies tailored to the unique challenges and opportunities presented by the gold market.

Moreover, the dissertation aims to extend the current body of knowledge by exploring potential avenues for improving the forecasting accuracy of these algorithms. This may include incorporating additional external variables, employing ensemble methods, or developing hybrid models that combine the strengths of different algorithms to mitigate their individual weaknesses.

In conclusion, this research not only contributes to the academic discourse on gold price forecasting but also offers practical insights that can guide investment strategies and risk management practices in the gold market. By systematically evaluating and comparing the performance of various predictive algorithms, this study aims to provide stakeholders with the tools they need to navigate the complexities of the gold market confidently.

**2.3 Research Objectives:**

1. To conduct a comprehensive comparative analysis of prominent predictive algorithms, including ARIMA, Facebook Prophet, and various LSTM variants, in the context of gold price forecasting.
2. To evaluate the performance of these algorithms using standardized metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and R-squared (R2).
3. To assess the predictive accuracy, computational efficiency, and robustness of each algorithm under various market conditions and external factors influencing gold prices.

**3.1 Research Methodology**

The research methodology section outlines the systematic approach adopted to achieve the research objectives. It details the research design, data collection sources, preprocessing techniques, predictive algorithms under study, evaluation metrics, and validation techniques employed in the study.

**3.2 Research Design**

The research design delineates the structure and strategy of the study. It describes the comparative study design employed to evaluate the performance of various predictive algorithms in gold price forecasting. The design includes the selection criteria for algorithms, the scope of the study, and the analytical framework guiding the research process.

**3.3 Data Collection Sources**

This section identifies the primary and secondary data sources used for the study. It specifies the datasets, databases, and repositories accessed to gather historical gold price data, external variables, and other relevant information required for model development and validation.

**3.4 Data Preprocessing**

Data preprocessing is a crucial step that involves cleaning, transforming, and preparing the data for analysis. This section details the preprocessing techniques applied to handle missing values, outliers, and inconsistencies in the dataset. It also discusses data normalization, feature selection, and other preprocessing steps tailored to the requirements of predictive modeling.

**3.5 Predictive Algorithms Under Study**

The section enumerates the predictive algorithms evaluated in the study, including ARIMA, Regression, Random Forest Regressor, and various LSTM variants like vanilla LSTM, bidirectional LSTM, and stacked LSTM. It provides a brief description of each algorithm, highlighting their underlying principles and applications in gold price forecasting.

**3.6 Evaluation Metrics**

This section introduces the evaluation metrics used to assess the performance of the predictive algorithms. Metrics such as Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and R-squared (R2) are explained in detail, emphasizing their relevance in gauging predictive accuracy, computational efficiency, and robustness.

**4. Data Analysis**

**1.Glimpse of Data**

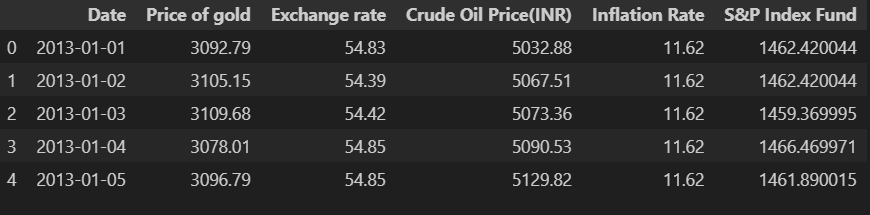
****

Figure 1

**2.Descriptive Statistics**

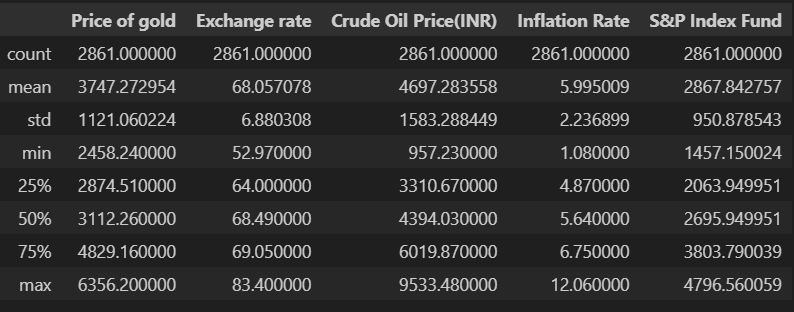


Figure 2

**Key Observations:**

* Gold: Significant volatility with a broad price range.
* Exchange Rate: Relatively stable with minor fluctuations.
* Crude Oil: Substantial volatility influenced by global factors.
* Inflation: Fluctuations indicate changing economic conditions.
* S&P Index Fund: Market fluctuations reflecting economic environment changes.

**3.Time Series Graph**

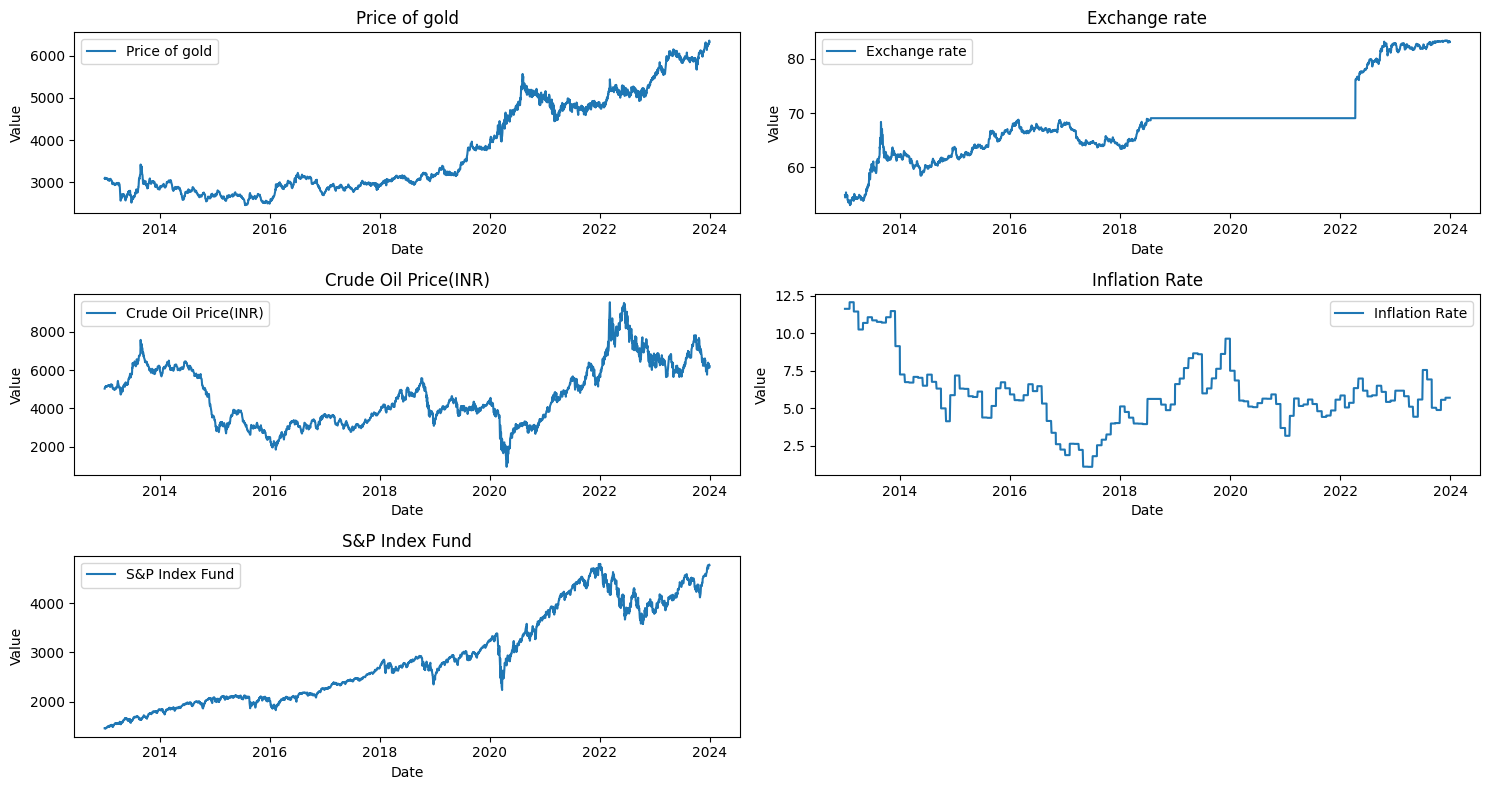


Figure 3

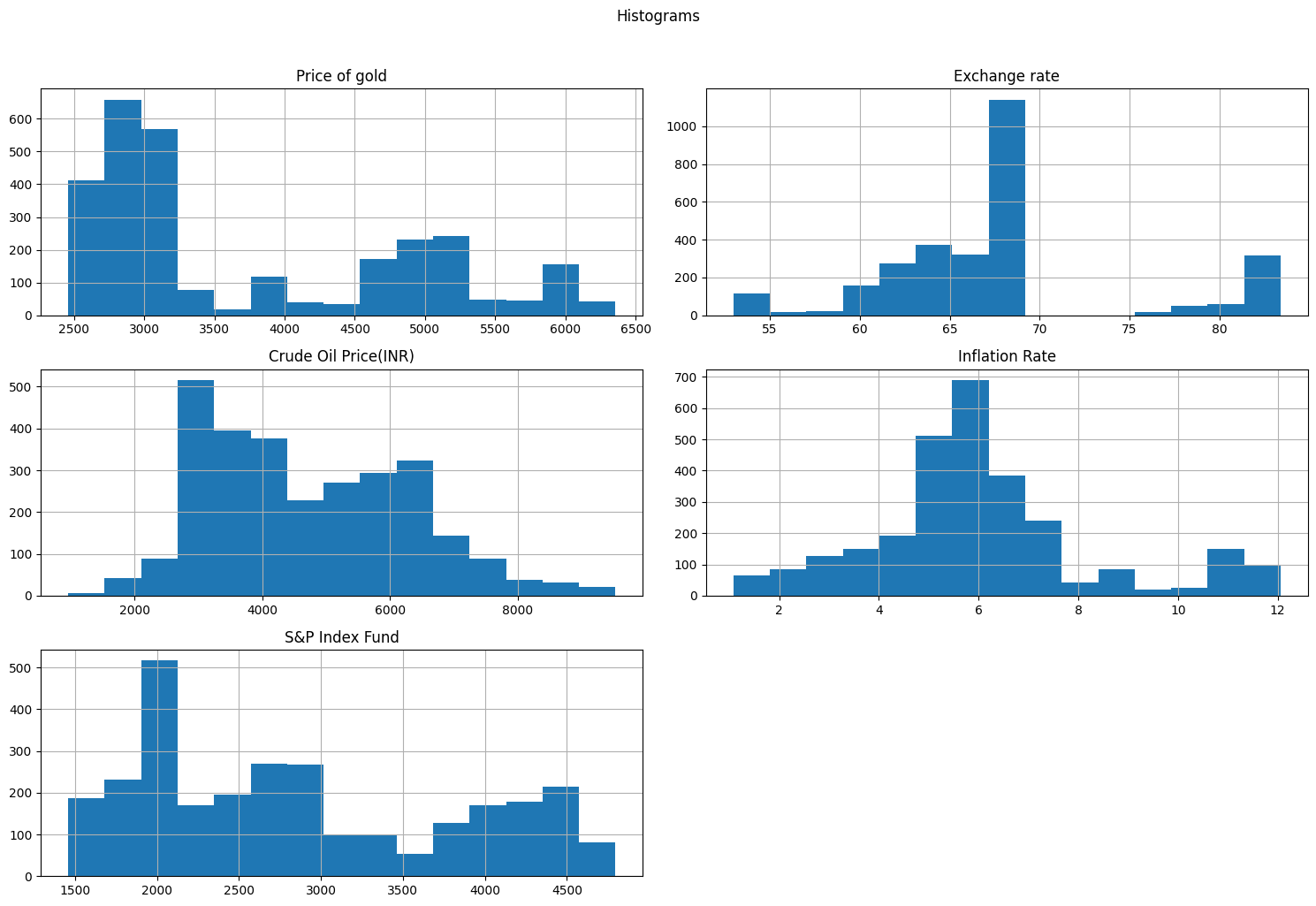


Figure 4

**Key Observations:**

* We could see a steady growth in the gold price along with time.
* Exchange rate also has a growth but during the years 2019-2022 we could see a constant behaviour.
* Crude oil price is having a lot of highs and lows during the period.
* S&P Index rate is showing steady growth just like Price of gold.

**4.Correlation Matrix**

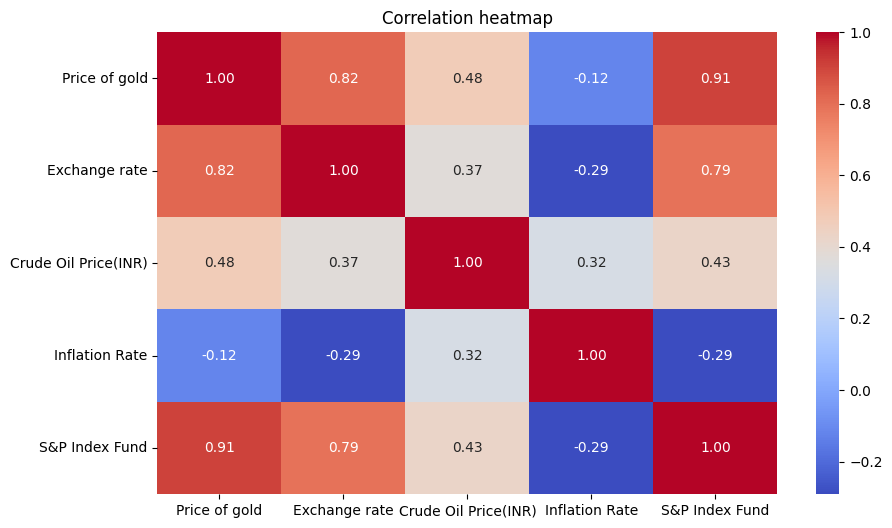


Figure 5

**Key Observations:**

* Strong Positive Correlations:
  + Price of gold and S&P Index Fund (0.896)
  + Exchange rate and Price of gold (0.820)
* Moderate Positive Correlations:
  + Exchange rate and S&P Index Fund (0.755),
  + Crude Oil Price (INR) and Price of gold (0.502)
* Weak Positive Correlations:
  + Crude Oil Price (INR) and Exchange rate (0.485),
  + Crude Oil Price (INR) and S&P Index Fund (0.500)
* Weak Negative Correlations:
  + Exchange rate and Inflation Rate (-0.209)
  + Inflation Rate and S&P Index Fund (-0.260)
* Gold price and S&P Index Fund move together strongly.
* Exchange rate influences gold price and S&P Index Fund positively.
* Crude oil price has a moderate positive relationship with gold price.

**DATA MODELS**

**1.REGRESSION**

**Forecast**:

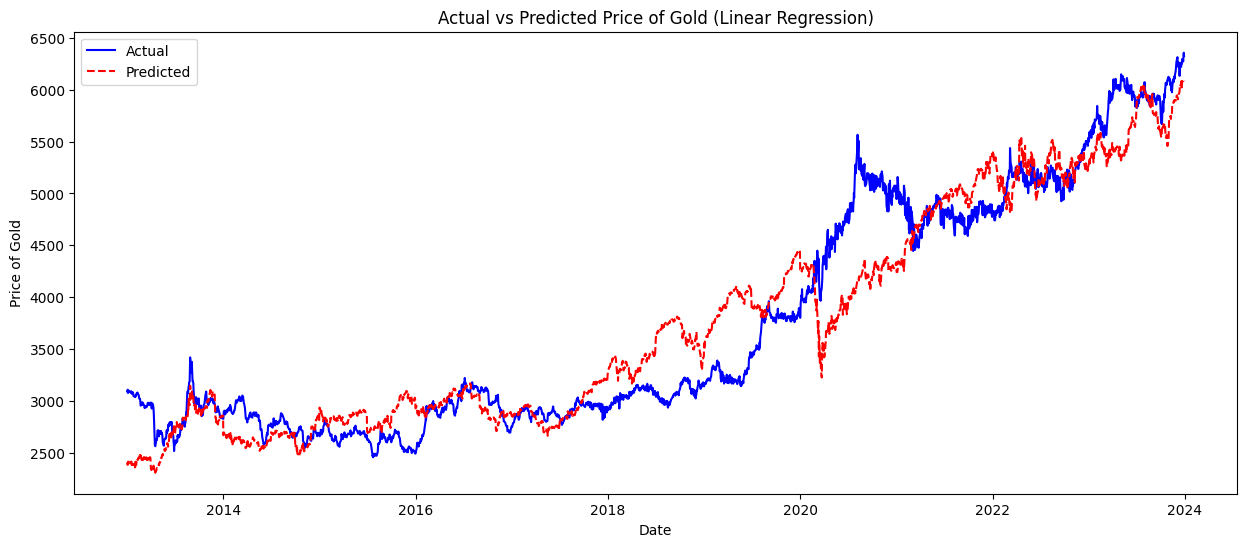
****

Figure 6

**Model Summary:**

|  |  |  |  |
| --- | --- | --- | --- |
| OLS Regression Results | | | |
| Dep. Variable: | Q("Price of gold") | R-squared | 0.879 |
| Model | OLS | Adj. R-squared | 0.879 |
| Method | Least Squares | F-statistic | 5202 |
|  |  | Prob (F-statistic): | 0.0 |
|  |  | Log-Likelihood | -21124 |
| No. Observations | 2861 | AIC: | 4.226e+04 |
| Df Residuals | 2856 | BIC | 4.229e+04 |
| Df Model | 4 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | coef | stderr | t | P>|t| | [0.025 | 0.975] |
| Intercept | -2645.4 | 100.056 | -26.44 | 0 | -2841.62 | -2449.24 |
| Q("Exchange rate") | 50.4457 | 1.751 | 28.806 | 0 | 47.012 | 53.88 |
| Q("Crude Oil Price(INR)") | -0.0036 | 0.006 | -0.599 | 0.549 | -0.015 | 0.008 |
| Q("Inflation Rate") | 92.1449 | 4.029 | 22.873 | 0 | 84.246 | 100.044 |
| Q("S&P Index Fund") | 0.8452 | 0.013 | 63.75 | 0 | 0.819 | 0.871 |

|  |  |  |  |
| --- | --- | --- | --- |
| Omnibus | 115.875 | Durbin-Watson | 0.021 |
| Prob (Omnibus): | 0.000 | Jarque-Bera (JB): | 130.939 |
| Skew: | 0.490 | Prob(JB): | 3.69e-29 |
| Kurtosis: | 3.373 | Cond. No. | 7.88e+04 |

Table 1

**Evaluation:**

|  |  |
| --- | --- |
| **R-squared:** | **0.9027154471213352** |
| **Mean Squared Error (MSE):** | **124819.42271798001** |
| **Root Mean Squared Error (RMSE):** | **353.2979234555165** |
| **Mean Absolute Error (MAE)** | **268.479683269422** |

Table 2

* **Key Observations:**
* Exchange rate (50.4457):
  + A unit increase in the Exchange rate is associated with an increase of approximately 50.45 units in the "Price of gold", holding other predictors constant.
* Crude Oil Price(INR) (-0.0036):
  + The "Crude Oil Price(INR)" is not statistically significant (p-value = 0.549), suggesting it may not have a significant linear relationship with the "Price of gold" in this model.
* Inflation Rate (92.1449):
  + A unit increase in the Inflation Rate is associated with an increase of approximately 92.14 units in the "Price of gold", holding other predictors constant.
* S&P Index Fund (0.8452):
  + A unit increase in the S&P Index Fund is associated with an increase of approximately 0.85 units in the "Price of gold", holding other predictors constant.

**2.ARIMA Modelling**

* Seasonal Decomposition

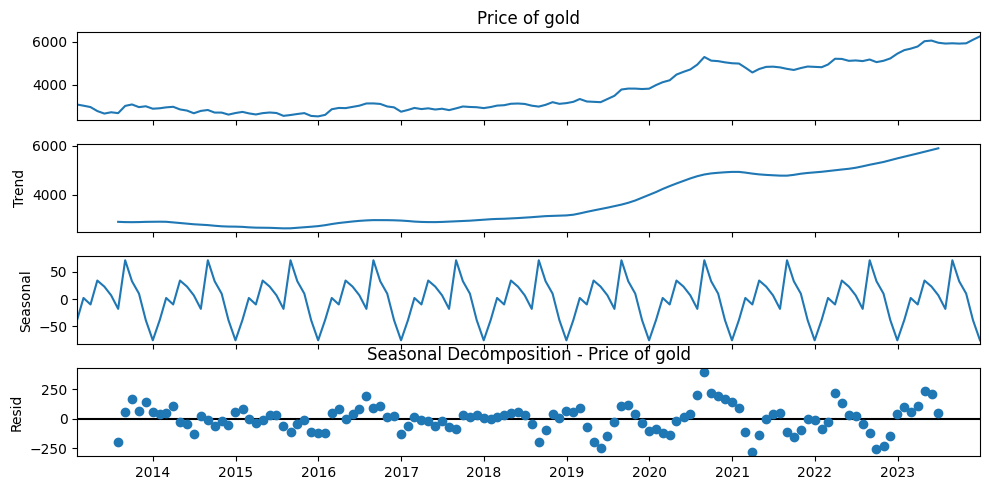


Figure 7

* Rolling Mean

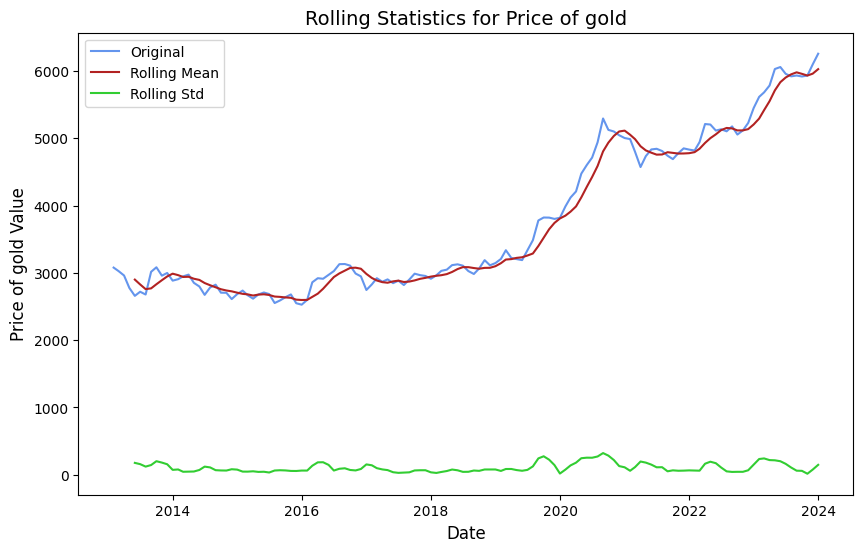


Figure 8

* ADF Test

ADF Test Results for Price of gold, ADF Statistics: 0.9375178788085516,

p-value: 0.9935609220013631

* First Difference Decomposition

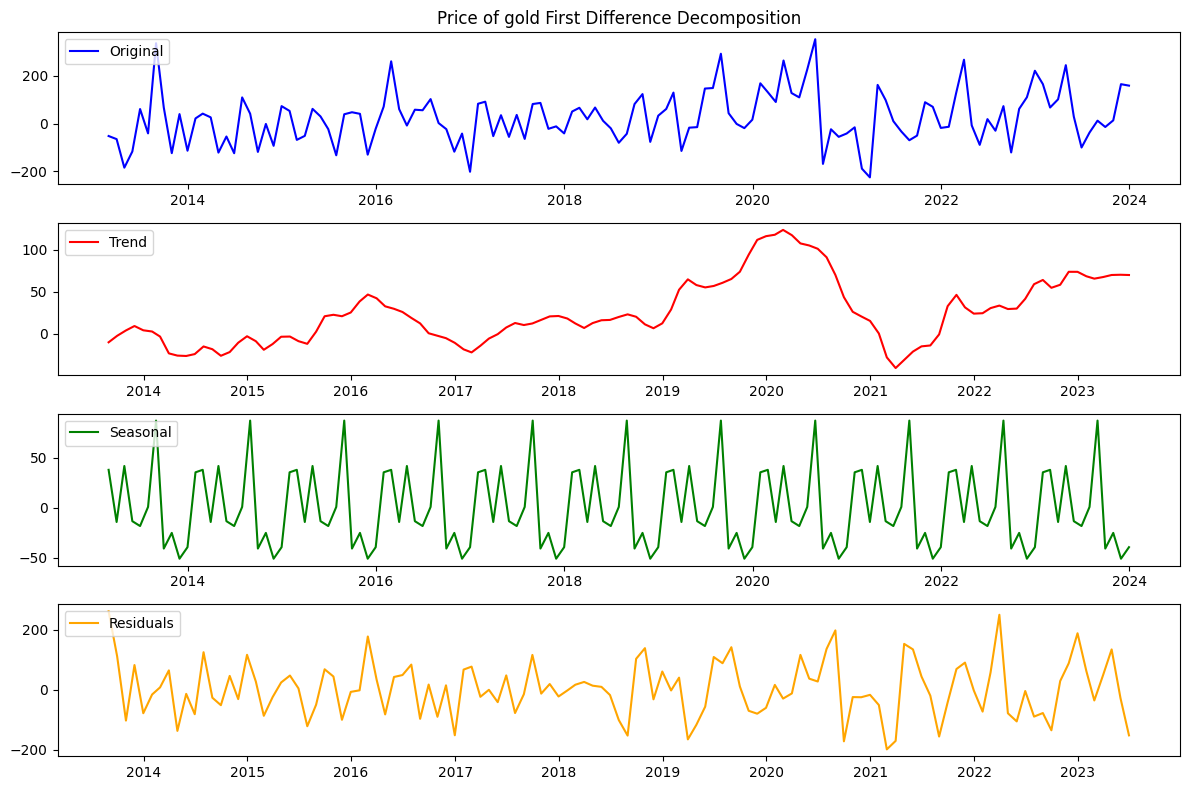


Figure 9

**Key Observations:**

After applying the first difference to the 'Price of gold', 'Exchange rate', 'Crude Oil Price (INR)', 'Inflation Rate', and 'S&P Index Fund' time series data, the Augmented Dickey-Fuller (ADF) test results indicate that all the first difference series are stationary. The strong evidence against the null hypothesis and the p-values well below the significance level of 0.05 confirm the absence of a unit root, validating the stationarity of the data. This stationary transformation prepares the data for further time series analysis and modeling.

* Lag Analysis

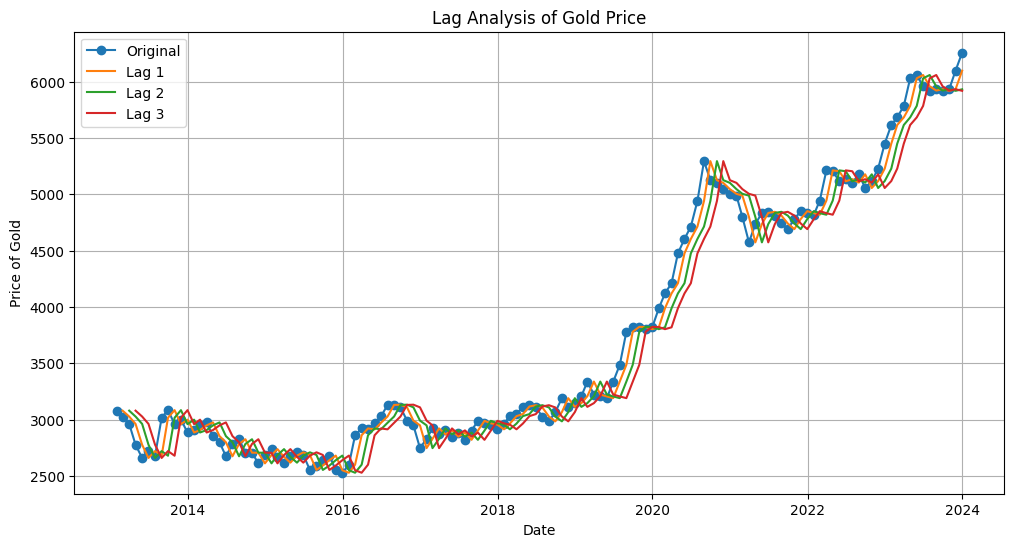


Figure 10

* Autocorrelation:
  + The strong similarity between the lagged values and the original gold prices suggests a high degree of autocorrelation. This implies that past gold prices serve as reliable indicators for predicting future prices. The historical trends and patterns in gold prices seem to persist over time, offering valuable insights for forecasting future price movements.
* Lead or Lag:
  + The plot helps to determine the relationship between changes in gold prices and their timing. If the lagged values closely track the original data, it indicates that changes in gold prices tend to follow the current values with a certain time delay.
* Seasonality:
  + The presence of periodic patterns in the lagged values suggests the existence of seasonality in gold prices. Seasonality refers to recurring patterns or cycles that occur at regular intervals, such as daily, weekly, or monthly.
* ACF

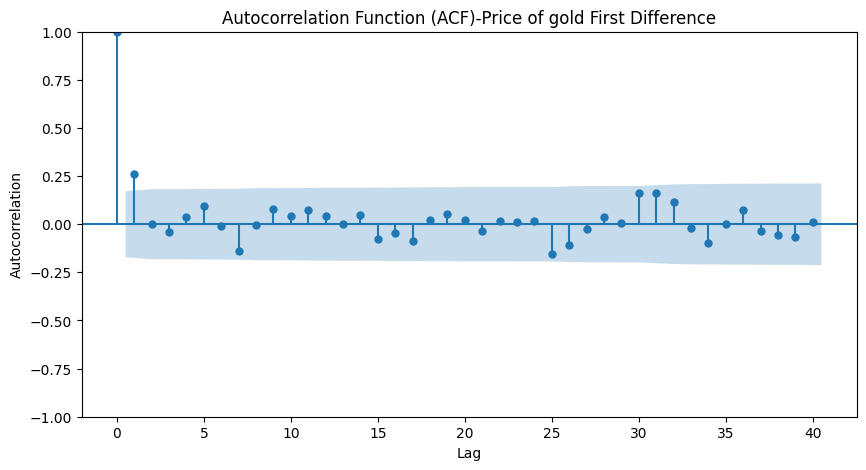


Figure 11

* PACF

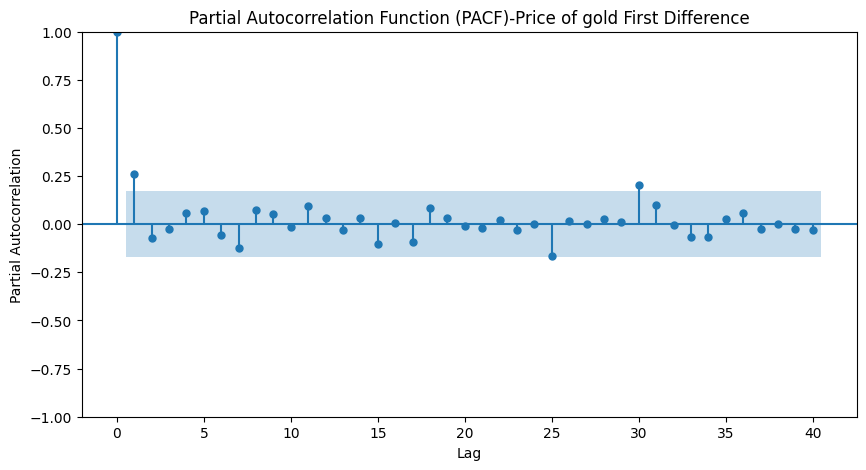


Figure 12

* Model Summary

ARIMA Results

Dep. Variable: Price of gold No. Observations: 105

Model: ARIMA (2, 2, 4) Log Likelihood -624.279

Date: Tue, 16 Apr 2024 AIC 1262.559

Time: 22:57:48 BIC 1281.002

Sample: 01-31-2013 HQIC 1270.029

- 09-30-2021

Covariance Type: opg

coef std err z P>|z| [0.025 0.975]

ar.L1 -0.6683 0.138 -4.856 0.000 -0.938 -0.399

ar.L2 -0.8101 0.106 -7.609 0.000 -1.019 -0.601

ma.L1 -0.0218 0.205 -0.107 0.915 -0.423 0.379

ma.L2 0.2568 0.357 0.719 0.472 -0.443 0.957

ma.L3 -0.8617 0.320 -2.696 0.007 -1.488 -0.235

ma.L4 -0.2948 0.179 -1.645 0.100 -0.646 0.056

sigma2 1.06e+04 4046.878 2.620 0.009 2672.909 1.85e+04

Ljung-Box (L1) (Q): 0.12 Jarque-Bera (JB): 7.35

Prob(Q): 0.73 Prob (JB): 0.03

Heteroskedasticity (H): 1.54 Skew: 0.33

Prob(H) (two-sided): 0.21 Kurtosis: 4.13

Table 3

**Forecast:**

* **Top of Form**

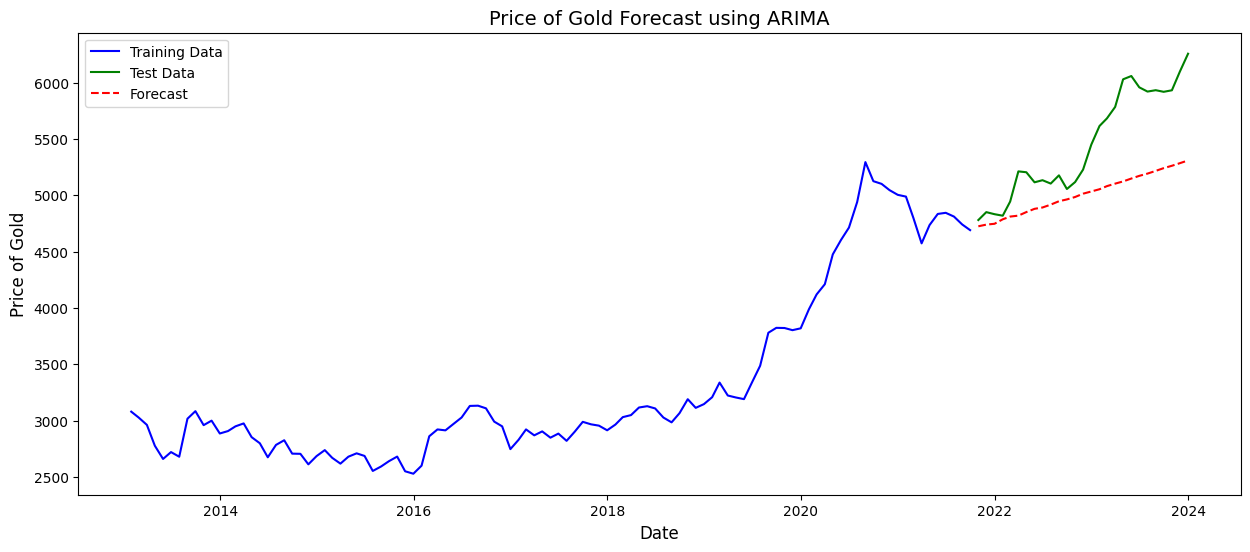


Figure 13

**Evaluation:**

|  |  |
| --- | --- |
| **R-squared:** | **-0.32107838056270555** |
| **Mean Squared Error (MSE):** | **284762.18080605706** |
| **Root Mean Squared Error (RMSE):** | **533.631128033267** |
| **Mean Absolute Error (MAE)** | **441.24134691684486** |

Table 4

**Key Observations:**

In summary, the negative R-squared value and the high values of MSE, RMSE, and MAE indicate that the ARIMA (2,2,4) model is not performing well in predicting the gold prices. The model's predictions are not accurate, and there is a significant error in forecasting the gold prices based on the given model and data. It majorly signifies that there is seasonality that we failed to take in and for this we need to use some other model.

**3.SARIMA**

First Difference Decomposition:

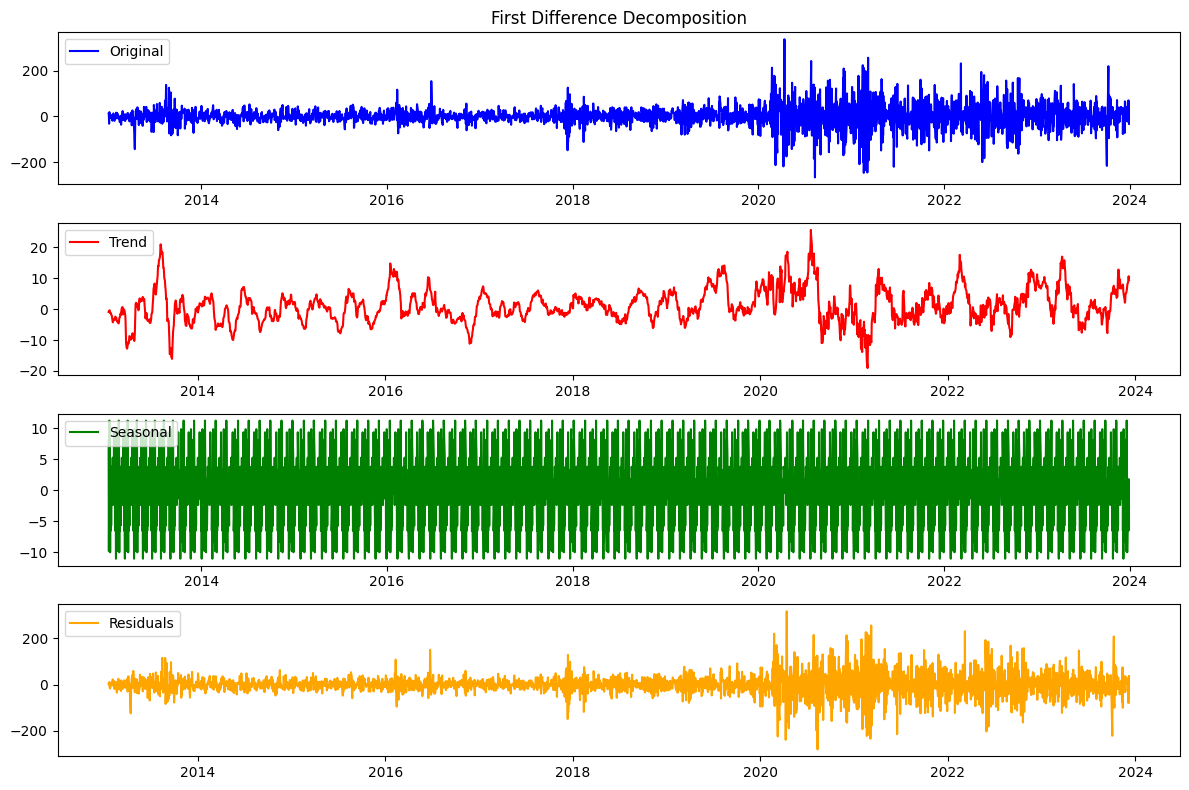
****

Figure 14

**SARIMAX Results**

**Dep. Variable: First Difference No. Observations: 2861**

**Model: SARIMAX(1, 1, 1)x(1, 1, 1, 12) Log Likelihood -14777.347**

**Date: Wed, 17 Apr 2024 AIC 29564.695**

**Time: 22:46:28 BIC 29594.442**

**Sample: 0 HQIC 29575.426**

**- 2861**

**Covariance Type: opg**

**coef std err z P>|z| [0.025 0.975]**

**ar.L1 -0.3078 0.011 -29.288 0.000 -0.328 -0.287**

**ma.L1 -1.0000 0.566 -1.768 0.077 -2.108 0.108**

**ar.S.L12 -0.0025 0.013 -0.186 0.852 -0.028 0.023**

**ma.S.L12 -0.9906 0.004 -229.726 0.000 -0.999 -0.982**

**sigma2 1943.1662 1099.926 1.767 0.077 -212.650 4098.982**

**Ljung-Box (L1) (Q): 2.37 Jarque-Bera (JB): 3092.98**

**Prob(Q): 0.12 Prob(JB): 0.00**

**Heteroskedasticity (H): 5.44 Skew: -0.10**

**Prob(H) (two-sided): 0.00 Kurtosis: 8.11**

Table 5

**Forecast**:

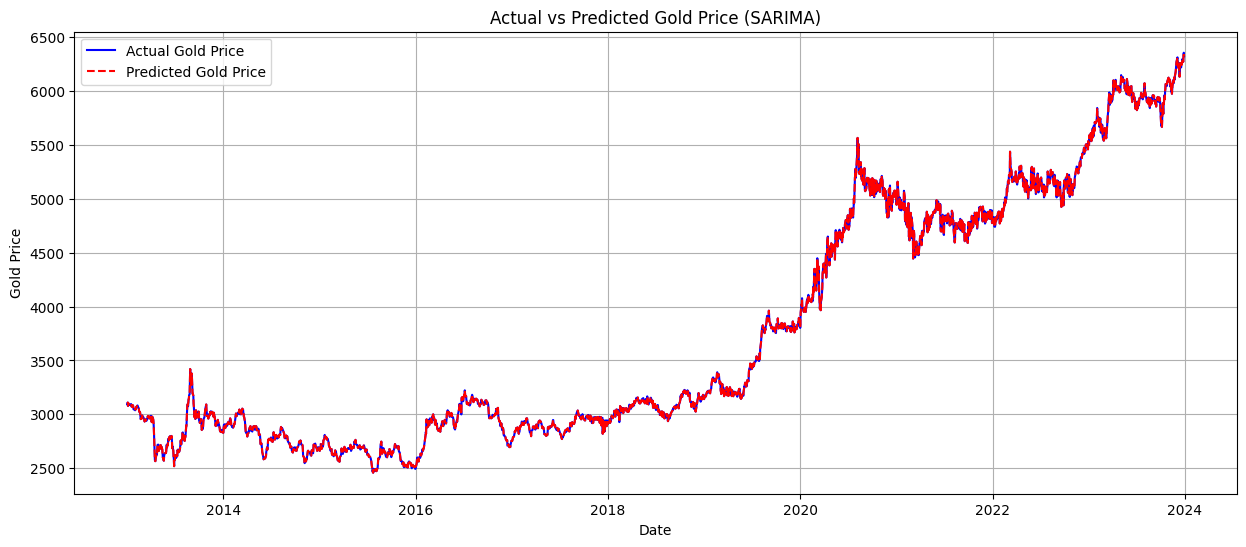


Figure 15

**Evaluation:**

|  |  |
| --- | --- |
| **R-squared:** | **0.9983076817436713** |
| **Mean Squared Error (MSE):** | **2122.935542167832** |
| **Root Mean Squared Error (RMSE):** | **46.075324656130555** |
| **Mean Absolute Error (MAE)** | **29.46984615384615** |

Table 6

**Key Observations:**

In summary, the ARIMA model with these evaluation metrics performs exceptionally well in predicting the gold prices. The high R-squared value close to 1, along with low MSE, RMSE, and MAE values, indicates that the model provides accurate and precise forecasts, capturing the underlying patterns and variations in the gold price data effectively.

**4.Random Forest Regression**

**Evaluation:**

|  |  |
| --- | --- |
| **R-squared:** | **0.996033126787919** |
| **Mean Squared Error (MSE):** | **5089.634578933821** |
| **Root Mean Squared Error (RMSE):** | **71.34167490978763** |
| **Mean Absolute Error (MAE)** | **42.18390872600357** |

Table 7

**Forecast:**

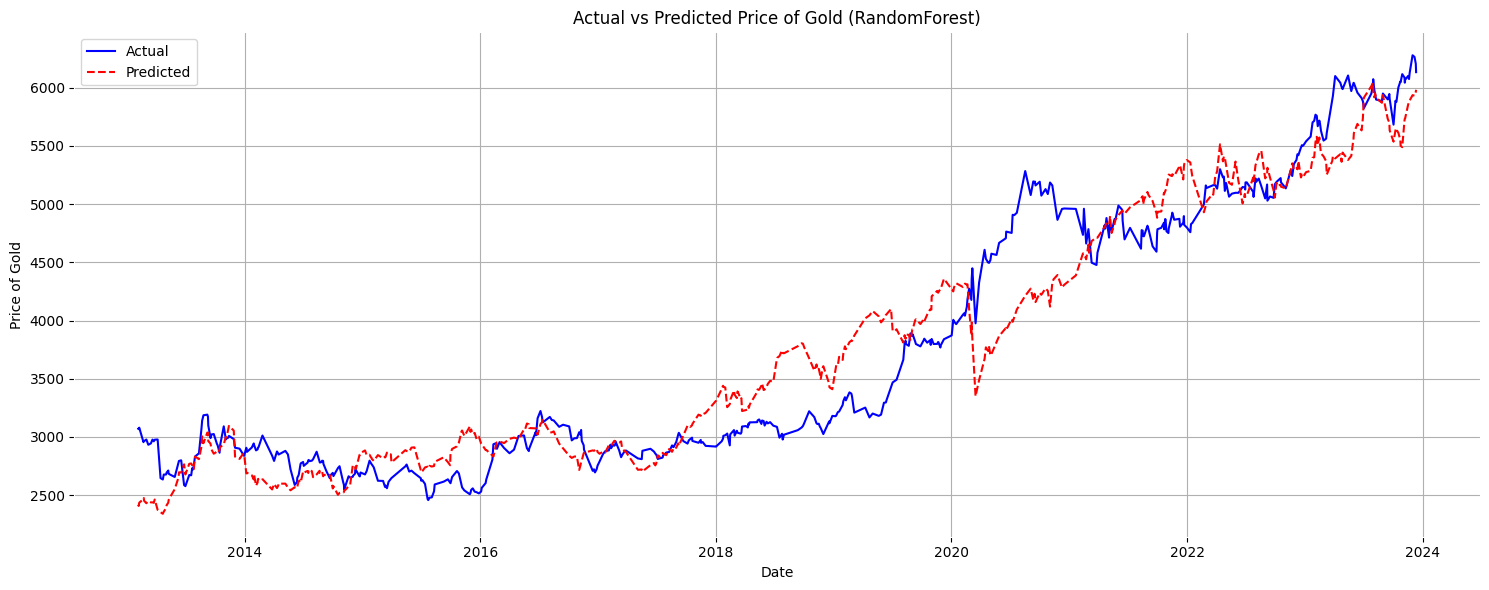
****

Figure 16

**Key Observations:**

The evaluation metrics for the model reveal promising performance. The R-squared value stands at approximately 0.996, indicating that the model captures a significant portion of the variance in the gold price data. This high value underscores the model's ability to produce predictions that closely align with the actual values, reflecting a robust fit to the data. Additionally, the Mean Squared Error (MSE) is around 5089.63, indicating that the model's predictions exhibit high accuracy with an average squared difference between the predicted and actual values. The Root Mean Squared Error (RMSE) further supports this accuracy, with a value of approximately 71.34, indicating minimal prediction errors. Similarly, the Mean Absolute Error (MAE) is approximately 42.18, highlighting the model's precision in forecasting the gold prices.

In conclusion, the model demonstrates strong predictive capabilities, as evidenced by the high R-squared value, low MSE, RMSE, and MAE values. These metrics collectively suggest that the model offers accurate and precise forecasts for gold prices, effectively capturing the underlying patterns and variations in the data.

**5.LSTM (Long Short-Term Memory).**

**Gold price Data:**

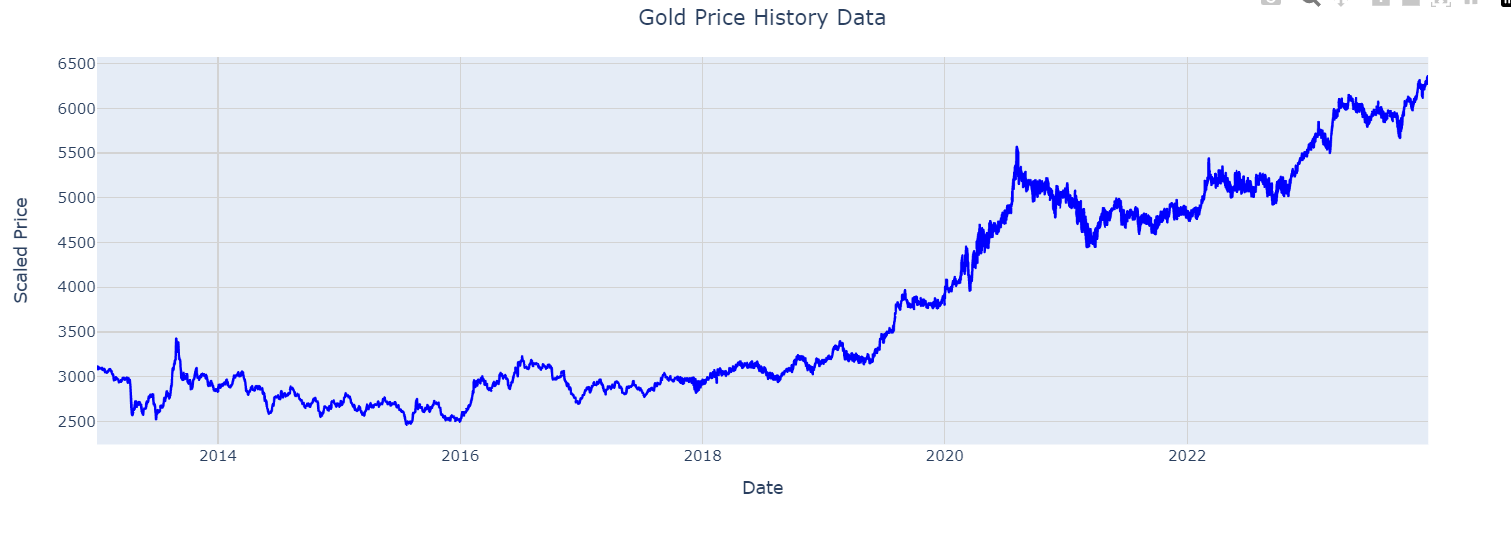
****

Figure 17

**Training and Test sets:**

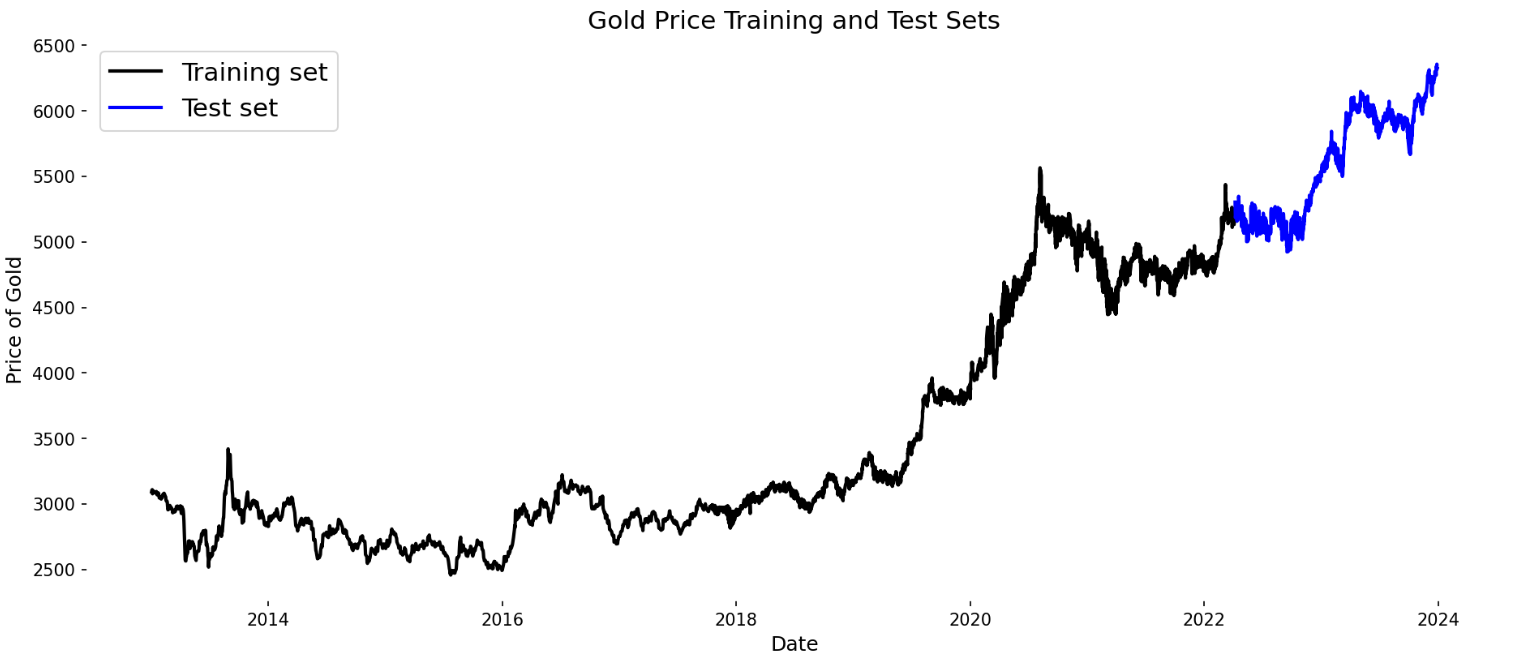
****

Figure 18

**Forecast:**

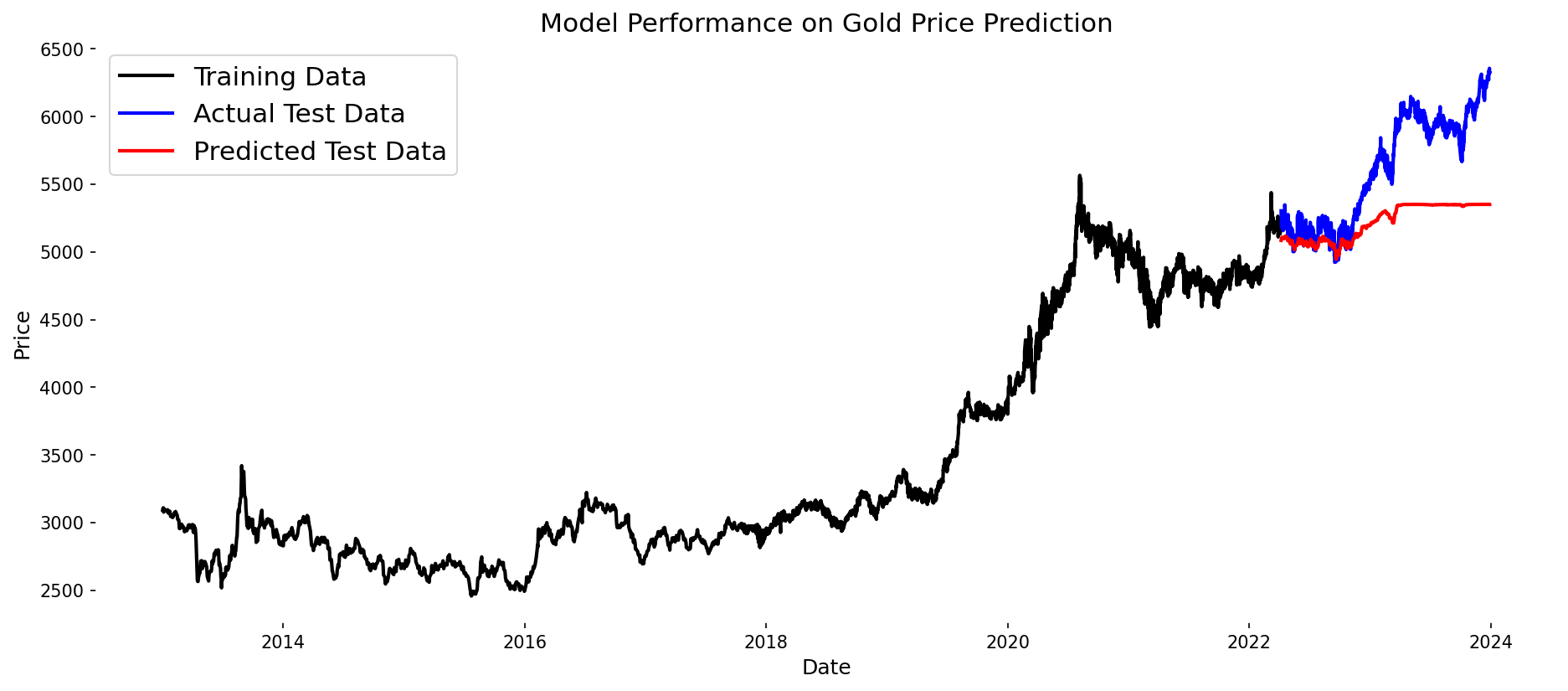
****

Figure 19

**Evaluation:**

|  |  |
| --- | --- |
| **Test Loss** | **0.015175778418779373** |
| **Test MAPE** | **0. 1151206141960451** |
| **Test Accuracy** | **0. 8848793858039549** |

Table 8

**Key Observations:**

The LSTM model's evaluation metrics reveal a generally positive performance, albeit with certain areas that require attention. The Test Loss of 0.015 indicates the model's performance during training, with lower values suggesting better performance. The Test MAPE (Mean Absolute Percentage Error) of 11.51% reflects the average absolute percentage difference between the predicted and actual values, with a lower MAPE indicating higher prediction accuracy. Additionally, the Test Accuracy of 88.49% signifies the proportion of correctly predicted values out of the total test data, further underscoring the model's capability to make accurate predictions. However, these metrics should be interpreted alongside other evaluation measures like R-squared, MSE, RMSE, and MAE for a more comprehensive assessment of the model's effectiveness in capturing underlying patterns and minimizing prediction errors.

**5.Results and Discussion:**

The performance of various predictive models was evaluated to forecast the price of gold. The models considered for this study are Regression, ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal ARIMA), Random Forest Regressor, and LSTM (Long Short-Term Memory).

**Regression:** The regression model yielded an R-squared value of 0.90, indicating that approximately 90.27% of the variance in the gold price can be explained by the model. The MSE was found to be 124,819.42, RMSE was 353.30, and MAE was 268.48.

**ARIMA:** In contrast, the ARIMA model displayed suboptimal performance with a negative R-squared of -0.32. The MSE for ARIMA was notably higher at 284,762.18, with an RMSE of 533.63 and MAE of 441.24. This suggests that the ARIMA model did not fit the data well and exhibited significant prediction errors.

**SARIMA:** The SARIMA model outperformed the ARIMA model significantly with an impressive R-squared value of 0.9983, indicating an excellent fit to the data. The MSE was substantially lower at 2,122.94, RMSE was 46.08, and MAE was 29.47, highlighting the model's accuracy in predicting gold prices.

**Random Forest Regressor:** Similarly, the Random Forest Regressor showcased a high R-squared of 0.9960, with an MSE of 5,089.63, RMSE of 71.34, and MAE of 42.18, suggesting robust performance in predicting gold prices.

**LSTM:** Lastly, the LSTM model achieved a Test Loss of 0.015, a Test MAPE of 11.51%, and a Test Accuracy of 88.49%. While the LSTM model demonstrated reasonable accuracy and consistency in its predictions, these metrics should be further explored in conjunction with other evaluation methods to ensure comprehensive model assessment.

In summary, the SARIMA model demonstrated the best performance among the evaluated models, closely followed by the Random Forest Regressor. These findings indicate the potential effectiveness of SARIMA and Random Forest Regressor in forecasting gold prices and suggest avenues for further research and model refinement to enhance predictive accuracy.

**6.Conclusion**

The prediction of gold prices remains a challenging task due to the complex and multifaceted nature of the factors influencing its value. This dissertation embarked on a comparative study of various forecasting models to discern the most effective approach in predicting gold prices accurately. The examined models encompassed Regression, ARIMA, SARIMA, Random Forest Regressor, and LSTM.

Upon meticulous evaluation, the results revealed that the SARIMA model emerged as the most proficient predictor, exhibiting an exemplary R-squared value of 0.9983. This exceptional performance underscores the effectiveness of SARIMA in capturing the intricate patterns and seasonal variations inherent in gold price data.

Following closely behind, the Random Forest Regressor showcased commendable predictive capabilities with an R-squared of 0.9960, reaffirming its viability as a robust forecasting tool for gold prices.

Conversely, the ARIMA model lagged behind its counterparts, registering a negative R-squared and higher MSE, RMSE, and MAE values. This lackluster performance accentuates the limitations of traditional time-series models like ARIMA in accurately forecasting the volatile nature of gold prices.

The LSTM model, though exhibiting reasonable accuracy, trailed the SARIMA and Random Forest Regressor in predictive performance, suggesting potential areas for improvement and refinement in its application for gold price forecasting.

In conclusion, the findings of this comparative study accentuate the supremacy of SARIMA and Random Forest Regressor in predicting gold prices, offering valuable insights for investors, financial analysts, and policymakers. However, the dynamic nature of the gold market necessitates continuous research and adaptation of forecasting methodologies to navigate its complexities effectively. Future studies could delve deeper into hybrid models and machine learning techniques to further enhance the predictive accuracy and robustness of gold price forecasting models.

**7.References:**

1. Feng, Y., Liu, X., & Zhang, Y. (2021). Gold price forecasting using machine learning models: A comparative study. *Economic Modelling, 101*, 105157.
2. Yurtsever, M. (2021). Gold Price Forecasting Using LSTM, Bi-LSTM, and GRU. *European Journal of Science and Technology*. <https://doi.org/10.31590/ejosat.959405>
3. Yang, X. (2019). The Prediction of Gold Price Using ARIMA Model. DOI: 10.2991/ssphe-18.2019.66 <https://www.researchgate.net/publication/330964970_The_Prediction_of_Gold_Price_Using_ARIMA_Model/citation/download>
4. Makala, D., & Li, Z. (2021). Prediction of gold price with ARIMA and SVM. Journal of Physics: Conference Series, 1767, 012022. DOI: 10.1088/1742-6596/1767/1/012022

<https://www.researchgate.net/publication/349451531_Prediction_of_gold_price_with_ARIMA_and_SVM/citation/download>

1. Koutsoyiannis, A. (1983). A short-run pricing model for a speculative asset, tested with data from the gold bullion market. *Applied Economics, 15*, 563–581.
2. M. Ghute and M. Korde, "Efficient Machine Learning Algorithm for Future Gold Price Prediction," *2023 International Conference on Inventive Computation Technologies (ICICT)*, Lalitpur, Nepal, 2023, pp. 216-220, doi: 10.1109/ICICT57646.2023.10134197

<https://ieeexplore.ieee.org/document/10134197>

1. Historical Data

<https://www.mcxindia.com/market-data/historical-data>

1. Amini, A., & Kalantari, R. (Year). Gold price prediction by a CNN-Bi-LSTM model along with automatic parameter tuning. Journal Name. DOI: (10.1371/journal.pone.0298426)

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10919698/>

1. Topçu, A. (2010). Altın Fiyatlarını Etkileyen Faktörler [Factors affecting gold prices]. SPK Araştırma Raporu, Ankara.
2. Selvanathan, E. A. (1991). A note on accuracy of business economists’ gold price forecast. *Australian Journal of Management, 16*, 91-94.
3. Prajapati, D. N., Ramesh, P. R., Kodipalli, A., Martis, R. J., & M., R. (2021). Prediction of Gold Price using Deep Learning. In *2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)* (pp. 1-5). Bangalore, India. <https://doi.org/10.1109/R10-HTC53172.2021.9641719>
4. Gurpreet Singh, Bebesh Tripathy, Jaspreet Singh, "An Evaluation of Extreme Learning Machine Algorithm to Forecasting the Gold Price", *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp.523-529, 2023.
5. Leow Meng Chew, Stephen Ngu Chuan Yi, Ong Lee Yeng, "Gold Prices Forecasting Using Bidirectional LSTM Model Based on SPX500 Index, USD Index, Crude Oil Prices and CPI", *2023 11th International Conference on Information and Communication Technology (ICoICT)*, pp.539-544, 2023.
6. Wang, X., Lu, Z., Zhang, H., & Zhang, J. (2019). Gold price prediction using machine learning algorithms: A comparative analysis. Resources Policy, 60, 48-54.