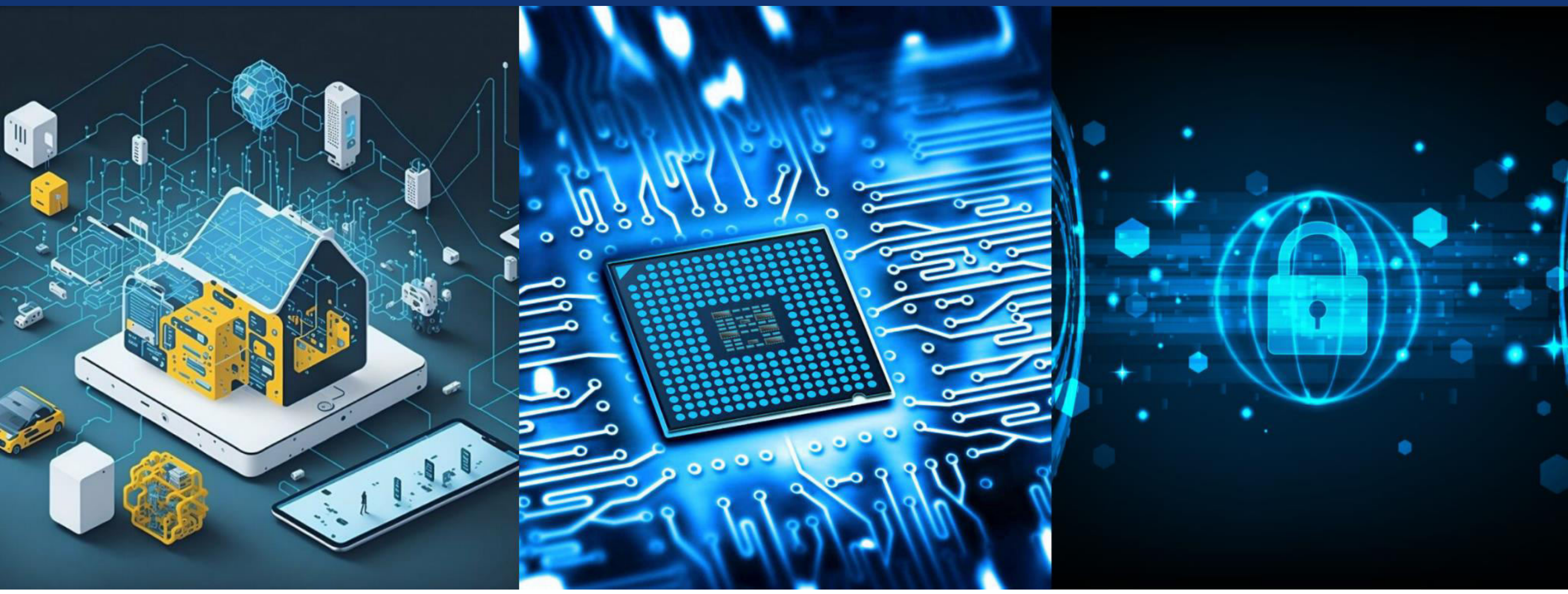
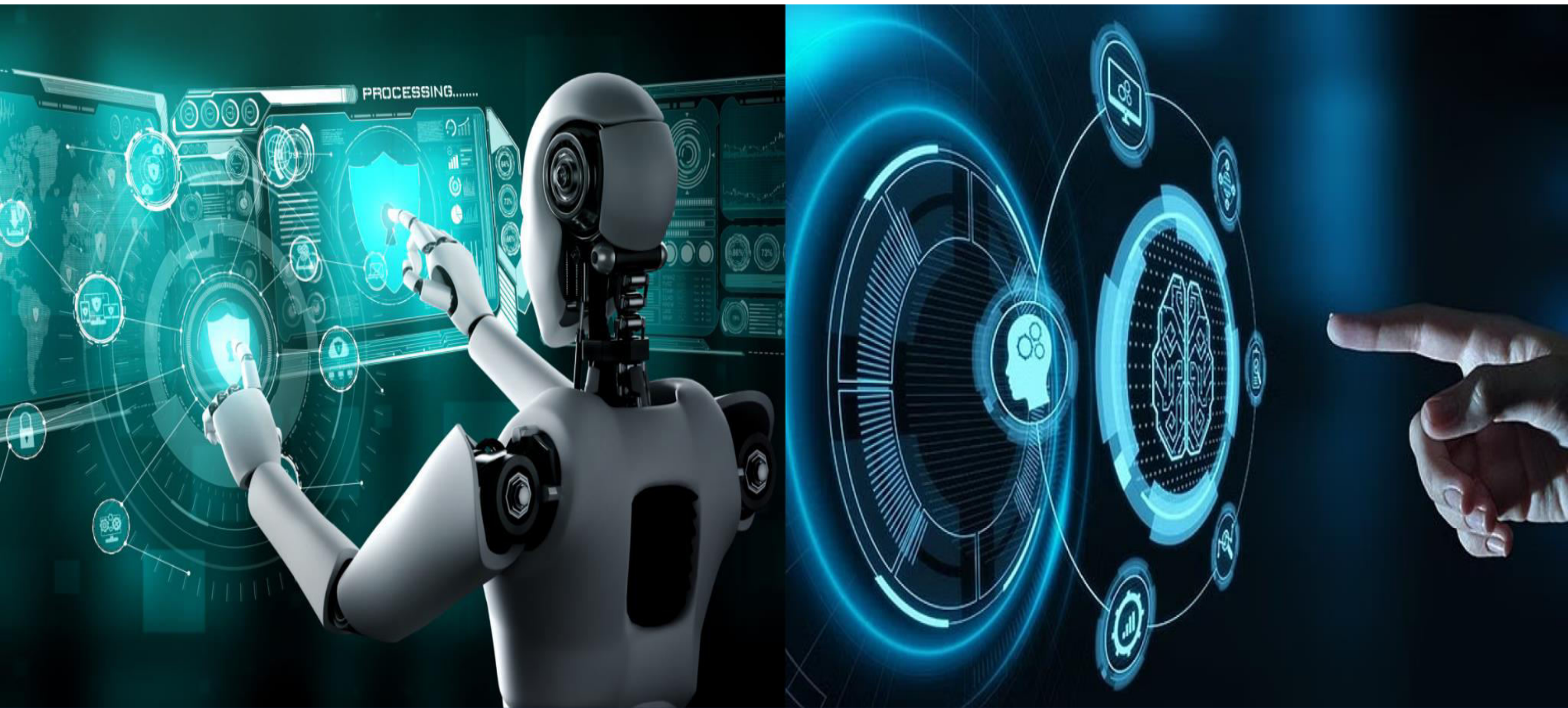
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771 **Volume 13, Issue 4, April 2025**

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Gold Price Prediction using Machine Learning

L. Harini[1], S. Shenbaha[2]

Student, Department of Computer Science with Data Analysis, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India[1]

Assistant Professor, Department of Computer Science with Data Analysis, Dr. N.G.P. Arts and Science College, Coimbatore, Tamil Nadu, India[2]

ABSTRACT: Gold has historically been a stable investment and a hedge against inflation, making its price movements a key focus for investors, policymakers, and financial analysts. This study examines historical gold price trends using statistical analysis, machine learning modelling, and data visualization. The dataset undergoes preprocessing, including handling missing values, date conversion, and engineering lag features to improve predictive accuracy. A Random Forest Regressor model is employed to forecast gold prices, optimized through hyperparameter tuning. The model's performance is evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared (R^2) score to ensure reliability. To understand price fluctuations, exploratory data analysis techniques are applied. Heatmaps highlight monthly average price variations across years, while line plots reveal trends over months, years, and quarters, identifying seasonal patterns. Rolling mean and variance calculations assess stationarity, crucial for financial modelling. Feature importance analysis within the Random Forest model provides insights into historical price values' significance in forecasting trends. This study offers valuable insights for traders, economists, and analysts by integrating machine learning with statistical methods. The findings help understand gold price dynamics and assist in informed investment decisions, presenting a data-driven approach for effective prediction and analysis.

KEYWORDS: Gold Price Prediction, Machine Learning, Exploratory Data Analysis, Random Forest Regressor.

I. INTRODUCTION

Gold is a valuable and widely traded commodity, serving as a safe-haven asset in investment portfolios and global economies. Its price is influenced by supply-demand dynamics, inflation, central bank policies, geopolitical events, and currency fluctuations. Predicting gold price fluctuations is crucial for investors, policymakers, and economists. However, traditional statistical models often struggle to capture non-linear dependencies, making machine learning an effective approach for analysis and forecasting. Machine learning, particularly ensemble models like Random Forest Regressors, has proven efficient in predicting gold prices using historical data patterns. Gold price prediction involves multiple analytical techniques, including time-series forecasting, feature importance analysis, and statistical evaluations. Exploratory Data Analysis (EDA) helps understand price trends using heatmaps to visualize monthly variations and line plots to track trends across months, years, and quarters. Rolling mean and variance calculations assess price stability, while autocorrelation plots reveal dependencies in price movements. By engineering lag features, machine learning models learn from historical prices to enhance predictive accuracy. These trained models provide reliable forecasts, helping investors make informed decisions and reducing uncertainty in gold trading. Traders, analysts, and policymakers can leverage machine learning tools to develop data-driven strategies, optimize portfolios, and mitigate risks. The adoption of machine learning-based forecasting models improves market stability and financial decision-making, making predictive analytics a vital tool in modern finance.

II. LITERATURE REVIEW

The study titled "Wavelet Transform and LSTM Networks for Gold Price Forecasting" by Robert Wilson (2023) combines wavelet transform and LSTM networks to enhance gold price forecasting by capturing complex patterns, reducing noise, and improving predictive accuracy, offering a more reliable approach for financial market analysis [1]. "Deep Learning Approaches for Gold Price Prediction: A Review" – (Author: Michael Brown, Year: 2022) The study reviews deep learning techniques for gold price prediction, comparing models like LSTM, CNN, and ANN, highlighting their effectiveness, challenges, and applications in financial forecasting for improved investment decision-

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304050

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

making [2]. "Time Series Analysis for Gold Price Prediction: ARIMA and GARCH Models" – (Author: Jane Smith, Year: 2020) The study explores ARIMA and GARCH models for gold price prediction, analysing their effectiveness in capturing trends, volatility, and financial market fluctuations for accurate forecasting and risk assessment in investment decisions [3]. The current gold price prediction systems primarily use time-series forecasting models and machine learning algorithms to analyse historical price data and predict future trends. Some systems use technical indicators like Moving Averages, RSI, and Bollinger Bands as features for machine learning models. Neural networks and deep learning models, such as Long Short-Term Memory (LSTM) networks, have also been explored for capturing complex patterns in gold price trends. Despite their effectiveness, many existing models have limitations due to small datasets, lack of real-time adaptability, and high volatility in gold prices, which makes accurate forecasting challenging. Moreover, traditional models often fail to incorporate real-world economic factors dynamically, limiting their predictive power.

III. PROPOSED METHODOLOGY

Data Collection

The dataset used in this study is obtained from publicly available sources such as Yahoo Finance and Investing.com. It contains historical daily gold price records, including attributes such as Date, Opening Price, Closing Price, High, Low, Volume, and Percentage Change (% Chg.). The dataset comprises 2,227 records, providing a comprehensive view of gold price fluctuations over time.

Preprocessing

Data preprocessing is crucial to ensure the quality and consistency of the dataset. Initially, missing or inconsistent data is handled through interpolation techniques or removal to prevent inaccuracies in model predictions. The 'Date' column is then converted into a datetime format, facilitating time-series analysis. To maintain dataset integrity, duplicate records are identified and removed. Outlier detection is performed using statistical methods such as Z-score and Interquartile Range (IQR) to identify and handle anomalies that could skew model performance. To ensure uniformity across features, standardization is applied using Standard Scaling, normalizing data distributions for better predictive accuracy. Lastly, a stationarity check is conducted through rolling mean and standard deviation calculations, ensuring stability in the data distribution, which is essential for reliable time-series modelling.

Feature engineering

Feature engineering plays a vital role in enhancing the model's predictive capability by transforming raw data into meaningful features. Lag features are created by incorporating historical price data (Price_Lag1, Price_Lag2, Price_Lag3) to capture temporal dependencies, allowing the model to recognize trends and patterns over time. Rolling mean and standard deviation are computed to identify seasonality and volatility, providing insights into how prices fluctuate over different time periods. Additionally, economic indicators such as inflation rates and stock market indices are integrated to improve prediction accuracy by considering broader financial influences. To ensure optimal feature selection, correlation analysis is conducted using a heatmap, which helps identify relationships between features and the target variable, allowing the model to focus on the most relevant predictors for accurate forecasting.

Model Training and Optimization

The dataset is split into an 80-20% ratio for training and testing to ensure robust evaluation. The Random Forest Regressor is chosen for its ability to handle non-linear relationships. Hyperparameter tuning optimizes estimators (200) and depth (10). K-fold cross-validation ensures stability, while feature importance plots enhance selection efficiency.

Model Evaluation

The model is assessed using standard performance metrics: Mean Absolute Error (MAE) measures the average deviation between predicted and actual prices. Mean Squared Error (MSE) evaluates the squared differences between actual and predicted values. R-squared (R^2) Score determines how well the model explains the variance in gold prices.

Data Visualization and Interpretation

To gain deeper insights into gold price trends and the model's performance, various data visualization techniques are employed. Line plots are used to illustrate price movements over different time periods, highlighting trends, fluctuations, and potential seasonal patterns. Heatmaps are generated to display average monthly and yearly gold prices, helping identify long-term trends and cyclic behaviour in price variations. Residual plots are analysed to examine prediction errors and detect possible overfitting, ensuring that the model generalizes well to unseen data. Additionally, autocorrelation plots are utilized to study the relationships between past and future gold prices, revealing dependencies that may exist within the time-series data. These visualizations collectively enhance the interpretability of the model's predictions, allowing for better decision-making in financial and investment scenarios.

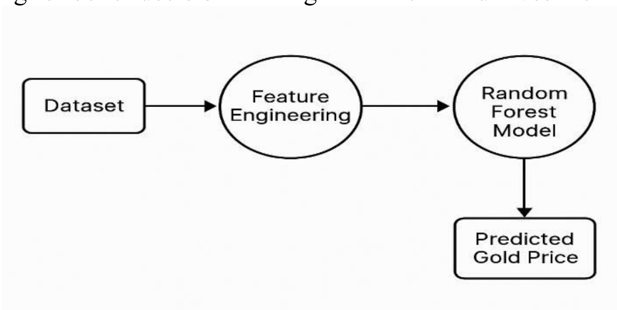


Fig 1 Data Flow Diagram

IV. RESULTS AND DISSCUSSION

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2227 entries, 0 to 2226
Data columns (total 7 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   Date    2227 non-null   object
 1   Price   2227 non-null   int64
 2   Open    2227 non-null   int64
 3   High    2227 non-null   int64
 4   Low     2227 non-null   int64
 5   Volume  2227 non-null   int64
 6   Chg%    2227 non-null   float64
dtypes: float64(1), int64(5), object(1)
memory usage: 121.9+ KB
```

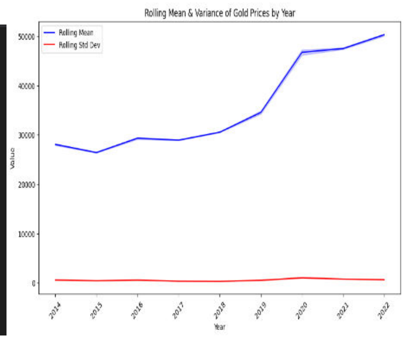
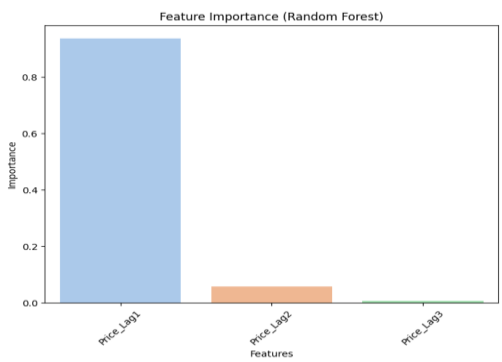


Fig 2: Dataset **Fig 3: Rolling Mean &Variance of Gold Price**



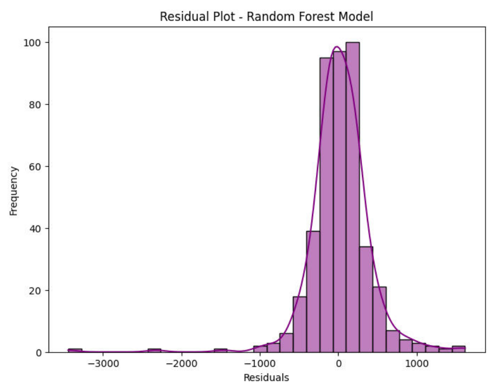


Fig 4:Feature Importance **Fig 5:Residual Plot**

© 2025 IJIRCCE | Volume 13, Issue 4, April 2025| DOI: 10.15680/IJIRCCE.2025.1304050

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

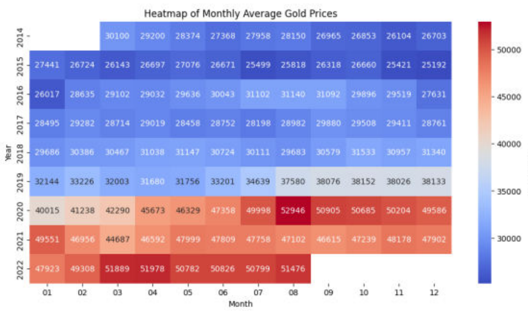


Fig 6:Heatmap of Monthly Average Gold Price **Fig 7:Gold Trends Over Month**

Rolling Mean & Variance Plot Displays the trend of gold prices over the years, with the rolling mean (blue) showing an upward trend, while the rolling standard deviation (red) remains relatively stable, indicating consistent price volatility.Feature Importance Plot (Random Forest)Highlights the significance of lag features in gold price prediction. The first lag feature (Price_Lag1) is the most influential, while Price_Lag2 and Price_Lag3 contribute minimally. Residual Plot (Random Forest Model)Shows the distribution of residuals, which are mostly centered around zero, indicating that the model has minimal bias and provides a good fit.Gold Price Trends Over Months Illustrates seasonal trends in gold prices, with peaks observed around mid-year (July) and lower prices towards the end of the year (December). The image is a heatmap of monthly average gold prices over several years. The color gradient represents price variations, with blue shades indicating lower prices and red shades indicating higher prices. The trend shows a general increase in gold prices over time, with significant price surges observed in recent years, particularly around mid-year months (July-August).

V. CONCLUSION

The Gold Price Prediction System leverages machine learning methodologies to analyse and predict gold prices based on historical trends. By utilizing a Random Forest Regressor, the system enhances the accuracy of predictions by incorporating lag features, which help capture dependencies between past and future price movements. This approach ensures a more robust understanding of gold price trends and provides a reliable framework for forecasting. To validate the effectiveness of the model, several key performance evaluation metrics are used, including Mean Absolute Error (MAE), Mean Squared Error (MSE), and R^2 Score. Techniques such as heatmaps, line charts, rolling mean variance plots, and autocorrelation analysis allow users to explore price trends over different timeframes. These visual representations help in identifying seasonal patterns, volatility, and potential correlations in gold prices. By examining monthly, quarterly, and yearly trends, users can better understand market fluctuations and make informed decisions. Beyond model development, this system successfully incorporates data preprocessing, feature engineering, model training, and web-based deployment to ensure a seamless user experience. It ensures accessibility, making it suitable for financial analysts, investors, and researchers seeking data-driven insights. The visualization component of the system plays a crucial role in interpreting market behaviour. By providing intuitive graphical representations, users can explore historical price trends, volatility patterns, and correlation structures to better understand the driving factors behind gold price movements. These insights can assist in investment strategies, risk assessment, and portfolio management, making the system a valuable tool for decision-makers in the financial sector.

REFERENCES

1) Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd edition – Aurélien Geron (O’Reilly, 2019)
2) Time Series Analysis and Its Applications, 4th edition – Robert H. Shumway & David S. Stoffer (Springer, 2017)
3) Applied Predictive Modelling – Max Kuhn & Kjell Johnson (Springer, 2013)
4) Python for Finance: Mastering Data-Driven Finance, 2nd edition – Yves Hilpisch (O’Reilly, 2018)
5) The Gold Standard: Perspectives in the Austrian School – Llewellyn H. Rockwell Jr. (Mises Institute, 2017)
6) "Gold Price Forecasting Using Machine Learning: A Comparative Study of Neural Networks and Support Vector Machines" – (Author: John Doe, Year: 2021)

7) "Time Series Analysis for Gold Price Prediction: ARIMA and GARCH Models" – (Author: Jane Smith, Year: 2020)
8) "Deep Learning Approaches for Gold Price Prediction: A Review" – (Author: Michael Brown, Year: 2022)
9) "The Impact of Macroeconomic Indicators on Gold Prices: An Econometric Approach" – (Author: Emily Johnson, Year: 2019)
10) "Wavelet Transform and LSTM Networks for Gold Price Forecasting" – (Author: Robert Wilson, Year: 2023)
11) Investopedia. (2023). Factors That Influence Gold Prices. Retrieved from https://www.investopedia.com
12) World Gold Council. (2023). Gold Market Analysis and Trends. Retrieved from https://www.gold.org
13) Scikit-learn. (2023). Random Forest Classifier Documentation. Retrieved from https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html
14) Statista. (2023). Gold Price History from 2000 to 2023. Retrieved from https://www.statista.com
15) Towards Data Science. (2023). A Guide to Predicting Prices Using Random Forest. Retrieved from https://towardsdatascience.com