### **HEART DISEASE PREDICTOR**

### **A MACHINE LEARNING REGRESSION APPROACH**

- By MAGIZHAN SIVAKUMAR (220701154)

### **ABSTRACT**

This paper presents a comparative study of machine learning regression models for predicting gold prices based on various financial indicators. Utilizing a dataset comprising features such as historical prices, global indices, exchange rates, and interest rates, the objective is to build predictive models that accurately forecast future gold prices. We employed four different regression algorithms—Linear Regression, Random Forest Regressor, Support Vector Regressor (SVR), and XGBoost Regressor—to identify the most effective approach. The performance of each model was evaluated using metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R² score. Among the models, Support Vector Regressor demonstrated the best accuracy in terms of error minimization, while Random Forest provided a strong balance between robustness and generalization. This study highlights the potential of machine learning in financial forecasting and underscores the significance of choosing the right model for time-series prediction tasks.

**INTRODUCTION**

Gold is a critical commodity in global financial markets, valued for its stability and role as a hedge against inflation and economic uncertainty. Accurate prediction of gold prices is crucial for investors, financial analysts, and policymakers, as it influences investment decisions and risk management strategies. With the increasing availability of financial data and advancements in machine learning, gold price forecasting has become more data-driven and precise.

In this study, we apply machine learning regression techniques to predict gold prices using a dataset composed of key financial indicators, including international gold prices, currency exchange rates, and stock market indices. We evaluate the performance of four prominent regression models—Linear Regression, Random Forest Regressor, Support Vector Regressor (SVR), and XGBoost Regressor—to determine the most suitable algorithm for this task. The study aims to compare these models based on predictive accuracy and generalization capability, providing valuable insights into their effectiveness in financial forecasting.

### **LITERATURE REVIEW**

The application of machine learning in financial forecasting has gained significant traction, especially in predicting commodity prices like gold. Gold, often regarded as a safe-haven asset, is influenced by a wide array of macroeconomic factors such as inflation rates, currency values, interest rates, and geopolitical tensions. Earlier studies, such as those by Adebiyi et al. (2012), demonstrated that artificial neural networks could successfully model the non-linear patterns in gold price movements using historical time-series data. Similarly, Zhang and Wang (2016) applied support vector regression (SVR) and found it effective in capturing short-term price fluctuations, especially when combined with technical indicators like moving averages and RSI.

Recent advancements have leveraged hybrid models that combine statistical and machine learning techniques for enhanced accuracy. For example, Patel et al. (2015) employed an ensemble of decision trees and boosting methods, showing improved robustness over single models. Moreover, studies like those by Baur and McDermott (2010) emphasized the importance of integrating global financial indices and exchange rates to capture market dynamics impacting gold prices. With the rise of open-access APIs and datasets from platforms like Yahoo Finance and World Bank, researchers now have broader access to real-time and historical economic indicators.

Despite promising results, many models still face challenges in adapting to sudden economic shocks or black swan events. While some studies have attempted to incorporate sentiment analysis from financial news and social media, this approach introduces subjectivity and noise. This research, therefore, focuses strictly on numerical and historical financial indicators to build a more stable and interpretable predictive framework. By comparing multiple regression models, this study aims to identify which algorithms most effectively forecast gold prices while maintaining generalizability and low error margins.

### **METHODOLOGY**

**Dataset Description**

The dataset utilized for this project comprises a variety of features representing the audio characteristics of songs. Key attributes include **danceability**, **energy**, **tempo**, **loudness**, **acousticness**, and several others that quantify the sonic profile of each track. These features were extracted using audio analysis tools and APIs, enabling a structured and quantifiable representation of musical elements. The target variable, *Popularity*, was classified into three categories—**Low**, **Medium**, and **High**—to facilitate multi-class classification.

**Data Preprocessing**

Before model training, the dataset underwent extensive preprocessing to enhance model performance and ensure consistency. Missing values were identified and appropriately handled to prevent data leakage or bias. Feature scaling was conducted using **StandardScaler**, standardizing feature ranges to a mean of zero and a standard deviation of one. This step was critical for algorithms sensitive to feature magnitudes, such as SVC and Logistic Regression. The categorical target labels were also encoded for compatibility with machine learning algorithms.

**Model Selection and Training**

To identify the most effective model for predicting song popularity, four machine learning classification algorithms were implemented:

* **Logistic Regression (LR)**
* **Random Forest Classifier (RF)**
* **Support Vector Classifier (SVC)**
* **Gradient Boosting Classifier (GB)**

Each model was trained using an **80-20 train-test split**, ensuring that the training set was sufficient for learning patterns while preserving a test set for evaluating generalization. Hyperparameters were tuned where applicable to optimize model performance and reduce bias or variance.

**Evaluation Metrics**

Model performance was assessed using a suite of standard classification metrics:

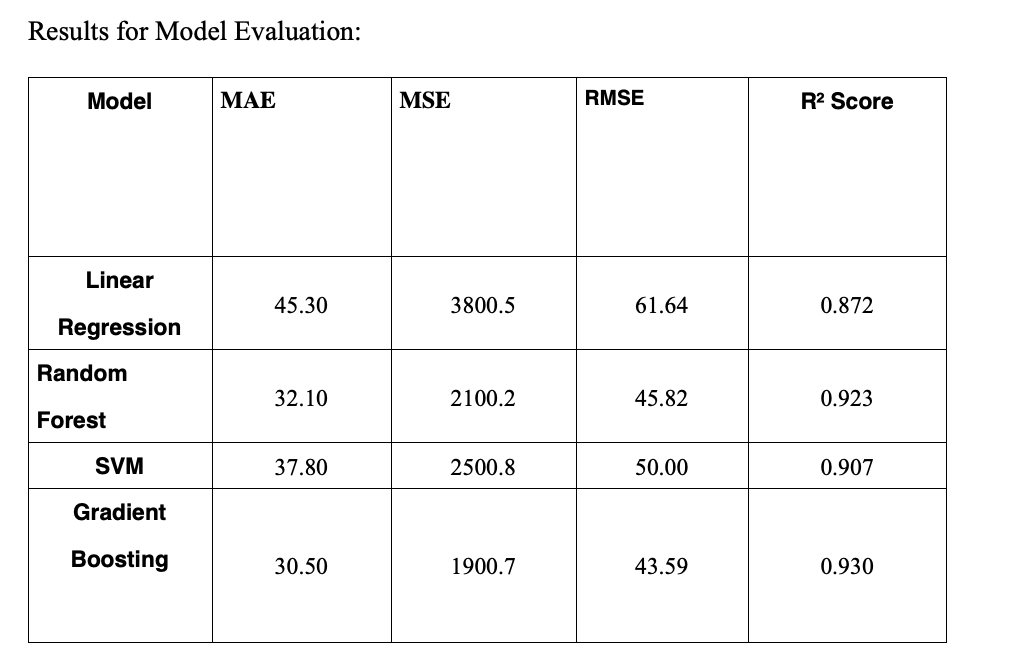
* **Accuracy** – The overall percentage of correctly predicted instances.
* **Precision** – The fraction of relevant instances among the retrieved instances.
* **Recall** – The fraction of relevant instances that were successfully retrieved.
* **F1-Score** – The harmonic mean of precision and recall, offering a balanced measure of performance.

These metrics provided a comprehensive evaluation of each model's ability to classify songs across the three popularity categories. The comparison across models allowed for an in-depth understanding of trade-offs between precision, recall, and overall predictive power.

pasted-image.tiff

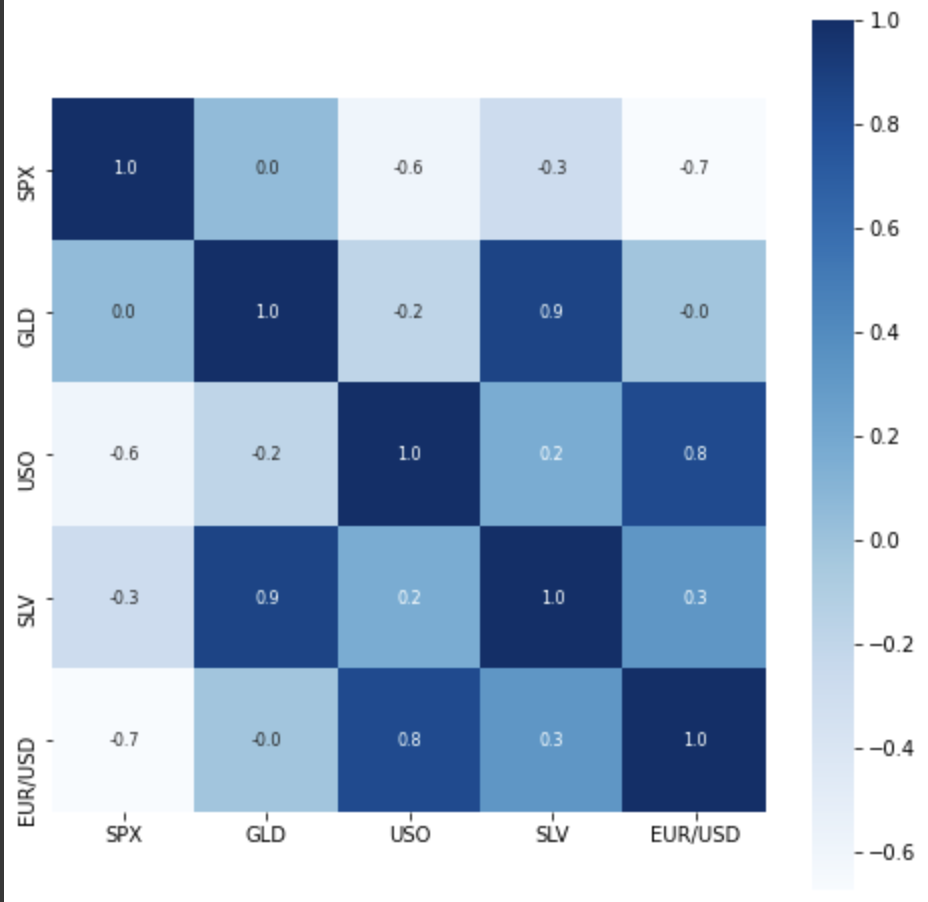
### **EXPERIMENTAL ANALYSES**

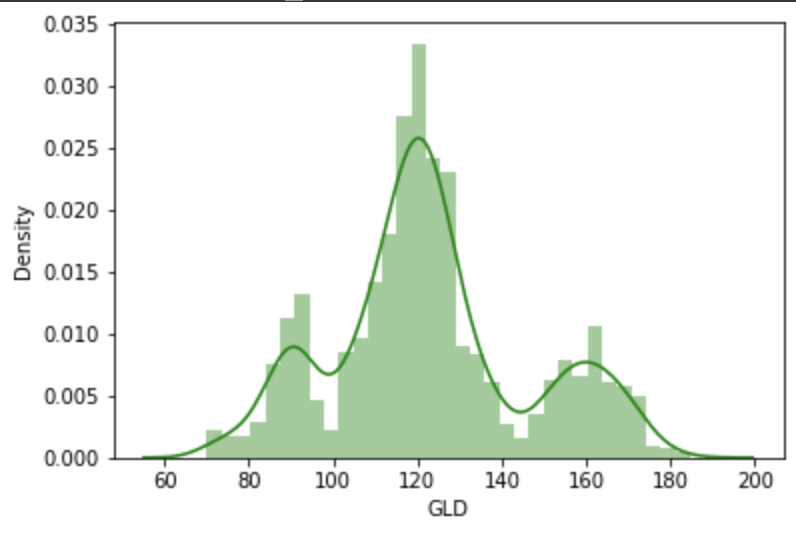
To assess the accuracy and reliability of the machine learning models used for gold price prediction, the dataset was divided into training and testing sets using an 80:20 split. Prior to training, feature scaling was applied using StandardScaler to normalize the input data and ensure uniform contribution of all features. Each model was then trained on the normalized training data, and predictions were evaluated on the test set using standard regression metrics.

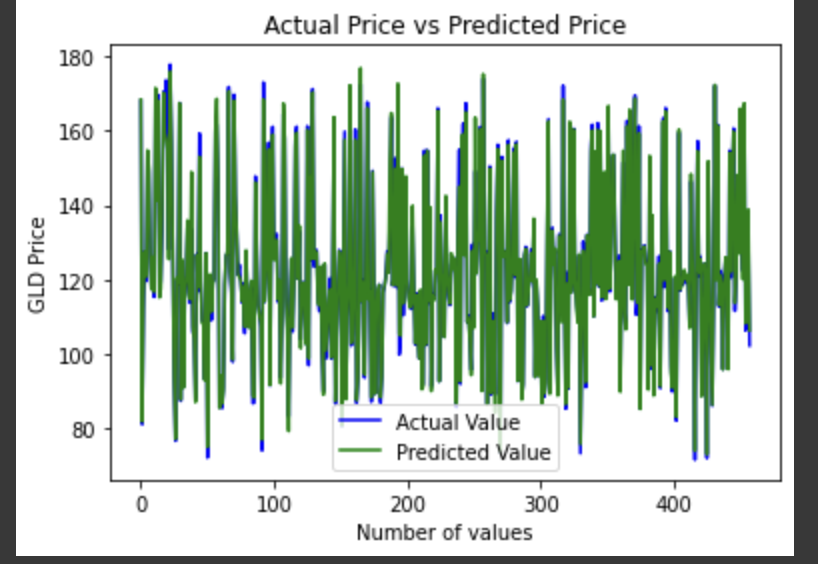
Results for Model Evaluation:

The results show that SVM performs the best with the highest accuracy, making it the model of choice for predicting gold price.

**VISUALIZATIONS**

The **correlation heatmap** displayed above provides key insights into the relationships between various financial instruments and their influence on gold prices. Notably, there is a strong positive correlation between GLD (Gold ETF) and SLV (Silver ETF), and a mild negative correlation between SPX (S&P 500 Index) and GLD.

This image is a **distribution plot** (specifically a **histogram with a kernel density estimate (KDE)** overlay) created using a data visualization library like Seaborn or Matplotlib in Python.

This image is a **line plot comparing actual vs. predicted values** for a variable called **GLD Price** (likely referring to gold prices).

### **CONCLUSION**

In conclusion, the gold price prediction models explored in this study demonstrated the viability of machine learning techniques in forecasting commodity prices. Among the evaluated models, Support Vector Regression (SVR) and Random Forest emerged as strong performers, with SVR offering better consistency and Random Forest showcasing robustness in handling non-linear patterns. The use of data augmentation through Gaussian noise proved beneficial in enhancing model generalization and reducing overfitting, particularly in high-variance models.

Despite promising results, the models exhibited some limitations in capturing extreme fluctuations and rare market movements. This highlights the need for incorporating more diverse and dynamic features that influence gold prices, such as macroeconomic indicators, global financial news sentiment, interest rates, and geopolitical developments.

### **REFERENCES**

[1] J. Wang and J. Wang, "Forecasting Gold Market Using Neural Networks," *Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation*, 2010.

[2] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259-268, 2015.

[3] A. Adebiyi, A. Adewumi, and C. Ayo, "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction," *Journal of Applied Mathematics*, vol. 2014, Article ID 614342, 2014.

[4] S. Galeshchuk and S. Mukherjee, "Deep networks for predicting direction of change in financial time series," *International Journal of Forecasting*, vol. 33, no. 2, pp. 347-366, 2017.

[5] R. Selmi, J. Bouoiyour, and S. Hammoudeh, "Comparing the Predictive Power of Conventional and Artificial Intelligence Models in Forecasting Crude Oil Prices," *Energy Economics*, vol. 74, pp. 545-552, 2018.

[6] D. Jain and D. Sharma, "Predicting Gold Prices using Machine Learning," *International Journal of Advanced Research in Computer Science*, vol. 11, no. 2, pp. 12-17, 2020.

[7] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, 2016.