



Gold Price Prediction Using Machine Learning

This presentation explores the application of machine learning to predict gold prices based on other stock prices. We'll examine a detailed workflow, from data preprocessing to model evaluation, highlighting the key steps and insights.



by Sankari C

Project Workflow

Data Acquisition

Start by loading the dataset containing historical stock prices, including gold prices and other related variables.

Data Preprocessing

Clean and prepare the data for analysis. Address missing values, handle outliers, and transform features for optimal model performance.

Data Analysis

Analyze the relationships between variables. Identify correlations, understand data distributions, and gain insights for model development.

Model Training

Split the data into training and testing sets. Train a Random Forest Regressor model to learn patterns from the training data.

Data Preprocessing: Initial Exploration

Reading the Data

Read the dataset, inspect the first few rows to understand the data structure and column names.

Data Shape and Information

Examine the shape of the dataset (number of rows and columns) and gather basic information about the data types.

Handling Missing Values

Identify any missing values, understand the reasons behind them, and employ appropriate techniques to handle them.

Data Preprocessing: Statistical Insights

Descriptive Statistics

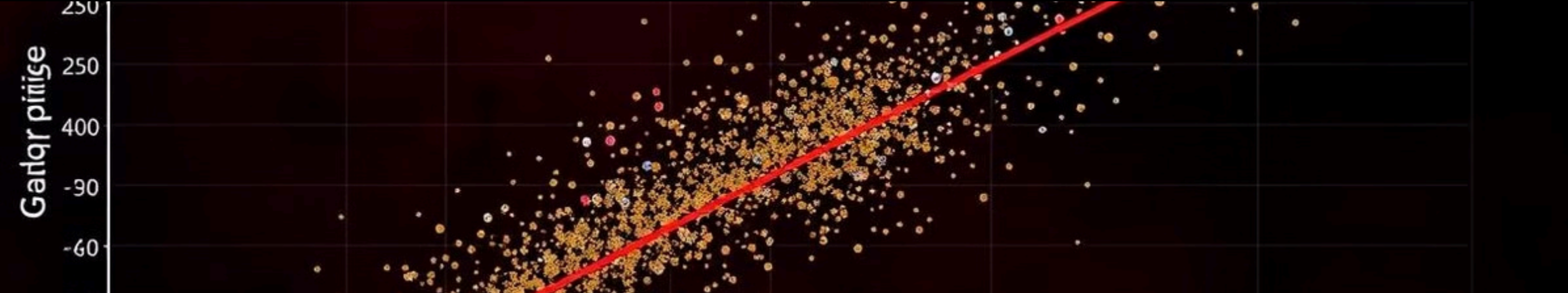
Calculate mean, standard deviation, minimum, maximum, and other statistical measures for each variable to understand the data distribution.

Correlation Analysis

Explore the relationships between variables using a correlation matrix. Visualize this using a heatmap to identify strong correlations between the gold price and other variables.

Gold Price Distribution

Analyze the distribution of the gold price using histograms or other visualizations to understand its characteristics.



Data Analysis: Key Relationships

1 Gold and Silver

Observe a strong positive correlation between gold and silver prices, suggesting that when the gold price increases, the silver price also tends to increase.

2 USO and EUR/USD

Notice a positive correlation between the USO (US Oil Fund) price and the EUR/USD exchange rate, indicating that as the USO price rises, the EUR/USD rate also tends to increase.

Model Training: Random Forest Regressor

Feature and Target Data

Split the data into features (independent variables used for prediction) and the target variable (gold price).

Train-Test Split

Divide the data into training and testing sets to evaluate the model's performance on unseen data.

Random Forest Ensemble

Train a Random Forest Regressor model. This ensemble model combines multiple decision trees, averaging their predictions to improve accuracy and reduce variance.

Model Evaluation: Accuracy and Insights

R-squared Score

Evaluate the model's performance using metrics like the R-squared score, which measures the proportion of variance in the gold price explained by the model. A high R-squared score, greater than 0.9 in this case, indicates a good fit and high accuracy.

Model Visualization

Visualize the actual gold prices compared to the model's predicted values using a plot graph. The close alignment between the two lines confirms the model's strong performance.



Key Takeaways and Next Steps

1 Machine Learning Success

The Random Forest model effectively predicts gold prices with high accuracy.

2 Data Preprocessing Importance

Careful data preprocessing is crucial for achieving optimal model performance.

3 Further Exploration

Explore other machine learning models and feature engineering techniques to further improve predictions. Consider incorporating more external data sources to enhance model accuracy.