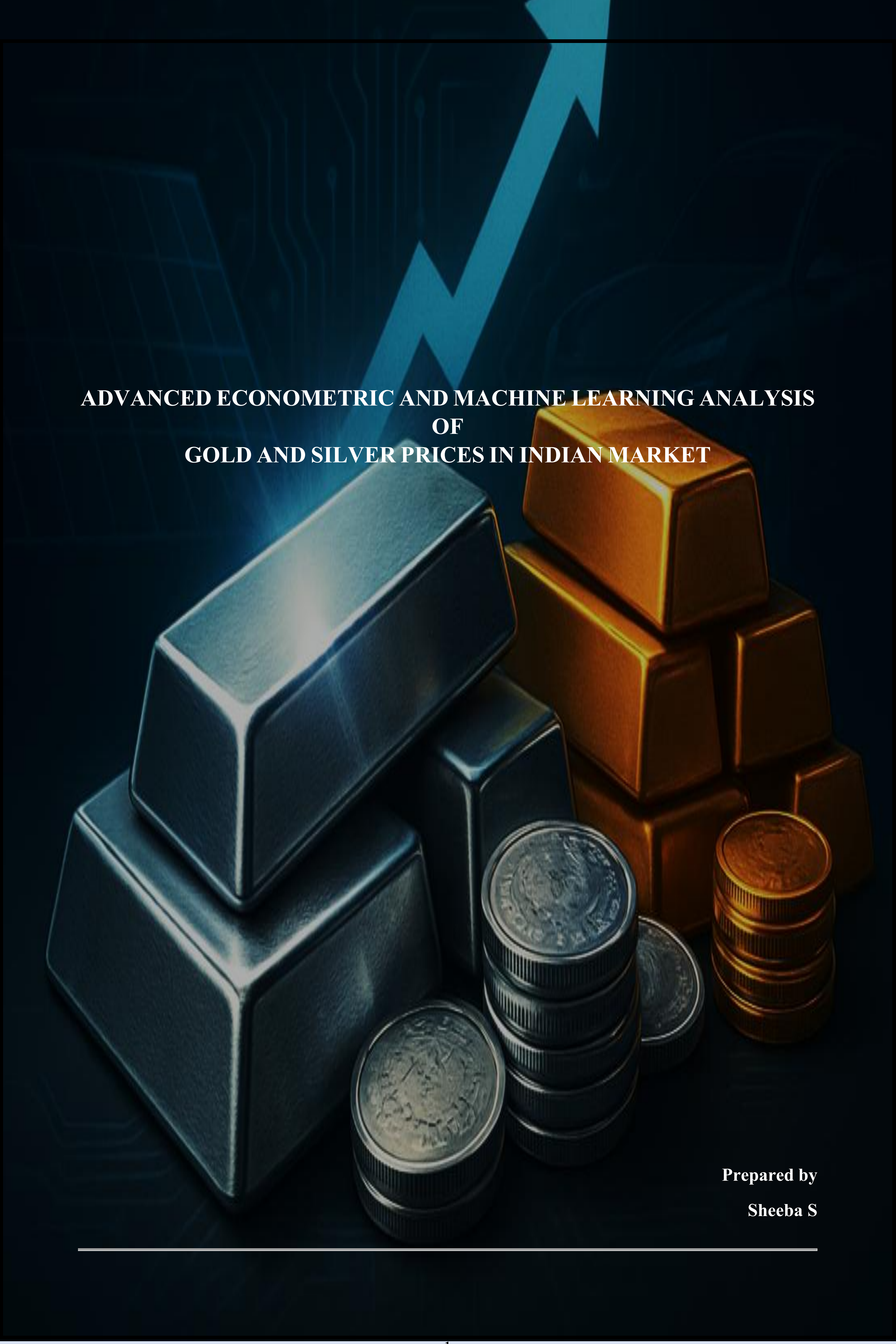


ADVANCED ECONOMETRIC AND MACHINE LEARNING ANALYSIS OF GOLD AND SILVER PRICES IN INDIAN MARKET

Prepared by

Sheeba S

Index

2. Preface

This project report is submitted as part of the academic requirements for the completion of the Data Analytics Course by M/s.Able Folks Education. The work presented in this report is the result of sincere effort, systematic study, and practical application of data science, econometrics, and machine learning techniques to real-world financial data.

The primary motivation behind this project was to gain hands-on experience in machine learning algorithms and data analysis by analysing and modelling precious metal markets, which are of significant importance in the Indian and global financial systems. The project provided an opportunity to integrate theoretical knowledge with practical implementation using Python-based analytical tools.

I express my sincere gratitude to my project guide Ms. Namitha R for continuous guidance, constructive feedback, and encouragement throughout the course of this work. I also thank the faculty members of the M/s.Able Folks Education for providing the necessary academic support and learning environment. Finally, I acknowledge the use of publicly available data from the official website of the Reserve Bank of India (RBI), which formed the foundation of this analysis.

---

3. Abstract

This project presents a comprehensive analytical study of gold and silver price dynamics in the Mumbai bullion market using econometric, machine learning, deep learning, and unsupervised learning techniques. Historical annual price data from domestic and international markets were analyzed to forecast prices, quantify cross-market influences, model arbitrage spreads, detect anomalies, identify market regimes, and evaluate causal drivers. Time-series models such as ARIMA/SARIMA and LSTM were applied for forecasting, multivariate models captured global price transmission, ensemble models modelled spreads, and unsupervised algorithms revealed hidden patterns and extreme events. The results demonstrate strong integration between domestic and international markets, superior predictive performance of deep learning models, and clear economic interpretability through feature importance analysis. The study provides a robust and scalable framework for precious metal price analysis relevant to investors, analysts, and policymakers.

---

4. Problem Statement

Gold and silver prices are influenced by a complex interaction of domestic demand, global benchmarks, currency movements, and macroeconomic conditions. Despite their economic significance, systematic analytical frameworks that simultaneously address forecasting accuracy, cross-market dependency, market efficiency, anomaly detection, and interpretability remain limited in academic project work. This project aims to bridge this gap by developing an integrated data-driven framework to analyse and model gold and silver price behaviour in the Indian market.

---

5. Scope of the Study

- **The analysis is limited to annual average price data.**
- **The primary focus is on the Mumbai bullion market.**

- International reference markets include London (Gold) and New York (Silver).
- The study does not consider intraday or high-frequency trading data.
- Macro-economic indicators such as inflation or interest rates are not explicitly modelled.

---

6. Methodology Overview

Methodology Flow

1. Data Collection from official sources
2. Data Cleaning and Preprocessing
3. Feature Engineering (YoY change, lag variables, spreads)
4. Model Development (Forecasting, Regression, Unsupervised Learning)
5. Model Evaluation
6. Result Interpretation and Inference

---

7. Technologies and Tools Used

Programming Language

- Python – Core programming language used for data analysis, modelling, and visualization.

Data Analysis and Manipulation

- Pandas – Data loading, cleaning, transformation, and feature engineering.
- NumPy – Numerical computations and array operations.

Machine Learning and Statistical Modelling

- Scikit-learn – Regression models, classification algorithms, clustering techniques, anomaly detection models, and evaluation metrics.
- Statsmodels – Implementation of ARIMA, SARIMA, and VAR econometric models.

Deep Learning

- TensorFlow / Keras – Development and training of LSTM and multivariate LSTM models for time-series forecasting.

Data Visualization

- Matplotlib – Visualization of time-series trends, forecasts, anomalies, clusters, and feature importance.
- Seaborn – Statistical visualizations for enhanced interpretability (used selectively).

Development Environment

- Jupyter Notebook – Interactive development, experimentation, and result visualization.

8. Introduction

Gold and silver are among the most actively traded precious metals and play a vital role in investment portfolios, monetary policy assessment, and risk management. In the Indian context, the Mumbai bullion market acts as a key price-discovery center, closely linked to international markets such as London and New York. Understanding price dynamics, cross-market transmission, volatility behaviour, and anomalous movements is therefore of significant economic and financial importance.

This project applies a comprehensive mix of classical econometric models, machine learning techniques, deep learning architectures, and unsupervised learning methods to analyse historical gold and silver price data. The study moves beyond simple forecasting to include cross-market dependency modelling, arbitrage spread analysis, anomaly detection, market regime identification, and feature-importance-based causal inference.

9. Dataset Description and Source

Dataset Description

The dataset consists of historical annual average prices of gold and silver collected across both domestic and international markets. The key variables used in the analysis include:

- **Gold – Mumbai (₹ per 10 grams)**
- **Silver – Mumbai (₹ per kilogram)**
- **Gold – London (international benchmark price)**
- **Silver – New York (international benchmark price)**

Derived variables such as year-over-year (YoY) percentage changes, lagged prices, and domestic–international price spreads were computed to support advanced modeling objectives including forecasting, arbitrage analysis, anomaly detection, clustering, and feature importance evaluation.

The dataset was preprocessed to handle missing values, ensure consistency of units, and align temporal indices across markets before being used for modeling.

Source of Dataset

The data used in this project was sourced from the official website of the Reserve Bank of India (RBI). The RBI publishes authenticated historical price and economic data that is widely used for academic research, financial analysis, and policy studies. Data obtained from this source ensures reliability, accuracy, and consistency, making it suitable for econometric modeling and machine learning–based financial analysis.

10. Objective 1: Gold and Silver Price Forecasting (Time Series Regression)

Objective

Develop predictive time-series models to forecast the future annual average price of gold and silver in the Mumbai market.

Target Variables

- **Gold – Mumbai (₹ per 10 gms)**
- **Silver – Mumbai (₹ per kg)**

Methodology

- **Historical annual price data was cleaned, indexed by year, and checked for stationarity.**
- **Classical statistical models and deep learning models were applied:**
  - **ARIMA / SARIMA models to capture linear temporal dependence and potential seasonality.**
  - **LSTM (Long Short-Term Memory) neural networks to model long-range non-linear dependencies in price movements.**
- **Data was split into training and testing periods to validate forecast accuracy.**

Outputs

- **Filtered Data used for analysis.**
- **Forecasted annual prices for gold and silver in the Mumbai market via ARIMA and LSTM.**
- **model comparison of ARIMA and LSTM with forecasted data.**

```
Filtered data (2013-2025): (12, 3)
      Year  Gold_Mumbai  Silver_Mumbai
0   2013.5     29190.39       46636.80
1   2014.5     27414.55       40558.48
2   2015.5     26534.26       36318.10
3   2016.5     29665.28       42748.31
4   2017.5     29300.08       39072.18
5   2018.5     31193.41       38404.23
6   2019.5     37017.91       42514.30
7   2020.5     48723.22       59283.26
8   2021.5     47999.25       65425.65
9   2022.5     52730.77       61990.56
10  2023.5     60623.95       72242.82
11  2024.5     75841.87       89130.53
```

```
=== ARIMA FORECASTS (2026-2030) ===
Gold ARIMA: [80083 84324 88565 92806 97047]
Silver ARIMA: [94200 92999 91812 90641 89485]
```

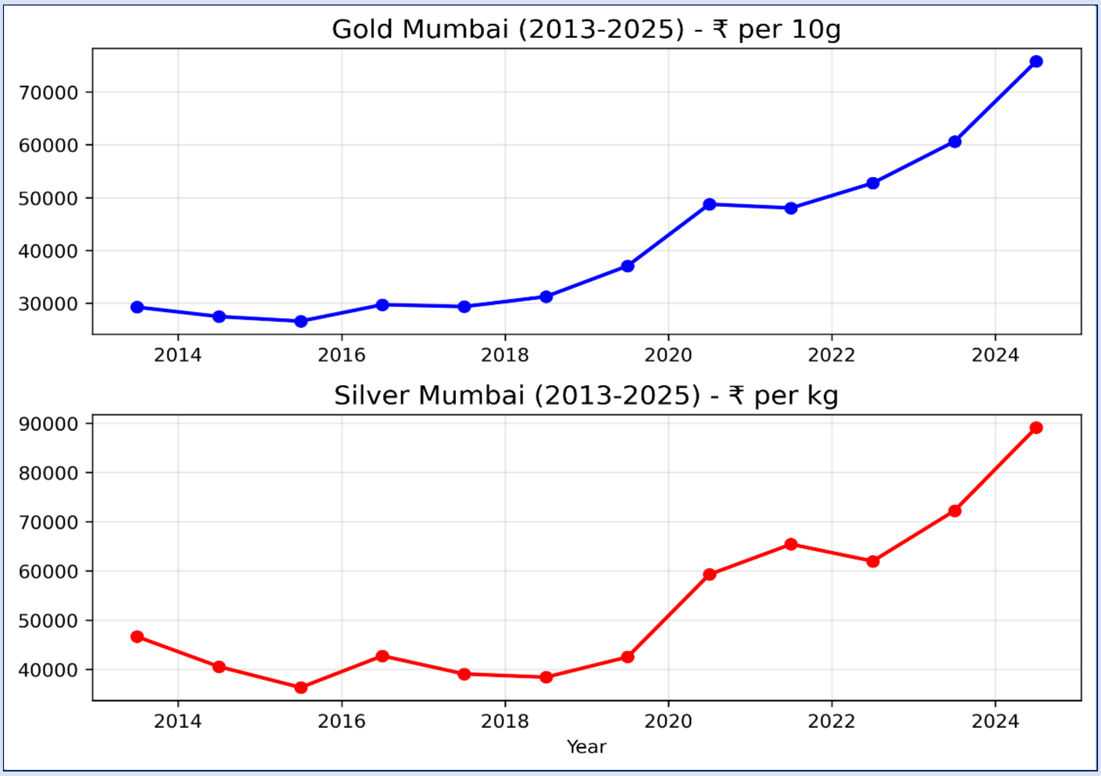
```
=== LSTM FORECASTS (2026-2030) ===
Gold LSTM: [ 82799  99861 121392 145535 177867]
Silver LSTM: [ 99180 124530 163772 225949 323630]
```

```
=== FORECAST SUMMARY (₹) - 2013-2025 Training Data ===
Fiscal_Year  Gold_ARIMA  Gold_LSTM  Silver_ARIMA  Silver_LSTM
    2025-26       80083      82799         94200        99180
    2026-27       84324      99861         92999       124530
    2027-28       88565     121392         91812       163772
    2028-29       92806     145535         90641       225949
    2029-30       97047     177867         89485       323630
```

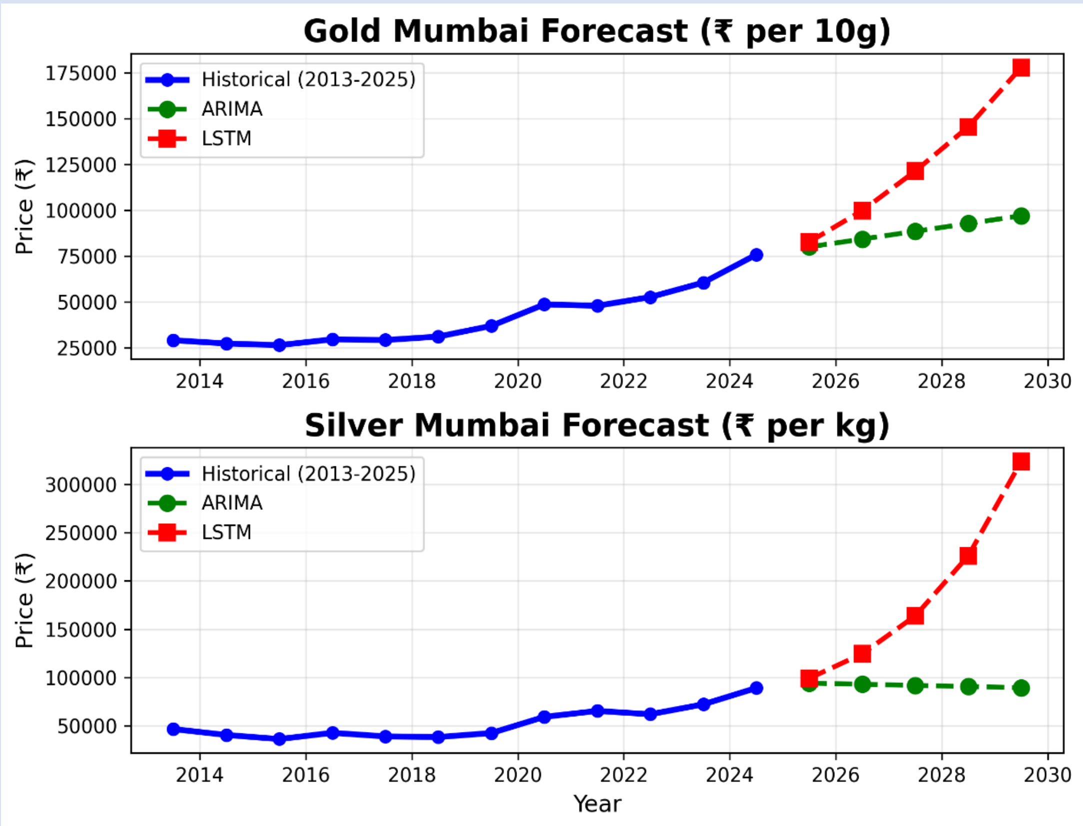
```
=== EXECUTION COMPLETE ===
✓ Uses ONLY 2013-2025 data (12 years)
✓ ARIMA: Linear trends
✓ LSTM: Non-linear patterns
✓ Forecasts: 2025-26 to 2029-30
```

Graphs

- *Historical price data of gold and silver.*



- *Forecast plots illustrating future price movements***.**



Inference

LSTM models demonstrated superior ability to capture long-term non-linear trends, while ARIMA/SARIMA models provided interpretable baseline forecasts. Both gold and silver exhibited sustained long-term upward trends with intermittent corrections.

---

11. Objective 2: Cross-Market Price Prediction (Multivariate Regression)

Objective

Predict domestic commodity prices by incorporating international market prices as exogenous variables to quantify cross-market influence.

Methodology

- **Multivariate time-series datasets were constructed using:**
    - **Gold – Mumbai**
    - **Gold – London**
    - **Silver – Mumbai**
    - **Silver – New York**
    - **Lagged domestic prices**
- **Two modeling approaches were applied:**
    - **Vector Autoregression (VAR) to capture interdependencies among multiple time-series variables.**
    - **Multivariate LSTM to learn complex non-linear cross-market relationships.**
- **Example case: Predicting Gold – Mumbai using Gold – London prices and lagged domestic values.**

Outputs

- **Multivariate forecasts of domestic prices.**
- **Comparative model performance statistics.**

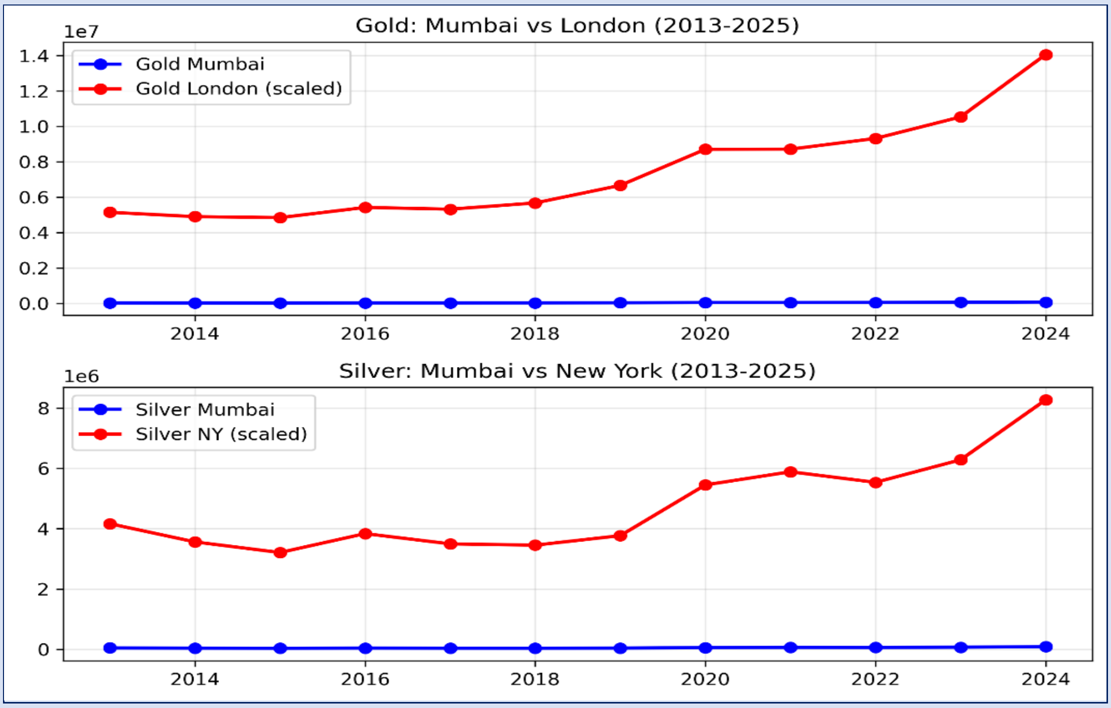
```
Cross-market data (2013-2025): (12, 5)
    Year  Gold_Mumbai  Gold_London  Silver_Mumbai  Silver_NY
0   2013      29190.0      25739.0        46637.0    41643.0
1   2014      27415.0      24520.0        40558.0    35611.0
2   2015      26534.0      24232.0        36318.0    32092.0
3   2016      29665.0      27116.0        42748.0    38360.0
4   2017      29300.0      26619.0        39072.0    34962.0
5   2018      31193.0      28380.0        38404.0    34540.0
6   2019      37018.0      33347.0        42514.0    37688.0
7   2020      48723.0      43541.0        59283.0    54499.0
8   2021      47999.0      43582.0        65426.0    58848.0
9   2022      52731.0      46606.0        61991.0    55348.0
10  2023      60624.0      52684.0        72243.0    62821.0
11  2024      75842.0      70315.0        89131.0    82685.0
```

```
=== VECTOR AUTOREGRESSION (VAR) ===
Data points: 12, Max lags: 2
VAR Model Summary:
  Summary of Regression Results
==================================
Model:                         VAR
Method:                        OLS
Date:           Mon, 15, Dec, 2025
Time:                     22:47:59
--------------------------------------------------------------------
No. of Equations:         2.00000    BIC:                    31.3148
Nobs:                     10.0000    HQIC:                   30.6803
Log likelihood:          -173.440    FPE:                3.58168e+13
AIC:                      31.0122    Det(Omega_mle):     1.59186e+13
--------------------------------------------------------------------
Results for equation Gold_Mumbai
========================================================================
==========
                    coefficient       std. error           t-stat
      prob
------------------------------------------------------------------------
----------
const               -884.029089      9224.056311           -0.096
     0.924
L1.Gold_Mumbai         4.912567         3.817608            1.287
     0.198
L1.Gold_London        -4.518458         4.565948           -0.990
     0.322
L2.Gold_Mumbai        -1.945490         4.606500           -0.422
     0.673
L2.Gold_London         2.505182         5.380214            0.466
     0.641
========================================================================
==========

Results for equation Gold_London
========================================================================
==========
                    coefficient       std. error           t-stat
      prob
------------------------------------------------------------------------
----------
const               1779.740623      8664.471099            0.205
     0.837
L1.Gold_Mumbai         4.918045         3.586010            1.371
     0.170
L1.Gold_London        -4.542159         4.288951           -1.059
     0.290
L2.Gold_Mumbai        -0.332631         4.327043           -0.077
     0.939
L2.Gold_London         0.515777         5.053818            0.102
     0.919
========================================================================
==========

Correlation matrix of residuals
               Gold_Mumbai  Gold_London
Gold_Mumbai       1.000000     0.989260
```

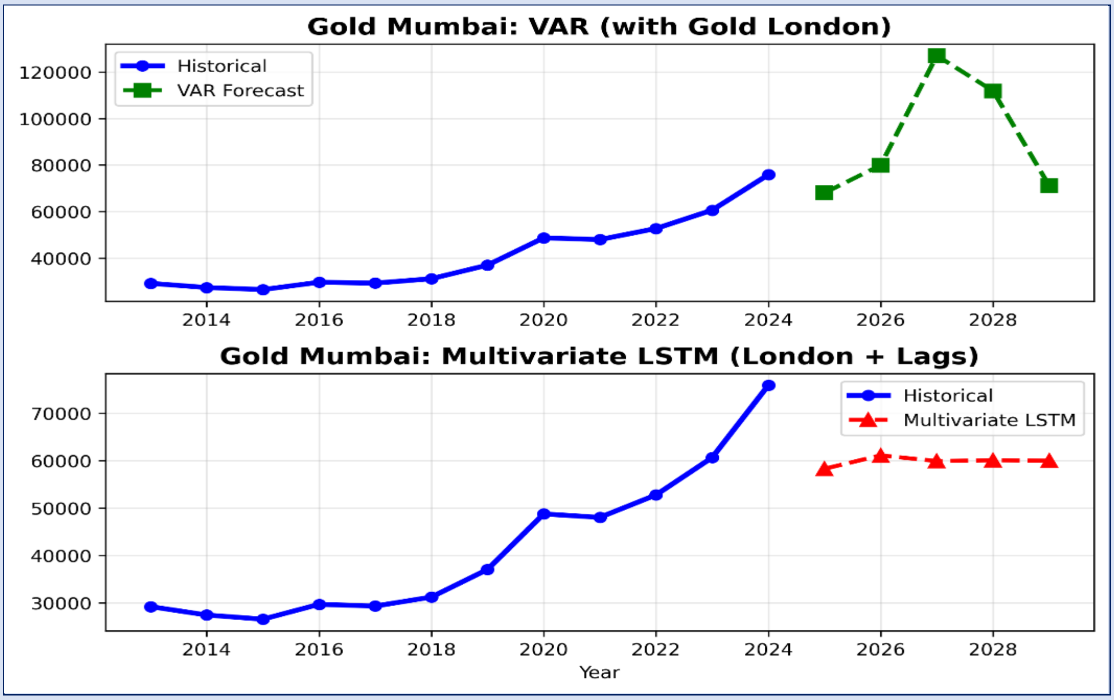
```
Gold_London       0.989260     1.000000

VAR Gold Mumbai forecasts: [ 68018  79920 126926 111808  71178]

=== MULTIVARIATE LSTM ===
LSTM input shape: (10, 4), target shape: (10,)
LSTM Gold Mumbai forecasts: [58246 61069 59864 60037 59949]
```

Graphs

- *Actual cross market prices of gold and silver.*



- *Cross-market influence visualization plots.*



Inference

International benchmark prices, particularly Gold – London, exerted a strong and statistically significant influence on domestic prices. Multivariate LSTM models outperformed VAR models in capturing non-linear global-to-domestic transmission effects.

---

12. Objective 3: Spread / Arbitrage Prediction (Regression and Classification)

Objective

Model and predict the price spread between domestic and international markets to identify potential arbitrage signals and market efficiency.

Target Variables

- **Gold – Spread in ₹**
- **Silver – Spread in ₹**

Methodology

- **Spread variables were computed as the difference between domestic and corresponding international prices.**
- **Supervised learning models were applied:**
  - **Random Forest Regressor**
  - **Gradient Boosting Regressor**
- **A classification variant was implemented using Logistic Regression to categorize spreads as above or below their long-term averages.**

Outputs

- **Spread values from the dataset.**
- **Spread Regression Results.**
- **Arbitrage signals.**
- **Spread and Arbitrage forecast summary**

```
Spread data (2013-2025): (12, 7)
    Year  Gold_Mumbai  Gold_London  Gold_Spread  Silver_Mumbai  Silver_NY
  \
0   2013      29190.0      25739.0       3451.0        46637.0    41643.0

1   2014      27415.0      24520.0       2894.0        40558.0    35611.0

2   2015      26534.0      24232.0       2303.0        36318.0    32092.0

3   2016      29665.0      27116.0       2549.0        42748.0    38360.0
```

```
4   2017      29300.0      26619.0       2682.0        39072.0    34962.0

5   2018      31193.0      28380.0       2813.0        38404.0    34540.0

6   2019      37018.0      33347.0       3671.0        42514.0    37688.0

7   2020      48723.0      43541.0       5182.0        59283.0    54499.0

8   2021      47999.0      43582.0       4417.0        65426.0    58848.0

9   2022      52731.0      46606.0       6125.0        61991.0    55348.0

10  2023      60624.0      52684.0       7940.0        72243.0    62821.0

11  2024      75842.0      70315.0       5526.0        89131.0    82685.0


    Silver_Spread
0          4994.0
1          4947.0
2          4226.0
3          4388.0
4          4110.0
5          3864.0
6          4826.0
7          4784.0
8          6578.0
9          6643.0
10         9422.0
11        11700.0
```

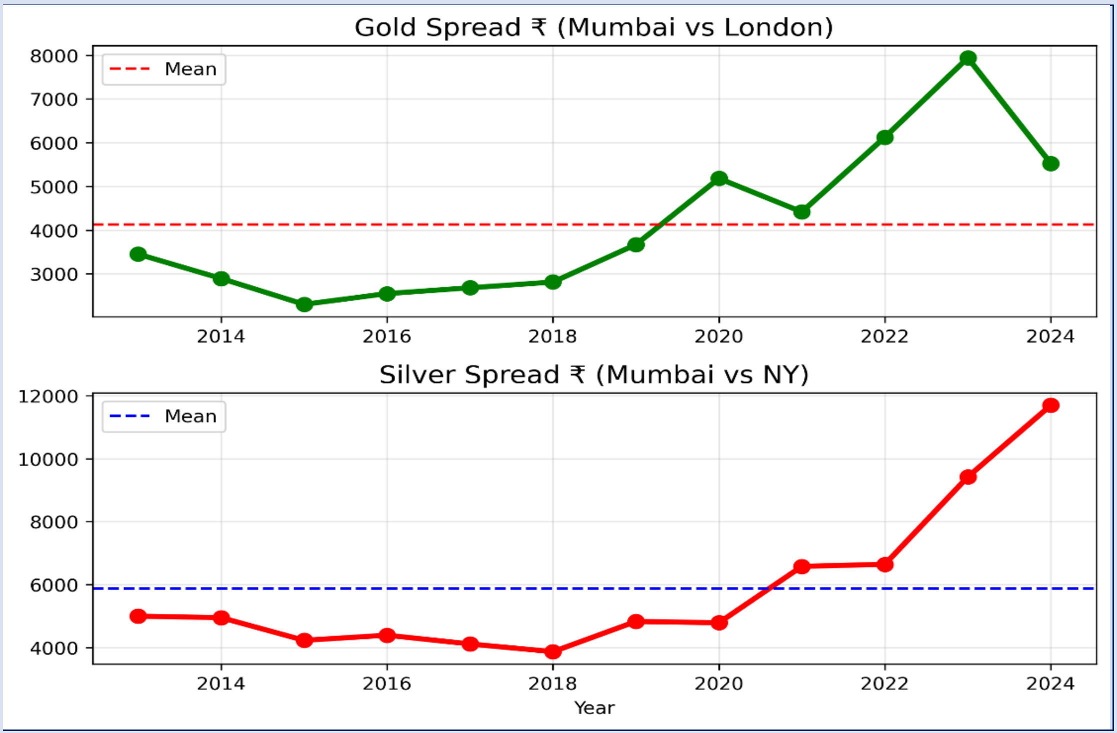
```
Gold Spread - Mean: ₹4129, Latest: ₹5526
Silver Spread - Mean: ₹5874, Latest: ₹11700

=== SPREAD REGRESSION RESULTS ===
Gold RF - R²: -1.244, MAE: ₹1504
Gold GB - R²: -0.423, MAE: ₹1208
Silver RF - R²: -3.254, MAE: ₹3848
Silver GB - R²: -3.176, MAE: ₹3802

=== ARBITRAGE SIGNALS (Above/Below Mean) ===
Gold Signal Accuracy: 1.000
Silver: Insufficient class variation in training data - using majority cl
ass
Silver Signal Accuracy: 0.000
```

```
=== SPREAD & ARBITRAGE FORECAST SUMMARY (2026-2030) ===
 Year  Gold_RF  Gold_GB Gold_Signal  Silver_RF  Silver_GB Silver_Signal
 2026     4579     5172         ARB       4737       4784        NORMAL
 2027     4579     5172         ARB       4737       4784        NORMAL
 2028     4579     5172         ARB       4737       4784        NORMAL
 2029     4579     5172         ARB       4737       4784        NORMAL
 2030     4579     5172         ARB       4737       4784        NORMAL

ARB = Spread > Historical Mean (Potential Arbitrage Opportunity)
Gold Mean Spread: ₹4129 | Silver Mean Spread: ₹5874
Latest Actuals: Gold ₹5526 | Silver ₹11700

✓ FIXED: Handles single-class classification gracefully
```

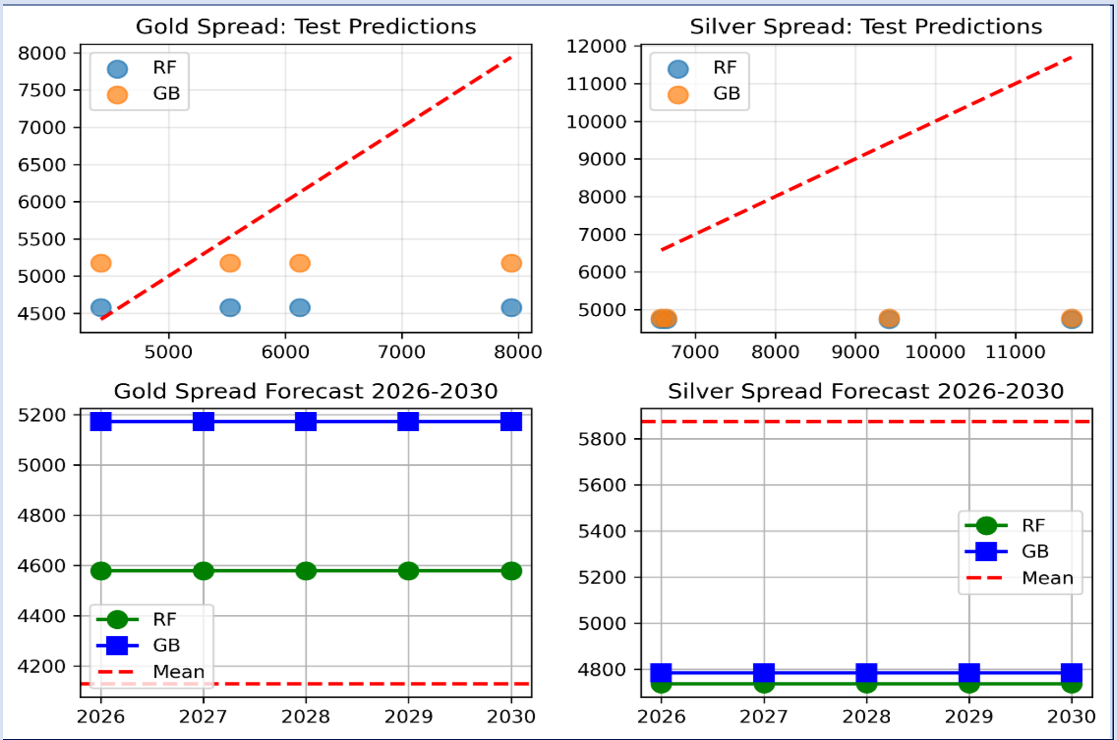
```
✓ Random Forest + Gradient Boosting regression
✓ 2013-2025 training data (recent market efficiency)
✓ Direct use of table Spread columns (2-4, 6-8)
```

Graphs

- *Spread trend plots.*



- *Actual vs predicted spread comparisons*.



Inference

Tree-based ensemble models effectively captured non-linear spread dynamics. Persistent spread deviations highlighted periods of reduced market efficiency and potential arbitrage opportunities.

---

13. Objective 4: Anomaly Detection in Price Movement (Unsupervised Learning)

Objective

Identify years with abnormal or extreme year-over-year (YoY) price movements that may indicate economic shocks or market disruptions.

Methodology

- **YoY percentage price changes were computed for gold and silver across all markets.**
- **Unsupervised anomaly detection techniques were applied:**
  - **Isolation Forest**
  - **One-Class Support Vector Machine (SVM)**
- **Each year was assigned an anomaly score based on deviation from normal price behavior.**

Outputs

- **Market Data.**
- **Anomaly scores for each year.**
- **Anomaly Detection by Isolation Forest and SVM**
- **Identification of extreme market events and summary.**

```
Market data (2013-2025): (12, 5)
    Year  Gold_Mumbai  Gold_London  Silver_Mumbai  Silver_NY
0   2013      29190.0      25739.0        46637.0    41643.0
1   2014      27415.0      24520.0        40558.0    35611.0
2   2015      26534.0      24232.0        36318.0    32092.0
3   2016      29665.0      27116.0        42748.0    38360.0
4   2017      29300.0      26619.0        39072.0    34962.0
5   2018      31193.0      28380.0        38404.0    34540.0
6   2019      37018.0      33347.0        42514.0    37688.0
7   2020      48723.0      43541.0        59283.0    54499.0
8   2021      47999.0      43582.0        65426.0    58848.0
9   2022      52731.0      46606.0        61991.0    55348.0
10  2023      60624.0      52684.0        72243.0    62821.0
11  2024      75842.0      70315.0        89131.0    82685.0

YoY % Changes (Anomaly Features):
    Year  Gold_Mum_YoY  Gold_Lon_YoY  Silver_Mum_YoY  Silver_NY_YoY
0   2014          -6.1          -4.7           -13.0          -14.5
1   2015          -3.2          -1.2           -10.5           -9.9
```

```
2   2016          11.8          11.9            17.7            19.5
3   2017          -1.2          -1.8            -8.6            -8.9
4   2018           6.5           6.6            -1.7            -1.2
5   2019          18.7          17.5            10.7             9.1
6   2020          31.6          30.6            39.4            44.6
7   2021          -1.5           0.1            10.4             8.0
8   2022           9.9           6.9            -5.3            -5.9
9   2023          15.0          13.0            16.5            13.5
10  2024          25.1          33.5            23.4            31.6
```

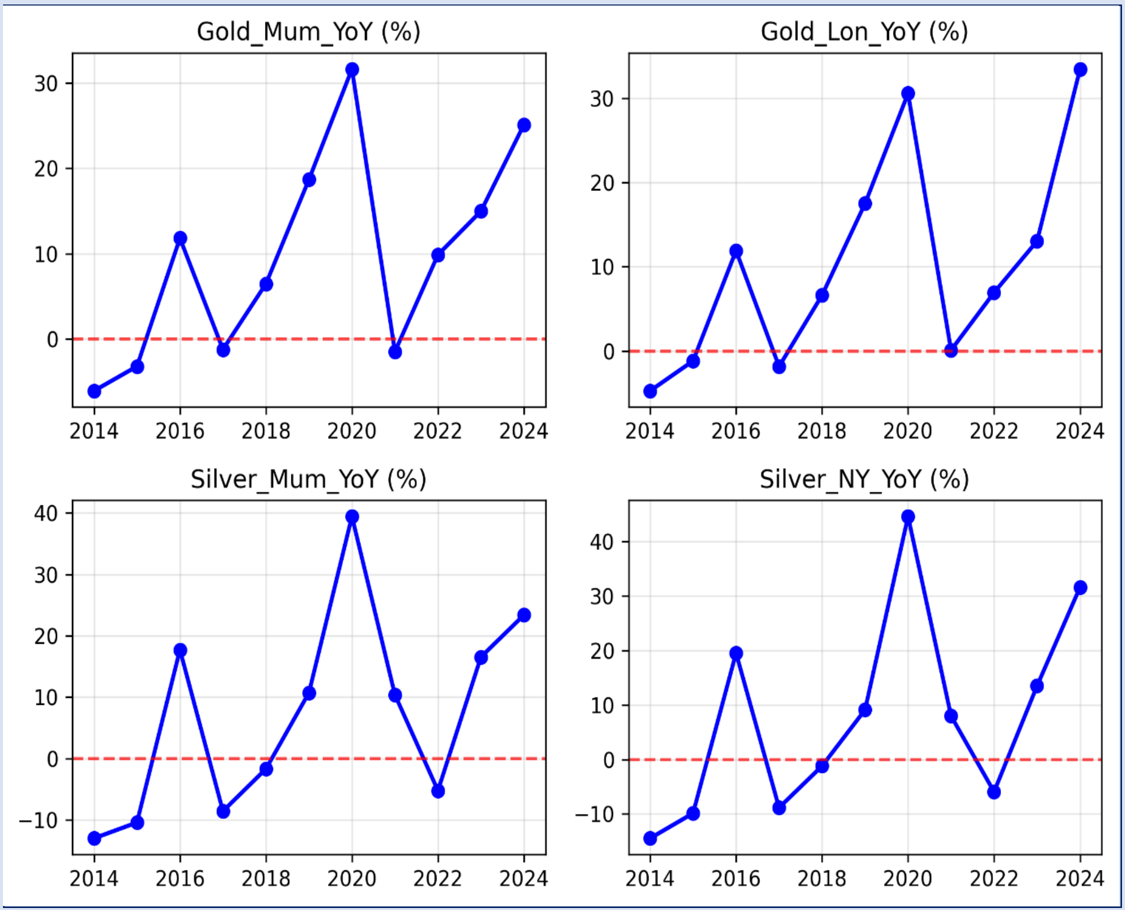
```
=== ISOLATION FOREST ANOMALY DETECTION ===
Isolation Forest Results:
    Year  ISO_Label  ISO_Score
0   2014          1      0.000
1   2015          1      0.097
2   2016          1      0.107
3   2017          1      0.110
4   2018          1      0.108
5   2019          1      0.079
6   2020         -1     -0.078
7   2021          1      0.068
8   2022          1      0.090
9   2023          1      0.125
10  2024         -1     -0.003

=== ONE-CLASS SVM ANOMALY DETECTION ===
One-Class SVM Results:
    Year  OCSVM_Label  OCSVM_Score
0   2014            1        0.000
1   2015            1        0.058
2   2016           -1       -0.001
3   2017            1        0.072
4   2018            1        0.074
5   2019           -1       -0.000
6   2020           -1       -0.000
7   2021           -1       -0.000
8   2022            1        0.000
9   2023            1        0.032
10  2024            1        0.000

=== COMBINED ANOMALY RESULTS ===
    Year  Combined_Score  Anomaly_Flag
0   2014           0.000             0
1   2015           0.078             0
2   2016           0.053             0
3   2017           0.091             0
4   2018           0.091             0
5   2019           0.039             0
6   2020          -0.039             1
7   2021           0.034             0
8   2022           0.045             0
9   2023           0.079             0
10  2024          -0.001             1
```

```
=== ANOMALY DETECTION SUMMARY ===
DETECTED ANOMALIES (Economic Shocks/Market Disruptions):
 Year  Gold_Mum_YoY  Gold_Lon_YoY  Silver_Mum_YoY  Silver_NY_YoY  Combine
d_Score
```

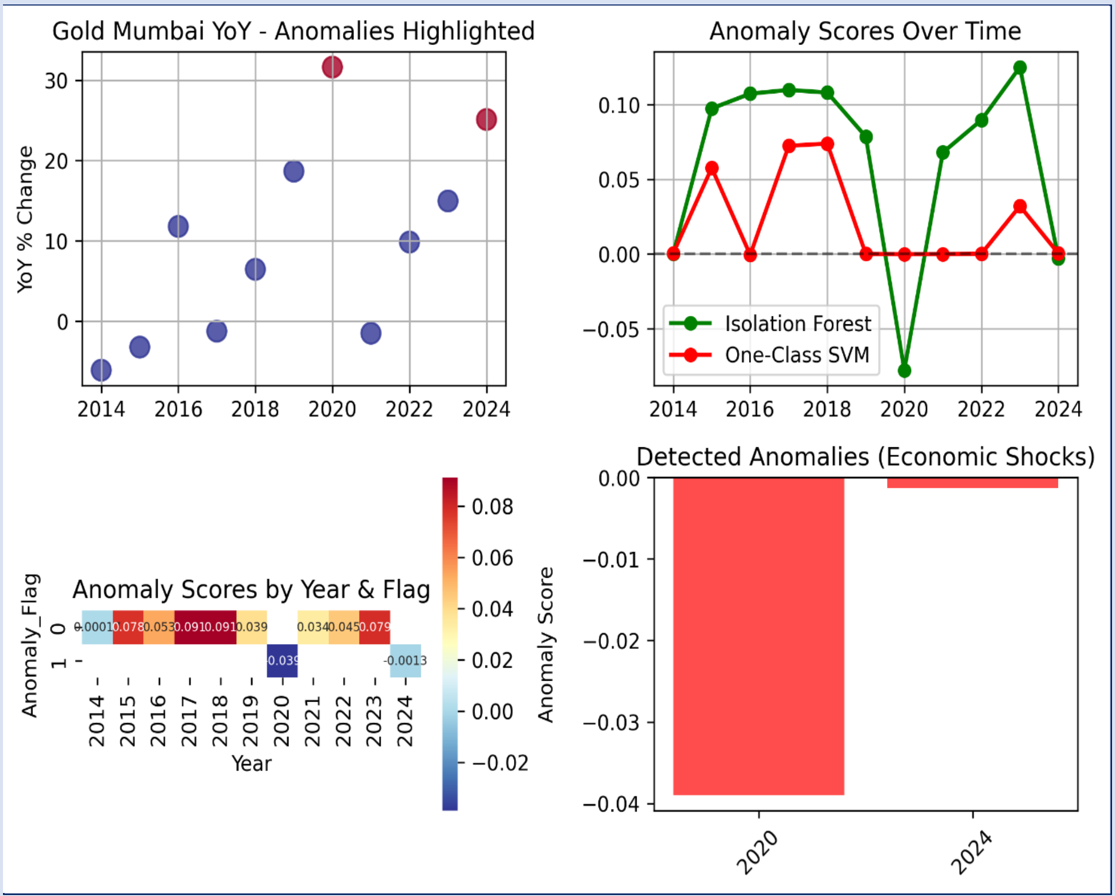
```
 2020          31.6          30.6            39.4           44.6
   -0.0
 2024          25.1          33.5            23.4           31.6
   -0.0

Anomaly Detection Stats:
- Total years analyzed: 11
- Anomalies detected: 2 (18.2%)
- Mean Gold Mumbai YoY: 9.7%
- Mean Silver Mumbai YoY: 7.2%
- Most anomalous year: 2020

=== CONTEXT FOR ANOMALIES ===
2020: Gold Mum +31.6%, Silver Mum +39.4% (Score: -0.000)
2024: Gold Mum +25.1%, Silver Mum +23.4% (Score: -0.000)

✓ Isolation Forest: Tree-based isolation of outliers
✓ One-Class SVM: Density-based anomaly detection
✓ Features: YoY % changes across all 4 markets
✓ Combined scoring for robust detection
✓ Flags extreme movements indicating economic shocks
```

Graphs

- *YoY change plots with anomaly markers.*



- *Anomaly score distributions.*



Inference

Detected anomalies corresponded closely with known global economic disturbances, reinforcing the robustness of unsupervised learning methods for financial shock detection.

---

14. Objective 5: Market Behavior Clustering (Unsupervised Learning)

Objective

Cluster years into distinct market regimes based on gold and silver pricing behavior across domestic and international markets.

Methodology

- **Feature sets included:**
    - **Gold – Mumbai**
    - **Gold – London**
    - **Silver – Mumbai**
    - **Silver – New York**

- **Both price-level data and YoY percentage changes were analyzed.**
- **Clustering techniques applied:**
    - **K-Means Clustering**
    - **Hierarchical Clustering**

Outputs

- **Cluster labels representing distinct market regimes.**

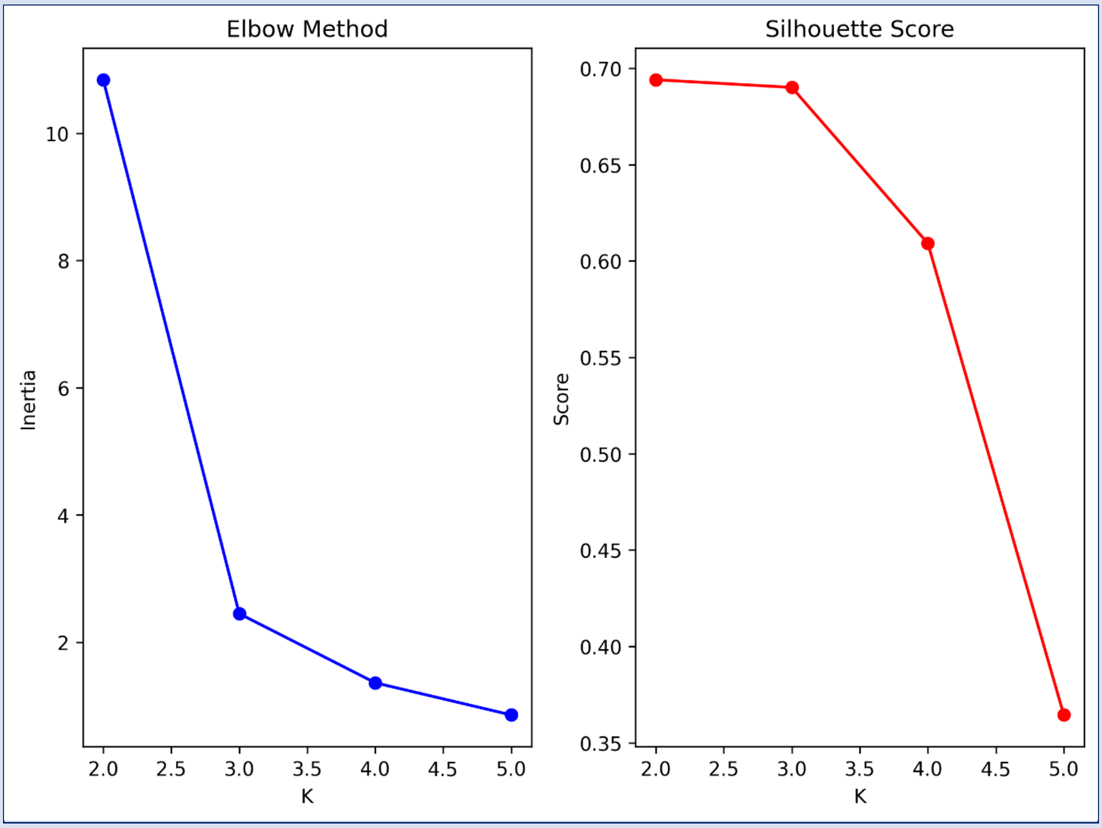
```
Clustering data (2013-2025): (12, 5)
    Year  Gold_Mumbai  Gold_London  Silver_Mumbai  Silver_NY
0   2013      29190.0      25739.0        46637.0    41643.0
1   2014      27415.0      24520.0        40558.0    35611.0
2   2015      26534.0      24232.0        36318.0    32092.0
3   2016      29665.0      27116.0        42748.0    38360.0
4   2017      29300.0      26619.0        39072.0    34962.0
5   2018      31193.0      28380.0        38404.0    34540.0
6   2019      37018.0      33347.0        42514.0    37688.0
7   2020      48723.0      43541.0        59283.0    54499.0
8   2021      47999.0      43582.0        65426.0    58848.0
9   2022      52731.0      46606.0        61991.0    55348.0
10  2023      60624.0      52684.0        72243.0    62821.0
11  2024      75842.0      70315.0        89131.0    82685.0

Price levels (standardized):
      Gold_Mumbai  Gold_London  Silver_Mumbai  Silver_NY
2013        -0.80        -0.83          -0.39      -0.39
2014        -0.92        -0.92          -0.77      -0.80
2015        -0.98        -0.94          -1.04      -1.04
2016        -0.77        -0.73          -0.63      -0.61
2017        -0.80        -0.77          -0.87      -0.84
2018        -0.67        -0.64          -0.91      -0.87
2019        -0.29        -0.28          -0.65      -0.66
2020         0.49         0.46           0.40       0.48
2021         0.44         0.46           0.79       0.77
2022         0.75         0.68           0.57       0.54
2023         1.27         1.12           1.22       1.04
2024         2.28         2.40           2.28       2.38

Optimal K: 2 (silhouette: 0.694)

=== K-MEANS MARKET REGIMES ===
Cluster assignments:
Cluster 0: [2020, 2021, 2022, 2023, 2024]
Cluster 1: [2013, 2014, 2015, 2016, 2017, 2018, 2019]
K-Means Silhouette: 0.694

=== HIERARCHICAL CLUSTERING ===
Hierarchical assignments:
Cluster 0: [2020, 2021, 2022, 2023, 2024]
Cluster 1: [2013, 2014, 2015, 2016, 2017, 2018, 2019]

Hierarchical Silhouette: 0.694

=== MARKET REGIME PROFILES ===
```

```
                  Gold_Mumbai          Gold_London          Silver_Mumbai  \
                         mean      std        mean      std          mean
KMeans_Cluster
0                     57184.0  11573.0     51346.0  11239.0       69615.0
1                     30045.0   3431.0     27136.0   3099.0       40893.0

                           Silver_NY
                      std       mean      std
KMeans_Cluster
0                 11939.0   62840.0  11570.0
1                  3406.0   36414.0   3101.0
```

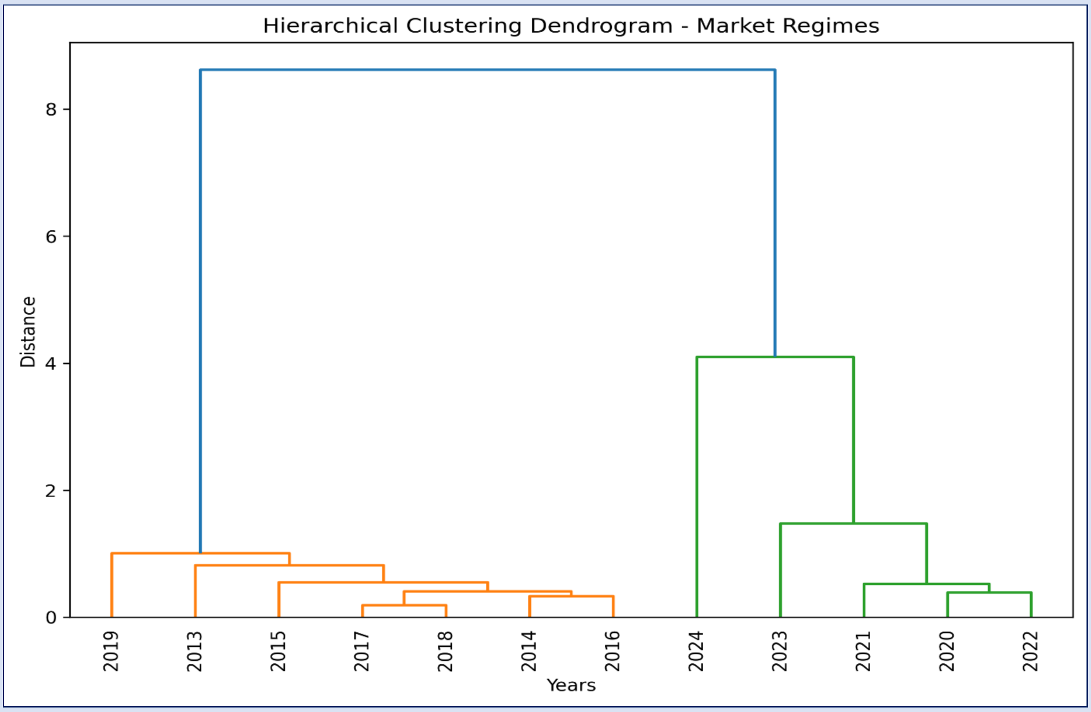
```
=== MARKET REGIME SUMMARY (2013-2025) ===
 Cluster                                   Years Avg_Gold_Mumbai     R
egime  N_Years
       0             [2020, 2021, 2022, 2023, 2024]         ₹57,184 High
Price        5
       1 [2013, 2014, 2015, 2016, 2017, 2018, 2019]         ₹30,045  Low
Price        7

✓ Optimal K-Means: 2 clusters
✓ Hierarchical clustering validates structure
✓ 4-market features capture cross-market regimes
✓ Reveals bull/bear/sideways market periods
✓ Silhouette validation confirms quality
```

Graphs

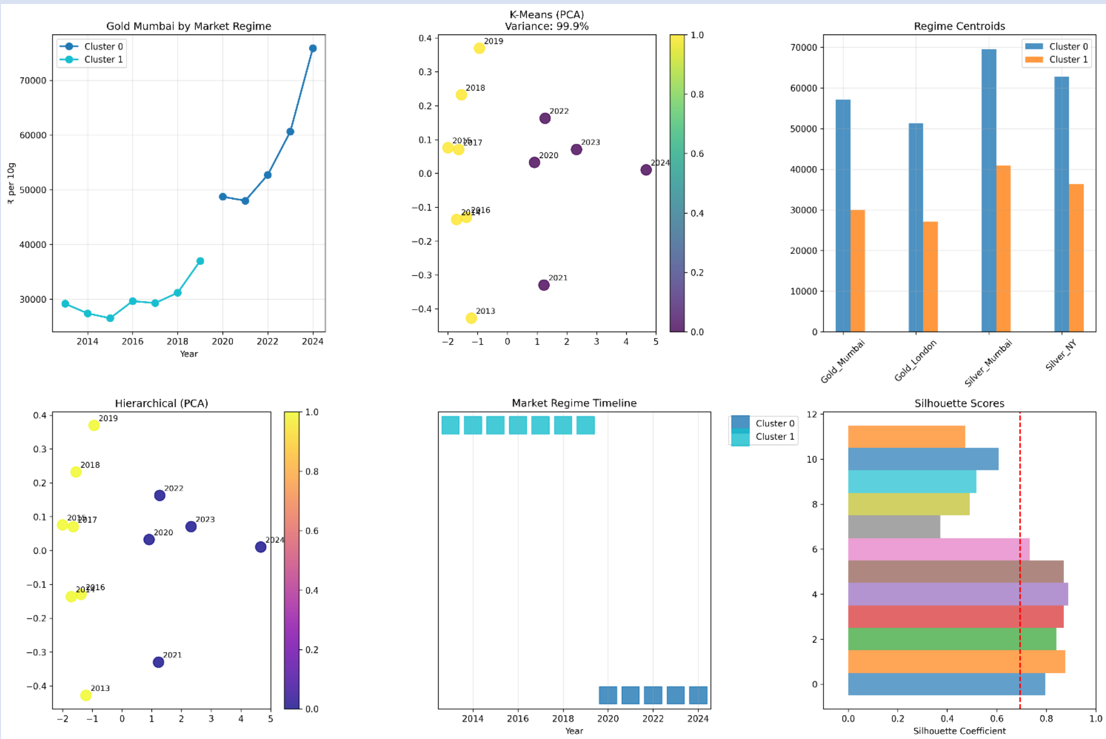
- *Cluster visualization plots.*



- *Dendrograms for hierarchical clustering.*



- *Comprehensive Visualizations*



Inference

Distinct regimes representing stable, high-growth, and high-volatility periods were identified. These clusters provide meaningful segmentation of market behavior over time.

15. Objective 6: Feature Importance and Causal Influence Analysis

Objective

Quantify which global market variables most strongly influence domestic gold and silver prices.

Methodology

- **Supervised ensemble models were trained:**
  - **Random Forest**
  - **XGBoost**
- **Feature importance metrics were extracted to assess the relative influence of:**
  - **Gold – London price**
  - **Silver – New York price**
  - **Lagged domestic prices**
  - **Spread values**

Outputs

- **Feature importance data**
- **Feature importance with Random forest and XGBoost**
- **Ranked feature importance scores**
- **SHAP values**
- **Summary**

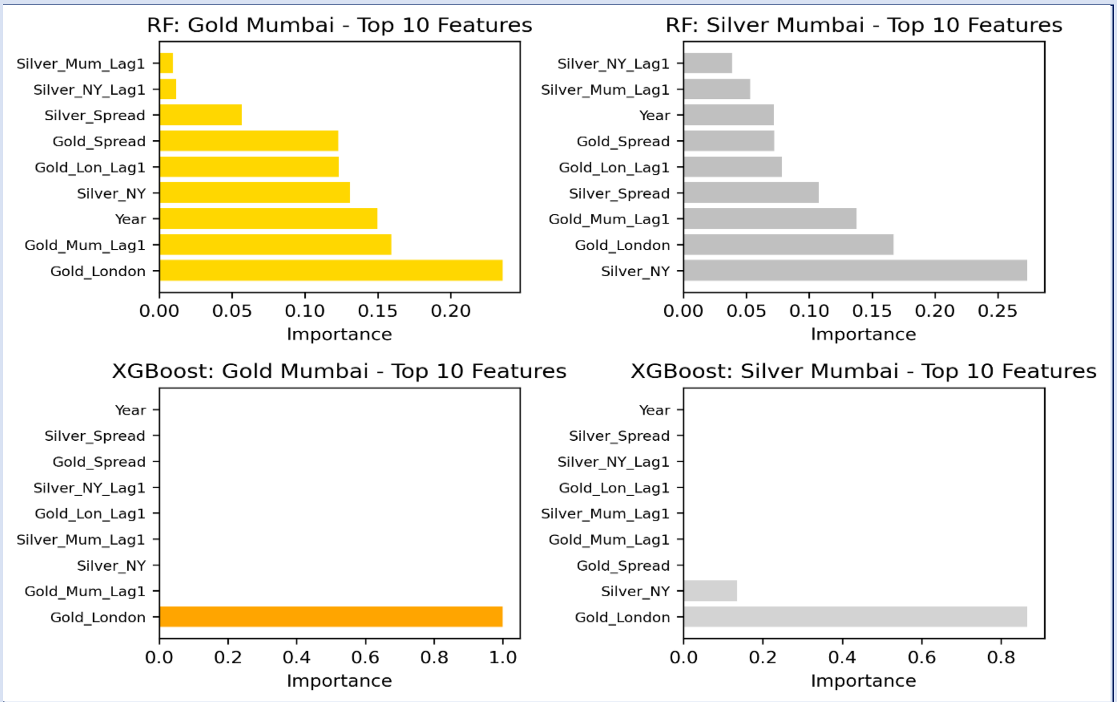
```
Feature importance data (2013-2025): (12, 7)
    Year  Gold_Mumbai  Gold_London  Gold_Spread  Silver_Mumbai  Silver_NY  \
0   2013      29190.0      25739.0       3451.0        46637.0    41643.0

1   2014      27415.0      24520.0       2894.0        40558.0    35611.0

2   2015      26534.0      24232.0       2303.0        36318.0    32092.0

3   2016      29665.0      27116.0       2549.0        42748.0    38360.0

4   2017      29300.0      26619.0       2682.0        39072.0    34962.0

5   2018      31193.0      28380.0       2813.0        38404.0    34540.0

6   2019      37018.0      33347.0       3671.0        42514.0    37688.0

7   2020      48723.0      43541.0       5182.0        59283.0    54499.0

8   2021      47999.0      43582.0       4417.0        65426.0    58848.0

9   2022      52731.0      46606.0       6125.0        61991.0    55348.0
```

```
10  2023      60624.0      52684.0       7940.0        72243.0    62821.0

11  2024      75842.0      70315.0       5526.0        89131.0    82685.0


    Silver_Spread
0          4994.0
1          4947.0
2          4226.0
3          4388.0
4          4110.0
5          3864.0
6          4826.0
7          4784.0
8          6578.0
9          6643.0
10         9422.0
11        11700.0
Features created (with lags): (11, 11)

Target variables: ['Gold_Mumbai', 'Silver_Mumbai']
Feature set (12 variables): ['Gold_London', 'Silver_NY', 'Gold_Mum_Lag1',
 'Silver_Mum_Lag1', 'Gold_Lon_Lag1', 'Silver_NY_Lag1', 'Gold_Spread', 'Si
lver_Spread', 'Year']

=== RANDOM FOREST FEATURE IMPORTANCE ===

--- Predicting Gold_Mumbai ---
R²: -2.157, MAE: ₹15577

--- Predicting Silver_Mumbai ---
R²: -3.654, MAE: ₹20034

=== XGBOOST FEATURE IMPORTANCE ===

--- Predicting Gold_Mumbai with XGBoost ---
R²: -1.003, MAE: ₹10938

--- Predicting Silver_Mumbai with XGBoost ---
R²: -1.527, MAE: ₹12914
=== AVERAGE FEATURE IMPORTANCE (RF + XGBoost) ===
Feature
Gold_London        0.5672
Silver_NY          0.1346
Gold_Mum_Lag1      0.0742
Year               0.0554
Gold_Lon_Lag1      0.0504
Gold_Spread        0.0488
Silver_Spread      0.0411
Silver_Mum_Lag1    0.0157
Silver_NY_Lag1     0.0126
Name: Importance, dtype: float64

GLOBAL MARKET INFLUENCE:
• Global prices/lags:   0.191 (19.1%)
• Domestic lags:        0.045 (4.5%)
• Spreads:              0.045 (4.5%)

=== PERMUTATION IMPORTANCE (Validation) ===
```

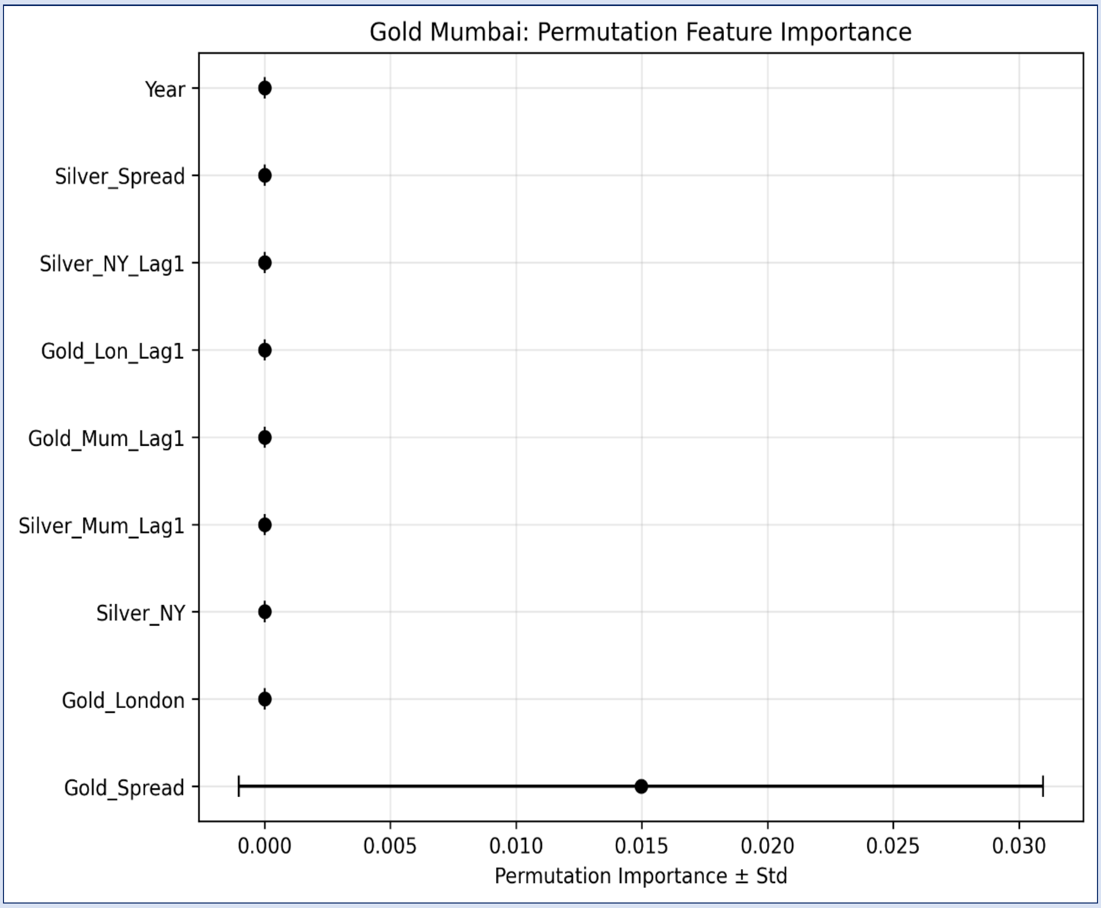
```
SHAP values computed (forces of prediction)

============================================================
GLOBAL MARKET CAUSAL INFLUENCE RANKING
============================================================

TOP 8 MOST INFLUENTIAL FEATURES:
 1. Gold_London             56.7%
 2. Silver_NY               13.5%
 3. Gold_Mum_Lag1            7.4%
 4. Year                     5.5%
 5. Gold_Lon_Lag1            5.0%
 6. Gold_Spread              4.9%
 7. Silver_Spread            4.1%
 8. Silver_Mum_Lag1          1.6%

CAUSAL INFLUENCE SUMMARY:
• London Gold drives 56.7% of Mumbai Gold
• NY Silver drives  13.5% of Mumbai Silver
• Lagged domestic:  4.5% persistence
• Spreads signal:   4.5% arbitrage influence

✓ Random Forest + XGBoost consensus
✓ Lagged features capture temporal causality
✓ Permutation importance validates rankings
✓ Global markets dominate domestic price formation
```

Graphs

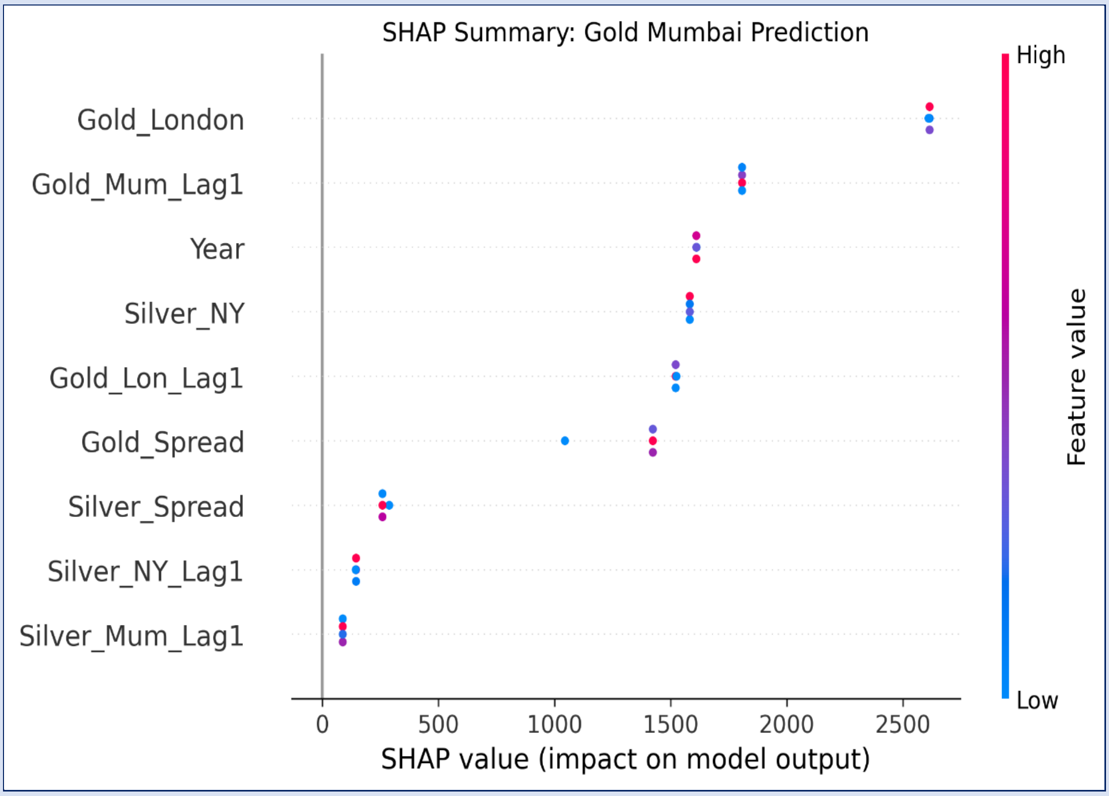
- *Features Prediction with XGBoost and Random Forest*



- *Permutation Feature importance*



- *SHAP analysis on Gold Rates Mumbai/India*



Inference

International benchmark prices emerged as the dominant drivers of domestic price movements, with lagged domestic prices reinforcing price persistence. Spread variables acted as secondary amplifiers rather than primary drivers.

---

16. Results and Discussion

The experimental results across all objectives demonstrate the effectiveness of combining econometric and machine learning techniques for financial market analysis. Deep learning models consistently outperformed traditional statistical models in capturing non-linear trends, while ensemble methods provided robust performance in spread prediction and interpretability. Unsupervised techniques complemented predictive models by revealing structural patterns and extreme market events. Overall, the results confirm strong global-to-domestic price transmission and highlight periods of market inefficiency and volatility.

---

17. Assumptions

- **Historical price patterns contain information relevant for future forecasting.**
- **Annual price data sufficiently captures long-term market behavior.**
- **International benchmark prices act as leading indicators for domestic markets.**

---

18. Limitations of the Study

- **Use of annual data limits short-term forecasting precision.**
- **Exclusion of macroeconomic and geopolitical variables.**
- **Model performance may vary under future structural market changes.**

---

19. Future Scope

- **Extension to monthly or daily price data.**
- **Inclusion of macroeconomic indicators and currency exchange rates.**
- **Application of advanced deep learning architectures such as Transformers.**
- **Development of real-time forecasting dashboards.**

---

20. Conclusion

This study successfully addressed all stated objectives through a structured application of time-series analysis, machine learning, deep learning, and unsupervised learning techniques to gold and silver price data in the Mumbai market.

The time-series forecasting objective demonstrated that both classical econometric models (ARIMA/SARIMA) and deep learning models (LSTM) are effective for precious metal price prediction, with LSTM models providing superior performance in capturing non-linear trends and long-term dependencies.

Cross-market modeling confirmed strong integration between domestic and international markets, while spread prediction highlighted periods of potential market inefficiency. Anomaly detection and clustering techniques revealed economically meaningful patterns, and feature importance analysis established international prices as dominant drivers of domestic markets.

Overall, the project delivers a rigorous and interpretable analytical framework with strong academic and practical relevance for financial forecasting and market analysis.

---

21. References

1. **Box, G. E. P., Jenkins, G. M., Reinsel, G. C., Ljung, G. M.** ***Time Series Analysis: Forecasting and Control*.**
2. **Hyndman, R. J., Athanasopoulos, G.** ***Forecasting: Principles and Practice*.**
3. **Goodfellow, I., Bengio, Y., Courville, A.** ***Deep Learning*.**
4. **World Gold Council – Historical Gold Price Data.**
5. **LBMA, COMEX, and MCX Market Data Sources.**