Predicting Gold Price (GLD) Using Financial Data

MUHAMMED SHAHAN P P

Contents

Project Objective & Use Case

01

Objective: Predict daily Gold Price (GLD)
The goal is to predict the daily price of gold (GLD) using related financial indicators.

02

Problem Type: Regression
This is a regression problem where the target variable is continuous.

03

Use Case: Trading support and financial forecasting
The model supports trading strategies and financial forecasting by estimating future gold prices based on indicators like SPX (S&P 500 Index), USO (United States Oil Fund), SLV (iShares Silver Trust), and other market metrics.

Data Preprocessing & Feature Engineering

Feature Engineering

```
# Feature Selection
# Adding an Extra Feature

gold_data['GLD_lag1'] = gold_data['GLD'].shift(1)
gold_data['GLD_lag4'] = gold_data['GLD'].shift(4)
# Price before 1 day and 4 day

gold_data.head()
```

[312] ✓ 0.0s

	Date	SPX	GLD	USO	SLV	EUR/USD	GLD_lag1	GLD_lag4
0	2008-01-02	1447.160034	84.860001	72.998754	15.180	1.471692	NaN	NaN
1	2008-01-03	1447.160034	85.570000	72.998754	15.285	1.474491	84.860001	NaN
2	2008-01-04	1411.630005	85.129997	72.998754	15.167	1.475492	85.570000	NaN
3	2008-01-07	1416.180054	84.769997	72.998754	15.053	1.468299	85.129997	NaN
4	2008-01-08	1390.189941	86.779999	72.998754	15.590	1.557099	84.769997	84.860001

```
# Dropping the null part
gold_data.dropna(inplace=True)
gold_data.head()
```

[313] ✓ 0.0s

```
# StandardScaler
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
gold_data[['SPX','USO','SLV']] = scaler.fit_transform(gold_data[['SPX','USO','SLV']])
```

[315] ✓ 0.0s

```
gold_data.head()
```

[316] ✓ 0.0s

	Date	SPX	GLD	USO	SLV	EUR/USD	GLD_lag1	GLD_lag4
4	2008-01-08	-0.509307	86.779999	2.64373	-0.651651	1.557099	84.769997	84.860001
5	2008-01-09	-0.472840	86.550003	2.64373	-0.662239	1.466405	86.779999	85.570000
6	2008-01-10	-0.451275	88.250000	2.64373	-0.580407	1.480100	86.550003	85.129997
7	2008-01-11	-0.488455	88.580002	2.64373	-0.577987	1.479006	88.250000	84.769997
8	2008-01-14	-0.459131	89.540001	2.64373	-0.547281	1.486900	88.580002	86.779999

Dataset

Used public financial dataset with features like SPX, USO, SLV

Feature Selection

Created lag features (GLD_lag1, GLD_lag4) to capture temporal patterns

Feature Scaling

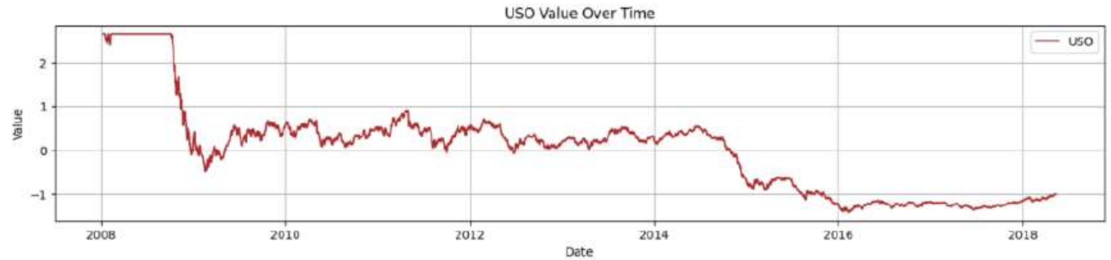
Scaled input features (except target) using Standardization

Trend Analysis & Outlier Handling

01 **Performed time series trend visualization**

02 **Outliers detected via boxplots**

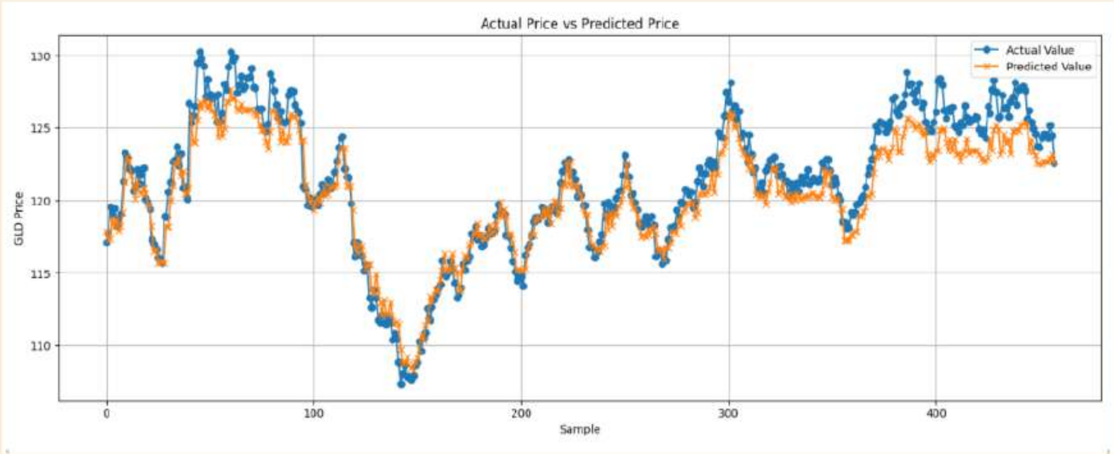
03 **Removed extreme outliers using IQR method**

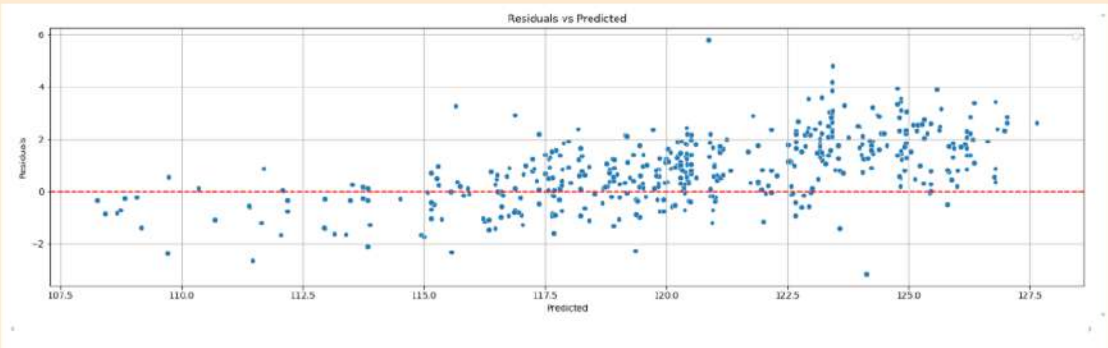


Modeling with LightGBM

- Chose LightGBM as final model for efficiency and accuracy
- Tuned hyperparameters using GridSearchCV
- Trained on historical financial indicators

Model Evaluation & Visualization



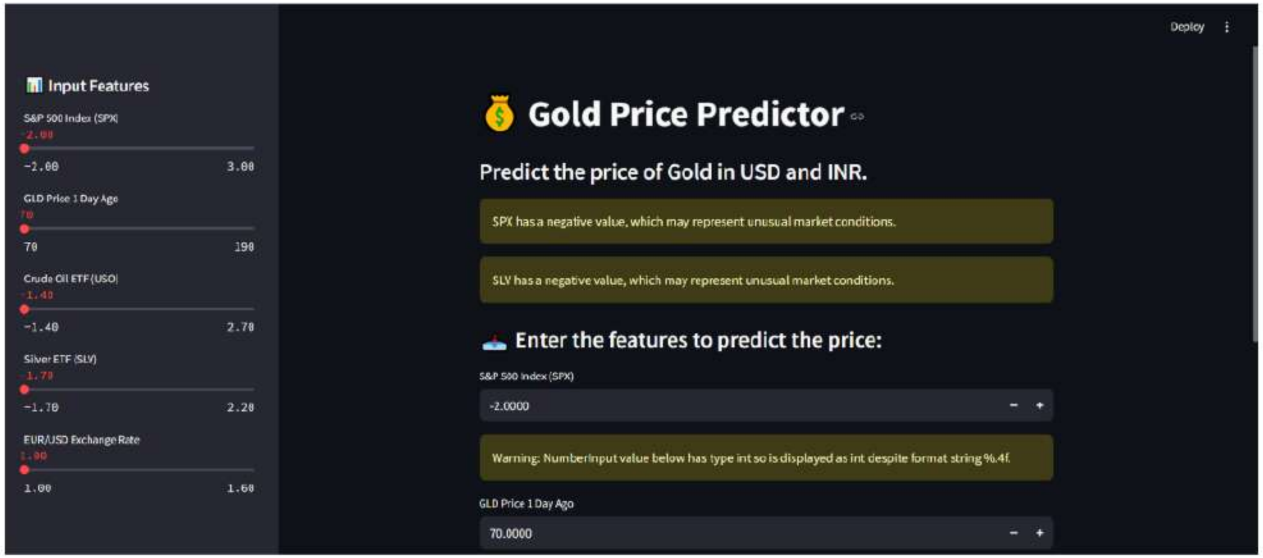


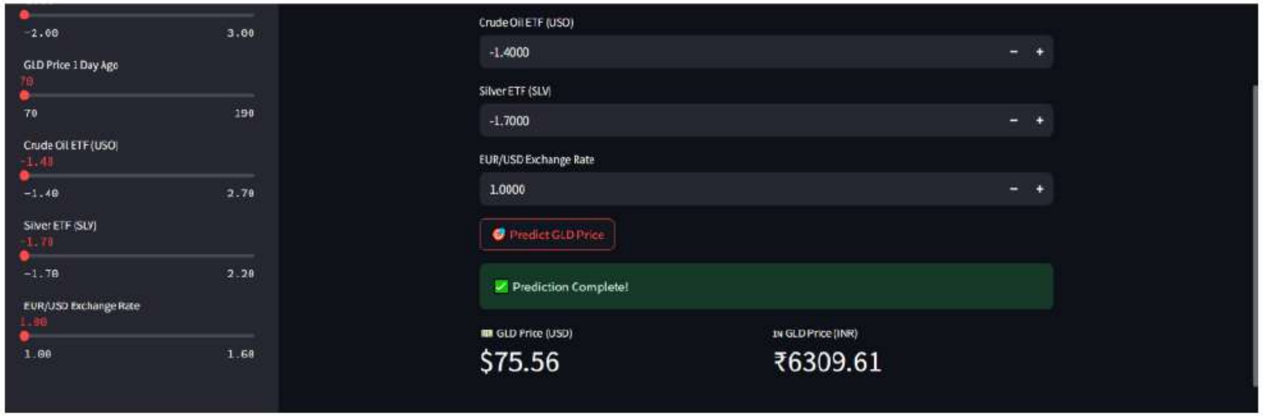
>> Evaluation using R^2, RMSE, MAE , MSE

>> Visualized Actual vs Predicted prices

>> Residual plot showed no major pattern (errors are random)

Deployment & Business Insights





Saved final model using Joblib

Built interactive frontend using Streamlit

Enables users to input new values and get real-time predictions

Business Insight: Strong price dependence on past GLD trends

Thank You

by MUHAMMED SHAHAN P P