رواد مصر الرقمية

Gold price prediction

This project uses machine learning to predict gold prices. We leverage sentiment analysis and diverse data sources for accurate predictions. Our goal is to create a reliable tool for investment decisions.

رواد مصر الرقمية

Team members

Kareem Fady

Reham Galal

Ameer Hesham

Mohamed Elmesery

Ola Mohamed

رواد مصر الرقمية

Table of Contents

رواد مصر الرقمية

What is our problem ?!

- ***High gold price volatility**:*

Influenced by economic and geopolitical factors.

- ***Impact of market sentiment:***

Difficult to analyze due to unstructured data.

- ***Traditional strategies fall short:***

Struggle to predict price fluctuations accurately.

- ***Investor uncertainty:***

Challenges in making informed buy/sell decisions.

رواد مصر الرقمية

Our Solution...

- *Predicted Gold Prices vs. Actual Prices:*

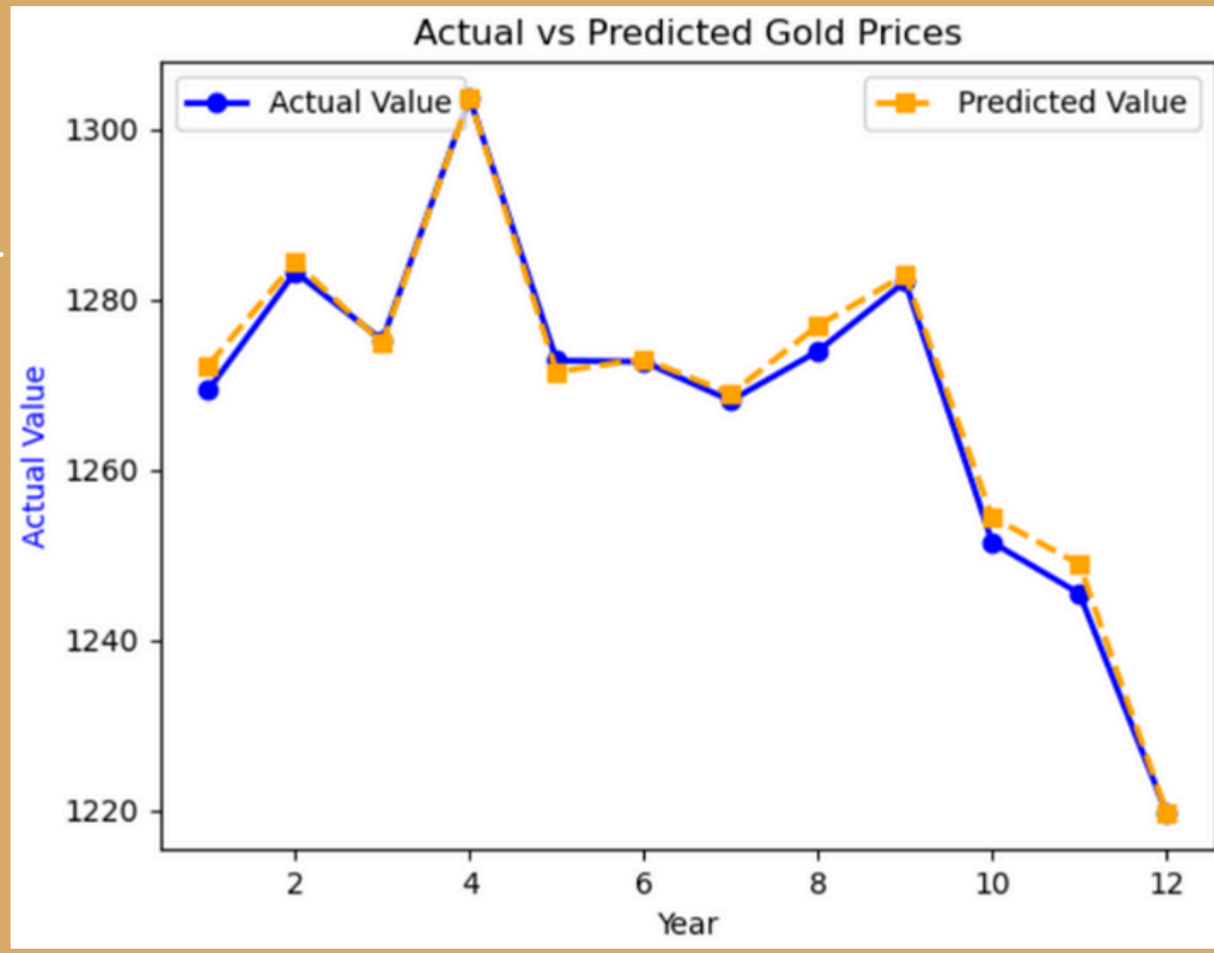
Present a graph comparing the predicted prices with actual prices.

- *Investment Insights:*

Demonstrate how your model can be used to provide actionable insights for investors, such as when to buy or sell gold.

- *Risk Analysis:*

Explain any steps taken to assess the risk associated with predictions and how the model can help in mitigating potential losses.



رواد مصر الرقمية

Data Gathering: Sourcing Gold Price and Sentiment Data

1 Gold Price Data

Gold Price Data from Yahoo Finance (YFinance)

2 Sentiment Data

Gold Sentiment Analysis from News Sources

رواد مصر الرقمية

1 Gold Price Data

Gold Price Data from Yahoo Finance (YFinance)

```
import yfinance as yf
```

```
gold_data = yf.download('GC=F', start='2000-01-01', end='2024-01-01')
```

```
[**********100%**********]  1 of 1 completed
```

```
gold_data.head()
```

Date	Open	High	Low	Close	Adj Close	Volume
2000-08-30	273.899994	273.899994	273.899994	273.899994	273.899994	0
2000-08-31	274.799988	278.299988	274.799988	278.299988	278.299988	0
2000-09-01	277.000000	277.000000	277.000000	277.000000	277.000000	0
2000-09-05	275.799988	275.799988	275.799988	275.799988	275.799988	2
2000-09-06	274.200012	274.200012	274.200012	274.200012	274.200012	0

NLP

```
df['News'].head(10)
```

	News
0	palladium rallies as gold falls
1	gold lower in quiet trading
2	gold steady on dollar's strength
3	gold gains on dollar weakness
4	gold slips for a second-straight day
5	gold slips amid selling overseas
6	gold quiet in low-volume trading
7	gold shares higher; futures prices fall
8	gold slightly lower on dollar moves
9	gold futures prices little changed

dtype: object

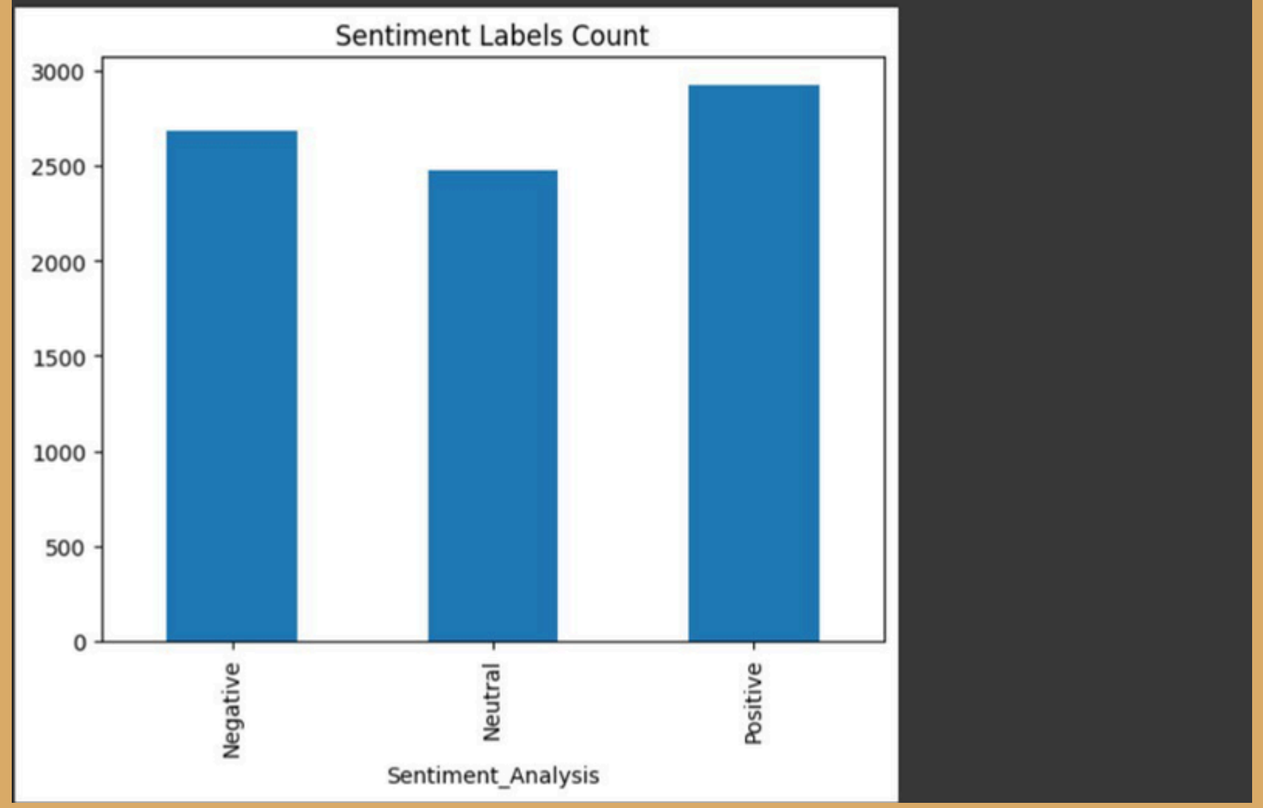
```
from transformers import BertTokenizer, BertForSequenceClassification
from transformers import pipeline

finbert = BertForSequenceClassification.from_pretrained('yiyanghkust/finbert-tone',num_labels=3)
tokenizer = BertTokenizer.from_pretrained('yiyanghkust/finbert-tone')

nlp = pipeline("sentiment-analysis", model=finbert, tokenizer=tokenizer,device=0)
```

```
def get_sentiment_label(news_text):
    result = nlp(news_text)[0]
    return result['label']  # Return only the sentiment label (e.g., 'positive', 'neutral', 'negative')
```

```
#Adding column with the sentiment label to the DataFrame
df['Sentiment_Analysis'] = df['News'].apply(get_sentiment_label)
```

	News	Sentiment_Analysis
0	palladium rallies as gold falls	Neutral
1	gold lower in quiet trading	Negative
2	gold steady on dollar's strength	Positive
3	gold gains on dollar weakness	Positive
4	gold slips for a second-straight day	Neutral
5	gold slips amid selling overseas	Negative
6	gold quiet in low-volume trading	Neutral
7	gold shares higher; futures prices fall	Positive
8	gold slightly lower on dollar moves	Positive
9	gold futures prices little changed	Neutral
10	gold prices fall; palladium rises	Negative
11	gold shares, prices decline	Negative
13	gold prices little changed; shares fall	Negative
14	gold shares climb as prices fall	Positive
15	gold closes lower, palladium rises	Positive

```
df['Sentiment_Analysis'].value_counts().sort_index().plot(kind="bar",title="Sentiment Labels Count");
```



رواد مصر الرقمية

Code

[Merge]

Result

```
gold_data['Date'] = pd.to_datetime(df['Date']).dt.date
```

```
gold_data_reset = gold_data.reset_index(drop=True)
```

```
gold_data.rename(columns={'Date': 'Date_Original'}, inplace=True)
gold_data_reset = gold_data.reset_index()
```

```
df['Date'] = pd.to_datetime(df['Date'])
gold_data_reset['Date'] = pd.to_datetime(gold_data_reset['Date'])
```

```
total = pd.merge(gold_data_reset, df, on='Date', how='inner')
```

```
total.head()
```

	Dates	URL	News	Price Direction Up	Price Direction Constant	Price Direction Down	Asset Comparision	Past Information	Future Information	Price Sentiment
0	28-01-2016	http://www.marketwatch.com/story/april-gold-do...	april gold down 20 cents to settle at $1,116.1...	0	0	1	0	1	0	negative
1	13-09-2017	http://www.marketwatch.com/story/gold-prices-s...	gold suffers third straight daily decline	0	0	1	0	1	0	negative
2	26-07-2016	http://www.marketwatch.com/story/gold-futures-...	Gold futures edge up after two-session decline	1	0	0	0	1	0	positive
3	28-02-2018	https://www.metalsdaily.com/link/277199/dent-r...	dent research : is gold's day in the sun comin...	0	0	0	0	0	1	none
4	06-09-2017	http://www.marketwatch.com/story/gold-steadies...	Gold snaps three-day rally as Trump, lawmakers...	0	0	1	0	1	0	negative

Column Name	Description
Date	The date of the record.
Open	The opening price of gold.
High	The highest price of gold during the day.
Low	The lowest price of gold during the day.
Close	The closing price of gold.
Adj Close	The adjusted closing price of gold (considering dividends and splits).
Volume	The volume of trades.
News	News headlines associated with gold on that day.
Price Direction Up	Indicator if the price went up.
Price Direction Constant	Indicator if the price remained constant.
Price Direction Down	Indicator if the price went down.
Asset Comparison	Whether gold was compared to other assets.
Past Information	Indicator of past trends being used.
Future Information	Indicator of future trends being used.
Price Sentiment	Sentiment analysis of the price (positive, negative, neutral).

رواد مصر الرقمية

Data cleaning

Read data

```
gold_data=pd.read_csv("total_gold data.csv")
```

```
gold_data.head()
```

Head of data

Date	Open	High	Low	Close	Adj Close	Volume	News	Price Direction Up	Price Direction Constant	Price Direction Down	Asset Comparision	Past Information	Future Information	Price Sentimen
2000-09-07	274.000000	274.000000	274.000000	274.000000	274.000000	125	palladium rallies as gold falls	0	0	1	1	1	0	negativ
2000-09-08	273.299988	273.299988	273.299988	273.299988	273.299988	0	gold lower in quiet trading	0	0	1	0	1	0	negativ
2000-09-14	272.399994	272.399994	272.399994	272.399994	272.399994	0	gold steady on dollar's strength	0	1	0	1	1	0	neutra
2000-10-10	272.399994	272.399994	272.399994	272.399994	272.399994	13	gold gains on dollar weakness	1	0	0	1	1	0	positiv
2000-10-16	271.500000	271.500000	271.500000	271.500000	271.500000	5	gold slips for a second-straight day	0	0	1	0	1	0	negativ

رواد مصر الرقمية

Data cleaning

Data shape & information

```
print(gold_data.shape)
print(gold_data.info())
```

```
(9846, 15)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9846 entries, 0 to 9845
Data columns (total 15 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Date                      9846 non-null   object
 1   Open                      9846 non-null   float64
 2   High                      9846 non-null   float64
 3   Low                       9846 non-null   float64
 4   Close                     9846 non-null   float64
 5   Adj Close                 9846 non-null   float64
 6   Volume                    9846 non-null   int64
 7   News                      9846 non-null   object
 8   Price Direction Up        9846 non-null   int64
 9   Price Direction Constant  9846 non-null   int64
 10  Price Direction Down      9846 non-null   int64
 11  Asset Comparision         9846 non-null   int64
 12  Past Information          9846 non-null   int64
 13  Future Information        9846 non-null   int64
 14  Price Sentiment           9846 non-null   object
dtypes: float64(5), int64(7), object(3)
```

The data consists of 9846 rows and 15 columns.

Data column , News & price sentiment is objects

رواد مصر الرقمية

Data cleaning

Null values in dataset

```
gold_data.isnull().sum()
```

```
Date                          0
Open                          0
High                          0
Low                           0
Close                         0
Adj Close                     0
Volume                        0
News                          0
Price Direction Up            0
Price Direction Constant      0
Price Direction Down          0
Asset Comparision             0
Past Information              0
Future Information            0
Price Sentiment               0
dtype: int64
```

The dataset does not contain missing values.

رواد مصر الرقمية

Data cleaning

Drop columns

```
gold_data = gold_data.drop(['Adj Close','News','Asset Comparision','Past Information','Future Information'], axis=1)
```

- **Adj Close:** Dropped to avoid redundancy since the regular close price is already included.
- **News:** Removed because it contains unstructured text data unsuitable for modeling without preprocessing.
- **Asset Comparison:** Excluded as it may introduce irrelevant data not directly related to gold price prediction.
- **Past Information:** Dropped to avoid redundancy with other historical data features.
- **Future Information:** Removed to prevent data leakage, as it contains information unavailable at prediction time.

رواد مصر الرقمية

Data cleaning

Remove duplicates

9843	2019-01-31
9844	2019-01-31

By observing the dataset, duplicate rows were observed.

```
gold_data = gold_data.drop_duplicates(subset=['Date'])
```

Data shape become

```
3278 rows × 10 columns
```

رواد مصر الرقمية

Data cleaning

```
gold_data['Price Sentiment'].value_counts()
```

```
Price Sentiment
positive    1441
negative    1227
none         496
neutral      114
Name: count, dtype: int64
```

Values and counts in column price sentiment

map price sentiment into numerical data

Maping Price Sentiment

```
sentiment_mapping = {
    'positive': 1,
    'neutral': 0,
    'none' : 0 ,
    'negative': -1
}

gold_data.loc[:, 'Price Sentiment'] = gold_data['Price Sentiment'].map(sentiment_mapping)
gold_data['Price Sentiment'] = gold_data['Price Sentiment'].astype(np.float64 )
gold_data.head()
```

Data cleaning

New data information

```
Data columns (total 10 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   Date                        3278 non-null   object
 1   Open                        3278 non-null   float64
 2   High                        3278 non-null   float64
 3   Low                         3278 non-null   float64
 4   Close                       3278 non-null   float64
 5   Volume                      3278 non-null   int64
 6   Price Direction Up          3278 non-null   int64
 7   Price Direction Constant    3278 non-null   int64
 8   Price Direction Down        3278 non-null   int64
 9   Price Sentiment             3278 non-null   float64
dtypes: float64(5), int64(4), object(1)
```

رواد مصر الرقمية

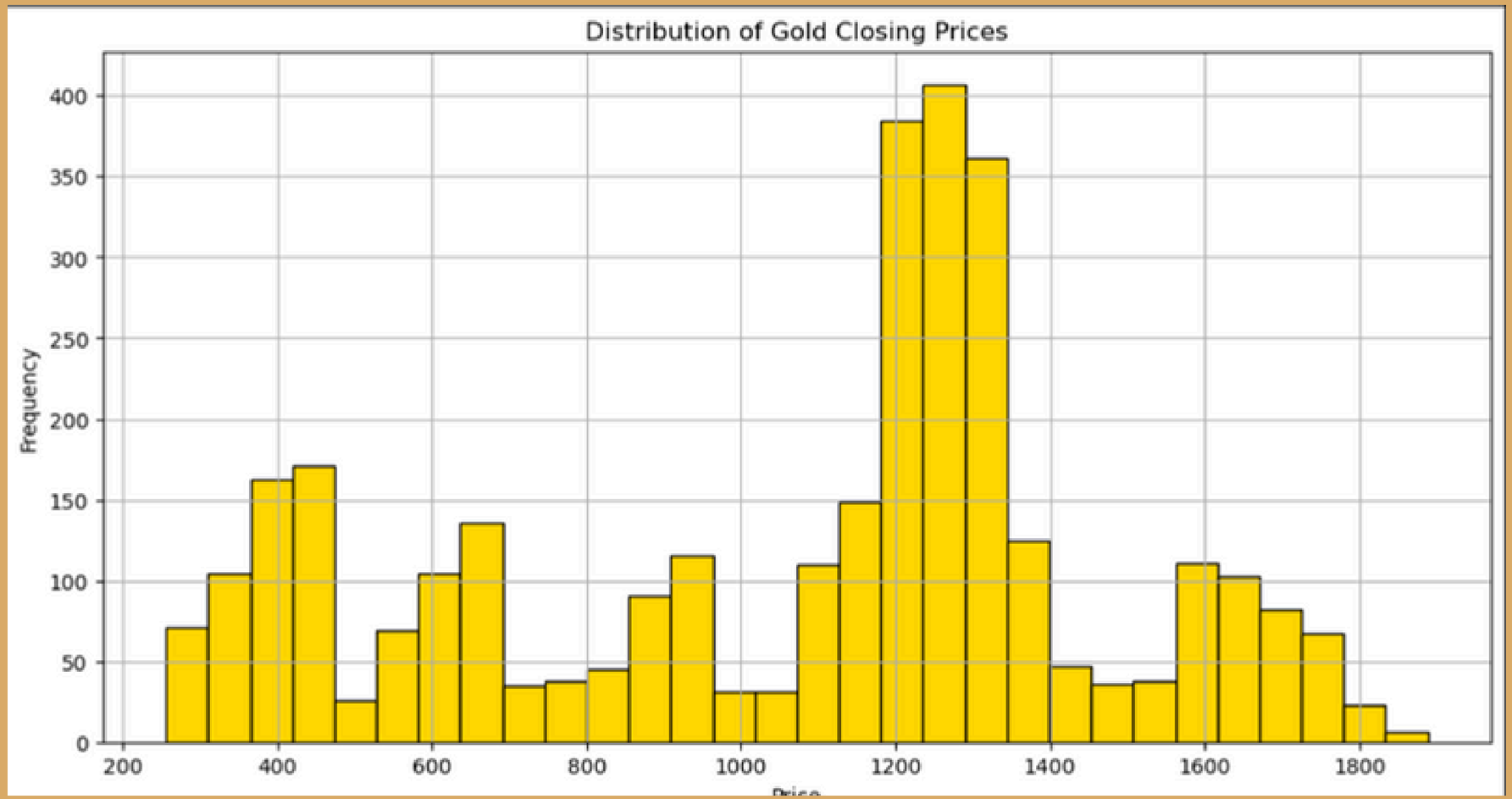
Data Visualization

Insight

The graph shows a significant increase in gold prices from 2000 to 2011, peaking around 2011 before experiencing a decline and stabilization over the next several years. This trend indicates a surge during economic uncertainty, followed by market corrections and relative stability in recent years.

رواد مصر الرقمية

Data Visualization



Insight

Highlighting a concentration of values around $1,200, with several peaks at lower and higher price ranges .This pattern could reflect different historical phases in the gold market, such as economic cycles or significant events influencing gold prices.

رواد مصر الرقمية

Data Visualization

Insight

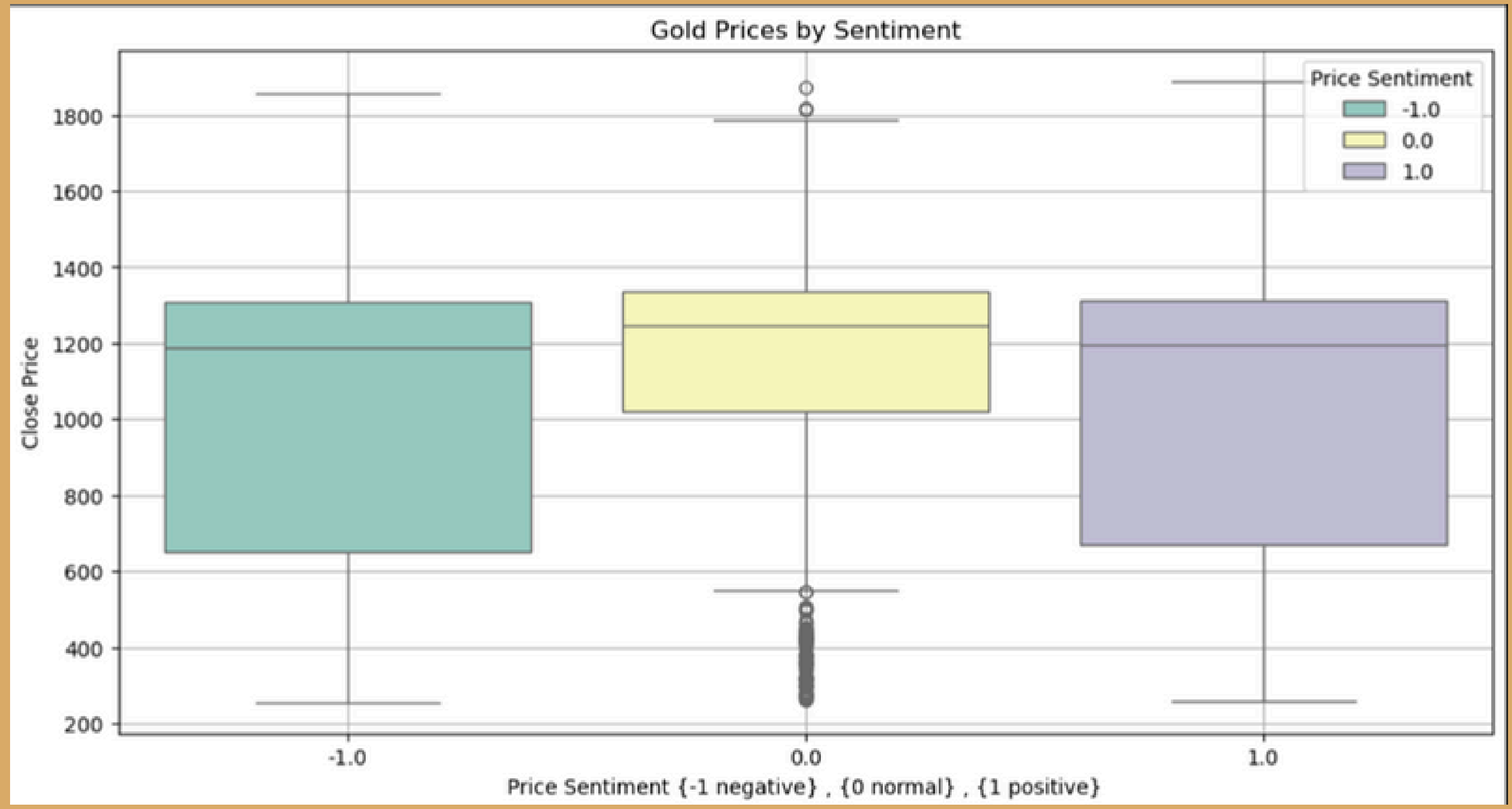
The scatter plot shows that there is no clear relationship between gold closing prices and trading volume

SO

```
gold_data = gold_data.drop(['Volume'], axis=1)
```

رواد مصر الرقمية

Data Visualization

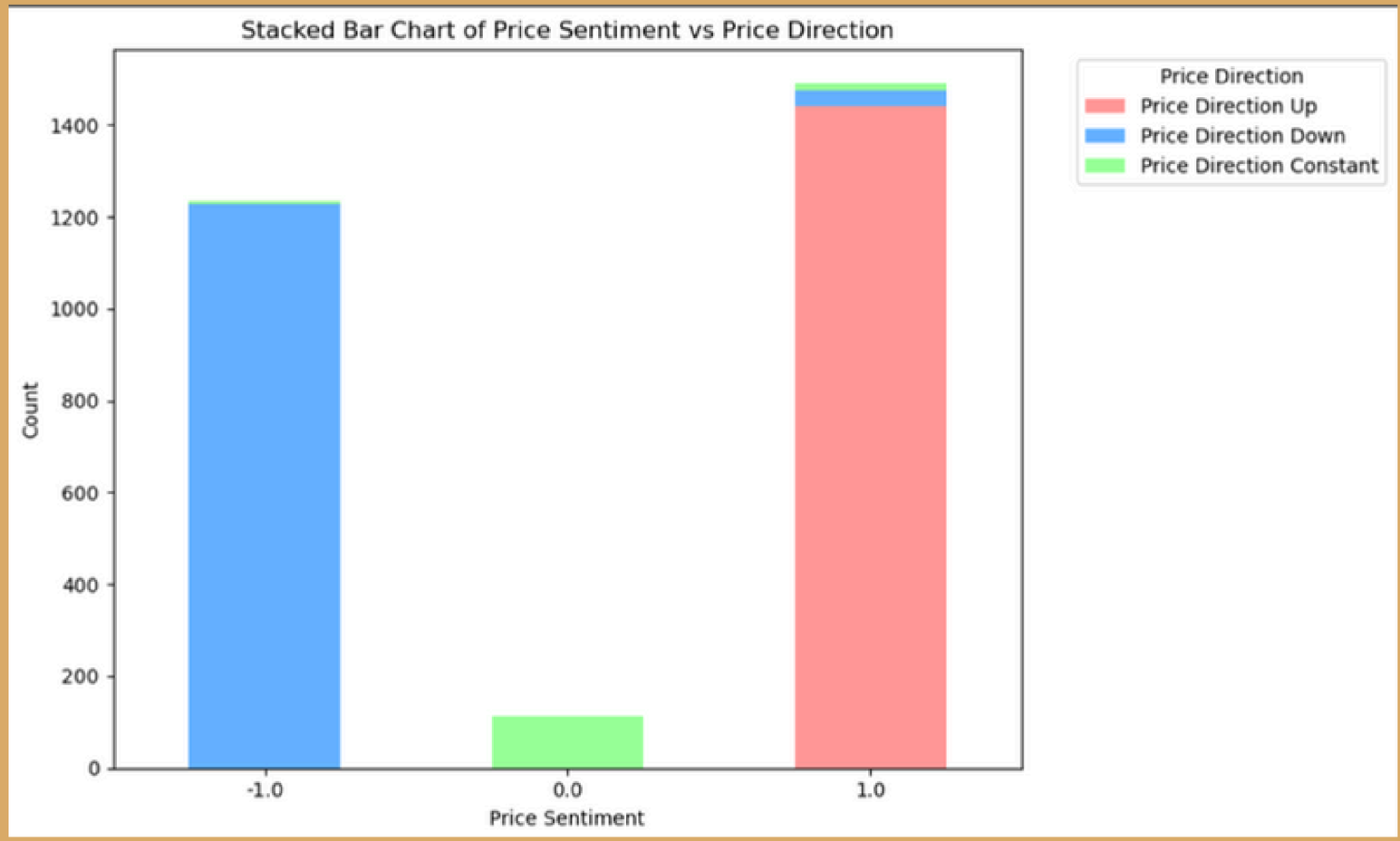


Insight

The box plot indicates that gold prices are generally higher during positive sentiment and lower during negative sentiment, with the median price rising as sentiment improves. There is also more variability in gold prices during negative sentiment, as shown by the wider range of values.

رواد مصر الرقمية

Data Visualization



Insight

The stacked bar chart reveals that during positive sentiment (1), the majority of gold price movements are upward, with very few downward or constant trends. In contrast, negative sentiment (-1) is associated with predominantly downward price movements, suggesting a strong correlation between sentiment and price direction

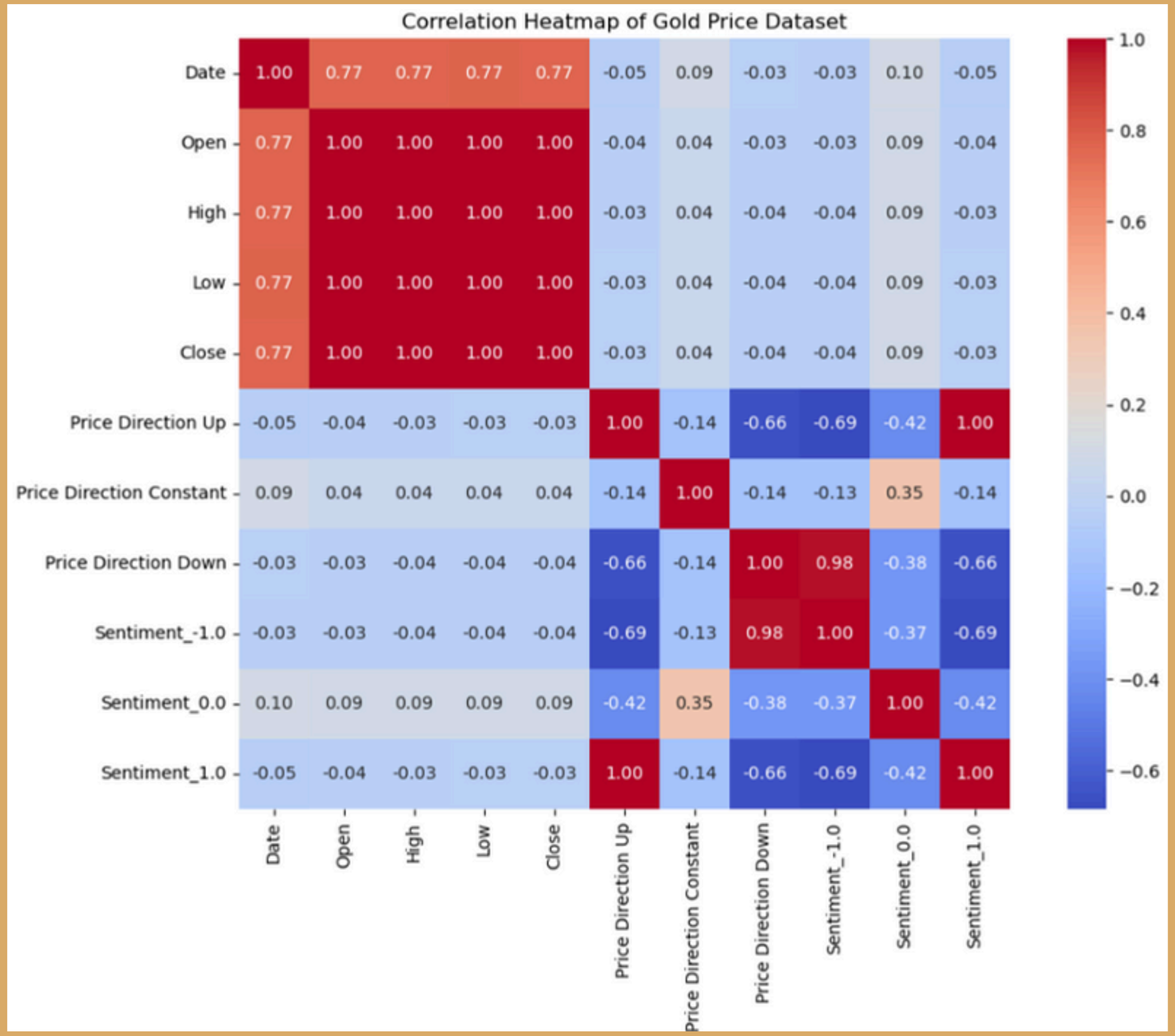
رواد مصر الرقمية

Data Visualization

Insight

Sentiment and Price:

The price movement of gold is closely tied to sentiment. Positive sentiment drives price increases, while negative sentiment leads to price declines.



Gold Price Dashboard

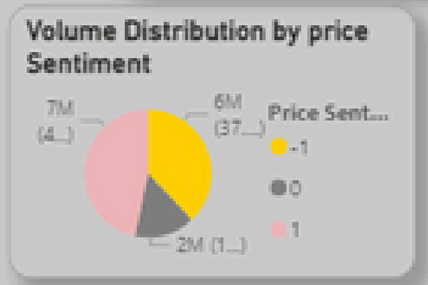
256.10
Min of Close

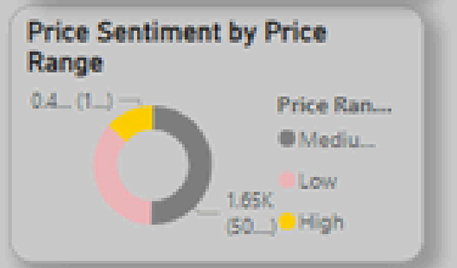
1.89K
Max of Close

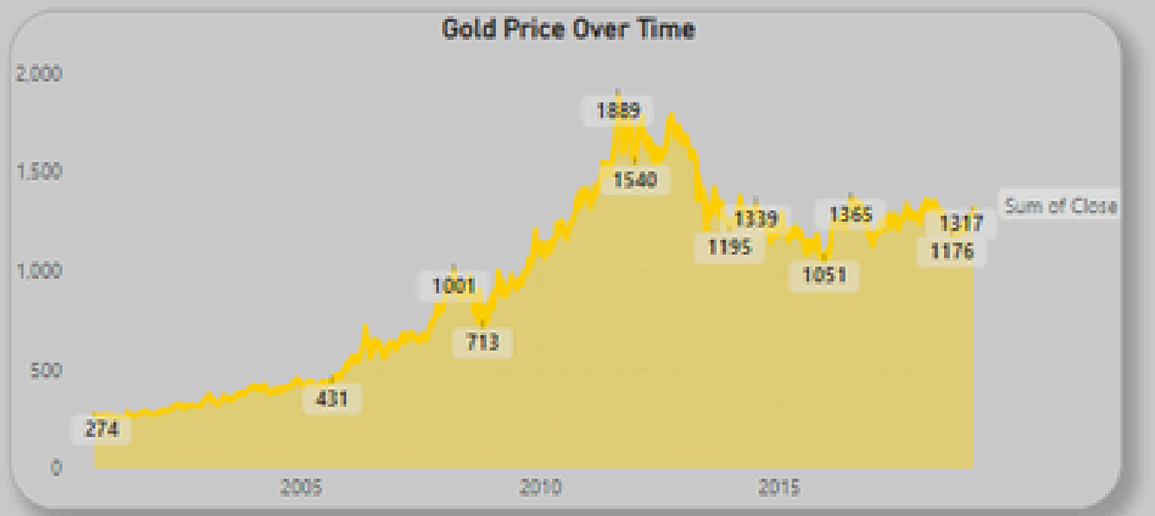
Date
07/09/2000 01/02/2019

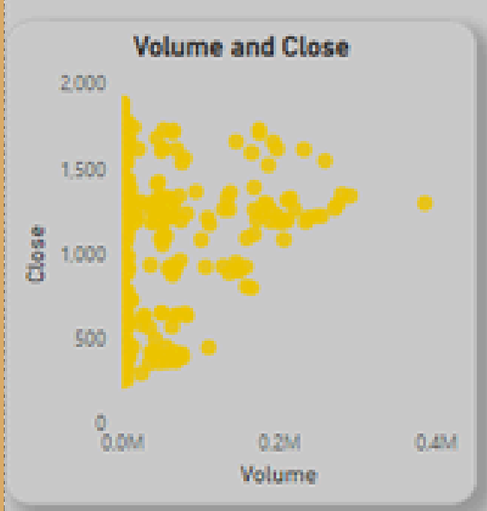
Price Sentiment
-1 1

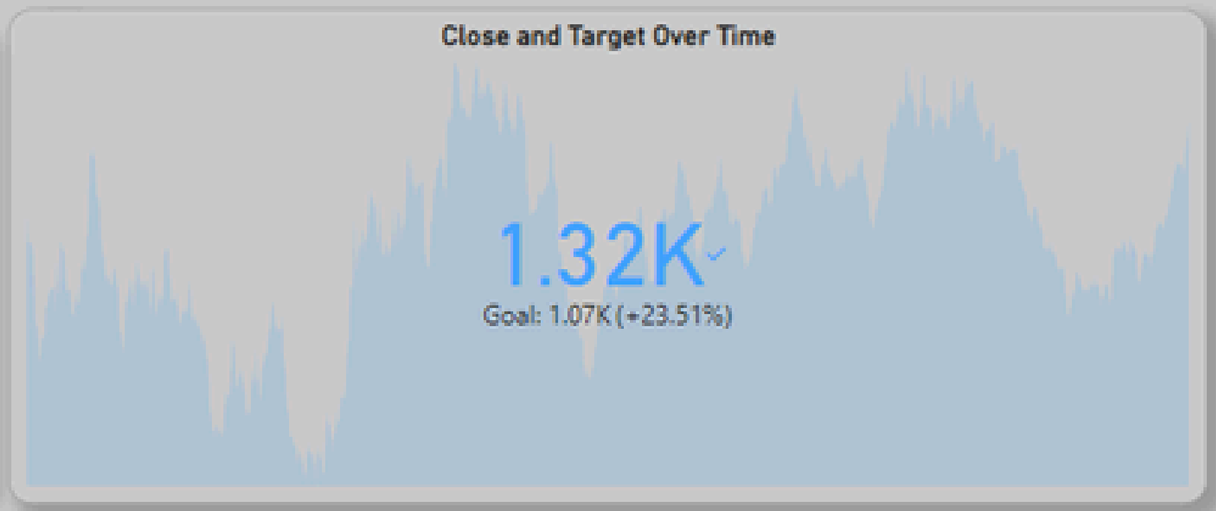
Close
256.10 1,888.70

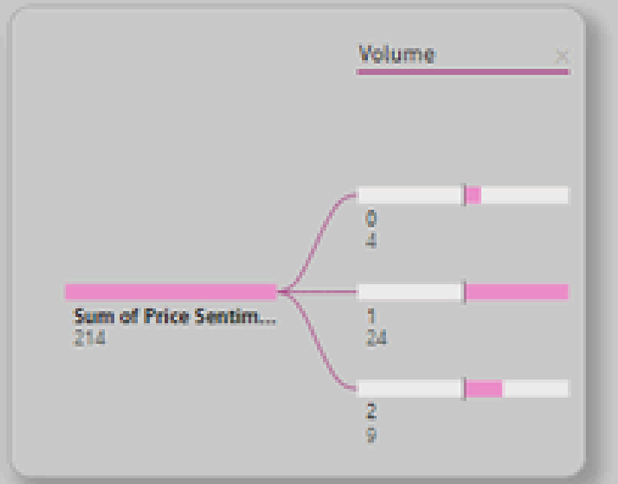












Regression

```
RandomForestRegressor
RandomForestRegressor(random_state=42)
```

```
y_pred = model.predict(X_test)
```

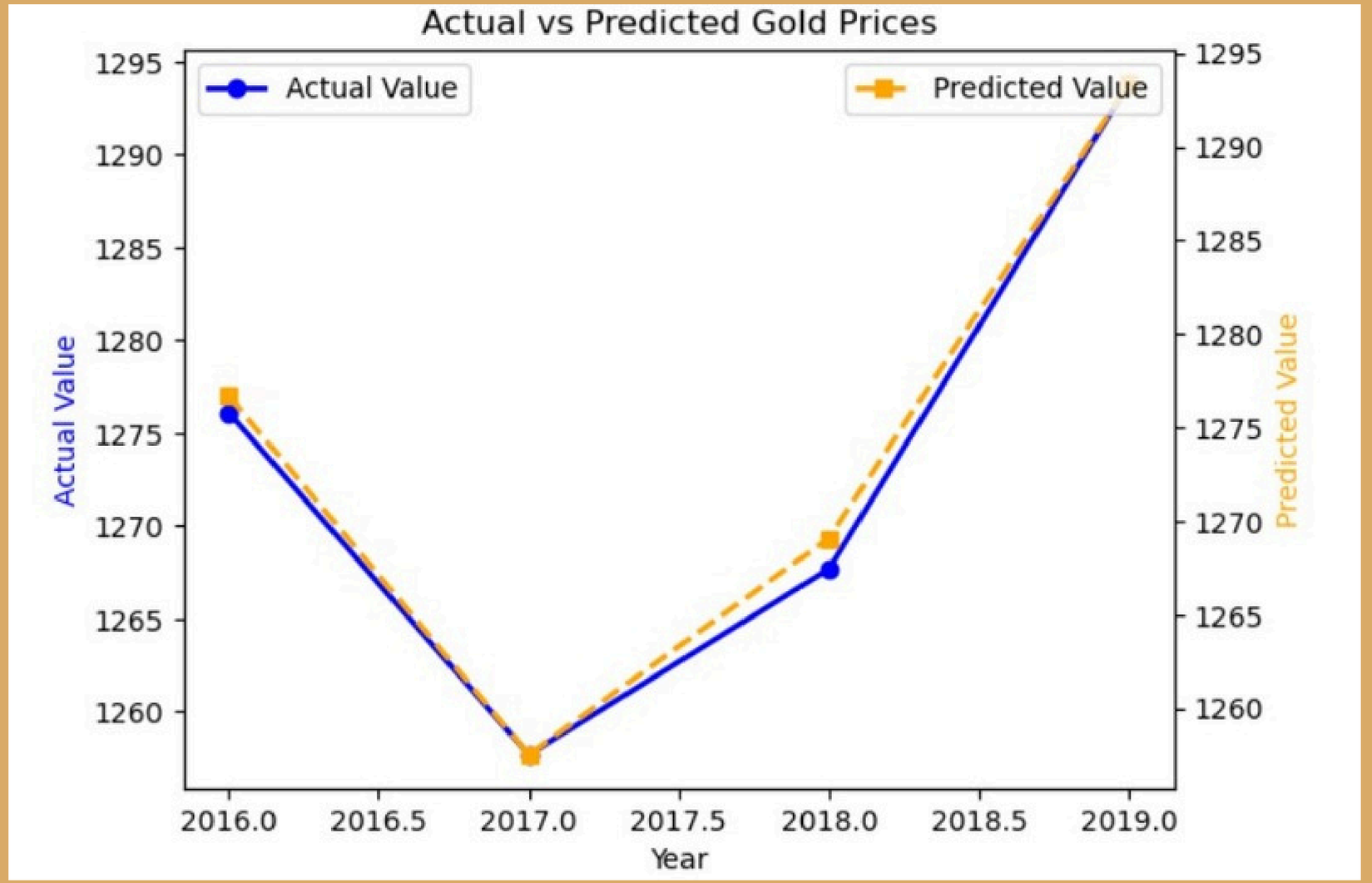
```
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
print(f'Mean Squared Error (MSE): {mse}')
print(f'Mean Absolute Error (MAE): {mae}')
```

```
Mean Squared Error (MSE): 104.23069125011506
Mean Absolute Error (MAE): 7.442910206492353
```

```
from sklearn.metrics import r2_score
r_squared = r2_score(y_test, y_pred)
print(f"R-squared: {r_squared * 100:.2f}%")
```

```
R-squared: 96.02%
```

Test & Predict Plot



Classification

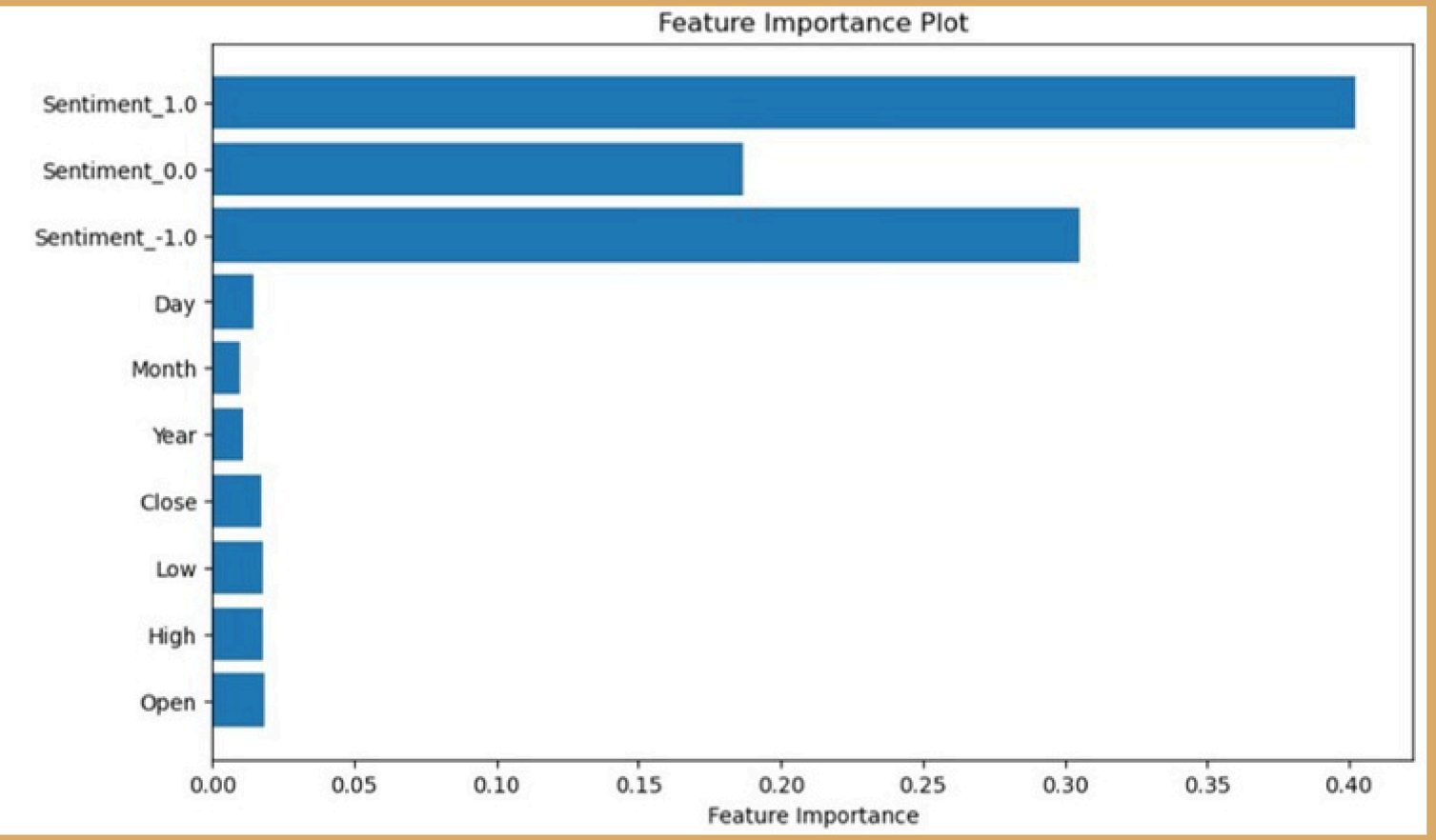
Classification

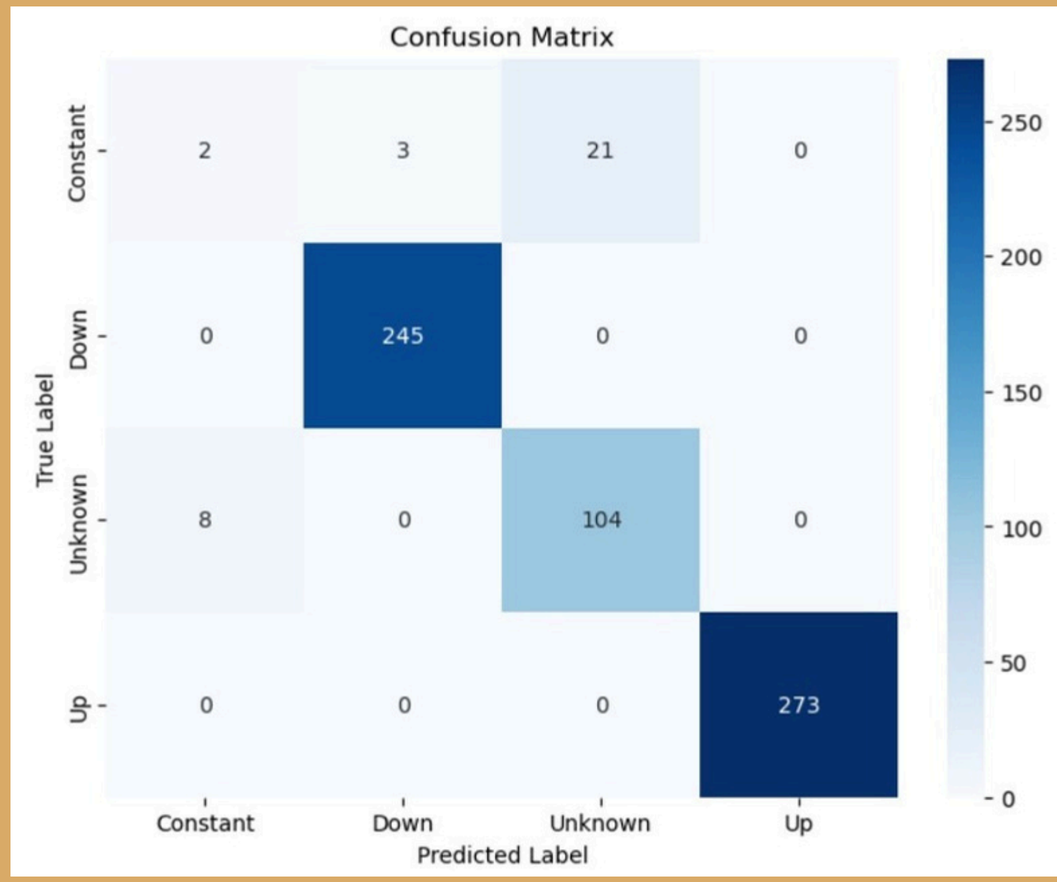
```
RandomForestClassifier
RandomForestClassifier(random_state=42)
```

```
y_pred = model.predict(X_test)

# Evaluate the model's performance
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy * 100:.2f}%')
```

```
Accuracy: 96.34%
```

Metrics


Feature Importance Plot
Sentiment_1.0
Sentiment_0.0
Sentiment_-1.0
Day
Month
Year
Close
Low
High
Open
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
Feature Importance


Confusion Matrix
Constant
Down
Unknown
Up
True Label
2
3
21
0
0
245
0
0
8
0
104
0
0
0
0
273
Constant
Down
Unknown
Up
Predicted Label
250
200
150
100
50
0

Time Series

ARIMA Model

SARIMA Model

ARIMA Model

1. ***Stationarity Check:***

ADF test was Applied

NON - Stationary

differencing was applied.

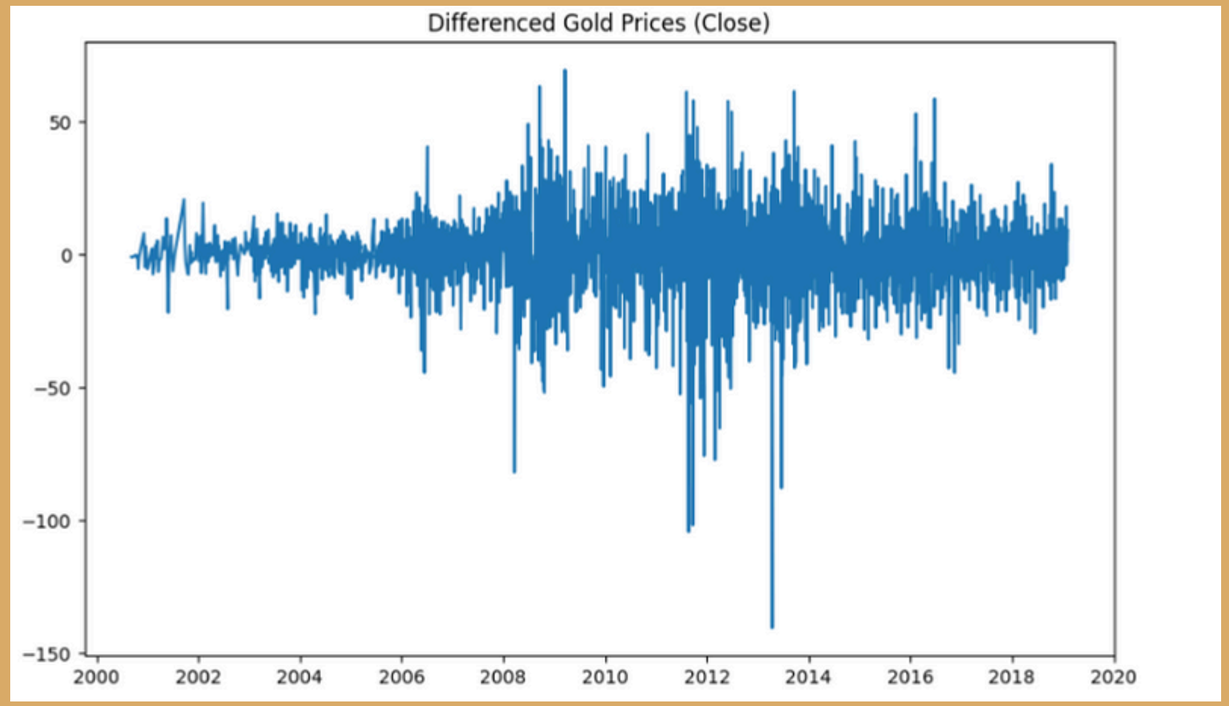
ADF Test Results:

Before differencing:
p-value: ***0.36200545419422614 (Non-stationary)***

After differencing: p-value is 0.0 (Stationary)

Before differencing:

After differencing:



ARIMA Model Overview

Fit the ARIMA Model

Predictions (Forecast)

Model Evaluation

ARIMA(1,1,1) Model Sumarry

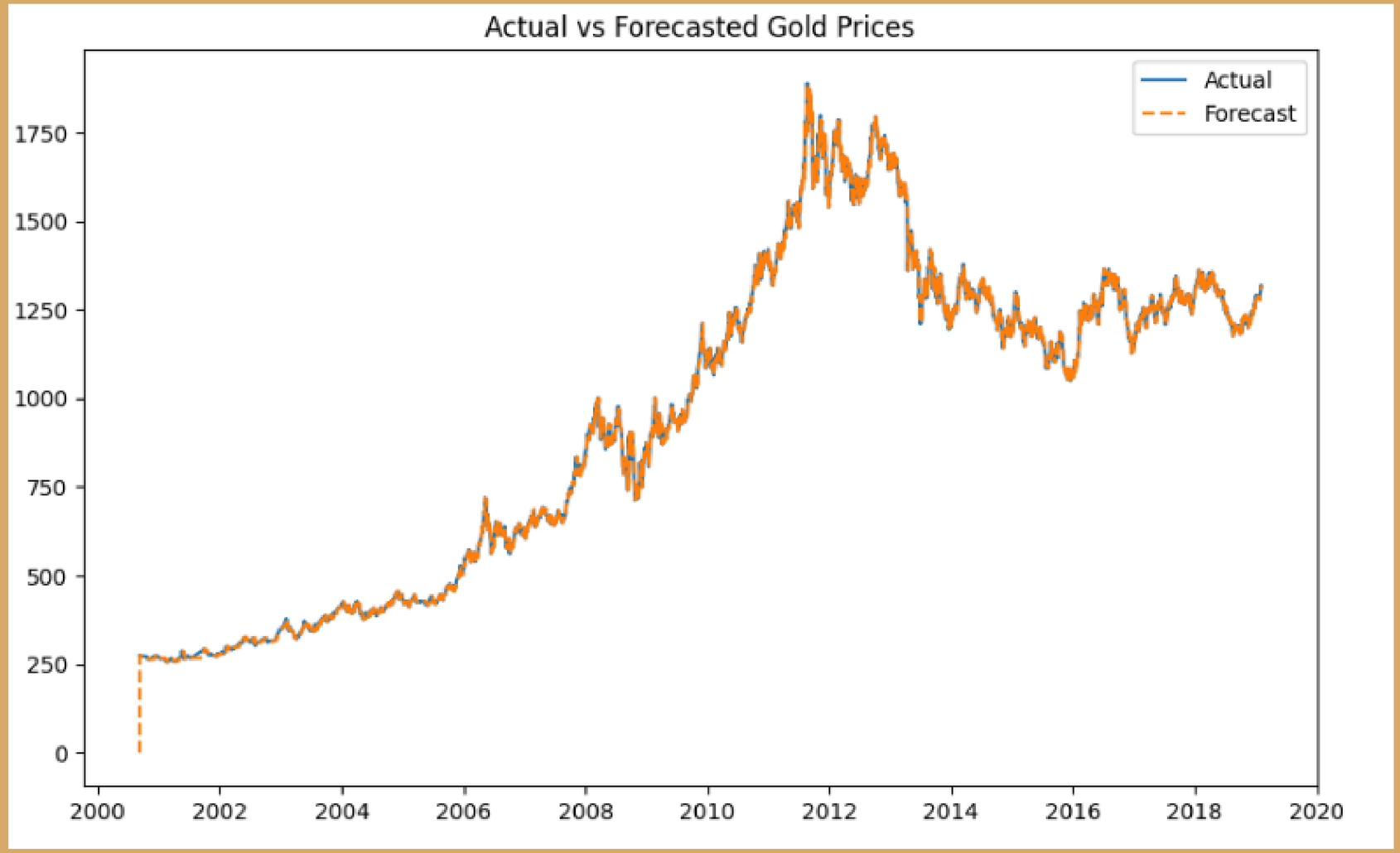
AIC : 68740.558

Predictions (Forecast)

The model is forecasting a slight downward trend in the Close prices over the next 10 periods.

```
Forecasted values for the next 10 periods:
3278    1316.419564
3279    1315.969450
3280    1315.547765
3281    1315.152716
3282    1314.782618
3283    1314.435896
3284    1314.111074
3285    1313.806768
3286    1313.521683
3287    1313.254604
```

Model Evaluation



Mean Squared Error: 213.93190396023306

ARIMA model Tuning

ARIMA(0, 1, 0)

The performance metrics for the ARIMA(0, 1, 0)
AIC :26517.40.
Mean Squared Error (MSE): 214.08

SARIMA Model Overview

Fit the SARIMA Model

Predictions (Forecast)

Model Evaluation

(ADF Test)

ADF Test Results:

Before differencing: p-value = 0.088 (Non-stationary)
After differencing: p-value = 1.17e-28 (Stationary)

SARIMA Model Selection

-SARIMA(1,1,1)(1,1,1,12) selected.

-Model chosen based on Auto-ARIMA’s search.

Model Evaluation & Forecasting

SARIMA AIC : 26518.926

ARIMA AIC: 71293

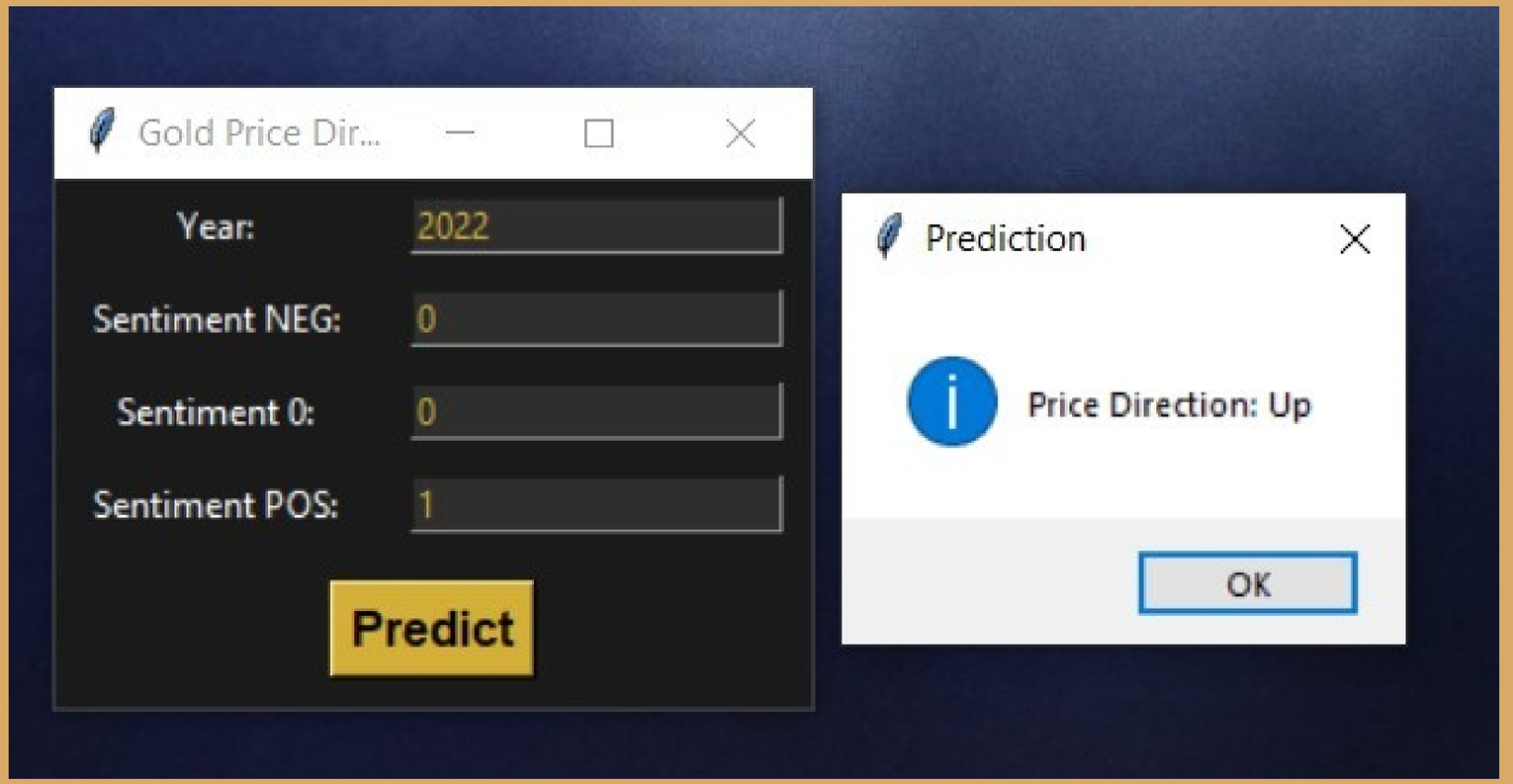
GUI

1. bash


Gold Price Direction Classifier
Open Price: 1500
Close Price: 1570
Year: 2021
Sentiment - NEG: 0
Sentiment - Neutral: 1
Sentiment - POS: 0
Predict

GUI

2. tkinter




Thanks