USING STATISTICAL MODEL AND MACHINE LEARNING FOR GOLD PRICE PREDICTION

# 0. ABSTRACT

This report presents a comparative analysis of three models, including Linear Regression, Gated Recurrent Unit (GRU), Autoregressive Integrated Moving Average (ARIMA), Long Short-Term Memory (LSTM) for predicting gold prices. The models are evaluated using MAE, MAPE, and RMSE metrics on historical gold price indicators data. The model with the lowest MAE, MAPE, and RMSE is recommended for gold price forecasting, contributing to improved understanding and accurate predictions in the gold market.

***Key words****: gold price, forecasting, linear regression, GRU, ARIMA, LSTM, ETS, random forest, BNN, GPR, and RNN*

# I. INTRODUCTION

Our research focused on forecasting gold prices, a commodity known for its historical stability and use as a currency and reserve asset. By analyzing key influencing factors, we aimed to generate precise predictions and insights into price fluctuations and trends. To accomplish this, we applied nine predictive algorithms, including GRU, ARIMA, LSTM. Our findings identify the model with the highest accuracy in gold price prediction, offering valuable guidance to investors and the public for making well-informed decisions about gold investments and purchases.

# II. RELATED WORKS

So…

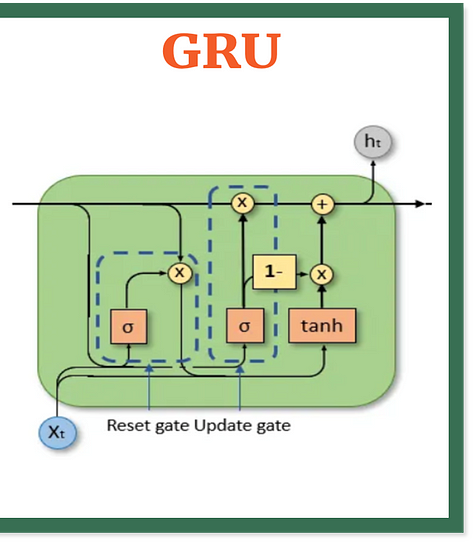
# III. ARTIFICIAL INTELLIGENCE MODELS

## 1. Gated Recurrent Unit (GRU)

The Gated Recurrent Unit (GRU) is a simplified type of Recurrent Neural Network (RNN) that is designed to handle the issue of long-term dependencies in time series data. It effectively deals with the challenge of modeling and predicting sequences with long time delays.

GRU employs two key gating mechanisms:

* **Update Gate**: Determines how much of the past information needs to be passed along to the future. It controls the extent to which the previous time step's data impacts the current time step.
* **Reset Gate**: Decides how much of the past information should be ignored or forgotten. It controls how much of the new input should be mixed with the previous state.



*Figure III.1: GRU Model*

The operations in a GRU are expressed as:

*Update Gate Calculation:*

*Reset Gate Calculation:*

*Current Memory Content:*

*Final Memory at Time t:*

Where:

* xₜ is the input at the current time step.
* hₜ₋₁ is the hidden state from the previous time step.
* W and U are the weight parameters.
* σ is the sigmoid activation function.
* tanh is the hyperbolic tangent function.
* ⊙ denotes element-wise multiplication.

GRUs are preferred for their ability to capture both short-term and long-term dependencies efficiently, with simpler architectures compared to LSTMs.

## 2. Autoregressive Integrated Moving Average (ARIMA)

ARIMA is a powerful model used for time series forecasting. It integrates three components: Autoregression (AR), Integration (I), and Moving Average (MA).

*AR (p): Autoregression*

The AR component looks at the relationship between the current value and past values (lags). It tries to predict the current value based on a set of previous observations.

* Where 𝜀ₜ represents random errors or "shocks" that cannot be predicted.

*I (d): Integration*

The I component deals with making the data stationary by finding differences between observations. It helps to remove trends or make a time series stable over time by comparing the current value with previous ones.

*MA (q): Moving Average*

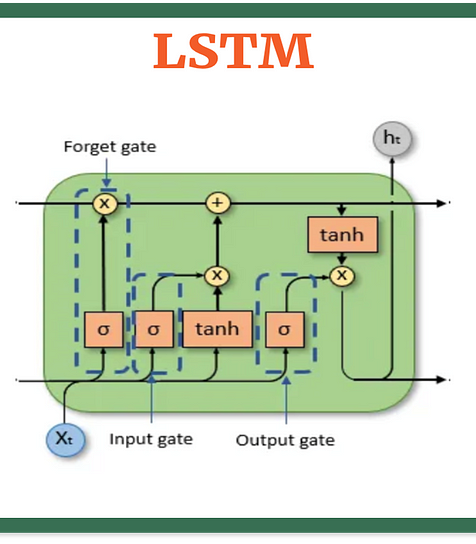
The MA component predicts the value by accounting for the past errors or random "shocks" in previous observations. It aims to smooth out the time series by combining the random variations that were not explained.

**Model Selection Process**

* ARIMA requires the time series data to be stationary, meaning the statistical properties (like mean and variance) are constant over time.
* Once the model is determined, it can be used to make predictions for future points in the time series.

## 3. Long Short-Term Memory (LSTM)

LSTM is a type of Recurrent Neural Network (RNN) used in deep learning that is capable of learning long-term dependencies. It is specifically designed to handle the vanishing gradient problem that occurs in traditional RNNs, making it useful for time series prediction.



*Figure III.2: LSTM Architecture*

**Components of LSTM**

*Forget Gate*: Determines how much of the past information should be forgotten.

*Input Gate:* Controls how much of the new information should be added to the cell state.

*Cell State Update:* Updates the cell state with new information.

*Output Gate:* Determines the output of the LSTM cell.

*Final Memory at Time Step:* Combines the updated cell state and output gate to produce the final output.

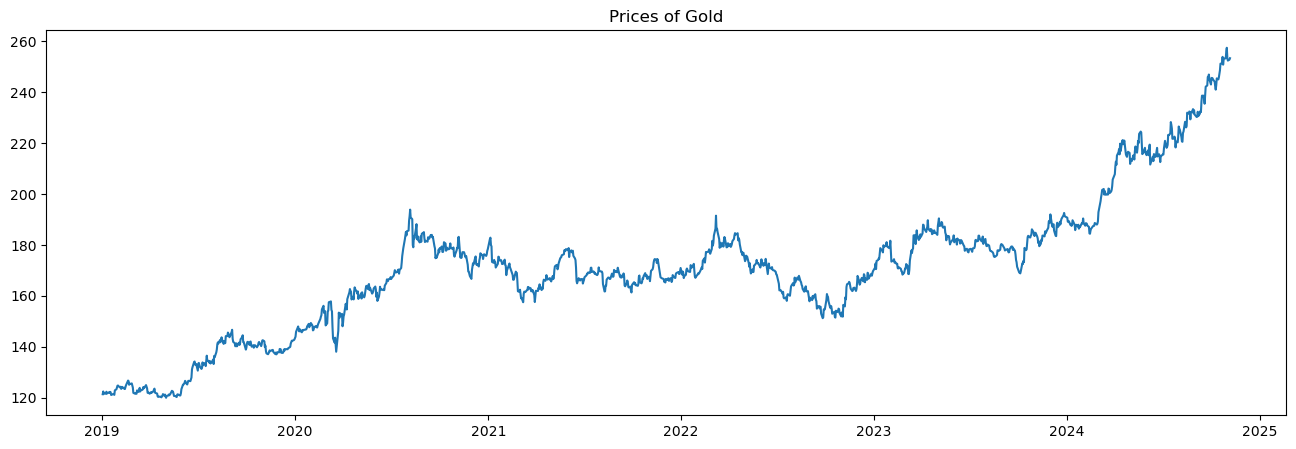
**How LSTM Works**

LSTM uses gates (forget, input, and output) to regulate the flow of information. This allows it to learn which information to keep, which to update, and which to forget over time, making it effective for capturing long-term dependencies in sequential data.

# IV. RESULT

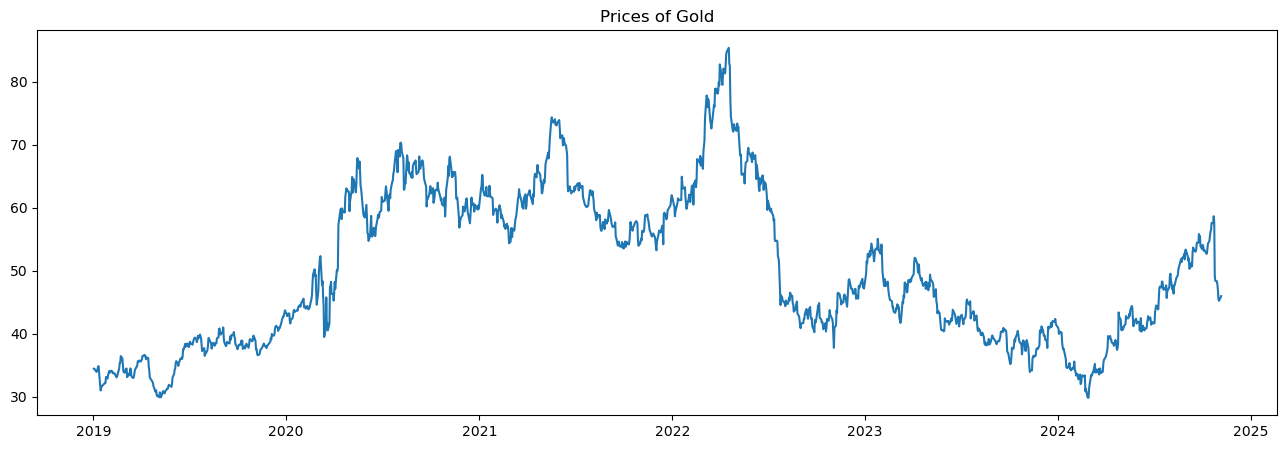
## A. Dataset

- GLD: The SPDR Gold Shares (GLD) is an exchange-traded fund (ETF) that tracks the price of gold. When you buy GLD, you are essentially buying a share of a basket of gold. GLD is traded on the New York Stock Exchange Arca (ARCA).Authors and Affiliations.



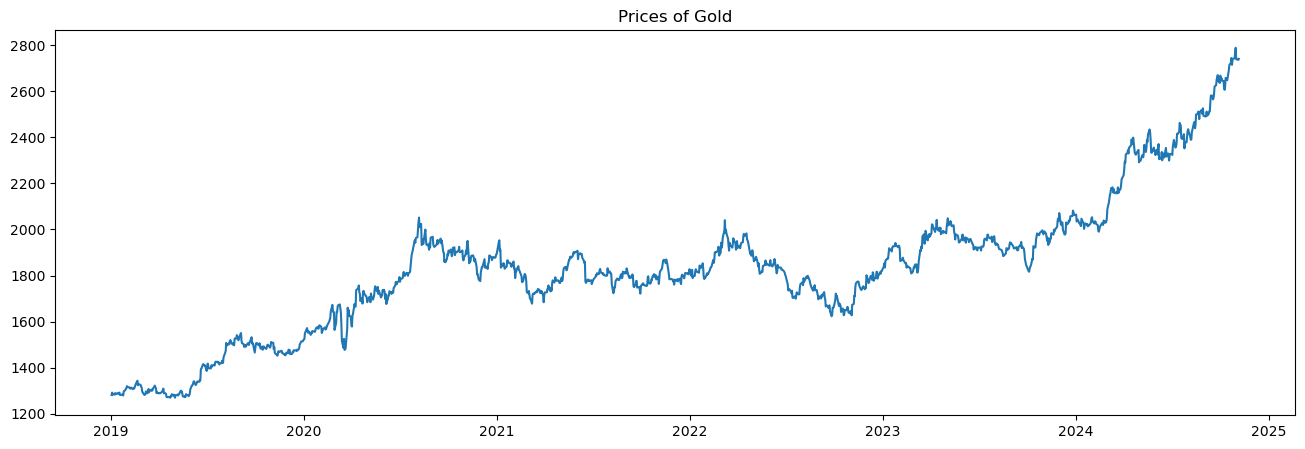
***Figure A.1: Close price by time of GLD dataset***

- NEM: Newmont Corporation is one of the largest gold mining companies in the world, with operations in the United States, Australia, Canada, Peru, and Ghana. The company's stock ticker symbol is NEM.



***Figure A.2: Close price by time of NEM dataset***

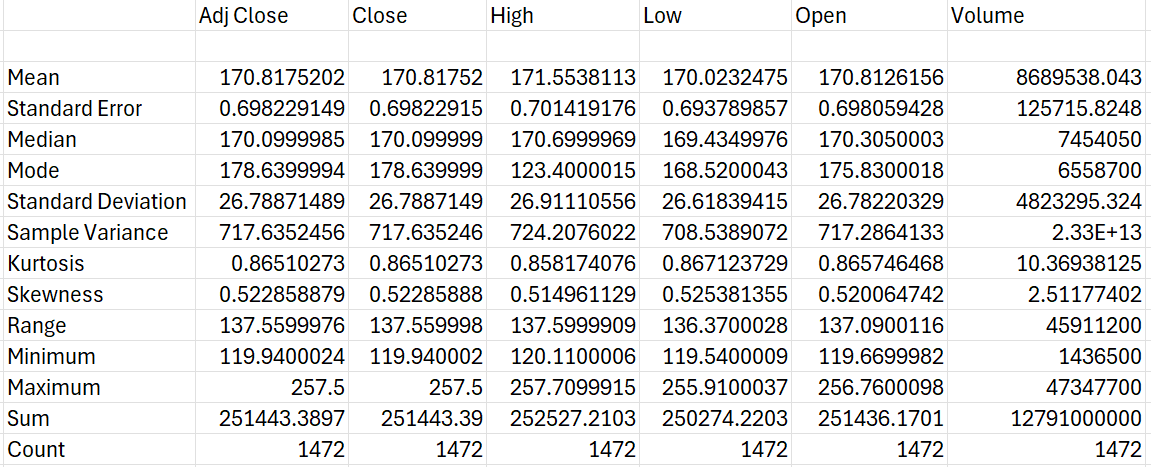
- GC=F: Gold Futures is another derivative investment that gives you the right to buy or sell a certain amount of gold at a predetermined price on a future date. The contract is also traded on the CME.



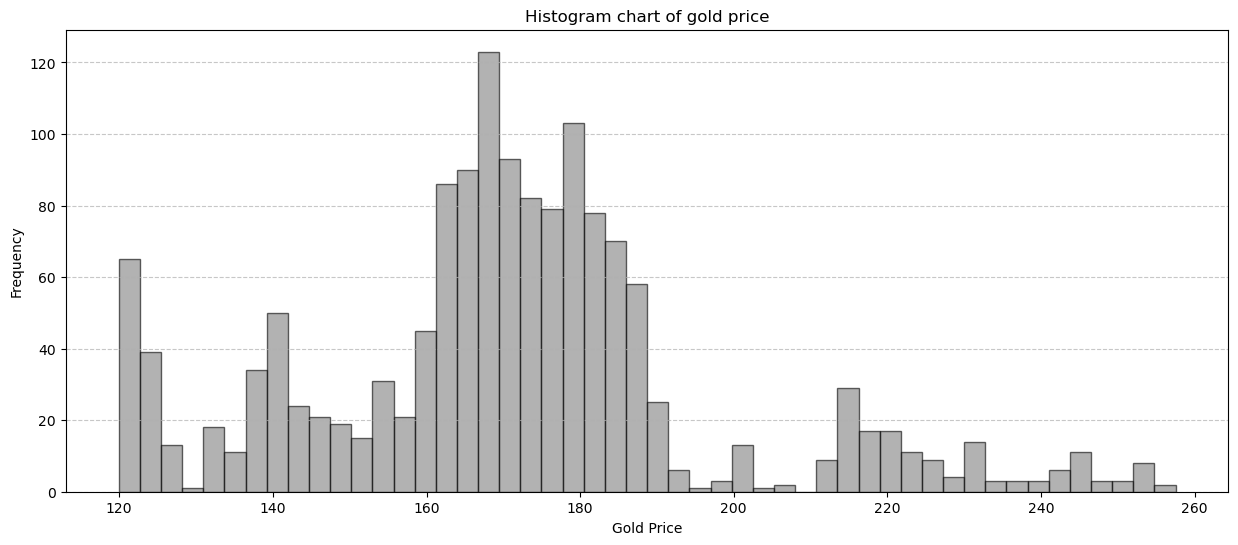
***Figure A.3: Close price by time of GC=F dataset***

## B. Descriptive statistic

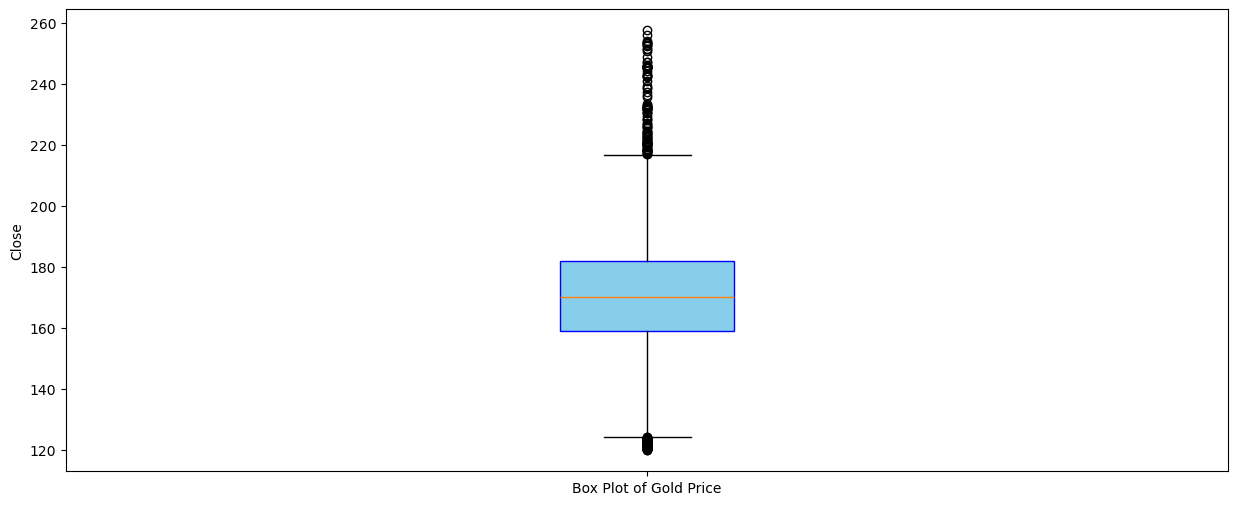
### B.1. GLD



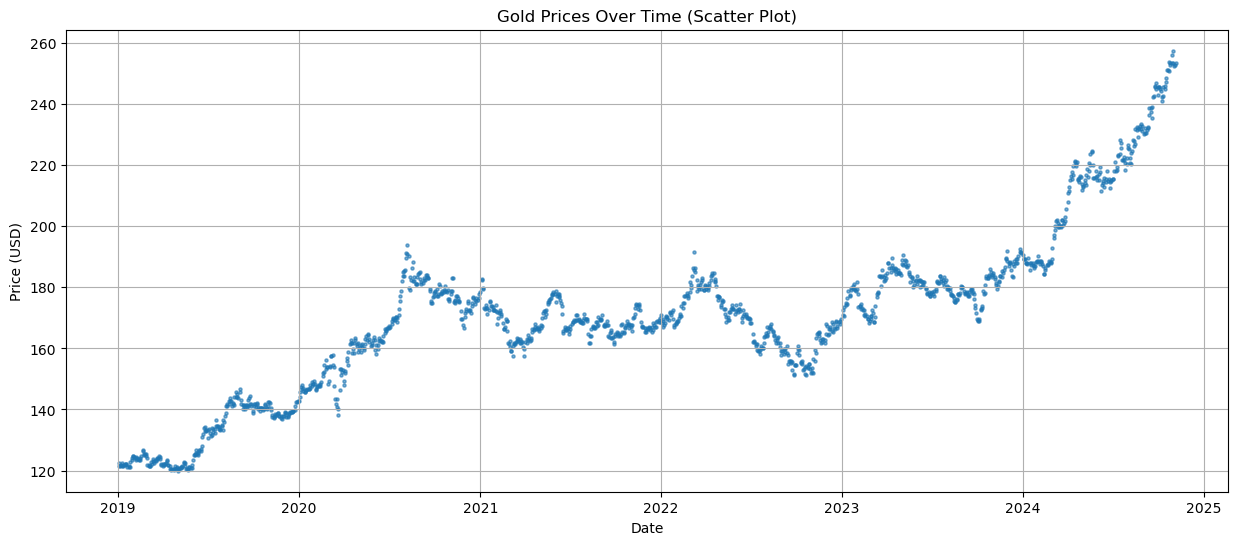
*Figure A.4: Statistic Description of GLD dataset*



*Figure A.5: Histogram char of GLD dataset*

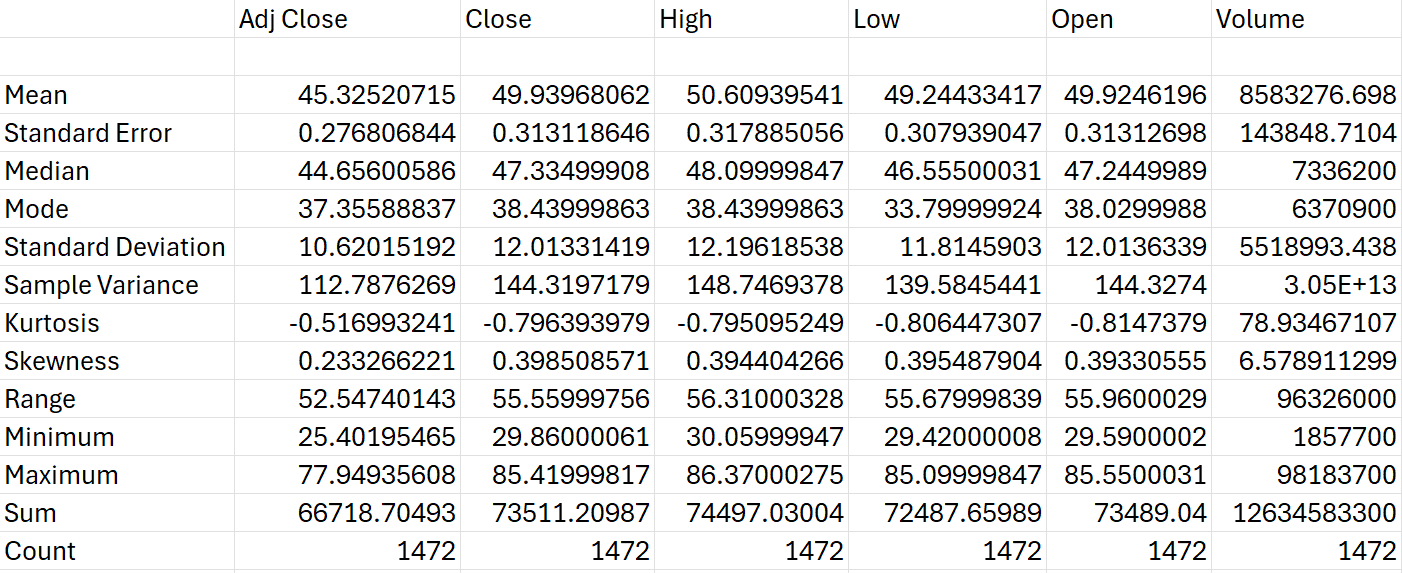


*Figure A.6: Box plot of GLD dataset*

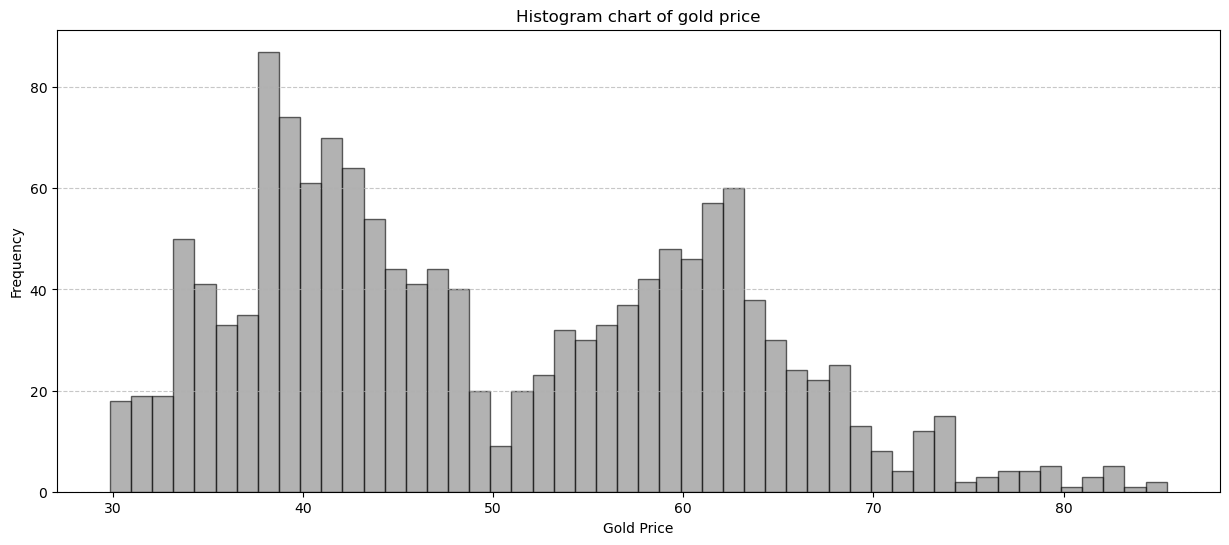


*Figure A.7: Scatter plot of GLD dataset*

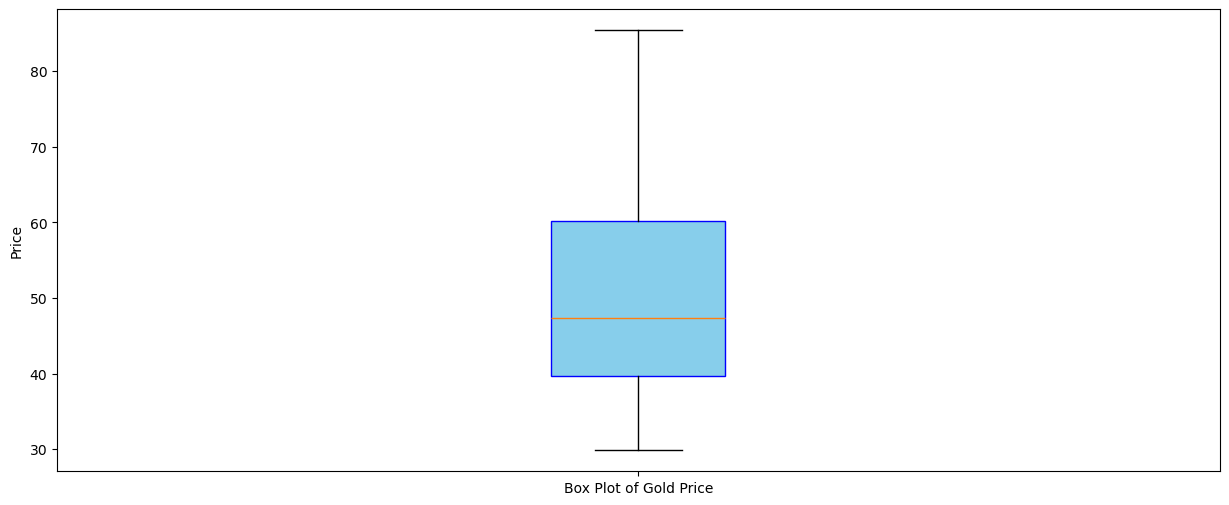
### B.2. NEM



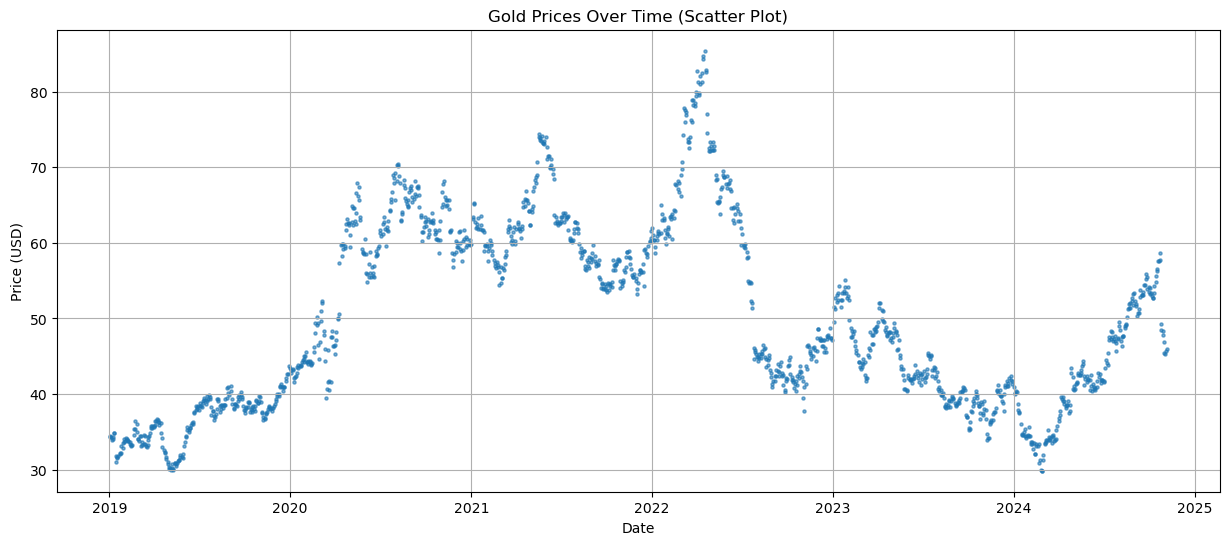
*Figure A.8: Statistic Description of NEM dataset*



*Figure A.9: Histogram chart of close price of NEM dataset*

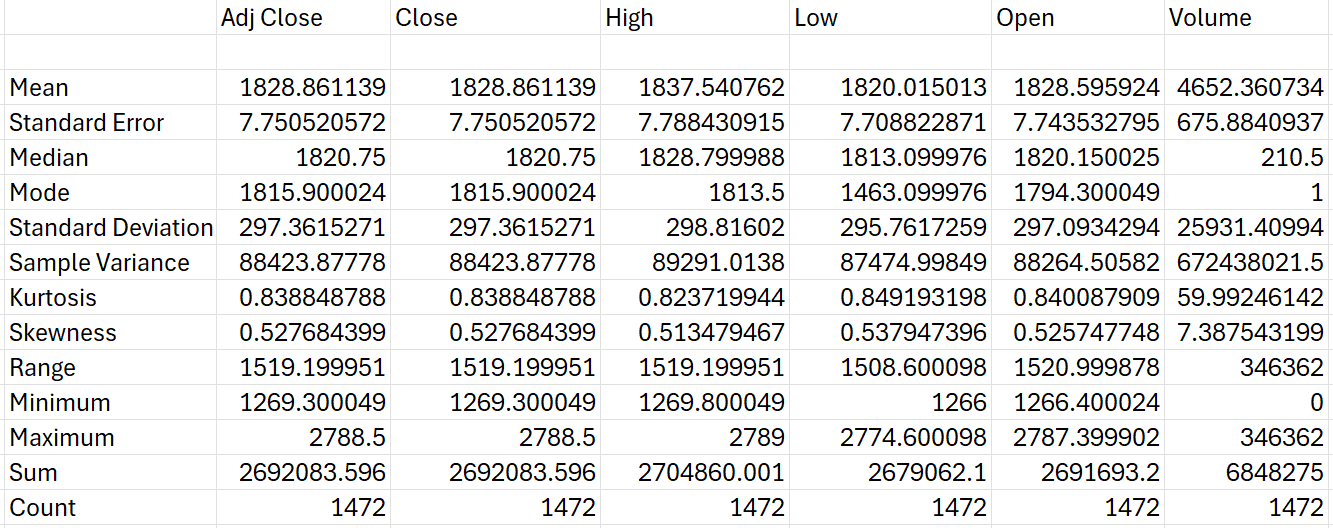


*Figure A.10: Box plot of close price of NEM dataset*

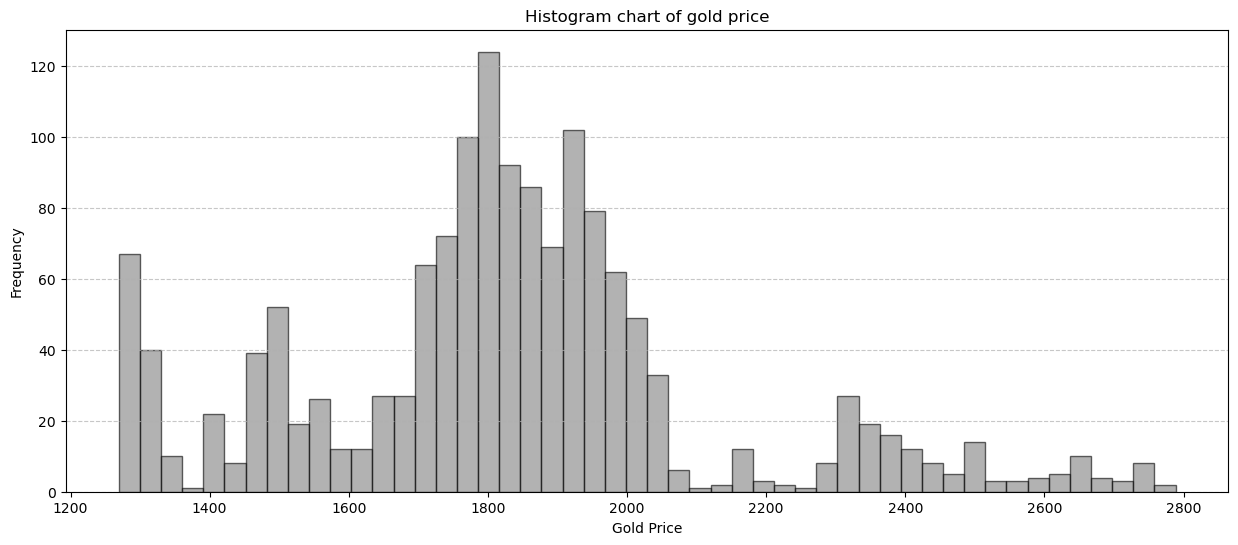


*Figure A.11: scatter plot of close price by time of NEM dataset*

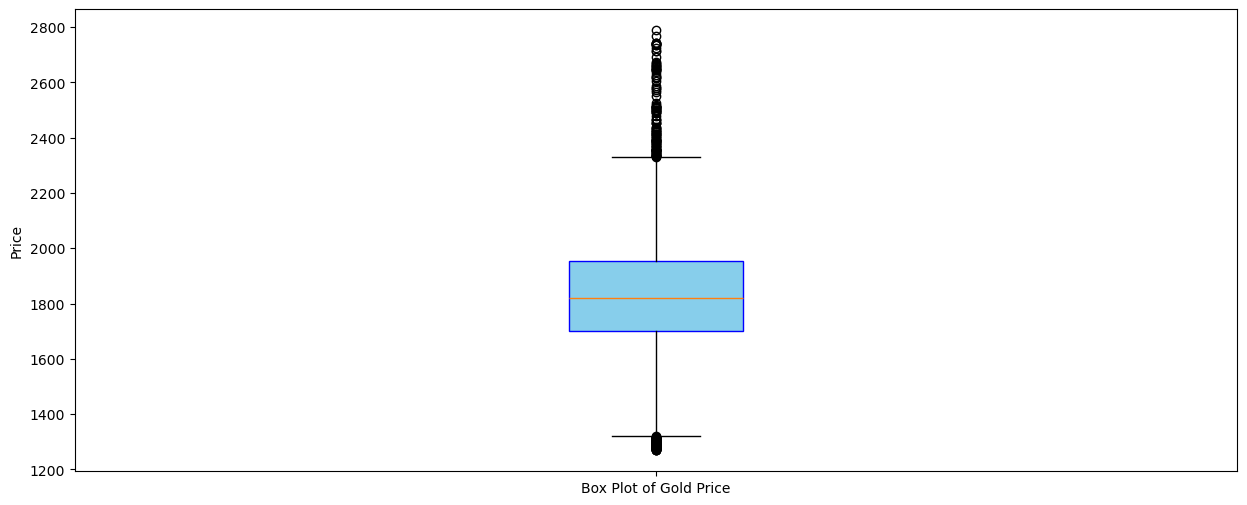
### B.3. GC=F



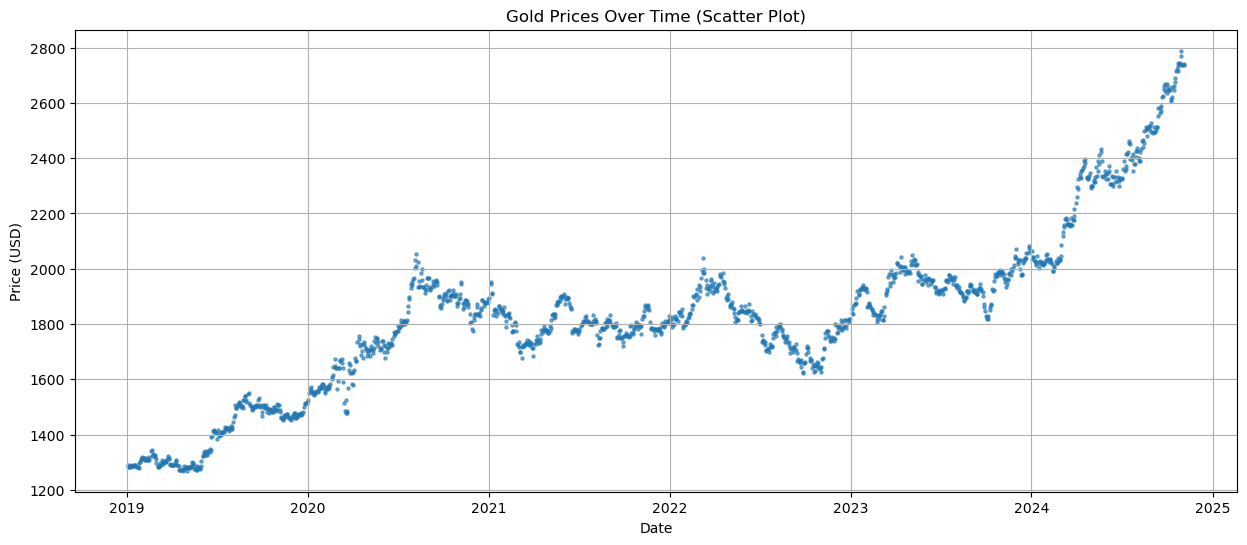
*Figure A.12: Statistic Description of GC=F dataset*



*Figure A.13: Histogram chart of close price of GC=F dataset*



*Figure A.14: Box plot of close price of GC=F dataset*



*Figure A.15: Scatter plot of close price of GC=F dataset*

## C. Inferential statistic

## D. Technology

### D.1. Tool and libraries

In this paper, various tools such as Python, Excel, R, and SPSS will be used for data analysis, each offering unique capabilities for different analytical needs:

* Python: A powerful, flexible language widely used in data analysis and machine learning. With a rich ecosystem of libraries like *Pandas, Matplotlib, TensorFlow, Keras, Sklearn*, and *Statsmodels*, Python is highly suitable for data manipulation, visualization, statistical modeling, and deep learning. Its open-source nature and extensive community support make it a preferred choice for researchers and data scientists.
* Excel: Known for its simplicity and accessibility, Excel is a popular tool for data management, basic analysis, and visualization. It allows users to handle data in a grid format, perform statistical calculations, and create charts, making it ideal for preliminary data exploration and straightforward analysis.
* R: A statistical programming language developed specifically for data analysis and visualization. R is renowned for its powerful statistical packages and visualizations, making it a staple for more complex statistical analyses. With libraries like *ggplot2, dplyr*, and *caret,* R is highly suitable for advanced statistical modeling, data wrangling, and machine learning.
* SPSS: A dedicated software for statistical analysis in social sciences, SPSS is user-friendly and equipped with an extensive suite of statistical tests, regression models, and reporting tools. Its point-and-click interface allows users with limited programming knowledge to conduct sophisticated analyses, making it ideal for academic research and survey data processing.

### D.2. Splitting data:

We will split the dataset into 3 ratios (Train: Test: Validate): 7:2:1, 6:3:1, 5:3:2. To build a better machine learning model that generalizes well on new data and avoids overfitting.

### D.3. Evaluation:

Evaluating a model is crucial to ensure that it is accurate, reliable, and generalizes well to new data. The following metrics will be used: MAE, MAPE and RMSE.

*MAE* (Mean Absolute Error): A measure of the average size of the mistakes in a collection of predictions, without taking their direction into account. It is measured as the average absolute difference between the predicted values and the actual values and is used to assess the effectiveness of a model.

*MAPE (Mean Absolute Percentage Error):* Calculates the accuracy of the forecasting method by finding the average of absolute percentage errors for each entry in the dataset. It is useful for understanding the accuracy in a relative context, particularly for larger datasets where direct value differences can be hard to interpret. MAPE is effective for comparisons across different scales, but it is less suitable when actual values are zero.

*RMSE (Root Mean Square Error):* Provides an estimate of the standard deviation of prediction errors. It considers the square of each error, giving more weight to larger differences between predicted and actual values. This helps to identify models that struggle with large outliers. RMSE is especially useful when dealing with normally distributed data and provides insight into how well the model predicts over the entire dataset.

### D.4. Results

#### D.4.1. GLD

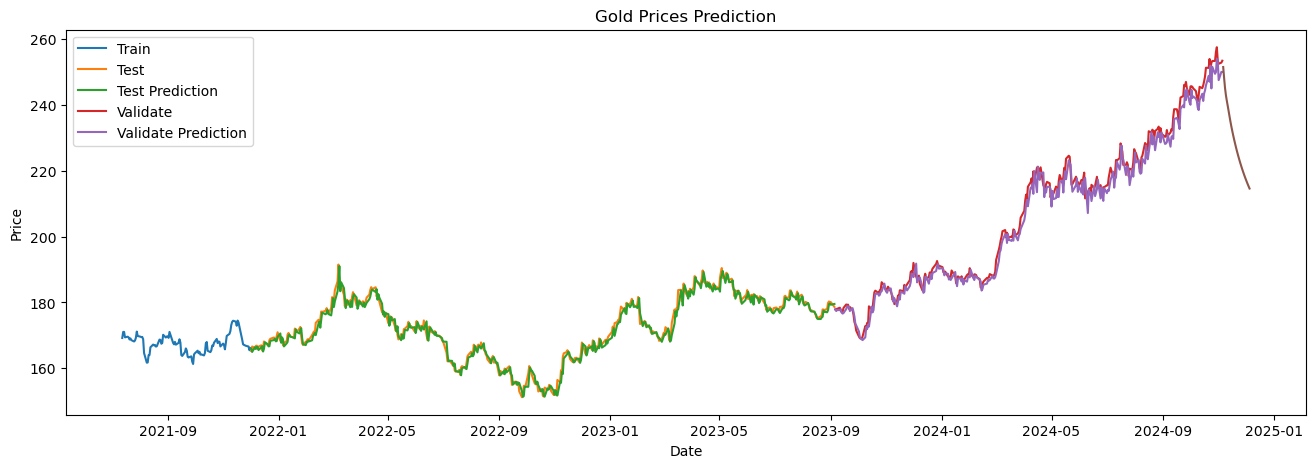
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Ratio** | **Testing** | | | **Validation** | | |
| **MAE** | **MAPE (%)** | **RMSE** | **MAE** | **MAPE (%)** | **RMSE** |
| ***GRU*** | *7-2-1* | 1.27 | 0.69 | 1.61 | 1.92 | 0.85 | 2.39 |
| *6-3-1* | 1.38 | 0.77 | 1.79 | 3.76 | 1.63 | 4.29 |
| *5-3-2* | 1.18 | 0.69 | 1.56 | 1.77 | 0.83 | 2.35 |
| ***ARIMA*** | *7-2-1* | 30.97 | 0.21 | 33.38 | 57.85 | 0.21 | 57.99 |
| *6-3-1* | 146.22 | 0.09 | 16.31 | 32.83 | 0.27 | 33.09 |
| *5-3-2* | 7.80 | 0.05 | 9.87 | 33.76 | 0.26 | 35.02 |
| ***LSTM*** | *7-2-1* | 1.52 | 0.82 | 1.99 | 6.41 | 2.77 | 7.12 |
| *6-3-1* | 1.23 | 0.70 | 1.62 | 3.14 | 1.36 | 3.86 |
| *5-3-2* | 1.35 | 0.78 | 1.74 | 2.79 | 1.28 | 3.59 |

*Table 1: Exchange-traded Fund Data Evaluation (GLD)*

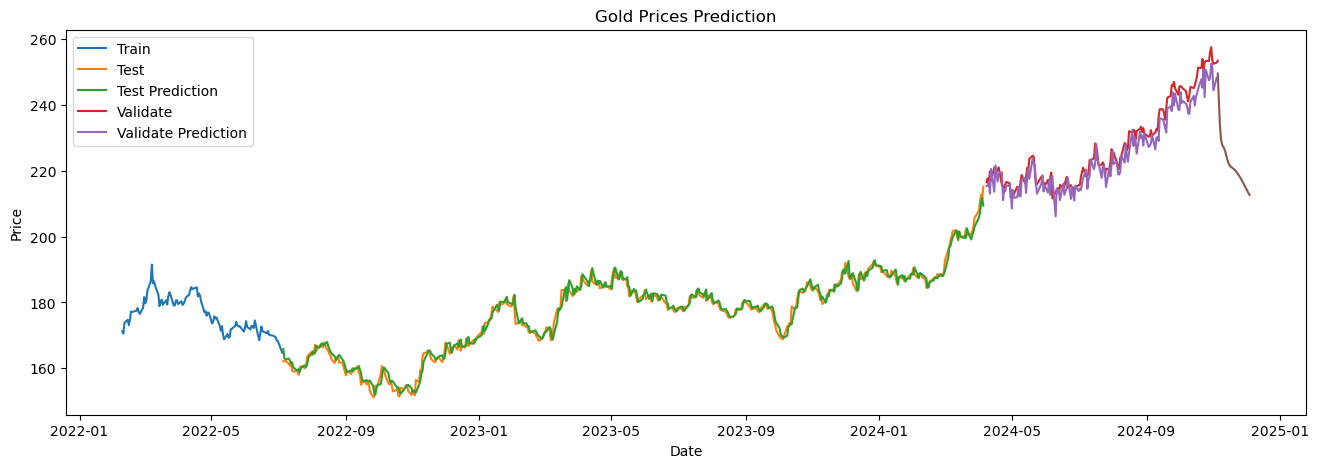
The GRU model with a 5-3-2 ratio is the most accurate, with consistently low values for both Testing and Validation.

The GRU model with a 7-2-1 ratio is a close second and can also be highlighted for good performance.

The LSTM model (6-3-1 ratio) shows reasonable accuracy for testing but exhibits greater discrepancies in validation.



*Figure D.1: GRU model with a 5-3-2 ratio of GLD*



*Figure D.3: LSTM model with a 6-3-1 ratio of GLD*

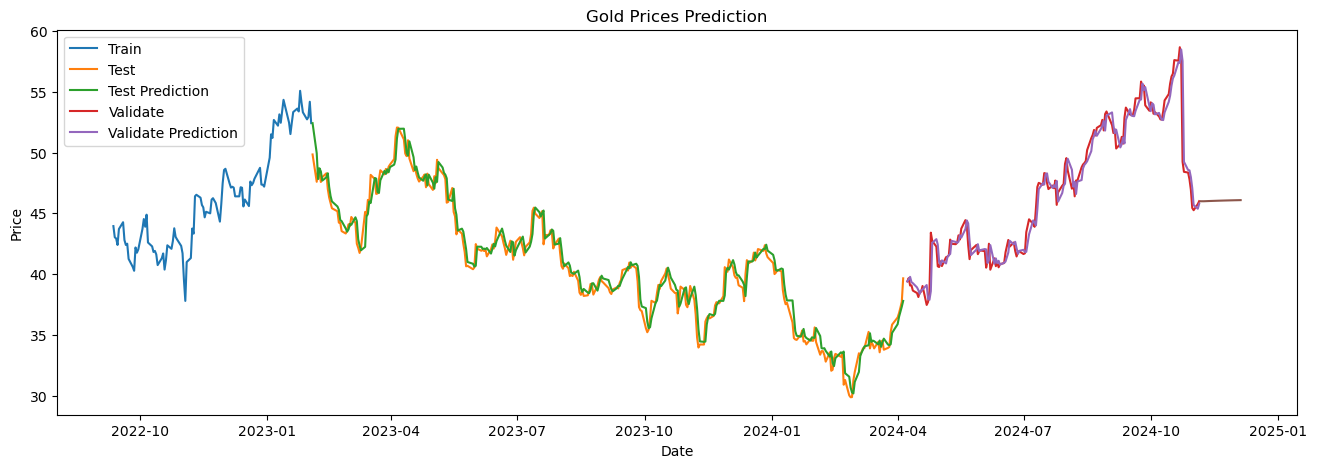
#### D.4.2. NEM

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Ratio** | **Testing** | | | **Validation** | | |
| **MAE** | **MAPE (%)** | **RMSE** | **MAE** | **MAPE (%)** | **RMSE** |
| ***GRU*** | *7-2-1* | 0.68 | 1.69 | 0.89 | 0.71 | 1.52 | 1.12 |
| *6-3-1* | 0.72 | 1.71 | 0.98 | 0.70 | 1.51 | 1.12 |
| *5-3-2* | 0.90 | 1.70 | 1.27 | 0.69 | 1.69 | 1.01 |
| ***ARIMA*** | *7-2-1* | 13.46 | 0.32 | 16.29 | 27.69 | 0.81 | 27.81 |
| *6-3-1* | 11.35 | 0.26 | 15.12 | 30.24 | 0.88 | 30.35 |
| *5-3-2* | 8.22 | 0.13 | 9.35 | 15.58 | 0.43 | 16.15 |
| ***LSTM*** | *7-2-1* | 0.74 | 1.87 | 0.97 | 0.76 | 1.63 | 1.18 |
| *6-3-1* | 0.86 | 2.08 | 1.14 | 0.91 | 1.93 | 1.30 |
| *5-3-2* | 0.97 | 1.86 | 1.32 | 0.84 | 2.11 | 1.19 |

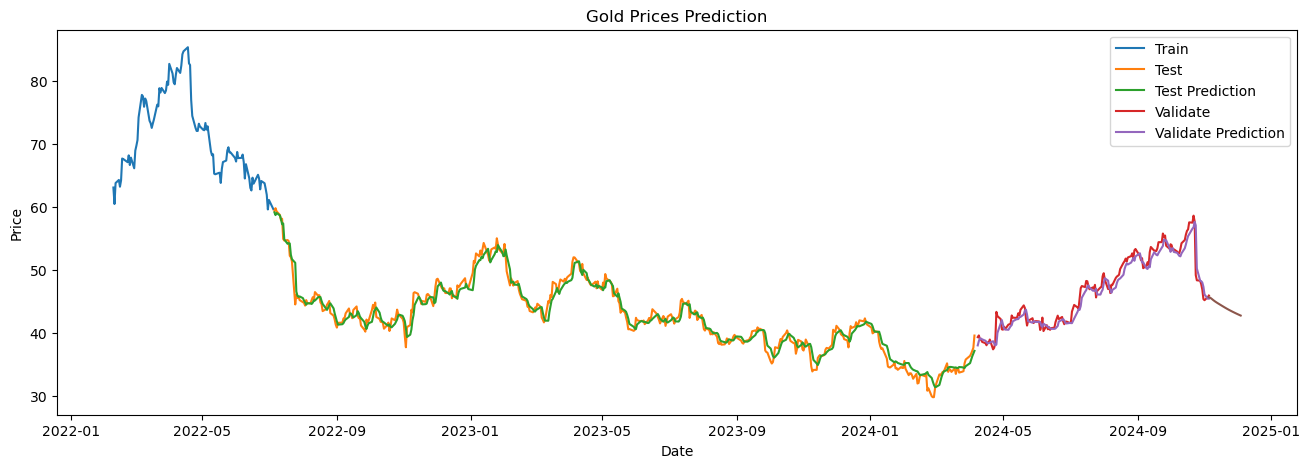
*Table 2: Newmont Corporation Data Evaluation (NEM)*

The GRU model with a 7-2-1 ratio has the most favorable accuracy metrics for both Testing and Validation. This would be the most suitable model to highlight.

LSTM (6-3-1) could be considered for its relatively strong performance, though it's slightly less accurate than GRU (7-2-1).



*Figure D.4: GRU model with a 7-2-1 ratio of NEM*



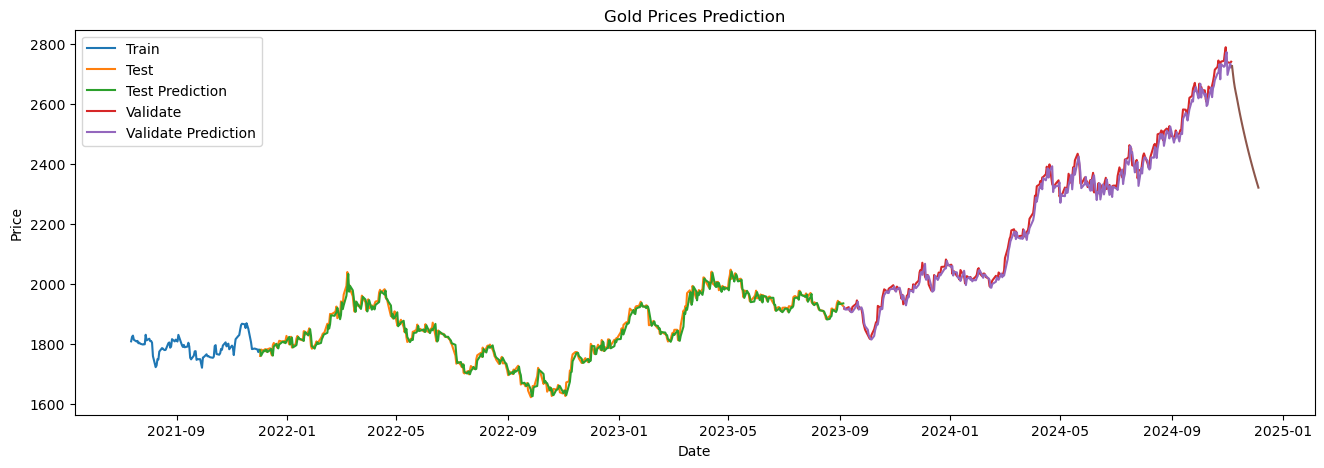
*Figure D.5: LSTM model with a 7-2-1 ratio of NEM*

#### D.4.3. GC=F

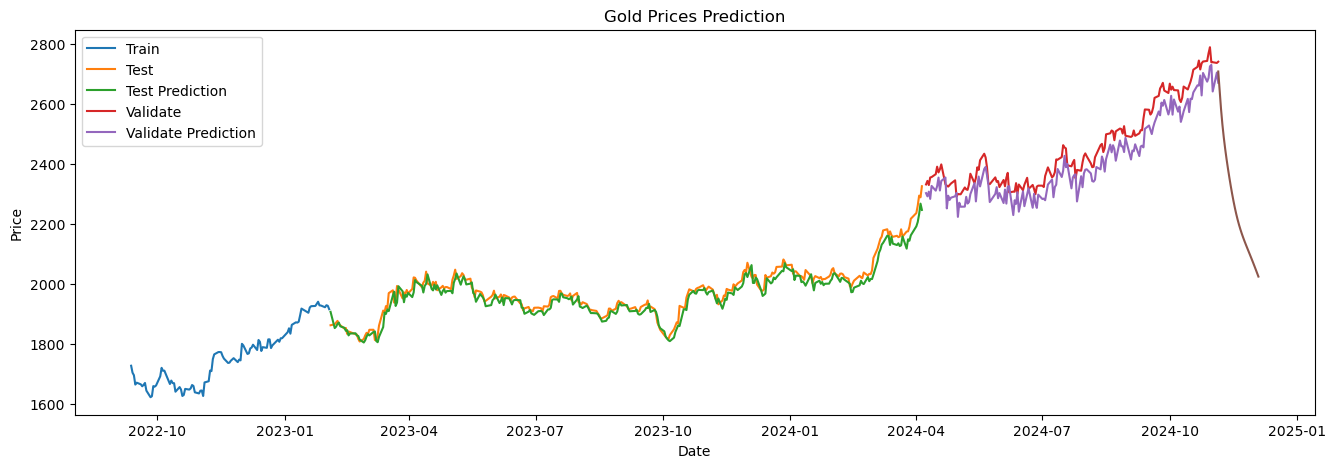
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Ratio** | **Testing** | | | **Validation** | | |
| **MAE** | **MAPE (%)** | **RMSE** | **MAE** | **MAPE (%)** | **RMSE** |
| ***GRU*** | *7-2-1* | 13.99 | 0.70 | 18.50 | 25.96 | 1.05 | 30.37 |
| *6-3-1* | 13.09 | 0.68 | 17.49 | 36.65 | 1.47 | 41.62 |
| *5-3-2* | 12.99 | 0.70 | 17.15 | 18.58 | 0.81 | 23.44 |
| ***ARIMA*** | *7-2-1* | 81.82 | 0.04 | 108.26 | 544.86 | 0.22 | 562.36 |
| *6-3-1* | 165.94 | 0.08 | 193.36 | 699.35 | 0.28 | 713.07 |
| *5-3-2* | 102.99 | 0.05 | 123.06 | 458.52 | 0.19 | 523.36 |
| ***LSTM*** | *7-2-1* | 13.30 | 0.66 | 17.95 | 45.28 | 1.81 | 53.28 |
| *6-3-1* | 15.52 | 0.80 | 20.37 | 45.71 | 1.84 | 51.52 |
| *5-3-2* | 15.58 | 0.84 | 20.23 | 52.23 | 2.19 | 65.42 |

*Table 3: Gold Futures Data Evaluation (GC=F)*

The GRU model with a 5-3-2 ratio stands out with the lowest metrics overall



*Figure D.6: GRU model with a 5-3-2 ratio of GC=F*



*Figure D.7: LSTM model with a 7-2-1 ratio of GC=F*