**Feature Engineering – Fall 2022**

**Project Increment-2**

**PREDICTION OF GOLD PRICES FOR FINANCIAL INSTITUTIONS**

**1.Team Members:**

1. Yash Zauwar
2. Goutham Pallapothu
3. Veeranjaneyulu Muppalla
4. Sai Manideep Reddy Pallerla

**2. Introduction:** To develop a machine learning model which will take the past prices data of Gold and predict the price for the future. To help investors to decide to buy or sell.

* **Motivation:** Predicting prices of commodities which are traded in the financial markets is very crucial. It plays a very important role in the success of any financial institution’s business.
* **Significance:** More and more investors are finding gold to be an appealing investment due to the rise in gold's value as well as the volatility and declining prices in other sectors, such as the capital and real estate markets. Developing a model to predict the price will help the institutions to attract clients and to generate the best returns.
* **Objectives:** To apply the best machine learning algorithms. This research is based on the applicability of the suggested machine learning methods, which have proven to be effective at forecasting gold prices more accurately**.**
* **Features:** We are using features like numerical features, discrete numerical features, correlation between the numerical features, from the data to applya Machine Learning Model using LSTM. With the help of data sources of price of gold, the data will be divided into training data and testing data then the Linear Regression Model will be applied to the data to generate few visualizations and forecast the price.

**Background:**

* **Related Work:** Numerous studies have been done on the price of gold. Gold prices are regressing against the US dollar and stock returns despite the incorporation of several different variables in this study. Numerous academics have also investigated the connection between other macroeconomic factors and gold prices. The correlation between the price of gold and the pricing of other commodities, particularly crude oil, has also been the subject of numerous studies. The results of these investigations, however, appear to be contradictory. To investigate the volatility of the gold price and its relationship to the variables that are thought to have an impact on it, researchers have used a variety of techniques.
* **Related Work References:**

1)<https://www.irjmets.com/uploadedfiles/paper//issue_5_may_2022/22138/final/fin_irjmets1651655991.pdf>

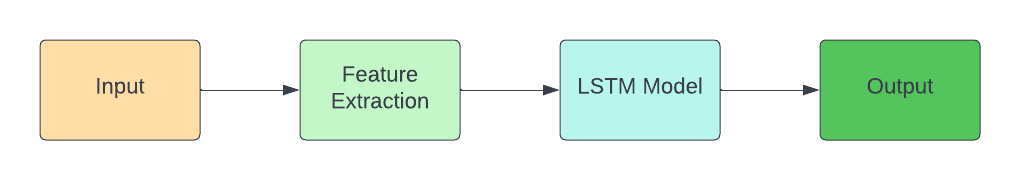
2) <https://www.semanticscholar.org/paper/Predicting-Future-Gold-Rates-using-Machine-Learning-Sami-Junejo/d1ddf3a4df034e89919df73b8c367c33cf5fb3b9>

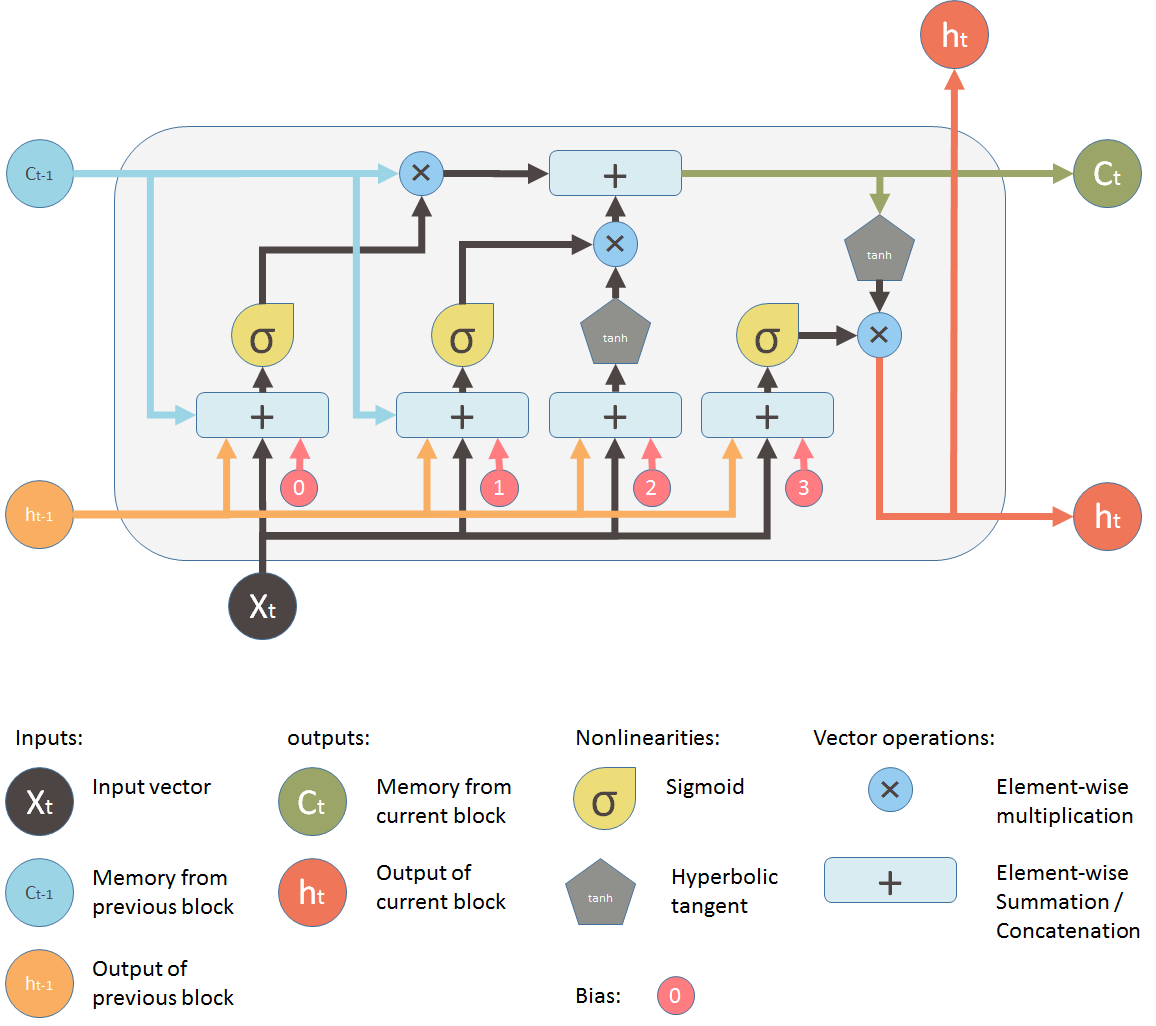
3)<https://www.researchgate.net/publication/336414732_Gold_Price_Prediction_using_Ensemble_based_Machine_Learning_Techniques>

**Model**:

**Architecture Diagram with explanation:**

We are using the LSTM Model.

****



**Reference**: <https://blog.mlreview.com/understanding-lstm-and-its-diagrams-37e2f46f1714>

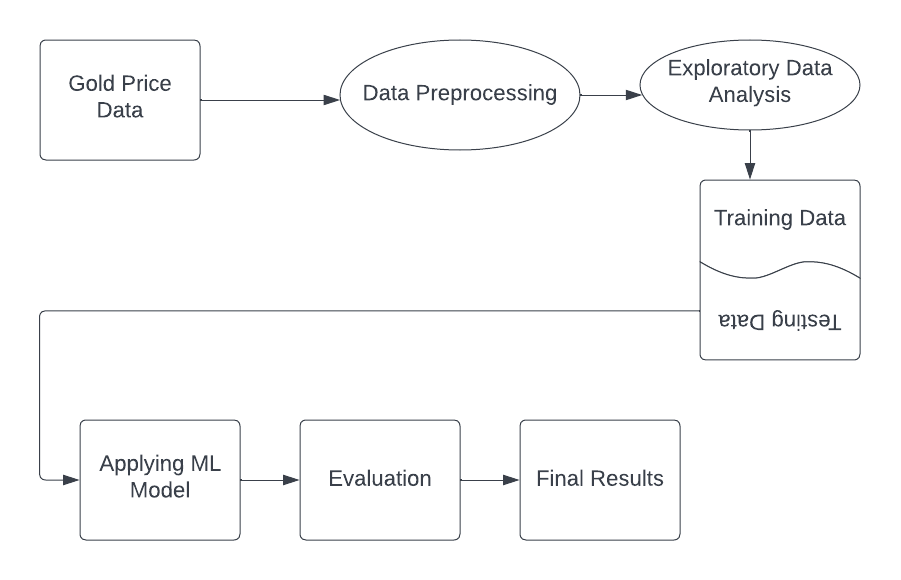
**Explanation:**

An artificial recurrent neural network is the LSTM network. The LSTM component is a three-layer model with end-to-end functionality that consists of the input layer, LSTM unit layer, and output layer. A standard RNN is analogized to have both "long-term memory" and "short-term memory" in the name of the LSTM. The activation patterns in the network change once per time-step, analogous to how physiological changes in synaptic strengths store short-term memories. The connection weights and biases in the network change once per episode of training, analogous to how physiological changes in synaptic strengths store long-term memories.

**Workflow diagram with explanation:**

As shown in the diagram below, we will first pull the gold price data from the dataset and preprocessing of the data will be done after that the Exploratory Data Analysis is done. Once the data is ready, it is divided into training and testing data to apply to the machine learning model which we have trained using LSTM algorithm and further evaluation is done to generate the results.

**Workflow diagram:**

****

**Dataset**: We are using the Open, Close, High, and Low price of Gold from the python module named as yfinance: <https://pypi.org/project/yfinance/>. We are fetching the data from yahoo by providing the start date and end date to the datareader using pandas\_datareader.

**Detail Design of Features:** We are working on the dataset by checking the missing values, checking for features, checking for categorical features in the data, checking for numerical features in the data, checking for Discrete Numerical features in the data, checking the Correlation between numerical features, Handling Missing Values, handling Categorical Values.

We have 0 discrete features and 6 continuous features as shown below:

Calendar

Description automatically generated

**Diagram:**

**Diagram

Description automatically generated**

**Analysis of data:**

**Pre-processing:**

Renaming the columns:

****

**Graphical user interface, text, email

Description automatically generated**

It is observed that the correlation between the close price and the other features is

close 1.000000

Adj Close 1.000000

low 0.999865

high 0.999850

open 0.999710

volume 0.256708

Name: close, dtype: float64

For the training dataset of size 4234, the Root Mean Square is 2.7746350225417964.

**Graph Model of Features:**

Chart, histogram

Description automatically generated

**Implementation:** We are reading the data from yahoo finance with the help of yfinance module. The code for reading the data using **pandas\_datareader** is written and the data is described. Then we are performing some Exploratory Data Analysis. Then we are performing feature scaling using the **MinMax** Scaler and transforming the dataset. Then we are dividing the dataset into training and testing dataset. After dividing the dataset, we are developing ML model using LSTM and apply the data to the model. Then we train the model with the training data. After this, we are applying the model to test data to predict the values. During this course, a few visualizations are generated for better understanding of the data.

**Pseudocode:**

Step 1: Start

Step 2: Fetch the data

Step 3: Preprocessing the data

Step 4: Exploratory Data Analysis

Step 5: Model Development

Step 6: Model Training

Step 7: Predictions using Model

Step 8: Visualizations of Results

Step 9: End

**Project Management:**

**Implementation Status Report:**

|  |  |  |  |
| --- | --- | --- | --- |
| S.No | Description of Task | Person | Contribution (%) |
| 1 | Understanding the dataset | Sai Manideep Reddy Pallerla | 50 |
| Veeranjaneyulu Muppalla | 50 |
| 2 | Fetching the dataset | Sai Manideep Reddy Pallerla | 50 |
| Veeranjaneyulu Muppalla | 50 |
| 3 | Exploratory Data Analysis | Yash Zauwar | 25 |
| Goutham Pallapothu | 25 |
| Sai Manideep Reddy Pallerla | 25 |
| Veeranjaneyulu Muppalla | 25 |
| 4 | Feature Engineering | Yash Zauwar | 25 |
| Goutham Pallapothu | 25 |
| Sai Manideep Reddy Pallerla | 25 |
| Veeranjaneyulu Muppalla | 25 |
| 5 | Machine Learning Model Training | Yash Zauwar | 50 |
| Goutham Pallapothu | 50 |
| 6 | Prediction using Model Developed | Yash Zauwar | 50 |
| Goutham | 50 |
| 7 | Documentation | Yash Zauwar | 25 |
| Goutham Pallapothu | 25 |
| Sai Manideep Reddy Pallerla | 25 |
| Veeranjaneyulu Muppalla | 25 |
| 8 | Visualizations of Predictions | Yash | 25 |
| Goutham | 25 |
| Sai | 25 |
| Veeranjaneyulu Muppalla | 25 |
| 9 | Final Documentation | Yash | 25 |
| Goutham | 25 |
| Sai | 25 |
| Veeranjaneyulu Muppalla | 25 |

**Results:** We can generate the output for predicted price of gold using the data available from the yfinance module. For the dataset we used, the loss is found to be 0.0010.

The below attached screenshot is the visualizations for the predicted price of gold with its actual price.

**Graphical user interface, chart

Description automatically generated**

**Github Link to the Project:** [**https://github.com/YashZauwar/Feature\_Engineering\_Fall\_2022/tree/main/Team\_Project**](https://github.com/YashZauwar/Feature_Engineering_Fall_2022/tree/main/Team_Project)

**References:**

1) chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.irjmets.com/uploadedfiles/paper//issue\_5\_may\_2022/22138/final/fin\_irjmets1651655991.pdf

2) https://www.semanticscholar.org/paper/Predicting-Future-Gold-Rates-using-Machine-Learning-Sami-Junejo/d1ddf3a4df034e89919df73b8c367c33cf5fb3b9

3) <https://www.researchgate.net/publication/336414732_Gold_Price_Prediction_using_Ensemble_based_Machine_Learning_Techniques>

4) <https://ieeexplore.ieee.org/document/7333120>

5) <https://www.kaggle.com/code/codestudiods/eda-and-model-building-gold-price-prediction>

6) https://towardsdatascience.com/how-to-model-gold-price-940e809ac665

----- End of Report ----