PROJECT REPORT
ON
GOLD PRICE PREDICTION
BY
APKESHA S KULKARNI

Course: MCA SY
University Seat No: -20694
Roll No: 18

Guide Name: Mr. Shubham Nagure

Kamala Education Society's
Pratibha Institute of Business Management
Chinchwad, Pune-411019
2021-2022

Index

CHAPTER 1

INTRODUCTION

1. Introduction to project:

The gold price prediction has been inevitable and the Billion market has always fluctuated. Investment in gold is essential to materialistic people who need to acquire large assets. It is also one of the liquid assets as exchange is quite easy. Stock market in terms of shares. crude oil and gold are the widely pronounced among the business people. Predicting the gold trend needs many models. Mathematical, Statistical, Random process theory, operational research gives people prior knowledge about the unforeseen data. Prediction is difficult because of the heavy change in the attributes and strategies followed by the business people according to the trend of the market. In foreign countries like US, UK, Australia the people believe only in real time news about the markets and accordingly incorporate new short-term strategies for investment.

1.1 Purpose(need) of System

- Many investors buy gold as a safe haven to protect themselves against a possible catastrophe, profit from these tremendous increases in the price of gold, diversify their portfolio and protect themselves against inflation.
- With recent fears of an economic recession looming in the distance, investors are looking to recession-proof their portfolios.
- Gold has historically been touted as an asset that booms in a recession because unlike fiat currencies such as the Dollar, Yen and the Pound, it has inherent value as a commodity currency.
- However, with the rapidly changing economic landscape, does this still hold true? Is gold a good commodity to buy now for later profits?

  These Problems Should Solved is the main purposed of the Project System.

CHAPTER 2:

PROPOSED SYSTEM

2.Proposed System

2.1 Objectives of proposed Project System

- The main objective of the Project on Gold Price Prediction System is to manage the details of Price of Gold, Gold Type, Gold Trends, Customer, Sale etc.
- **Improvement in control: -** The system was developed to overcome the current problem occurred while Analysis of data.
- It manages all the information about Gold Price, Trades, Buying, Selling. The project is totally built at administrative end and thus only the administrator is guaranteed the access.
- The purpose of the project is to build an application system to reduce the manual work for managing the Trades.

2.2 Project Overview. (Functionality of Models)

- ✓ **ARIMA model:** (**Autoregressive Integrated Moving Average)** is a class of statistical models for analyzing and forecasting time series data. ARIMA models take into account the past values to predict the future values.
- ✓ **Multiple Regression:** We try to create a simple regression model. It is a multiple regression model with input parameters as the moving average of the past 1 month and the past 2 months. We can clearly observe overfitting in this model. This overfitting can be attributed to the data being non-stationary.
- ✓ Check Stationarity: To check the stationarity of the data we plot the data along with the dates.Just by looking at the plot we can conclude that the data is non-stationary.We can also see from the histograms(in the code) that the data has seasonality and some component of trends.We also performed the dickey-fuller test to confirm the stationarity.We can see the ADF statistic is higher than any of the critical values, and the p value is much greater than 0.05, so we cannot reject the null hypothesis that the data is non-stationary.

- ✓ **Make Data Stationary To make the data stationary, we use simplest technique of taking a log transform:** We can observe that there is no change and the data is still non-stationary. We now try to perform differencing on this data.

  We perform differencing of order 2 and observe the following results. We can see that the ADF statistic is less than 1% critical value hence we can reject the null hypothesis and conclude with a confidence level of 99% that the data is stationary. We can now use this data for further modelling.

  Regression Model Again We use the old regression model again for this stationary data. We see the following results. We obtain R square value of 30% which is below par. And the root mean square error is also very high. Though the RMSE is an absolute statistic and it cannot be used to judge the goodness of fit, we will use this value for further comparison with other models

- ✓ **ACF and PACF Plots:** We now try to plot ACF i.e., Autocorrelation and PACF i.e., Partial Autocorrelation plots for this data to find the p, q, d values for creating an ARIMA model.

- ✓ **SARIMA model:** We model this data using a SARIMA model. A SARIMA model stands for a Seasonal ARIMA model. SARIMA Model is better over a simple ARIMA model when there is seasonal data. i.e., the timeseries data has repeating cycles. We observe that the model fits much better than any of the previous models. Below are the results and the diagnostics of the model. We see that the R square value is 73% which is acceptable and the RMS Error has reduced to 1715 from 5000, which is a good sign.

- ✓ **LSTM: Long Short-Term Memory** is a kind of recurrent neural network. In **RNN** output from the last step is fed as input in the current step. It tackled the problem of long-term dependencies of **RNN** in which the **RNN** cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases **RNN** does not give an efficient performance. **LSTM** can by default retain the information for a long period of time. It is used for processing, predicting, and classifying on the basis of time-series data.
  Time Series Prediction: The time series prediction analyzes and uses the features of data from specific past time periods to predict the characteristics of future data. The construction of the time series models is closely related to the order that the happening time of events follows, and it is more complicated than that of normal regression predictions. Time series problems can be partitioned into several types according to the following factors: whether the time series is repetitive, whether the change factors are independent, whether the dimensionality reduction or the dimensionality ascending is used, whether there are complex historical dependencies, and so on

2.3 Detail Description of Technology/Tool Used

A]. Jupyter Notebook :

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more. Jupyter has support for over 40 different programming languages and Python is one of them. Python is a requirement (Python 3.3 or greater, or Python 2.7) for installing the Jupyter Notebook itself. Install Python and Jupyter using the Anaconda Distribution, which includes Python, the Jupyter Notebook, and other commonly used packages for scientific computing and data science. You can download Anaconda's latest Python3 version from here. Now, install the downloaded version of Anaconda.

- **Installing Jupyter Notebook using pip:**

  python3 -m pip install --upgrade pip

  python3 -m pip install jupyter.

B] Colaboratory:

Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing access free of charge to computing resources including GPUs.

If you have used **Jupyter** notebook previously, you would quickly learn to use Google Colab. To be precise, Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries which can be easily loaded in your notebook.

2.5 Scope of the Project

- ✓ The scope of this project is at the Gold Price Prediction System of the Buyer and at the Seller desk.
- ✓ To provide a user-friendly environment where User/Company can be serviced better.
- ✓ It is only concerned with active use of resources of the gold trade and does not have long term storage of information. It only deals with past data, gold sold years and trends of shares, and does not handle other aspects, such as online applications, design etc.
- ✓ To utilize resources in an efficient manner by increasing their productivity through automation.
- ✓ The system generates types of information that can be used for various purposes.
- ✓ To satisfy the **Investors** requirement.
- ✓ The scope of this project is focused on Gold Rs (Indian) price forecasting, trends and 1-year predictions.
- ✓ The goal of this project is to understand and apply time-series models like ARIMA, SARIMA and LSTM in forecasting the price of gold.

CHAPTER 3:

OPERATING ENVIRONMENT

3.Hardware & Software Requirements

3.1 At Server Side:

Hardware Configuration

Name	Details
Processor	**Intel Pentium IV or more**
RAM	**At least 128 MB RAM is required (formally 4 GB RAM)**
Hard Drive	**1 TB**

Software Environment

Name	Details
Operating System	**Windows 10,11**
Language Used	**Python Programing**
Browser	**Localhost (Chrome)**
Tool	**Jupyter Notebook, Colab**

CHAPTER 4:

IMPLEMENTATION

4.1 MODEL 1

A] MODEL 1 (LINEAR REGRESSION, ARIMA, SARIMAX)

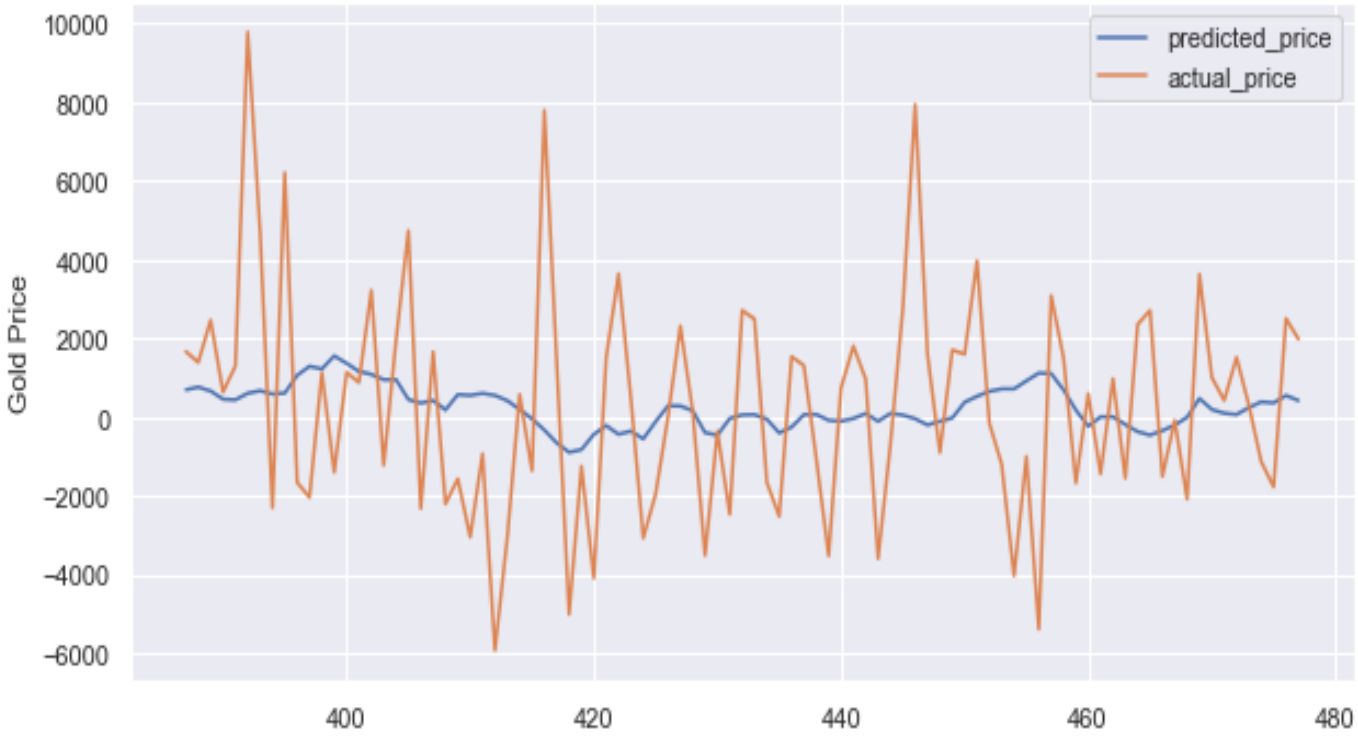
1. Linear Regression

Description: In First Model We Performing Linear Regression. To Predict the actual & predicted prices.

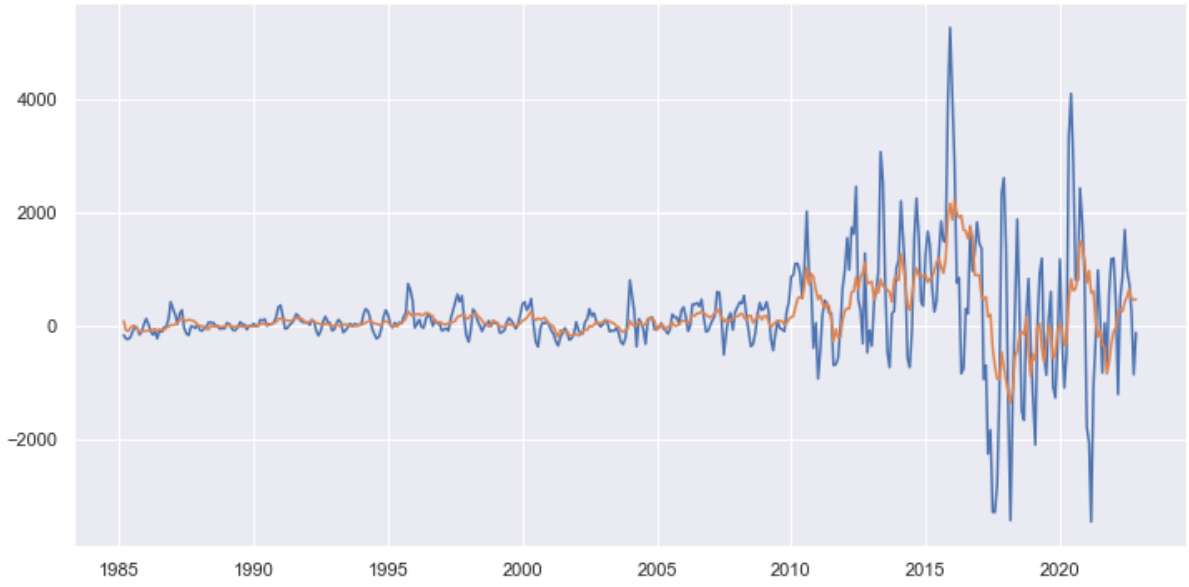
LINEAR REGRESSION

2. Again regression: 1ST ORDER DIFFRENCING

Steps: Finding moving average of past 3 days and 9 days >> Finding dependent variable >> Split into train and test >> Performing linear regression >> Predict prices >> Calculate R square and rmse to check goodness of fit.



- **Till Year 2020: -**

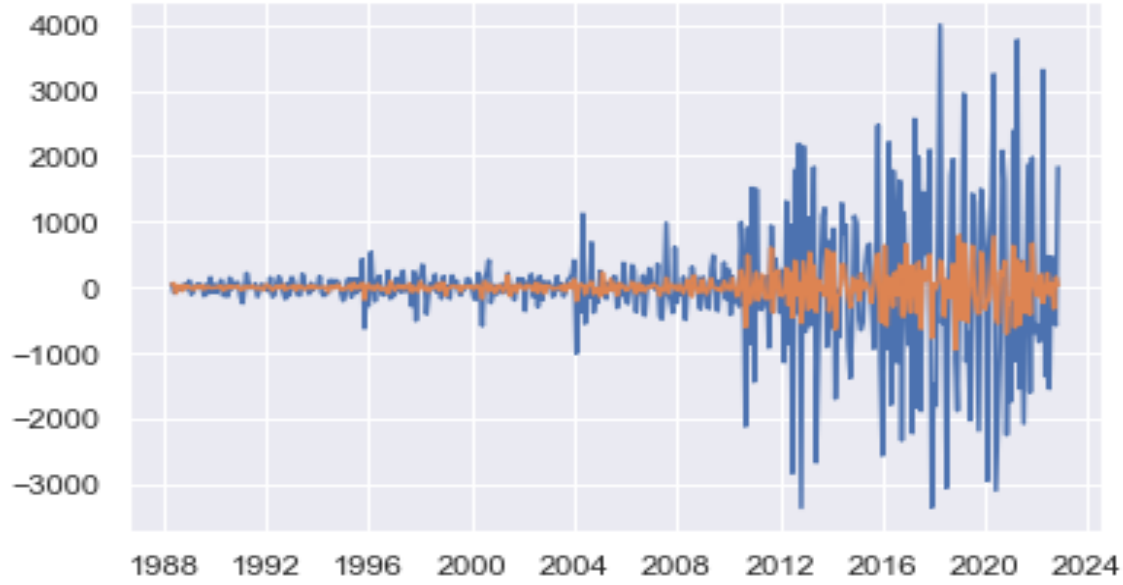


- **Again Regression 2ND ORDER DIFFRENCING**

Steps: Finding moving average of past 3 days and 9 days >> Finding dependent variable >> Split into train and test >> Performing linear regression >> Predict prices >> Calculate R square and rmse to check goodness of fit.

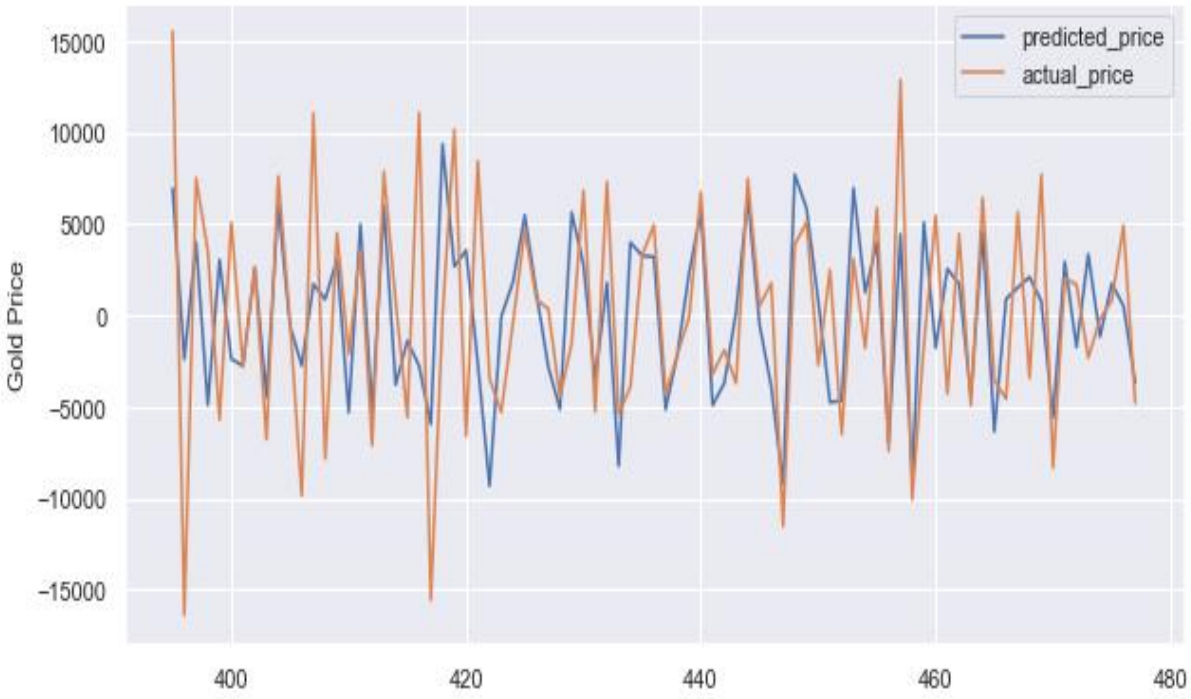
Till Year 2024 :-

X_test




```
Gold Price = -1.94 * 2 Month Moving Average -3.73 * 1 Month Moving Aver
age + -10.91.
```

Y_test:


predicted_price
actual_price
Gold Price
15000
10000
5000
0
−5000
−10000
−15000
400
420
440
460
480

3. SARIMAX Result

1] Covariance matrix calculated using the outer product of gradients (complex-step).
2] Covariance matrix is singular or near-singular, with condition number 4.66e+29. Standard errors may be unstable.

Dep. Variable:	y	**No. Observations:**	439
Model:	SARIMAX(2, 1, 2) x(2, 1, 2, 12)	**Log Likelihood**	-3633.014
Date:	Sun, 05 Mar 2023	**AIC**	7284.028
Time:	10:03:56	**BIC**	7319.929
Sample:	0	**HQIC**	7298.247
	- 439		
Covariance Type:	opg		

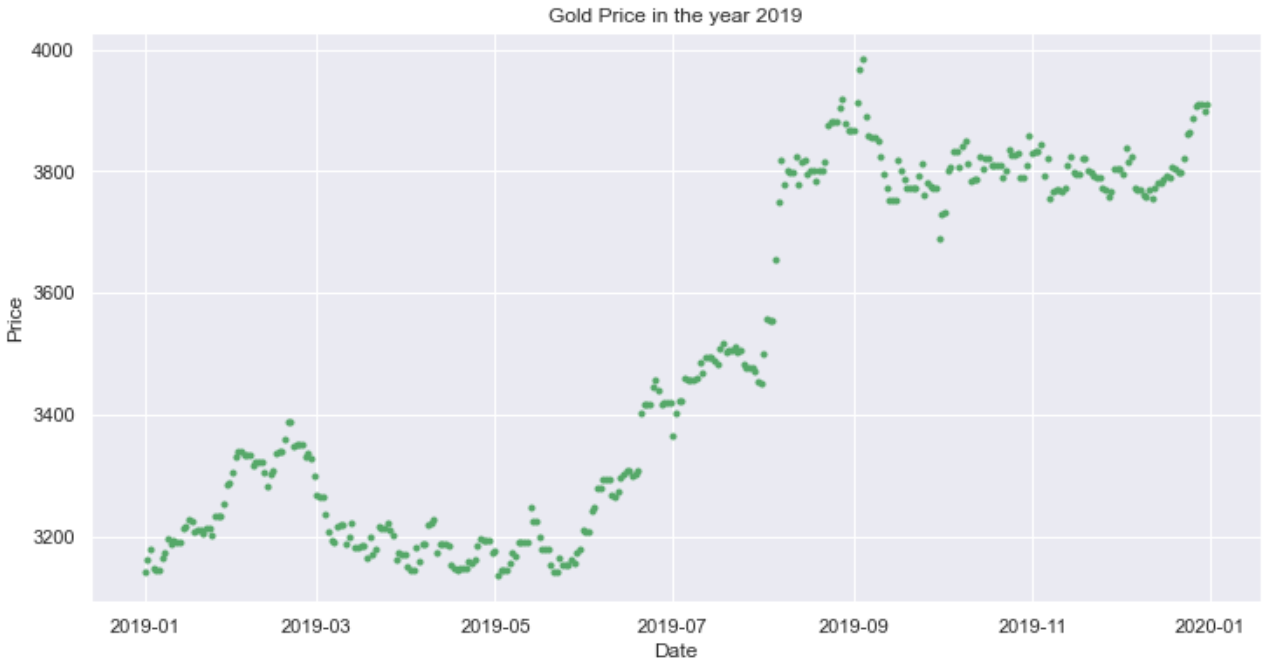
Reached Our Model
Getting Satisfied Result:
RMSE: 1701.3108826842893
R2 SCORE: 0.7379329718997134.

4. Finding Trend Between Years.

1] Year 2017: Which Predict the Value of Next Year i.e., Year 2018 (Month Jan)

2] Year 2018: Which Predict the Value of Next Year i.e., Year 2019(Month Jan).

3] Year 2019: Which Predict the Value Of Next Year i.e. Year 2020(Month Jan).

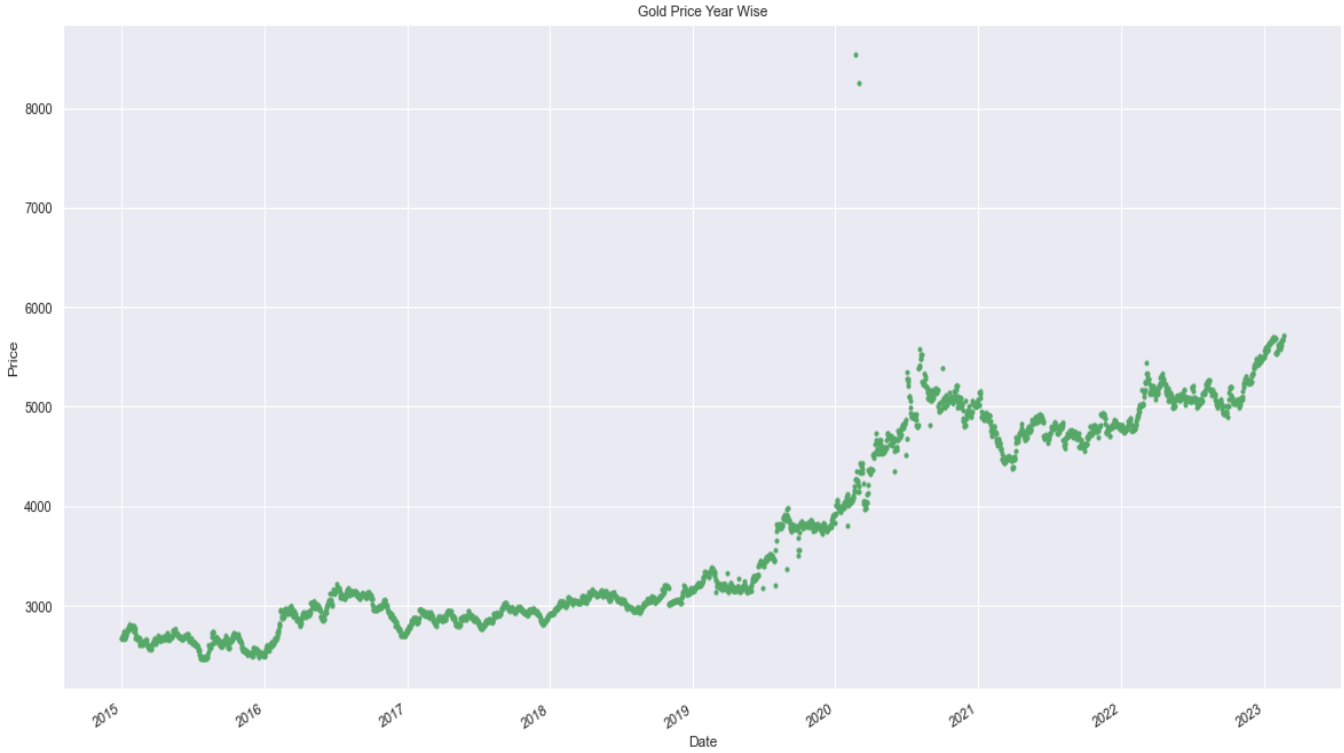


4] Year 2020: Which Predict the Value Of Next Year i.e. Year 2021(Month Jan)

5] Year 2021: Which Predict the Value Of Next Year i.e. Year 2022(Month Jan)

6] Year 2022: Which Predict the Value of Next Year i.e., Year 2023(Month Jan)

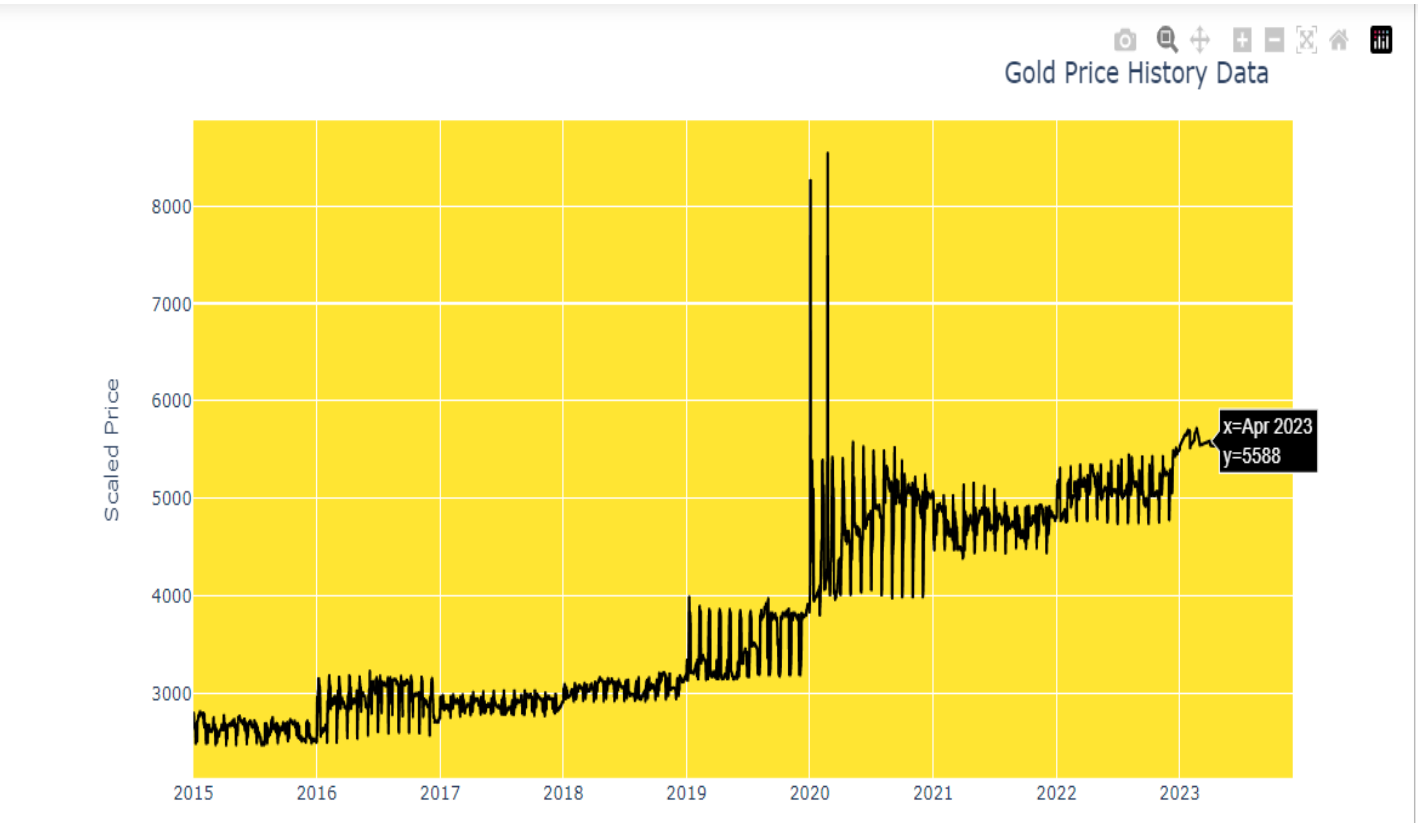
7. Trend All Years (2015 To 2023)

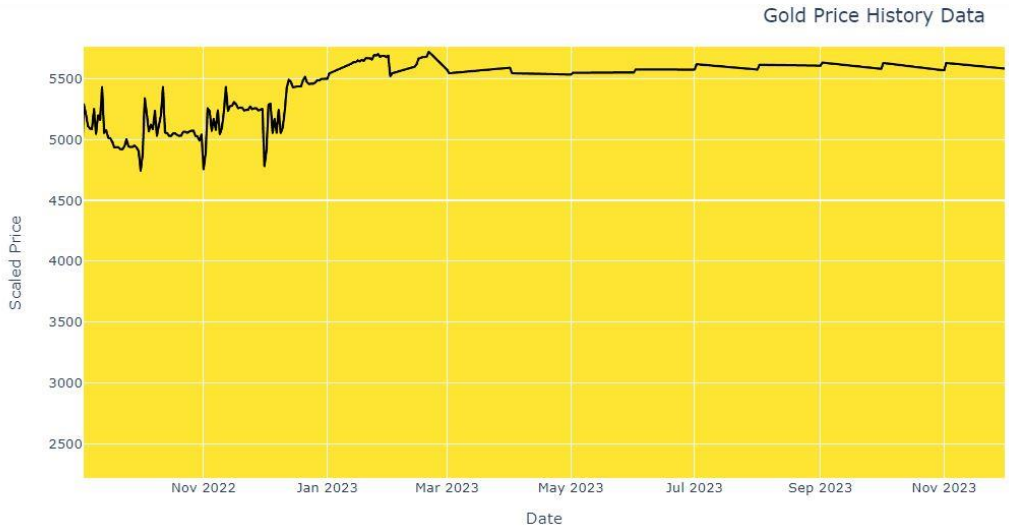


B] MODEL LSTM Graphs

1] BEFORE IMPLAMENTING LSTM MODEL

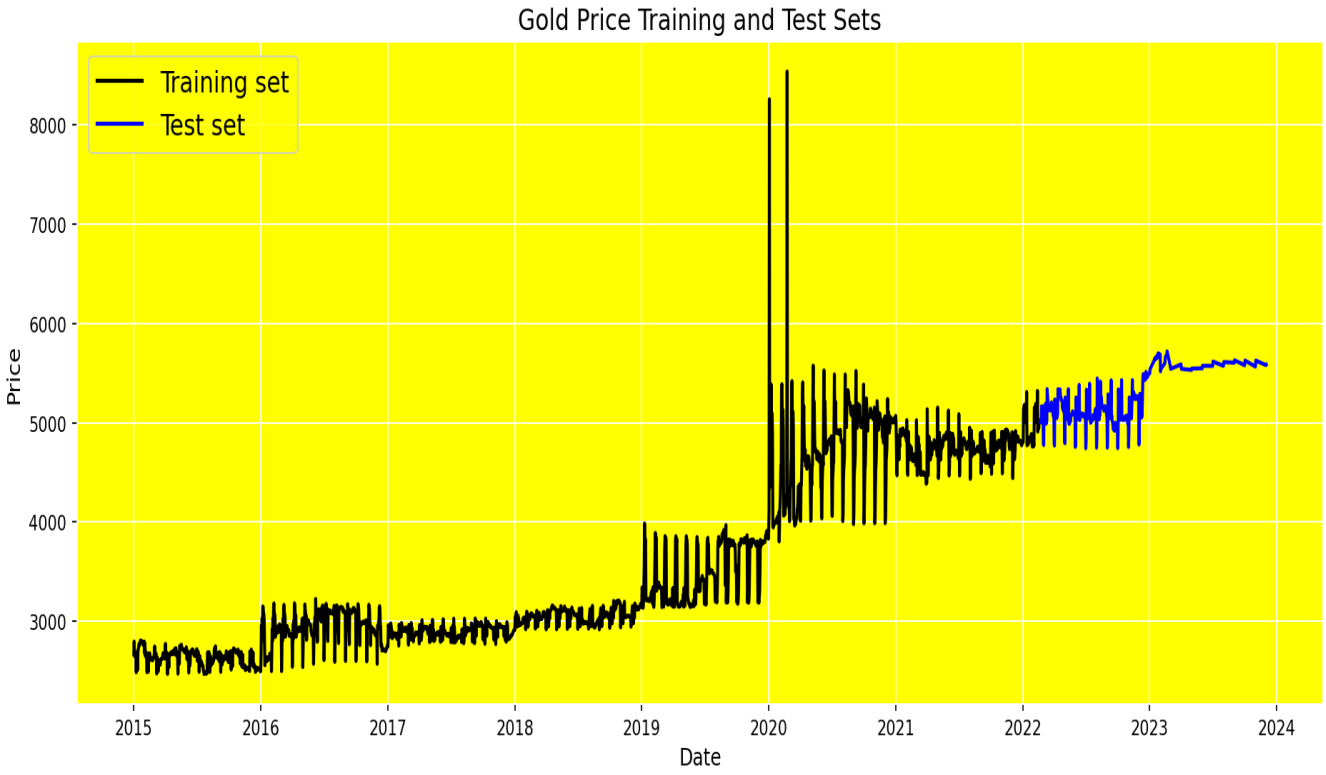
Interactive Chart





Month Wise Prediction Through Interactive Chart:

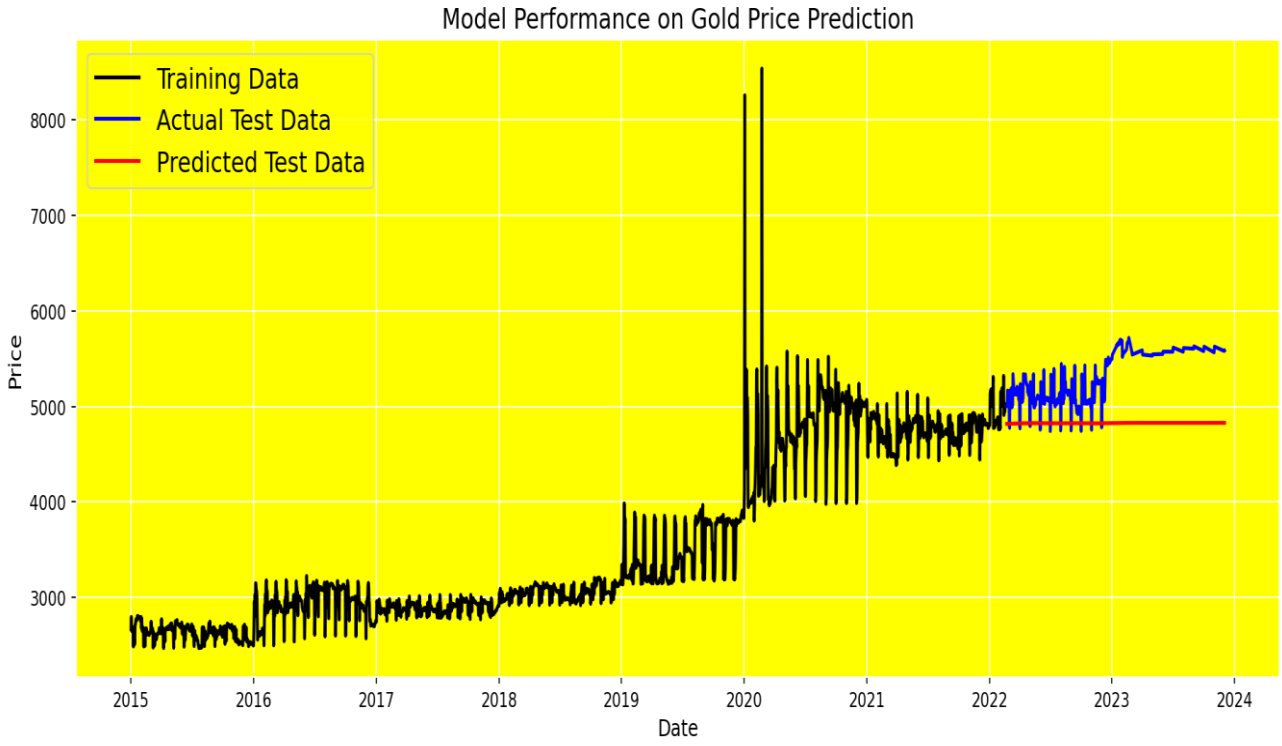
Gold Price Training and Test Sets Plot:



Train and Test Dataset >> Display Graph

3] Investigating the closeness of the prices predicted by the model to the actual prices:

- **Finally Reached the LSTM NETWORK MODEL**



CHAPTER 5:

DATASET

Dataset: We now use the other (Indian market dataset) for trying and finding any interesting trends in the price fluctuation. Interestingly for various years the prices of gold have been maximum during the wedding season corresponding to that year. We can also see a cyclic trend in the price, there is 6-8 years of bullish growth followed by 6-8 years of bearish market. Although other than that there are no obvious trends in the data. The maximum price has always been March-April or September-October which falls during or just before the wedding season.

- **Website Used for Data collecting: https://www.goldpriceindia.com/**
- **Dataset Years : 2017, 2018 , 2019 , 2020 , 2021 , 2022 .**

CHAPTER 6:

FUTURE SCOPE

6. Future scope of project

The future enhancements that may be brought about in the project may pertain to the limitations of the project.

- Now I can use the other (Indian market dataset) for trying and finding any interesting trends in the price fluctuation.
- Interestingly for various years the prices of gold have been maximum during the wedding season corresponding to that year.
- We can also see a cyclic trend in the price, there is 6-8 years of bullish growth followed by 6-8 years of bearish market.
- Although other than that there are no obvious trends in the data.
- The maximum price has always been March-April or September-October which falls during or just before the wedding season.

CHAPTER 7 , 8:

CONCLUSION & BIBLIOGRAPHY

Conclusion:

Currently I have a model which can predict the gold price with almost Good accuracy and have found an interesting correlation between the market price of gold and the wedding season in India. **Loss: 0.004 & Accuracy: 85%-90%**

I can safely conclude that the price of gold in the world market and the regional Indian market are very volatile and depend a lot of external factors which cannot be modelled so easily.

For future work, we can use and build upon our existing model to build a recommendation system suggesting the users the right time to buy and sell gold for people who take interest in investing in gold.

Bibliography

We take references from these following websites to build our project and ideas only for study material purpose.

1] For Dataset: https://www.goldpriceindia.com/

2] https://www.google.com

3] https://github.com/

4] https://stackoverflow.com/