GOLD PRICE PREDICTION

A PROJECT REPORT

submitted By

ANUSHKA SANJEEV

TVE21MCA2019

to

the APJ Abdul Kalam Technological University
in partial fullfilment of the requirements for the award of the degree

of

Master of Computer Applications

Department of Computer Applications

College of Engineering

Trivandrum-695016

NOVEMBER 2022

Declaration

I undersigned hereby declare that the project report titled **"Gold Price Prediction"** submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Applications of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Prof. Shine S, Asst.Professor. This submission represents my ideas in my words and where ideas or words of others have been included. I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity as directed in the ethics policy of the college and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and/or University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title.

Place : Trivandrum **Anushka Sanjeev**

Date : 14/11/2022

DEPARTMENT OF COMPUTER APPLICATIONS

COLLEGE OF ENGINEERING
TRIVANDRUM

CERTIFICATE

This is to certify that the report entitled **Automated essay scoring using machine learning** submitted by **Anushka Sanjeev** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of the project work carried out by her under my guidance and supervision. This report in any form has not been submitted to any University or Institute for any purpose.

Internal Supervisor External Supervisor

Head of the Dept

Acknowledgement

First and for most I thank **GOD** almighty and to my parents for the success of this project. I owe a sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my project.

I am extremely thankful to **Dr. Suresh Babu V**, Principal, College of Engineering Trivandrum for providing me with the best facilities and atmosphere which was necessary for the successful completion of this project.

I am extremely grateful to **Prof. Deepa S S**, HOD, Dept of Computer Applications, for providing me with best facilities and atmosphere for the creative work guidance and encouragement.

I express our sincere thanks to **Prof. Shine S**, Asst. Professor, Department of Computer Applications, College of Engineering Trivandrum for her valuable guidance, support and advice that aided in the successful completion of my project.

I profusely thank other Asst. Professors in the department and all other staffs of CET, for their guidance and inspirations throughout my course of study.

I owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this project. No words can express my humble gratitude to my beloved parents and relatives who have been guiding me in all walks of my journey.

Anushka Sanjeev

Abstract

Gold has always occupied a predominant place in country's economies, and among populations. Owing to its characteristics, it is used as a hedging tool or can act as a safe haven in turmoil conditions. The aim of this project is to explore the relationship of gold price with various explanatory variables that tend to be considered as indicators of financial and geopolitical crises.

The project also investigates the possibility of predicting gold price based on these variables. Hence, it might prove to be helpful in investment of gold if one knows the extent of fluctuation of prices.

For this study, I shall be performing the analysis on dataset and using Linear Regression and LSTM for the purpose of prediction. The linear regression technique is used for the training process. I aim to train classifiers in the training set, make it familiar with the downloaded dataset, and then calculate the performance of the dataset by performing proper comparison. The system is implemented using Python with Machine Learning and Deep Learning.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Savings and investments form an integral part of everyone's life. Investments refer to the employment of present funds with an objective of earning a favourable return on it in future. In an economic sense, an investment can be considered as the purchase of assets that are not consumed today but are used in the future to create wealth. The Indian economy being one of the fastest growing in the world has resulted in higher disposable income level and a plethora of investment avenues. There are a number of investment avenues available for investors, which includes stocks, deposits, commodities and real estate. Each of them differs in terms of risk and return characteristics.

Gold is an asset which is being considered as an attractive investment avenue by many investors due to its increasing value and the area of usage. Investors preference for gold as a protective asset increases due to their negative expectations concerning the situation in the developed foreign exchange markets and the capital markets.

Gold is additionally thought-about to be the asset investors rely on, once the developed world capital markets aren't capable to produce desirable gain. So it may be said that investors see gold as a tool to hedge against the fluctuations in different markets. Gold is a precious metal, thus like several different goods, gold's value ought to rely upon supply and demand. But, since gold is storable and the supply is accumulated over centuries, this year's production has very little influence on its costs. Gold is used as a commodity and as a financial asset. Gold behaves less sort of a commodity than long-lived assets like stocks or bonds.

The raising worth of gold as well as the volatilities and fall in costs of different markets like capital markets and property markets has attracted a lot of and a lot of investors towards gold as

a pretty investment. But, currently worth of gold is additionally witnessing high volatility and investments in gold ar turning to be riskier. there's a worry on whether or not these high costs ar property and once the costs would reverse. Understanding the extent of fluctuation of gold costs are useful not only to financial policymakers however conjointly to investors, fund managers and portfolio managers to require higher investment selections within the market.

Chapter 2

Problem Defnition and Motivation

The main goal of this project is to forecast the rise and fall in the daily gold rates, that can help investors to decide when to buy or sell the gold. Inventory forecasting plays a crucial role in the financial success of the business. The price of gold is calculated by looking at the dataset that contains the previous years' gold price, of varying frequencies of time intervals. The proposed research provides the highest likelihood of achieving high training rate prediction precision for the considered gold price. Generally speaking, this work is performed to suggest suitable predictor models to effectively show the deemed gold in the different scenario with the datasets deemed from their respective databases of previous years.

The major motivation behind choosing the project was the drawbacks of the existing system which are as follows:

- Datasets only include a limited timeframe. They are also static.
- There is only an overlap found between present and predicted values. There is no forecast for days in the future.
- Lack of user-friendly interfaces to use prediction tools

The major objective the project is to use a dynamic dataset that gets updated to the most recent dates to deliver an accurate prediction of the future gold prices with help of machine learning and deep learning models, and presenting it in a user-friendly and efficient manner.

Chapter 3

Literature Review

The gold values are often directly linked to other resources. Future gold price forecast is the investors' alert mechanism because of unpredictable market risk. Therefore, precise forecasting of gold prices is needed to predict the market patterns. Several computational intelligent techniques for gold forecasting applications have been noted over the past decade, and the review of different models applied for these applications. As a part of my literature review I went through various papers and presentations on this topic. The quick summary of my findings are specified in this chapter.

3.1 Relationship with other variables

The relationship between other macroeconomic variable and gold prices has also been studied by many researchers. The relationship between gold price and prices of other commodities especially crude oil has also been extensively studied. But the results from these studies are found to be contradicting. Some of the studies on the factors influencing gold price and various techniques used for studying these relationships are discussed in the following sections. In 2003, C. Lawrence has found that there is no significant correlations between returns on gold and changes in certain macroeconomic variables such as inflation and GDP. He has also found that that gold returns are less correlated with returns on equity and bond indices than returns of other commodities. But, Sjaastad and Scacciavillani(1986) reported that gold is a store of value against inflation and Baker and Van Tassel(1985) had reportedly found that the price of gold depends on the future inflation rate.

3.2 Data and Methodology

Based on the review of literature five major factors that is considered to have influence on the gold price were identified. The factor that is considered for this study are Historical Gold Prices Values.

Machine learning algorithms were used to train and model the collected data. From the data collected, eighty percentage of the data was used for training and remaining twenty percentages for testing the model. The machine learning algorithms used in this study are linear regression and LSTM Model.

The statistical process for estimating the relationship between different variables is called regression analysis. Regression analysis is used to understand how the value of the dependent variable changes when one of the independent variables changes, while other variables are fixed.

The Long Short-Term Memory network is a RNN that is trained using Backpropagation. It takes care of the disappearing gradient problem encountered earlier. LSTM networks have their own memory and so they prove to be efficient in creating large RNNs and handle time specific scheduling problems. The memory blocks in LSTM network are connected through recurrent layers rather than having neurons.

3.3 The social factor

Not only economic and political factors influence the movement of the gold price, but also social celebrations at which gold gifts are offered. In fact, Schmidbauer and Rösch (2018) examine the effects of festivals on the volatility of spot gold prices on a daily basis. The tests were based on 13 different holiday events around the world, such as Ramadan Eid, Eid Al Adha, Christmas, New Years Eve, and Chinese New Year.You build a model that combines regression with the GARCH model. The results of Schmidbauer and Rösch show that various celebrations correlate with fluctuations in the price of the yellow metal.

From the above data, it would suffice to say the price of gold is affected by a myriad of factors. Keeping this in mind, I would be employing means of both - machine learning and deep learning models - for the purpose of prediction.

Chapter 4

Requirement Analysis

4.1 Purpose

The forecasting of gold prices in the future depend on several factors. In order to achieve a working project in the python environment some major requirements are needed. Software Development require some important steps to be taken. In developing our web application we also follow similar steps.

4.2 Overall Description

Price forecasting is calculating the price of a commodity / product / service by examining different factors such as the existence, pricing, seasonal trends, the costs of other products (i.e. fuel), several manufacturer offers etc. Gold spot rates are decided twice a day based on supply and demand in gold market. Fractional change in gold price may result in huge profit or loss for these investors as well as government banks. Forecasting rise and decline in the daily gold rates, can help investors to decide when to buy (or sell) the commodity. Price forecasting can be conceived as a regression function. Linear Regression model is used on the features, which in turn is compared with LSTM model for accuracy.

4.2.1 Product Functions

- Data collection and preparation
- Training and testing of model

- Develope a user interface
- connect the UI with the model.

4.2.2 Hardware Requirements

- Processor : Intel Core i3
- Storage : 512 GB Hard Disk space
- Memory : 4 GB RAM

4.2.3 Software Requirements

- Operating System : Linux/Windows
- Platform : Python
- Libraries used : keras, pandas, tensorflow, matplotlib, numpy, sklearn, plotly

4.3 Functional Requirements

The functional requirements includes all the activities or processes that should be achieved by the proposed system. It includes

- **TensorFlow:** TensorFlow is an open-sourced end-to-end platform, a library for multiple machine learning tasks, while Keras is a high-level neural network library that runs on top of TensorFlow. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple backend neural network computation.
- **sklearn:** sk learn (formerly sci-kit learn and sometimes called sk learn) is a machine learning library can be used in python programming language. By using this library, we can implement various regression, classification and clustering algorithms such as random forest, support vector machine, k-means and DBSCAN. And the sk learn library is built in a way that it can work with various scientific and numeric libraries of python such as scipy and numpy.

- **matplotlib:** It's used for the visualisation of data in python programming language. It's implemented to work with the wider scipy stack and it's built on numpy arrays. It's a multi platform data visualization technique. It was developed in 2002 by John Hunter. Visualization is the most efficient way to understand the data. Using this library, we can represent our data in various plots such as line, bar, histogram, scatter etc.

4.4 Non Functional Requirements

4.4.1 Performance Requirements

- Accuracy : Accuracy in functioning and the nature of user-friendly should be maintained by the system.
- Speed : The system must be capable of offering speed.
- Low cost: This system is very cheap to implement and is also user-friendly.
- Less Time consuming: It uses very less time comparing to the existing system .
- User Friendly: This proposed system is highly user friendly they enables to create a good environment.

4.4.2 Quality Requirements

- Scalability : The software will meet all of the functional requirements.
- Maintainability : The system should be maintainable. It should keep backups to atone for system failures, and should log its activities periodically.
- Reliability : The acceptable threshold for down-time should be large as possible. i.e. mean time between failures should be large as possible. And if the system is broken, time required to get the system backup again should be minimum.
- Availability: This system is easily available as the core equipments in building the software is easily obtained.
- High- Functionality: This system is highly functional in all environment since, They are highly adaptable.

Chapter 5

Design And Implementation

The proposed system is used to forecast gold prices in the near future using a pre-trained model. The model is trained using Linear Regression and LSTM and then compared in terms of accuracy and fitting of data, to find the most optimum method of prediction.

5.1 Overall Design

The proposed system follows client server architecture. That is the automated essay scoring system has a client part and a server part as well. The client part is used by the user to input the essay which is to be evaluated. The input is passed to server and the evaluated result is given back to the client. The server side is developed in Python and the client side is built using HTML and Python.

5.1.1 System Design

The system is web based. The input is taken from the user through a web page and the input is passed to the python program running in the server side. The server program perform tasks such as pre-processing and feature extraction on the input data. The results of these processes are used to evaluate the input using the pre-trained model.

The model is created using the data obtained from Yahoo finance. yfinance is a popular open source library developed by Ran Aroussi as a means to access the financial data available on Yahoo Finance.

Yahoo Finance offers an excellent range of market data on stocks, bonds, currencies and

cryptocurrencies. It also offers market news, reports and analysis and additionally options and fundamentals data.

5.1.2 Methodology

There are two parts in this project. The first part is the creation of the model and the second one is the creation of user program which will work with the pre-trained model.

The main process of the gold price prediction is the creation of the trained model. The major steps in the model creation Feature extraction, training, testing and model evaluation. The major steps in the model creation are mentioned below.

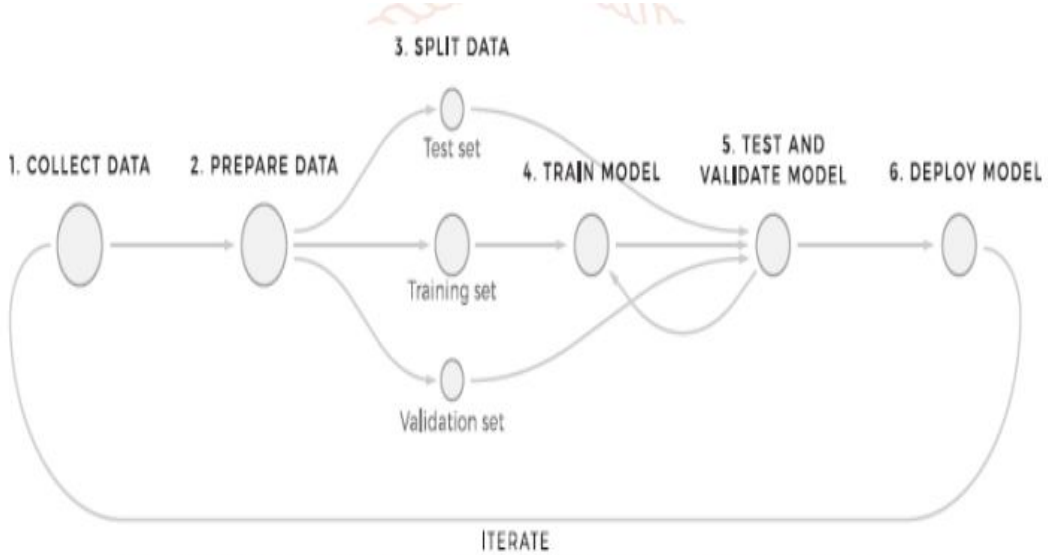

Figure 1. Methodology used for price prediction using machine learning.

Figure 5.1: Architecture of model Creation

- **Training:**Linear Regression is the technique used for the learning of the system.It's one of the widely used supervised learning techniques It uses the labelled dataset to generate the model.

  The second model used is Long Short-Term Memory. The LSTM component enables to harness the sequential order of daily gold price. A portion of our dataset is used for training. The remaining portion is for testing the model.

- **Testing:** In testing phase we test the generated model with the remaining portion of the dataset. The data set is fed into the generated model and their results are recorded for the next stage which is the model evaluation and error analysis.

- **Model evaluation and error analysis:** The results of the testing data along with their original values are used for the error calculation. Various statistical measures can be adopted for calculating the efficiency of the model. The various measures are accuracy, precision, recall, kappa score etc. If the results of these quantitative analyses are acceptable, then we move forward with the generated model. If the results are poor, the model is to be regenerated with more features so that the best result is obtained.

The second part of the project is to build the user interface. The user interface is build using HTML and Python. This is the part of project which deals with the user. The input text is fed into the server through this. And the results returned are also displayed in the user program. The interface is built in a way such that it is easy and understandable for the person who uses it. For that we uses responsible HTML designs which uses CSS and JavaScript also to provide the better user experience. Python Flask is also used for the development of user interface.

5.2 Data Flow Diagram

DFD is one of the graphical representation techniques used in a project to show the flow of the data through a project. DFD helps us to obtain an idea about the input, output, and process involved. The things absent in a DFD are control flow, decision rules, and loops. It can be described as a representation of functions, processes that capture, manipulate, store, and distribute data between a system and the surrounding and between the components of the system. The visual representation helps for good communication.

It shows the journey of the data and how will it be stored in the last. It does not provide details about the process timings or if the process shall have a parallel or sequential operation. It is very different from a traditional flow chart or a UML that shows the control flow or the data flow.

In level 0 the basic data flow of the application is showcased. It does not show the flow of data much deeper. It will be evaluated in the higher levels of Data Flow Diagram. The Data Flow Diagram of Automated essay scoring system is shown below.

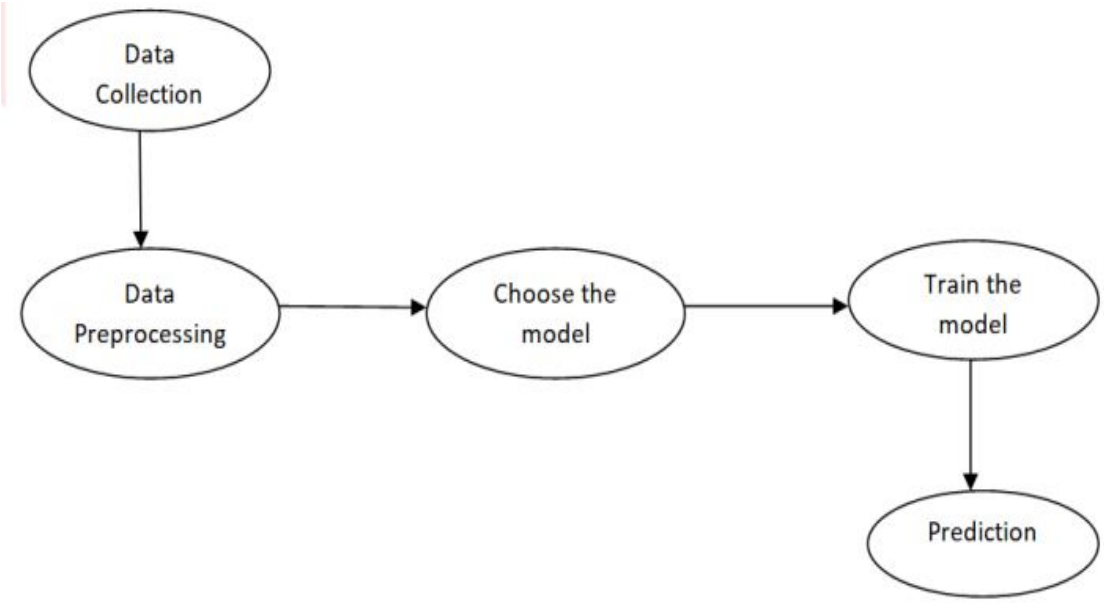

Figure 5.2: Level 0 DFD

The diagram shows Level 0 Data flow diagram of the Gold Price Prediction System. As the diagram indicates there is a user part and a backend. The input of the project is the date selected by the user and which is given to the Gold Price Prediction system. Then the date is passed to the backend part for the evaluation of the prediction. The evaluation of the parsed date and prediction computation occurs on the admin side. The value of the evaluation is passed back to the user through the application. This is how the data flows through the application. Since there is no database in the application, the data is not stored anywhere. The data is lost after the evaluation.

5.3 Screenshots of user interface - **TO BE REPLACED**

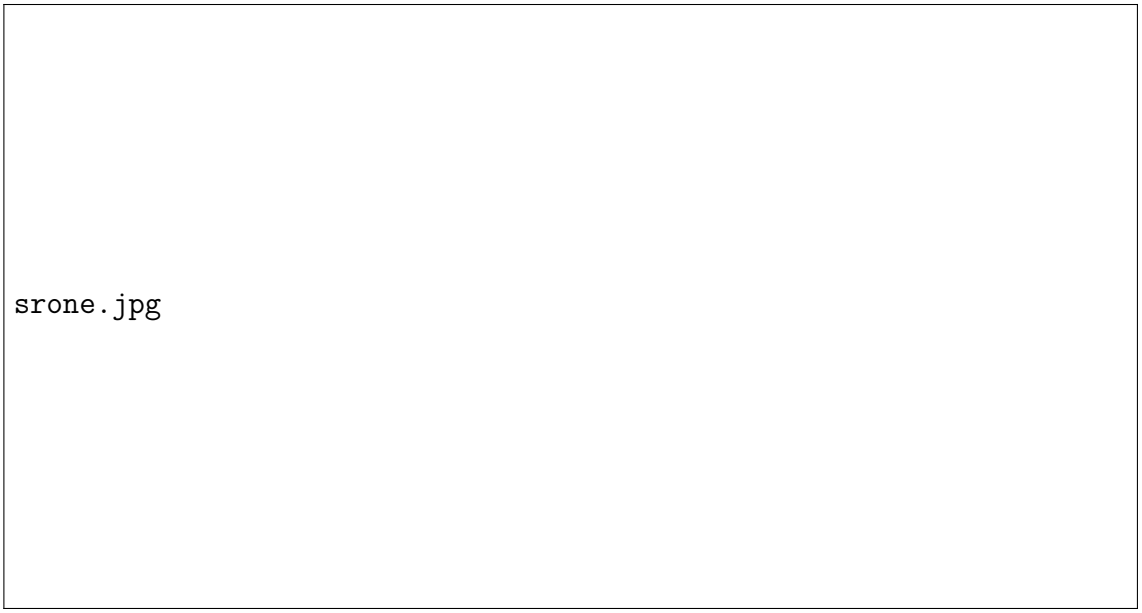

Figure 5.3: input

Figure 5.4: output

Chapter 6

Coding

Algorithm 1 Algorithm for Creating the model:

1: Split the data set into training data set and testing dataset. 80% of the dataset is used for training and the remaining 20% is used for testing to obtain better result.

2: The training dataset is used for the preprocessing stage.

3: The feature with high dependency on the result of prediction is used to create the model.

4: The model is created using Linear Regression and LSTM, with the selected features. The linear regression model evaluates how much the dependent values depend upon the independent values. The LSTM model uses a series of 'gates' which control how the information in a sequence of data comes into, is stored in and leaves the network. There are three gates in a typical LSTM; forget gate, input gate and output gate. These gates can be thought of as filters and are each their own neural network.

5: The testing dataset is fed into the created models and their results are noted down.

6: The result of testing dataset evaluated using the created models is then compared with the actual values of the testing dataset to evaluate the efficiency of the models. Various statistical measures such as Accuracy, Precision, Recall, Kappa score etc. can be used to evaluate the model.

7: Further tuning is performed upon the created model to improve the efficiency of the model.

Algorithm 2 Algorithm for web Application and prediction:

1: Read the input date from the user through the user interface.

2: On button click the value in the web page is passed to the server program for the parsing of the input.

3: From the server program, access the input date and perform the pre-processing tasks on it.

4: Using the pretrained model, we evaluate the date against the price forecast on that day.

5: The price is evaluated by the results of the model and the prediction is passed to the web page.

6: The prediction is displayed in the web application.

Chapter 7

Testing and Implementation

7.1 Testing and various types of testing used.

Once a software is developed, the major activity is to test whether the actual results match with the experimental results. This process is called testing. It's used to make sure that the developed system is defect free. The main aim of testing is to find the errors and missing operations by executing the program. It also ensure that all of the objectives of the project are met by the developer. The objective of testing is not only to evaluate the bugs in the created software but also finding the ways to improve the efficiency, usability and accuracy of it. It aims to measure the functionality, specification and performance of a software program. Tests are performed on the created software and their results are compared with the expected documentation. When there are too much errors occurred, debugging is performed. And the result after debugging is tested again to make sure that the software is error free. The major testing processes applied to this project are unit testing, integration testing and system testing. In unit testing, our aim is to test all individual units of the software. It makes sure that all of the units of the software works as it intended. In integration testing, the combined individual units are tested to check whether it met the intended function or not. It helps us to find out the faults that may arise when the units are combined. In system testing the entire software is tested to make sure that it satisfies all of the requirements. The tables shown below describes the testing process occurred during the development of this project "Gold Price Prediction". This defines the various steps took to create the project error free.

7.1.1 Unit Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	create the user interface	To load the web page with required fields	Same as expected	Pass
2	pre-processing	clean the dataset for feature extraction	same as expected	Pass
3	extract features from dataset	extract various features from dataset	extracted successfully	Pass
4	training and testing of model	create the model and store it in a pickle file	pickle file generated	Pass
5	prediction	predict the result accurately	same as expected.	Pass
6	python server program	set up a python flask server to run the program	Same as expected	Pass

Table 7.1: Unit test cases and results

7.1.2 Integration Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	load the user interface from python	the user interface is loaded when we run the flask program	Same as expected	Pass
2	pass input date from web page to server	To pass the date entered by the user to the python program to and receive it there.	Same as expected	Pass
3	mark evaluation	load the previously generated file to the server and predict the result with it and extracted features.	Same as expected	Pass
4	display results	pass the result to web page and display it there	Same as expected	Pass

Table 7.2: Integration cases and result

7.1.3 System Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	to run python server	Server program executed successfully, hence the entire program worked without any crash	Same as expected	Pass
2	date evaluation	allow user to input date and output generated according to the input date.	Same as expected	Pass

Table 7.3: System test cases and results

Chapter 8

Results and Discussion

The main aim of the project was to predict gold price for the near future with a suitable machine learning or deep learning model. And it is observed that the system performs all the functionalities as expected. By using the appropriate model the computer can compute the forecast price against the input date.

8.1 Advantages and Limitations

The proposed system is a machine learning model to evaluate the input date and predict the price on that day. The proposed system poses more advantages over the existing system. The proposed system provides decent insight into the future, and is easier to use. Like every other system, this system also has certain disadvantages. But they are negligible while comparing with the advantages and can be overcome in future.

8.1.1 Advantages

- The high accuracy model can help with near-precise predictions
- Can save investors from making wrong investments.
- Helps with budget-planning.
- Encourages people to plan their assets better
- Can be used by everyone

8.1.2 Limitations

- Additional features could be implemented into the dataset
- The training time could be improved using systems with better performance

Chapter 9

Conclusion and Future Scope

In this study, we used machine learning algorithms to predict the gold rates with maximum possible accuracy. Through this project, I tried to develop a comparative study between a machine learning and deep learning model to evaluate and predict the cost of gold.

LSTM Model is found to have better prediction accuracy for the entire period is found to give better accuracy for the two period taken separately. It is concluded that machine learning algorithms are very useful in such analysis, but the characteristics of the data influences their accuracy.

The results obtained by the created model seems encouraging and can be improved in future. The rate of errors in the machine learning model is very minimum . The majority of the project was built in python. It uses a flask server to connect to the user interface built using HTML, JavaScript and CSS. The project was built with the help of various python libraries such as keras, tensorflow, sklearn, pandas, numpy etc.

The feature scope of this particular machine learning model can be extended to multiple dimensions, including additional variables. Future development on this topic may lead to create a more detailed prediction, including the reasons for said fluctuations in price. This empowers users to make informed decisions based on their resources. This technology could give us meaningful insights, not only about precious but also about underlying factors that sometimes go overlooked at first glance while analysing the rise or drop in prices.

Bibliography

[1] V. K. F. B. Rebecca Davis, "Modeling and Forecasting of Gold Prices on Financia Markets," American International Journal of Contemporary Research, pp. Vol 4, No 3, 2014.

[2] Chen, H

[3] Chang, T.-H., Tsai,

[4] Lawrence, "Why is gold different from other assets? An empirical investigation," Lond. UK World Gold Council., 2003.

[5] L. A. Sjaastad and F. Scacciavillani, "The price of gold and the exchange rate," J. Int. Money Finance, vol. 15, no.6, pp. 879–897, 1996.

[6] S. A. Baker and R. C. Van Tassel, "Forecasting the price of gold: A fundamentalist approach," Atl. Econ. J., vol. 13, no. 4, pp. 43–51, 1985.