A

Major Project

On

GOLD PRICE PREDICTION USING ENSEMBLE BASED MACHINE LEARNING

(Submitted in partial fulfillment of the requirements for the award of Degree)

BACHELOR OF TECHNOLOGY

In

COMPUTER SCIENCE AND ENGINEERING

By

CHETTI ABHINAY (197R1A0510)

KOMMU BINDHU (197R1A0521)

KATAKAM ANJALI (197R1A0518)

Under the Guidance of

Dr. PUNYABAN PATEL

(Professor)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CMRTECHNICALCAMPUS

UGC AUTONOMOUS

(Accreditedby NAAC, NBA, Permanently Affiliatedto JNTUH, Approved by AICTE, NewDelhi)

Recognized Under Section 2(f) & 12(B) of the UGCAct.1956, Kandlakoya (V), Medchal Road, Hyderabad-501401.

2019-2023

DEPARTMENT OFCOMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the project entitled **"GOLD PRICE PREDICTION USING ENSEMBLE BASED MACHINE LEARNING"** being submitted by **CHETTI ABHINAY (197R1A0510),KOMMU BINDHU (197R1A0521) & KATAKAM ANJALI (197R1A0518)** in partial fulfillment of the requirements for the award of the degree of B.Tech in Computer Science and Engineering to the Jawaharlal Nehru Technological University Hyderabad, is a record of bonafide work carried out by them under our guidance and supervision during the year 2022-23.

The results embodied in this thesis have not been submitted to any other University or Institute for the award of any degree or diploma.

Dr. PUNYABAN PATEL
(Professor)
INTERNAL GUIDE

Dr.A. Raji Reddy
DIRECTOR

Dr. K. Srujan Raju
HOD

EXTERNAL EXAMINER

Submitted for viva voice Examination held _____

ACKNOWLEDGEMENT

Apart from the efforts of us, the success of any project depends largely on the encouragement and guidelines of many others. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of this project.

We take this opportunity to express my profound gratitude and deep regard to my guide **Dr. Punyaban Patel,** Professor for his exemplary guidance, monitoring, and constant encouragement throughout the project work. The blessing, help, and advice given by him shall carry us a long way in the journey of life on which we are about to embark.

We also take this opportunity to express a deep sense of gratitude to the Project Review Committee (PRC) **Dr. Punyaban Patel, Ms. K. Shilpa, Dr. M . Subha Mastan Rao & J. Narasimharao** for their cordial support, valuable information and guidance, which helped us in completing this task through various stages.

We are also thankful to **Dr. K. Srujan Raju,** Head of the Department of Computer Science and Engineering, **Dr. Ashuthosh Saxena,** Dean R&D, and **Dr. D T V Dharmajee Rao,** Dean Academics for providing encouragement and support for completing this project successfully.

We are obliged to **Dr. A. Raji Reddy,** Director for being cooperative throughout the course of this project. We also express our sincere gratitude to Sri. **Ch. Gopal Reddy,** Chairmanfor providing excellent infrastructure and a nice atmosphere throughout the course of this project.

The guidance and support received from all the members of **CMR Technical Campus** contributed to the completion of the project. We are grateful for their constant support and help.

Finally, we would like to take this opportunity to thank our family for their constant encouragement, without which this assignment would not be completed. We sincerely acknowledge and thank all those who gave support directly and indirectly in the completion of this project.

CHETTI ABHINAY (197R1A0510)

KOMMU BINDHU (197R1A0521)

KATAKAM ANJALI (197R1A0518)

ABSTRACT

This article is based on a study conducted to understand the relationship between gold price and selected factors influencing it, namely stock market, crude oil price, rupee dollar exchange rate, inflation and interest rate. Monthly price data for the period January 2000 to December 2018 was used for the study. The data was further split into two periods, period I from January 2000 to October 2011 during which the gold price exhibits a raising trend and period II from November 2011 to December 2018 where the gold price is showing a horizontal trend. Four machine learning algorithms, linear regression, random forest regression and gradient boosting regression and Voting regression were used in analyzing these data. It is found that the correlation between the variables is strong during the period I and weak during period II. While these models show good fit with data during period I. the fitness is not good during the period II. While random forest regression is found to have better prediction accuracy for the entire period, gradient boosting regression is found to give better accuracy for the two periods taken separately.

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
Figure 3.1.1	System architecture	7
Figure 3.1.2	Dataflow analysis	8
Figure 3.2.1	Usecase diagram	12
Figure 3.2.2	Class diagram	13
Figure 3.2.3	Activity diagram	14
Figure 3.2.4	Sequence diagram	15
Figure 3.2.5	Collaboration diagram	16
Figure 3.2.6	Component diagram	17
Figure 3.2.7	Deployment diagram	18
Figure 6.1.1	Static Testing	33
Figure 6.1.2	Structural Testing	33
Figure 6.1.3	Black Box testing	34
Figure 7.1	Home Screen	35
Figure 7.2	User Registration	36
Figure 7.3	User Login	36
Figure 7.4	Accuracy Graph	37
Figure 7.5	Main Page	37
Figure 7.6	User Input	38
Figure 7.7	Prediction Result	38

TABLE OF CONTENTS

Page No

ABSTRACT i

LIST OF FIGURES ii

1. INTRODUCTION **1**

1.1 Project Scope 1

1.2 Project Purpose 2

1.3 Project Features 2

2. System analysis **3**

2.1 Existing system 3

2.1.1 Disadvantages of existing system 3

2.2 Proposed System 3

2.2.1 Advantages of proposed system 3

2.3 Functional requirements 4

2.4 Non-Functional requirements 4

2.5 Feasibility Study 5

2.5.1 Economical Feasibility 5

2.5.2 Technical Feasibility 5

2.5.3 Social Feasibility 6

2.6 Hardware & Software Requirements 6

2.6.1 Hardware Requirements 6

2.6.2 Software Requirements 6

3. System design **7**

3.1 System architecture 7

3.2 Data Flow Diagram 8

3.3 UML diagrams 9

4. Implementation **17**

4.1 Modules 17

4.2 Algorithms 17

4.3 Sample code 18

5. Software environment **26**

6. System testing **32**

6.1 Software Testing Strategies 32

6.1.1 Static Testing 32

6.1.2 Structural Testing 33

6.1.3 Behavioral Testing 34

6.2 Test Cases 34

7. Screenshots **35**

8. Conclusion & Feature Scope **39**

9. Bibilography **40**

9.1 GITHUB LINK **41**

1. INTRODUCTION

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. In finance, an investment is purchase of a monetary asset with the idea that the asset will provide income in the future or will later be sold at a higher price for a profit. 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 another 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[1]. Gold is also considered to be "the asset of final instance" i.e. is the asset investors rely on, when the developed world capital markets are not capable to provide desirable profitability[2]. Thus it can be said that investors see gold as a tool to hedge against the fluctuations in other markets. Gold is a precious metal, so like any other goods, gold's price should depend on supply and demand. But, since gold is storable and the supply is accumulated over centuries, this year's production has little influence on its prices. Gold is used both as a commodity and as a financial asset. Gold behaves less like a commodity than long-lived assets such as stocks or bonds. Price of gold depends on a myriad of interrelated variables, including inflation rates, currency fluctuation and political turmoil[3].

1.1 PROJECT SCOPE

The gold price prediction to average around Rs.43,000,and to Rs,42800 for the years 2023,2024,2025, respectively.

1.2 PROJECT PURPOSE

Price prediction uses an algorithm to analyze a product or service basesd on its characteristics, demand, and current market trends. The software sets a price at a level it predicts will both attract customers and maximize sales. In some circles , the practice is call price forecasting or predictive pricing.

1.3 PROJECT FEATURES:

Gold is often used by investors as a hedge against inflation or adverse economic times. Consequently, it is important for investors to have accurate forecasts of gold prices.

2. SYSTEM ANALYSIS

2. SYSTEM ANALYSIS

2.1 EXISTING SYSTEM:

Lawrence[5] 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[6] reported that gold is a store of value against inflation and Baker and VanTassel[7] also have found that the price of gold depends on the future inflation rate. With respect to the relationship between gold price and inflation, based on the review of literature Hanan Naser[8] is of the opinion that historical studies with regards to the effectiveness of gold as a hedge against inflation are contradicting.

2.1.1 DISADVANTAGES OF EXIXTING SYSTEM:

1. But, of late price of gold is also witnessing high volatility and investments in gold are turning to be riskier.

2.2 PROPOSED SYSTEM:

This paper is aimed at studying the relationship between gold price and selected economic and market variable. Understanding such relationship will be helpful not only to monetary policymakers but also to investors, fund managers and portfolio managers to take better investment decisions in the market. Further this study uses three machine learning algorithms, linear regression, random forest regression and gradient boosting regression in analyzing these data. Comparison of these three methods will help us in identifying the accuracy of these methods under various conditions.

2.2.1 ADVANTAGES OF PROPOSED SYSTEM:

1. Better prediction accuracy

2. Comparison of these three methods will help us in identifying the accuracy of these under various conditions.

2.3 FUNCTIONAL REQUIREMENTS:

1. Data prediction accuracy
2. Data Preprocessing
3. Training and Testing
4. Modiling
5. Predicting

2.4 NON FUNCTIONAL REQUIREMNETS:

NON-FUNCTIONAL REQUIREMENT (NFR) specifies the quality attribute of a software system. They judge the software system based on Responsiveness, Usability, Security, Portability and other non-functional standards that are critical to the success of the software system. Example of non-functional requirement, "how fast does the website load?" Failing to meet non-functional requirements can result in systems that fail to satisfy user needs. Nonfunctional Requirements allows you to impose constraints or restrictions on the design of the system across the various agile backlogs. Example, the site should load in 3 seconds when the number of simultaneous users are > 10000. Description of non-functional requirements is just as critical as a functional requirement.

- Usability requirements
- Serviceability requirements
- Manageability requirements
- Security requirements
- Data Integrity requirements
- Capacity requirement
- Availability requirement
- Scalability requirement
- Interoperability requirement
- Reliability requirement
- Maintainability requirement
- Regulatory requirement
- Environmental requirement

2.5 FEASIBILITY STUDY

The feasibility of the project is analyzed in this phase and business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis the feasibility study of the proposed system is to be carried out. This is to ensure that the proposed system is not a burden to the company. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- ECONOMICAL FEASIBILITY
- TECHNICAL FEASIBILITY
- SOCIAL FEASIBILITY

2.5.1 ECONOMICAL FEASIBILITY:

This study is carried out to check the economic impact that the system will have on the organization. The amount of fund that the company can pour into the research and development of the system is limited. The expenditures must be justified. Thus the developed system as well within the budget and this was achieved because most of the technologies used are freely available. Only the customized products had to be purchased.

2.5.2 TECHNICAL FEASIBILITY:

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Any system developed must not have a high demand on the available technical resources. This will lead to high demands on the available technical resources. This will lead to high demands being placed on the client. The developed system must have a modest requirement, as only minimal or null changes are required for implementing this system.

2.5.3 SOCIAL FEASIBILITY

The aspect of study is to check the level of acceptance of the system by the user. This includes the process of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system.

2.6 HARDWARE & SOFTWARE REQUIREMENTS

2.6.1 HARDWARE REQUIREMENTS

Hardware interface specifies the logical characteristics of each interface between the software product and the hardware components of the system. The following are the some hardware requirements.

1. Processor : i5 and above
2. Hard disk : 1TB
3. RAM : 8GB and above
4. GPU : 2GB(optional)

2.6.2 SOFTWARE REQUIREMENTS

Software Requirements specifies the logical characteristics of each interface and software components of the system. The following are some software requirements.

1. Operating System : Windows only
2. Languages : Python
3. Tools : Anaconda- Jupyter, Spyder

3. SYSTEM DESIGN

3.SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE

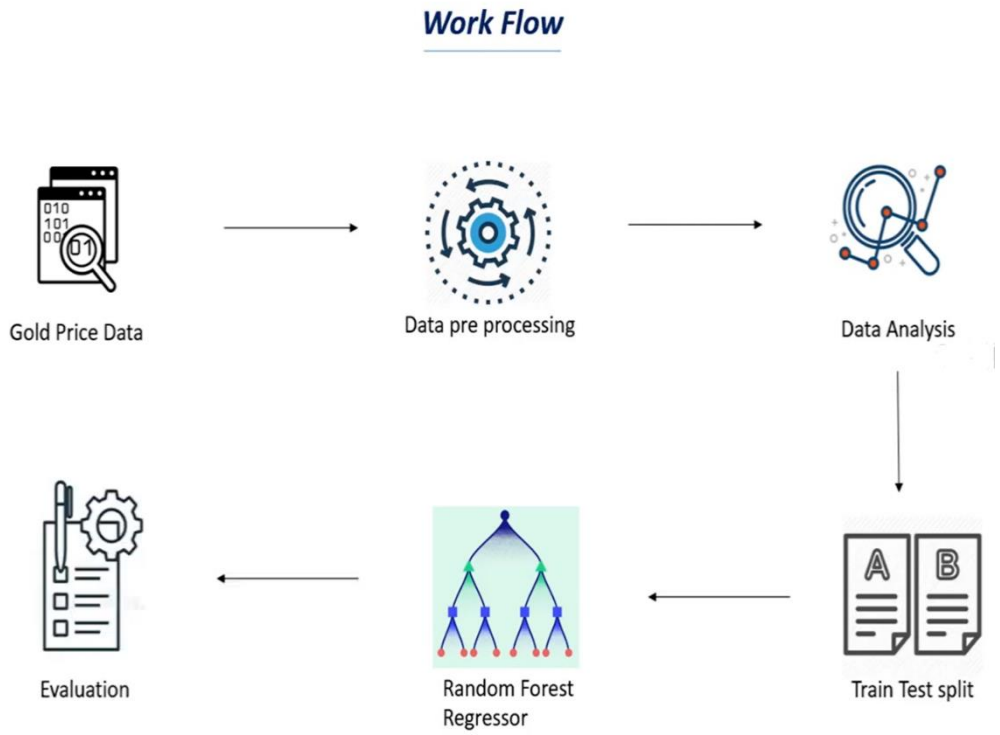

Figure 3.1.1: System architecture

3.2 DATA FLOW DIAGRAM

1. The DFD is also called as bubble chart. It is a simple graphical formalism that can be used to represent a system in terms of input data to the system, various processing carried out on this data, and the output data is generated by this system.
2. The data flow diagram (DFD) is one of the most important modeling tools. It is used to model the system components. These components are the system process, the data used by the process, an external entity that interacts with the system and the information flows in the system.
3. DFD shows how the information moves through the system and how it is modified by a series of transformations. It is a graphical technique that depicts information flow and the transformations that are applied as data moves from input to output.
4. DFD is also known as bubble chart. A DFD may be used to represent a system at any level of abstraction. DFD may be partitioned into levels that represent increasing information flow and functional detail.

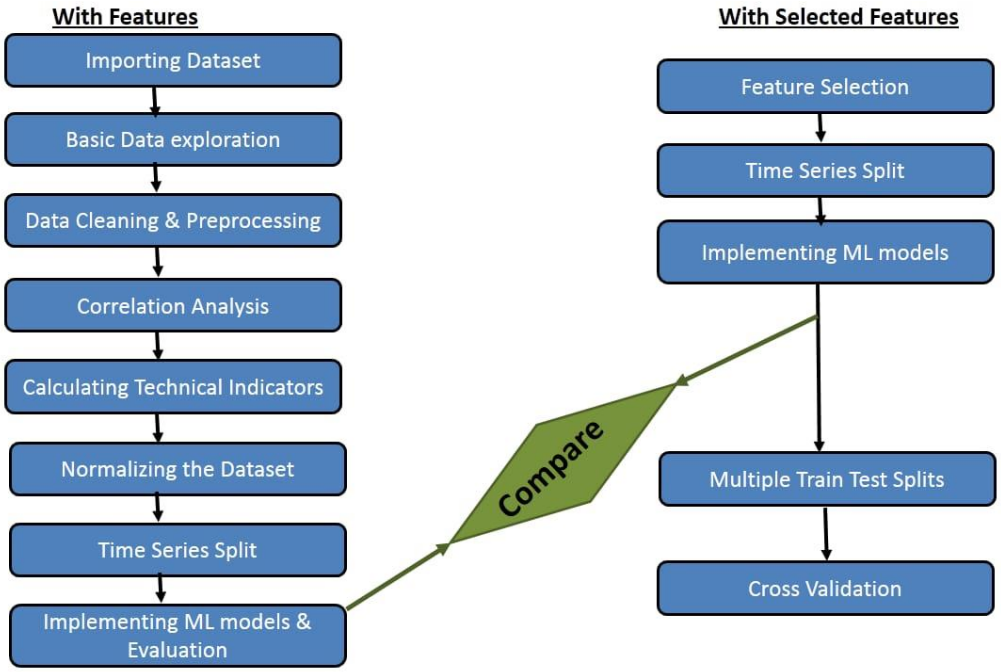

Figure 3.2.1: Data flow Diagram

3.3 UML DIAGRAMS

UML stands for Unified Modeling Language. UML is a standardized general-purpose modeling language in the field of object-oriented software engineering. The standard is managed, and was created by, the Object Management Group.

The goal is for UML to become a common language for creating models of object oriented computer software. In its current form UML is comprised of two major components: a Meta-model and a notation. In the future, some form of method or process may also be added to; or associated with, UML.

The Unified Modeling Language is a standard language for specifying, Visualization, Constructing and documenting the artifacts of software system, as well as for business modeling and other non-software systems.

The UML represents a collection of best engineering practices that have proven successful in the modeling of large and complex systems.

The UML is a very important part of developing objects oriented software and the software development process. The UML uses mostly graphical notations to express the design of software projects.

3.3.1 USE CASE DIAGRAM

A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

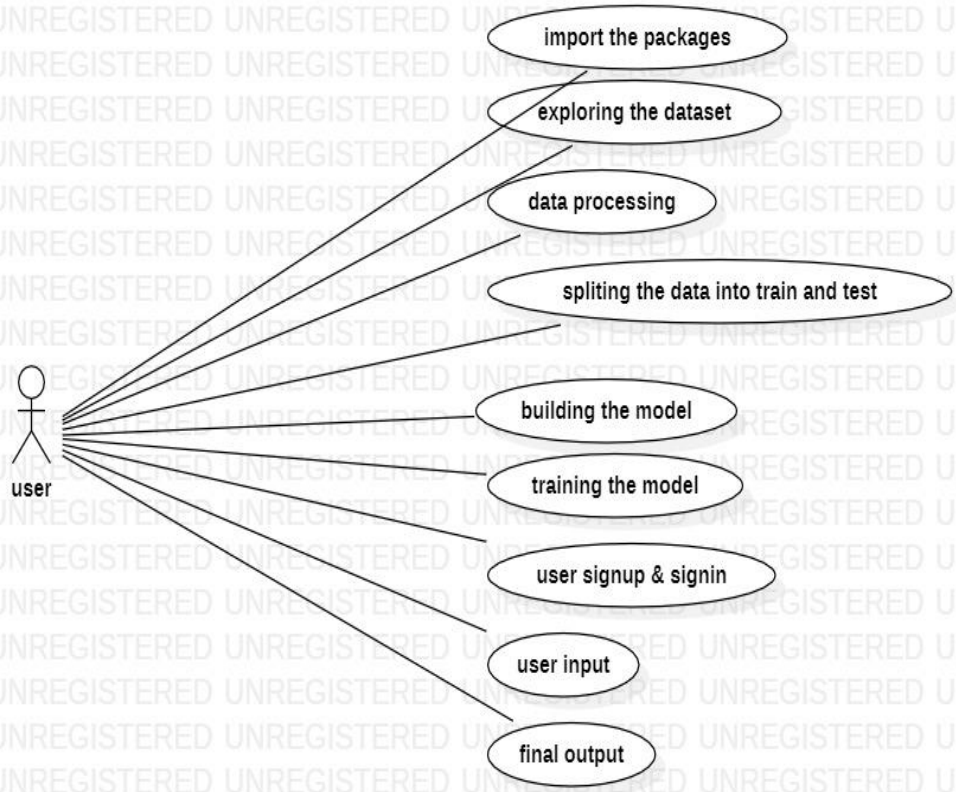

Figure 3.3.1: Usecase Diagram

3.3.2 CLASS DIAGRAM

The class diagram is used to refine the use case diagram and define a detailed design of the system. The class diagram classifies the actors defined in the use case diagram into a set of interrelated classes. The relationship or association between the classes can be either an "is-a" or "has-a" relationship. Each class in the class diagram may be capable of providing certain functionalities. These functionalities provided by the class are termed "methods" of the class. Apart from this, each class may have certain "attributes" that uniquely identify the class.

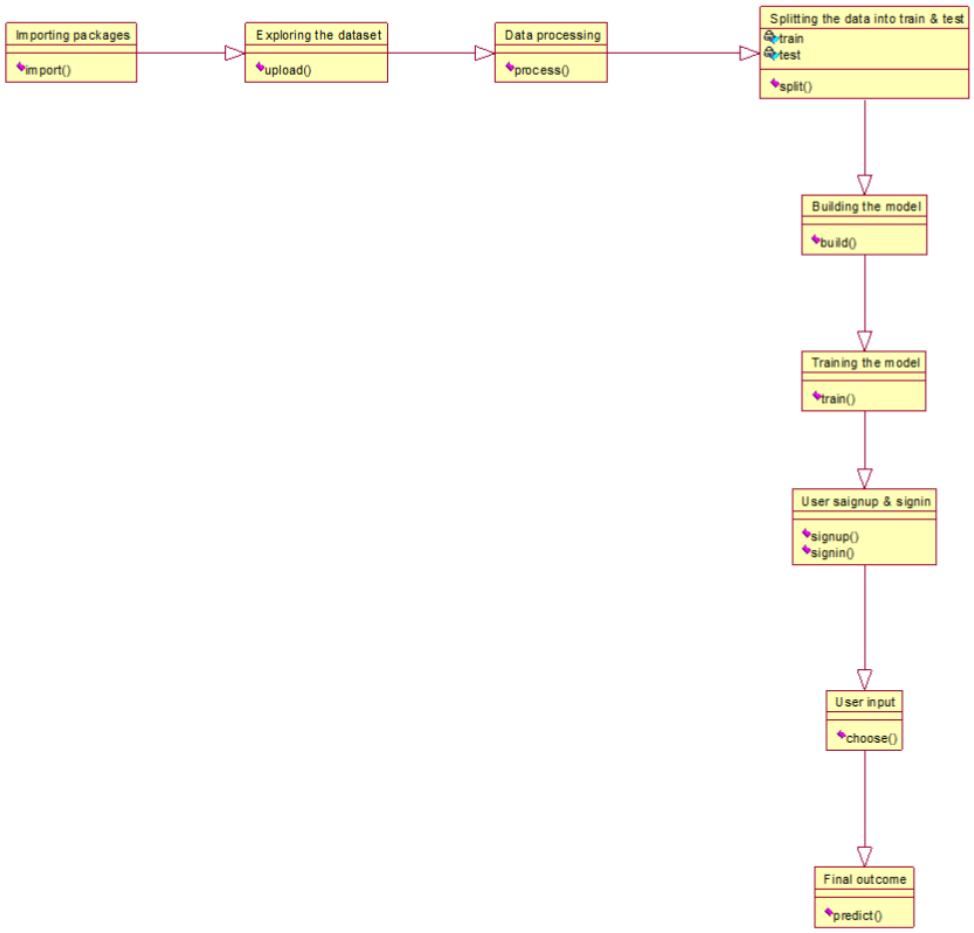

Figure 3.3.2 :Class Diagram

3.3.3 ACTIVITY DIAGRAM

The process flows in the system are captured in the activity diagram. Similar to a state diagram, an activity diagram also consists of activities, actions, transitions, initial and final states, and guard conditions.

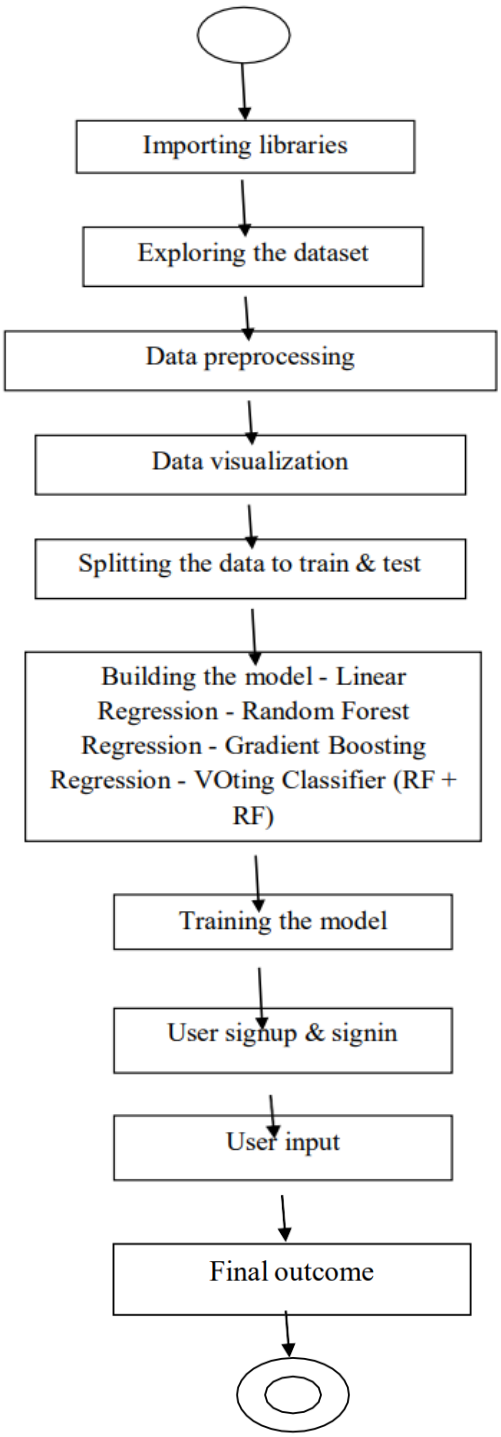

Figure 3.3.3: Activity Diagram

3.3.4 SEQUENCE DIAGRAM

A sequence diagram represents the interaction between different objects in the system. The important aspect of a sequence diagram is that it is time-ordered. This means that the exact sequence of the interactions between the objects is represented step by step. Different objects in the sequence diagram interact with each other by passing "messages".

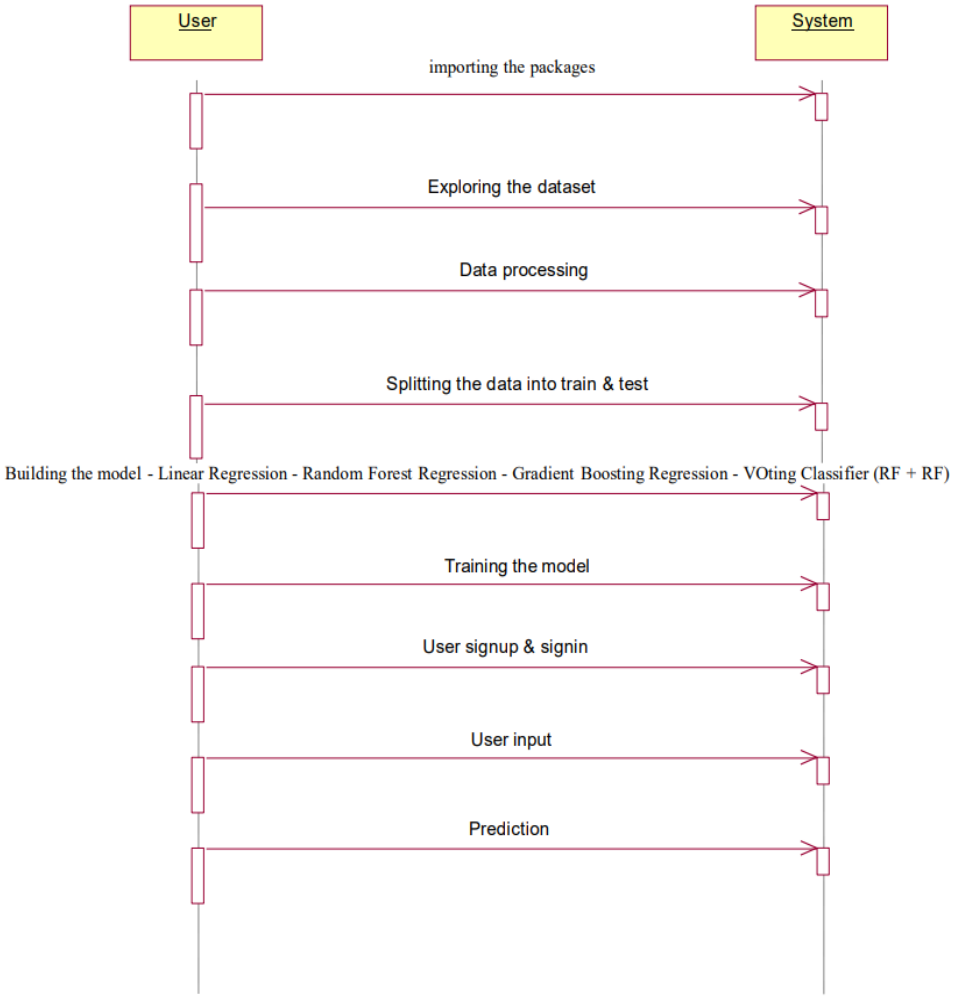

Figure 3.3.4: Sequence Diagram

3.3.5 COLLABORATION DIAGRAM

A collaboration diagram groups together the interactions between different objects. The interactions are listed as numbered interactions that help to trace the sequence of the interactions. The collaboration diagram helps to identify all the possible interactions that each object has with other objects.

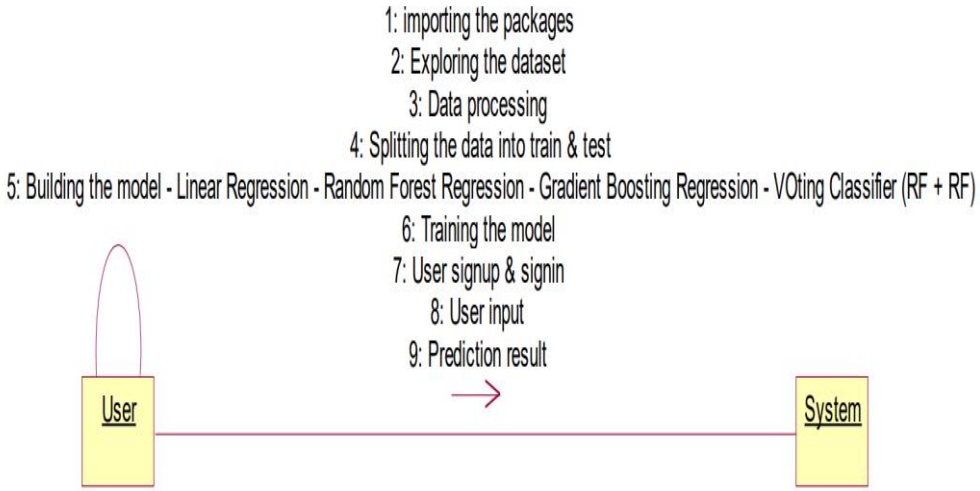

Figure 3.3.5: Collaboration Diagram

3.3.6 COMPONENT DIAGRAM

The component diagram represents the high-level parts that make up the system. This diagram depicts, at a high level, what components form part of the system and how they are interrelated. A component diagram depicts the components culled after the system has undergone the development or construction phase.

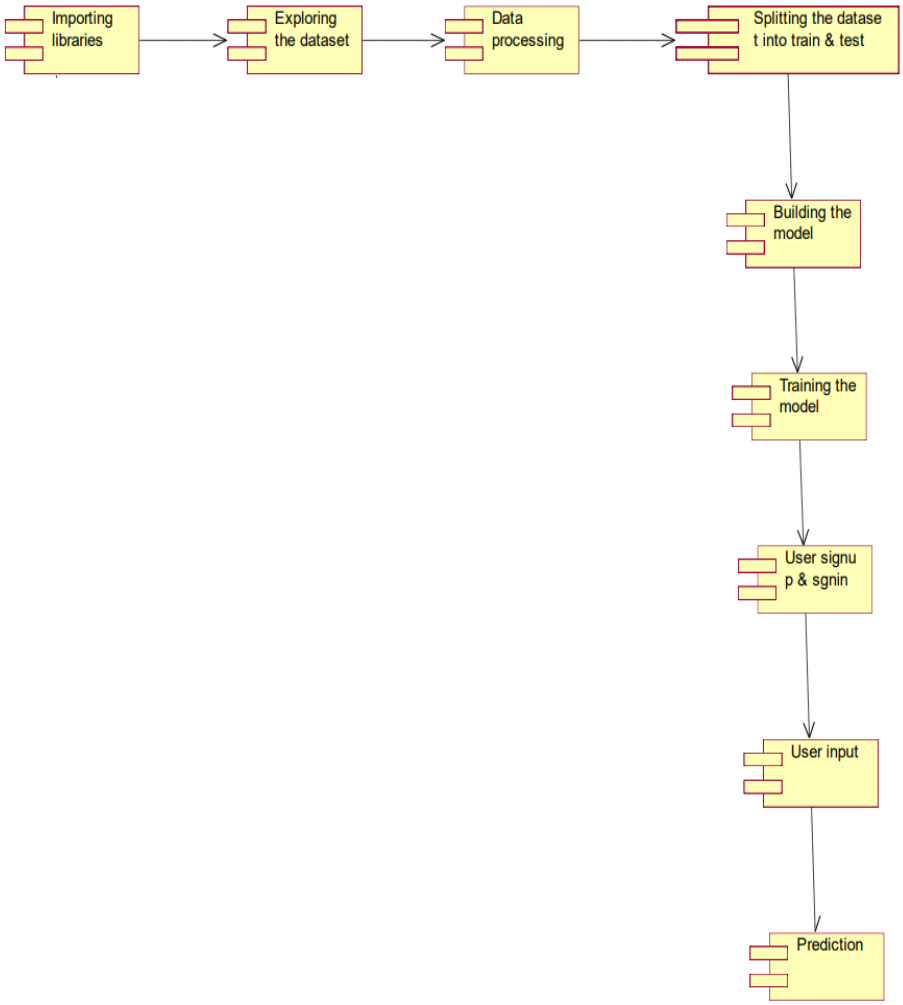

Figure 3.3.6: Component Diagram

3.3.7 DEPLOYMANT DIAGRAM

The deployment diagram captures the configuration of the runtime elements of the application. This diagram is by far most useful when a system is built and ready to be deployed.

Figure 3.3.7: Deployment Diagram

4. IMPLEMENTATION

4. IMPLEMENTATION

4.1 MODULES

- Data exploration: using this module we will load data into system
- Processing: Using the module we will read data for processing
- Splitting data into train & test: using this module data will be divided into train & test
- Building the model - Building the model - Linear Regression - Random Forest Regression – Gradient Boosting Regression- Voting Classifier(RF+RF)
- User signup & login: Using this module will get registration and login
- User input: Using this module will give input for prediction
- Prediction: final predicted displayed

4.2 ALGORITHMS

Linear Regression – Linear regression is an algorithm that provides a linear relationship between an independent variable and a dependent variable to predict the outcome of future events. It is a statistical method used in data science and machine learning for predictive analysis.

Random Forest Regression – A Random Forest Algorithm is a supervised machine learning algorithm that is extremely popular and is used for Classification and Regression problems in Machine Learning. We know that a forest comprises numerous trees, and the more trees more it will be robust.

Gradient Boosting Regression – Gradient boosting is a type of machine learning boosting. It relies on the intuition that the best possible next model, when combined with previous models, minimizes the overall prediction error. The key idea is to set the target outcomes for this next model in order to minimize the error.

Voting Classifier (RF + RF) -A voting classifier is a machine learning estimator that trains various base models or estimators and predicts on the basis of aggregating the findings of each base estimator. The aggregating criteria can be combined decision of voting for each estimator output.

4.3 SAMPLE CODE:

```
# -*-
Coding
:
utf-8
-*-
# drive.mount("/content/drive")
# import os
# os.chdir("/content/drive/My Drive/Colab Notebooks/data")
# print(os.getcwd())
# Change directory to the directory above "data"
# LinearRegression is a machine learning library for linear regression
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error
# pandas and numpy are used for data manipulation
import pandas as pd
import numpy as np
from math import sqrt
from numpy import log
from pandas import Series
from statsmodels.tsa.arima_model import ARMA
from statsmodels.tsa.arima_model import ARIMA
from statsmodels.tsa.stattools import adfuller, arma_order_select_ic
from statsmodels.graphics.tsaplots import plot_acf
from statsmodels.graphics.tsaplots import plot_pacf
import statsmodels as sm
# matplotlib and seaborn are used for plotting graphs
import matplotlib.pyplot as plt
from matplotlib.dates import date2num
import seaborn as sns
from datetime import datetime
import subprocess
# fix_yahoo_finance is used to fetch data
# import fix_yahoo_finance as yf
ds_gold = 'Indian rupee'
ds_etf = 'Close'
date_format = '%Y-%m-%d'
df = pd.read_csv("data_inr.csv")
df = df[['Name', ds_gold]]
df['Name'] = [datetime.strptime(i, date_format) for i in df['Name']]
df.set_index('Name')
# df.index = pd.to_datetime(df.index, format=date_format)
print(df.columns)
dd =df
```

```
"""*  Drop rows with missing values"""
df = df.dropna()
df[ds_gold].hist()
plt.show()
log_transform = log(df[ds_gold])
print(min(log_transform), max(log_transform))
sns.set()
sns.distplot(df[ds_gold])
plt.show()
plt.plot(df['Name'], df[ds_gold])
plt.show()
plt.plot(df['Name'], log_transform)
plt.show()
# Can be used to show non stationary
# Define exploratory variables
# Finding moving average of past 3 days and 9 days
df['S_1'] = df[ds_gold].shift(1).rolling(window=3).mean()
df['S_2'] = df[ds_gold].shift(1).rolling(window=12).mean()
df = df.dropna()
X = df[['S_1', 'S_2']]
X.head()
plt.plot(df['Name'], df['S_1'])
plt.plot(df['Name'], df["S_2"])
plt.show()
# dependent variable
y = df[ds_gold]
y.head()
# Split into train and test
t = 0.2
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=t, shuffle=False)
# Performing linear regression
linear = LinearRegression().fit(X_train, y_train)
print("Gold Price =", round(linear.coef_[0], 2), "* 2 Month Moving
 Average", round( linear.coef_[1], 2), "* 1 Month Moving Average +"
 , round(linear.intercept_, 2))
# Predict prices
predicted_price = linear.predict(X_test)
predicted_price = pd.DataFrame(
    predicted_price, index=y_test.index, columns=['price'])
predicted_price.plot(figsize=(10, 5))
y_test.plot()
plt.legend(['predicted_price', 'actual_price'])
plt.ylabel("Gold Price")
plt.show()
# Calculate R square and rmse to check goodness of fit
```

```
r2_score = linear.score(X_test, y_test)*100
print("R square for regression", float("{0:.2f}".format(r2_score)))
sqrt(mean_squared_error(y_test,predicted_price))
# We observe significantly different accuracies for same dataset in
 USD and INR.
# The reason for this difference could be attributed to the
# Check stationarity
X = df[ds_gold]
split = len(X) // 2
X1, X2 = X[0:split], X[split:]
mean1, mean2 = X1.mean(), X2.mean()
var1, var2 = X1.var(), X2.var()
print('mean1=%f, mean2=%f' % (mean1, mean2))
print('variance1=%f, variance2=%f' % (var1, var2))
result_of_adfuller = adfuller(df[ds_gold])
print('ADF Statistic: %f' % result_of_adfuller[0])
print('p-value: %f' % result_of_adfuller[1])
print('Critical Values:')
for key, value in result_of_adfuller[4].items():
    print('\t%s: %.3f' % (key, value))
# we can conclude it has time dependent structure and cannot reject
     null hypothesis.
# from statsmodels.tsa.seasonal import seasonal_decompose
# print(df.index.dtype)
# print(df[ds_gold].dtype)
# decomposition = seasonal_decompose(df[ds_gold], freq = 200)
# trend = decomposition.trend
# seasonal = decomposition.seasonal
# residual = decomposition.resid
# plt.subplot(411)
# plt.plot(df[ds_gold], label='Original')
# plt.legend(loc='best')
# plt.subplot(412)
# plt.plot(trend, label='Trend')
# plt.legend(loc='best')
# plt.subplot(413)
# plt.plot(seasonal,label='Seasonality')
# plt.legend(loc='best')
# plt.subplot(414)
# plt.plot(residual, label='Residuals')
# plt.legend(loc='best')
# plt.tight_layout()
# print(df.head())
# Now taking log transform
log_transform = log(df[ds_gold])
result_of_adfuller = adfuller(log_transform)
```

```
print('ADF Statistic: %f' % result_of_adfuller[0])
print('p-value: %f' % result_of_adfuller[1])
print('Critical Values:')
for key, value in result_of_adfuller[4].items():
    print('\t%s: %.3f' % (key, value))
# To remove trends, differencing of order 1
k = df[ds_gold].diff()
plt.plot(df['Name'], k)
plt.show()
# print(k.head())
k = k.dropna()
# check stationarity after differencing
result_of_adfuller = adfuller(k)
print('ADF Statistic: %f' % result_of_adfuller[0])
print('p-value: %f' % result_of_adfuller[1])
print('Critical Values:')
for key, value in result_of_adfuller[4].items():
    print('\t%s: %.3f' % (key, value))
# So now we can say with 1 % confidence level that its stationary
# We can do other stuff now
# Again regression
df[ds_gold] = k
# Finding moving average of past 3 days and 9 days
df['S_1'] = df[ds_gold].shift(1).rolling(window=3).mean()
df['S_2'] = df[ds_gold].shift(1).rolling(window=12).mean()
df = df.dropna()
X = df[['S_1', 'S_2']]
X.head()
print(X.head())
plt.plot(df['Name'], df['S_1'])
plt.plot(df['Name'], df["S_2"])
plt.show()
df['S_1'] = df[ds_gold].shift(1).rolling(window=3).mean()
df['S_2'] = df[ds_gold].shift(1).rolling(window=12).mean()
# dependent variable
y = df[ds_gold]
y.head()
# print(y.head())
# Split into train and test
t = 0.2
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=t, shuffle=False)
# Performing linear regression
linear = LinearRegression().fit(X_train, y_train)
print("Gold Price =", round(linear.coef_[0], 2), "* 2 Month Moving
 Average", round( linear.coef_[1], 2), "* 1 Month Moving Average +",
```

```
 round(linear.intercept_, 2))
# Predict prices
predicted_price = linear.predict(X_test)
predicted_price = pd.DataFrame(
    predicted_price, index=y_test.index, columns=['price'])
predicted_price.plot(figsize=(10, 5))
y_test.plot()
plt.legend(['predicted_price', 'actual_price'])
plt.ylabel("Gold Price")
plt.show()
# Calculate R square and rmse to check goodness of fit
r2_score = linear.score(X_test, y_test)*100
print("R square for regression", float("{0:.2f}".format(r2_score)))
sqrt(mean_squared_error(y_test,predicted_price))
# Trying 2nd order differencing
k = df[ds_gold].diff().diff()
plt.plot(df['Name'], k)
plt.show()
# print(k.head())
k = k.dropna()
# check stationarity after differencing
result_of_adfuller = adfuller(k)
print('ADF Statistic: %f' % result_of_adfuller[0])
print('p-value: %f' % result_of_adfuller[1])
print('Critical Values:')
for key, value in result_of_adfuller[4].items():
    print('\t%s: %.3f' % (key, value))
# Again regression
df[ds_gold] = k
# Finding moving average of past 3 days and 9 days
df['S_1'] = df[ds_gold].shift(1).rolling(window=3).mean()
df['S_2'] = df[ds_gold].shift(1).rolling(window=12).mean()
df = df.dropna()
X = df[['S_1', 'S_2']]
X.head()
print(X.head())
plt.plot(df['Name'], df['S_1'])
plt.plot(df['Name'], df["S_2"])
plt.show()
# dependent variable
y = df[ds_gold]
y.head()
# print(y.head())
df['S_1'] = df[ds_gold].shift(1).rolling(window=3).mean()
df['S_2'] = df[ds_gold].shift(1).rolling(window=12).mean()
# Split into train and test
```

```
t = 0.2
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=t, shuffle=False)
# Performing linear regression
linear = LinearRegression().fit(X_train, y_train)
print("Gold Price =", round(linear.coef_[0], 2), "* 2 Month Moving
Average", round( linear.coef_[1], 2), "* 1 Month Moving Average +",
    round(linear.intercept_, 2))
# Predict prices
predicted_price = linear.predict(X_test)
predicted_price = pd.DataFrame(
    predicted_price, index=y_test.index, columns=['price'])
predicted_price.plot(figsize=(10, 5))
y_test.plot()
plt.legend(['predicted_price', 'actual_price'])
plt.ylabel("Gold Price")
plt.show()
# Calculate R square and rmse to check goodness of fit
r2_score = linear.score(X_test, y_test)*100
print("R square for regression", float("{0:.2f}".format(r2_score)))
print("RMSE: ",sqrt(mean_squared_error(y_test,predicted_price)))
#Now after 2nd order differencing the results make sense, but are
still unacceptably innacurate
#Let us now try to make an ARMA model with the new non-stationary data
# order = arma_order_select_ic(df[ds_gold])
# print(df[ds_gold])
#ACF and PACF plots
series = df[ds_gold]
plt.figure()
plt.subplot(211)
plot_acf(series, ax=plt.gca())
plt.subplot(212)
plot_pacf(series, ax=plt.gca())
plt.show()
del df['S_1']
del df['S_2']
# data = pd.Series(df['Indian rupee'], index=df['Name'])
# model = ARMA(data, order=(5,1))
# data
# df
# arma_model = ARMA(df,order = (2,3))
# df['Name'] = df['Name'].values.astype(float)
# ts = pd.Series(df[ds_gold], index = df.index)
# print(ts.head())
# model = ARIMA(df[ds_gold].values, order=(1, 1, 1))
# results_ARIMA = model.fit(disp=-1)
```

```
# # np.asarray(dd)
# plt.plot(df)
# plt.plot(results_ARIMA.fittedvalues, color='red')
# plt.title('RSS: %.4f'% sum((results_ARIMA.fittedvalues-ts_log_diff)
**2))
# from statsmodels.tsa.statespace import SARIMAX
import statsmodels.api as sm
mod = sm.tsa.statespace.SARIMAX(df[ds_gold].values,order=(2, 1, 2)
,seasonal_order=(2, 1, 2, 12),enforce_stationarity=False,enforce_
invertibility=False)
results = mod.fit()
results.summary()
df['sarimax_predict'] = results.predict()
# del df['S_1']
# del df['S_2']
df.plot(y = ['Indian rupee','sarimax_predict'], x = df['Name'])
results.summary()
results.plot_diagnostics(figsize=(15, 12))
plt.show()
print("RMSE: ",sqrt(mean_squared_error(df[ds_gold],df['sarimax_predict'])))
from sklearn.metrics import r2_score
print("R2 SCORE: ",r2_score(df[ds_gold],df['sarimax_predict']))
#we have finally reached a good model. GG
#Finding trends
import datetime as dt
data = pd.read_csv("dataset2018.csv")
data.head()
x_18 = data.iloc[:,0]
y_18 = data.iloc[:,1]
new_18 = [dt.datetime.strptime(d,'%d-%m-%Y').date() for d in x_18]
plt.plot(new_18,y_18, '.r',color='g')
plt.xlabel('date')
plt.ylabel('price')
plt.title('gold price in the year 2018')
plt.show()
data2017 = pd.read_csv("dataset2017.csv")
x_17 = data2017.iloc[:,0]
y_17 = data2017.iloc[:,1]
new_17 = [dt.datetime.strptime(d,'%d-%m-%Y').date() for d in x_17]
plt.plot(new_17,y_17, '.r',color='g')
plt.xlabel('date')
plt.ylabel('price')
plt.title('gold price in the year 2017')
plt.show()
data2016 = pd.read_csv("dataset2016.csv")
x_16 = data2016.iloc[:,0]
```

```
y_16 = data2016.iloc[:,1]
new_16 = [dt.datetime.strptime(d,'%d-%m-%Y').date() for d in x_16]
plt.plot(new_16,y_16, '.r',color='g')
plt.xlabel('date')
plt.ylabel('price')
plt.title('gold price in the year 2016')
plt.show()
data2015 = pd.read_csv("dataset2015.csv")
x_15 = data2015.iloc[:,0]
y_15 = data2015.iloc[:,1]
new_15 = [dt.datetime.strptime(d,'%d-%m-%Y').date() for d in x_15]
plt.plot(new_15,y_15, '.r',color='g')
plt.xlabel('date')
plt.ylabel('price')
plt.title('gold price in the year 2015')
plt.show()
data2014 = pd.read_csv("dataset2014.csv")
x_14 = data2014.iloc[:,0]
y_14 = data2014.iloc[:,1]
new_14 = [dt.datetime.strptime(d,'%d-%m-%Y').date() for d in x_14]
plt.plot(new_14,y_14, '.r',color='g')
plt.xlabel('date')
plt.ylabel('price')
plt.title('gold price in the year 2014')
plt.show()
data2014 = pd.read_csv("dataset2013.csv")
x_14 = data2014.iloc[:,0]
y_14 = data2014.iloc[:,1]
new_14 = [dt.datetime.strptime(d,'%d-%m-%Y').date() for d in x_14]
plt.plot(new_14,y_14, '.r',color='g')
plt.xlabel('date')
plt.ylabel('price')
plt.title('gold price in the year 2013')
plt.show()
data2014 = pd.read_csv("datasetfull.csv")
x_14 = data2014.iloc[:,0]
y_14 = data2014.iloc[:,1]
new_14 = [dt.datetime.strptime(d,'%d-%m-%Y').date() for d in x_14]
plt.plot(new_14,y_14, '.r',color='g')
plt.xlabel('date')
plt.ylabel('price')
plt.title('gold price')
plt.show()
```

5. SOFTWARE ENVIRONMENT

5. SOFTWARE ENVIRONMENT

5.1 PYTHON LANGUAGE

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed. Often, programmers fall in love with Python because of the increased productivity it provides. Since there is no compilation step, the edit-test-debug cycle is incredibly fast. Debugging Python programs is easy: a bug or bad input will never cause a segmentation fault. Instead, when the interpreter discovers an error, it raises an exception. When the program doesn't catch the exception, the interpreter prints a stack trace. A source level debugger allows inspection of local and global variables, evaluation of arbitrary expressions, setting breakpoints, stepping through the code a line at a time, and so on. The debugger is written in Python itself, testifying to Python's introspective power. On the other hand, often the quickest way to debug a program is to add a few print statements to the source: the fast edit-test-debug cycle makes this simple approach very effective.

Python is a dynamic, high-level, free open source, and interpreted programming language. It supports object-oriented programming as well as procedural-oriented programming. In Python, we don’t need to declare the type of variable because it is a dynamically typed language. For example, x = 10 Here, x can be anything such as String, int, etc.

5.2 FEATURES OF PYTHON

There are many features in Python, some of which are discussed below as follows:

5.2.1 Free and Open Source

Python language is freely available at the official website and you can download it from the given download link below click on the Download Python keyword. Download Python Since it is open-source, this means that source code is also available to the public. So you can download it, use it as well as share it.

5.2.2 Easy to code

Python is a high-level programming language. Python is very easy to learn the language as compared to other languages like C, C#, Javascript, Java, etc. It is very easy to code in the Python language and anybody can learn Python basics in a few hours or days. It is also a developer-friendly language.

5.2.3 Easy to Read

As you will see, learning Python is quite simple. As was already established, Python's syntax is really straightforward. The code block is defined by the indentations rather than by semicolons or brackets.

5.2.4 Object-Oriented Language

One of the key features of Python is Object-Oriented programming. Python supports object-oriented language and concepts of classes, object encapsulation, etc.

5.2.5 GUI Programming Support

Graphical User interfaces can be made using a module such as PyQt5, PyQt4, wxPython, or Tk in python. PyQt5 is the most popular option for creating graphical apps with Python.

5.2.6 High-Level Language

Python is a high-level language. When we write programs in Python, we do not need to remember the system architecture, nor do we need to manage the memory.

5.2.7 Extensible feature

Python is an Extensible language. We can write some Python code into C or C++ language and also we can compile that code in C/C++ language.

5.2.8 Easy to Debug

Excellent information for mistake tracing. You will be able to quickly identify and correct the majority of your program's issues once you understand how to interpret Python's error traces. Simply by glancing at the code, you can determine what it is designed to perform.

5.2.9 Python is a Portable language

Python language is also a portable language. For example, if we have Python code for windows and if we want to run this code on other platforms such as Linux, Unix, and Mac then we do not need to change it, we can run this code on any platform.

5.2.10 Python is an Integrated language

Python is also an Integrated language because we can easily integrate Python with other languages like C, C++, etc.

5.2.11 Interpreted Language

Python is an Interpreted Language because Python code is executed line by line at a time. like other languages C, C++, Java, etc. there is no need to compile Python code this makes it easier to debug our code. The source code of Python is converted into an immediate form called bytecode.

5.2.12 Large Standard Library

Python has a large standard library that provides a rich set of modules and functions so you do not have to write your own code for every single thing. There are many libraries present in Python such as regular expressions, unit-testing, web browsers, etc.

5.2.13 Dynamically Typed Language

Python is a dynamically-typed language. That means the type (for example- int, double, long, etc.) for a variable is decided at run time not in advance because of this feature we don't need to specify the type of variable.

5.2.14 Frontend and backend development

With a new project py script, you can run and write Python codes in HTML with the help of some simple tags <py-script>, <py-env>, etc. This will help you do frontend development work in Python like javascript. Backend is the strong forte of Python it's extensively used for this work cause of its frameworks like Django and Flask.

5.2.15 Allocating Memory Dynamically

In Python, the variable data type does not need to be specified. The memory is automatically allocated to a variable at runtime when it is given a value. Developers do not need to write int y = 18 if the integer value 15 is set to y. You may just type y=18.

5.3 LIBRARIES/PACKGES

5.3.1 Tensorflow

TensorFlow is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks. It is used for both research and production at Google.

TensorFlow was developed by the Google Brain team for internal Google use. It was released under the Apache 2.0 open-source license on November 9, 2015.

5.3.2 Numpy

Numpy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, Numpy can also be used as an efficient multi-dimensional container of generic data. Arbitrary data-types can be defined using Numpy which allows Numpy to seamlessly and speedily integrate with a wide variety of databases.

5.3.4 Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

5.3.5 Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

5.3.6 Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use.

6. SYSTEM TESTING

6. SYSTEM TESTING

System testing, also referred to as system-level tests or system-integration testing, is the process in which a quality assurance (QA) team evaluates how the various components of an application interact together in the full, integrated system or application. System testing verifies that an application performs tasks as designed. This step, a kind of black box testing, focuses on the functionality of an application. System testing, for example, might check that every kind of user input produces the intended output across the application.

Phases of system testing

A video tutorial about this test level. System testing examines every component of an application to make sure that they work as a complete and unified whole. A QA team typically conducts system testing after it checks individual modules with functional or user-story testing and then each component through integration testing.

If a software build achieves the desired results in system testing, it gets a final check via acceptance testing before it goes to production, where users consume the software. An app-dev team logs all defects, and establishes what kinds and amount of defects are tolerable.

6.1 Software Testing Strategies

Optimization of the approach to testing in software engineering is the best way to make it effective. A software testing strategy defines what, when, and how to do whatever is necessary to make an end-product of high quality. Usually, the following software testing strategies and their combinations are used to achieve this major objective:

6.1.1 Static Testing

The early-stage testing strategy is static testing: it is performed without actually running the developing product. Basically, such desk-checking is required to detect bugs and issues that are present in the code itself. Such a check-up is important at the pre-deployment stage as it helps avoid problems caused by errors in the code and software structure deficits as shown in fig.

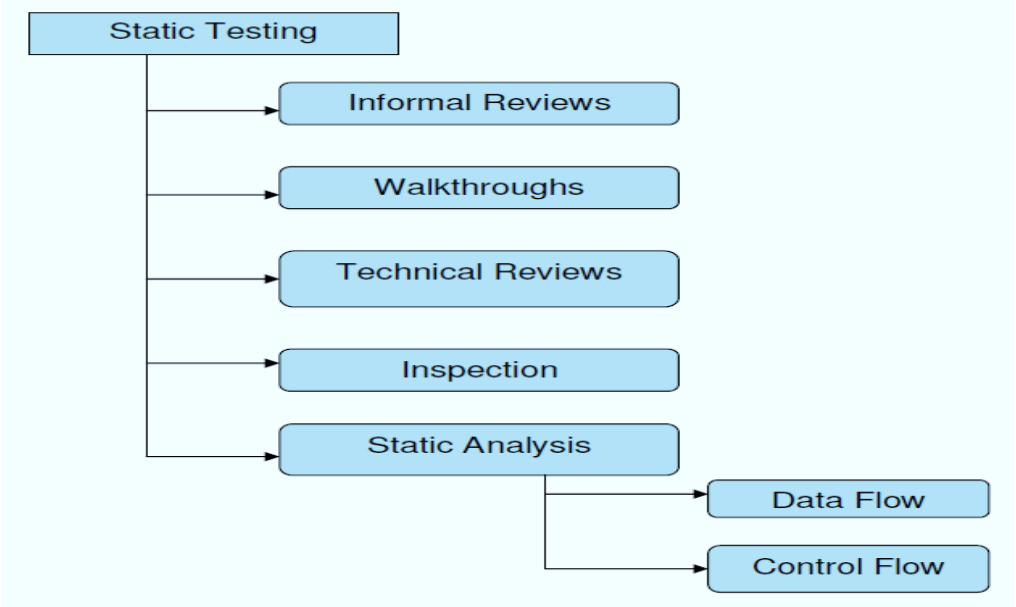

Figure 6.1.1: Static Testing

6.1.2 Structural Testing

It is not possible to effectively test software without running it. Structural testing, also known as white-box testing, is required to detect and fix bugs and errors emerging during the pre-production stage of the software development process. At this stage, unit testing based on the software structure is performed using regression testing. In most cases, it is an automated process working within the test automation framework to speed up the development process at this stage. Developers and QA engineers have full access to the software's structure and data flows (data flows testing), so they could track any changes (mutation testing) in the system's behavior by comparing the tests' outcomes with the results of previous iterations (control flow testing).

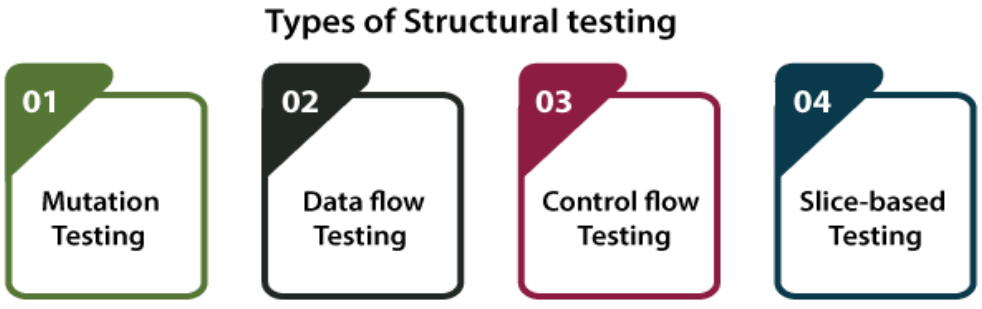

Figure 6.1.2: Structural Testing

6.1.3 Behavioral Testing

The final stage of testing focuses on the software's reactions to various activities rather than on the mechanisms behind these reactions. In other words, behavioral testing, also known as black-box testing, presupposes running numerous tests, mostly manual, to see the product from the user's point of view. QA engineers usually have some specific information about a business or other purposes of the software ('the black box') to run usability tests, for example, and react to bugs as regular users of the product will do. Behavioral testing also may include automation (regression tests) to eliminate human error if repetitive activities are required. For example, you may need to fill 100 registration forms on the website to see how the product copes with such an activity, so the automation of this test is preferable.

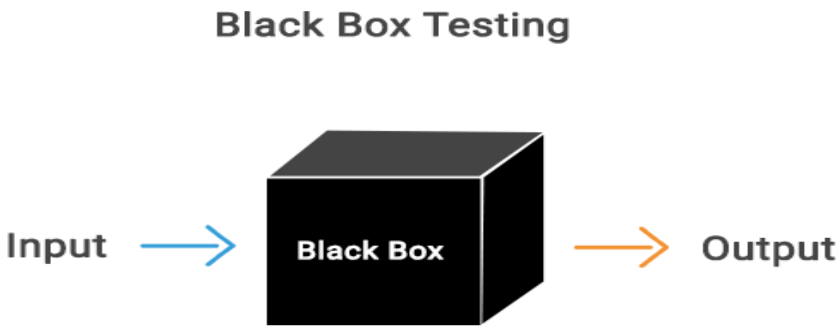

Figure 6.1.3: Black Box Testing

6.2 TEST CASES

SI.NO.	Input	If available	If not available
1	User signup	User get registered into the application	There is no process
2	User signin	User get login into the application	There is no process
3	Enter input for prediction	Prediction result displayed	There is no process

7. SCREENSHOTS

7.SCREENSHOTS

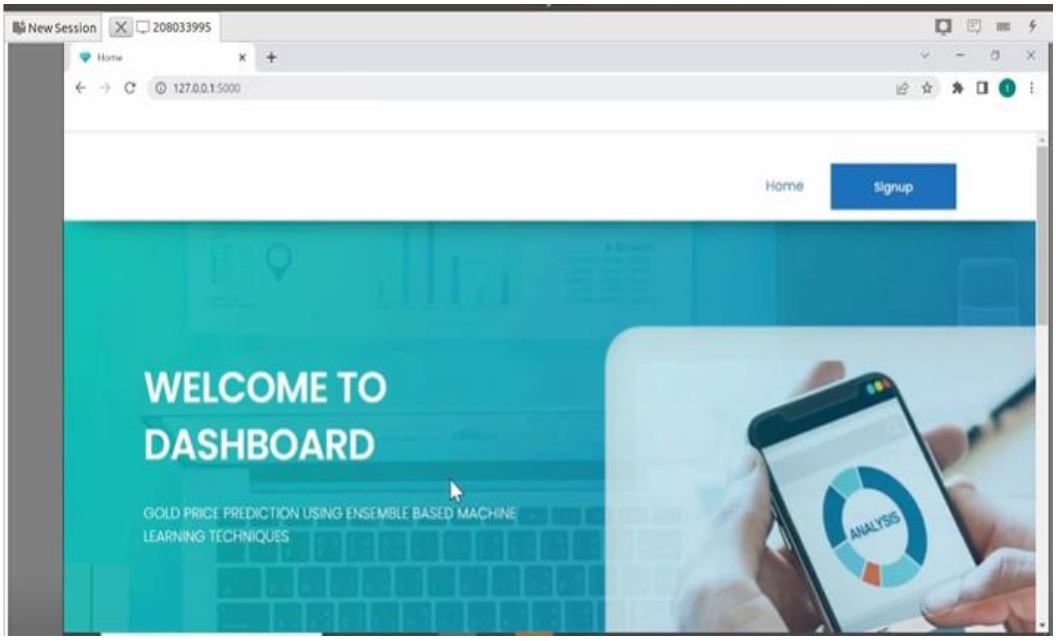

Figure 7.1: Home page

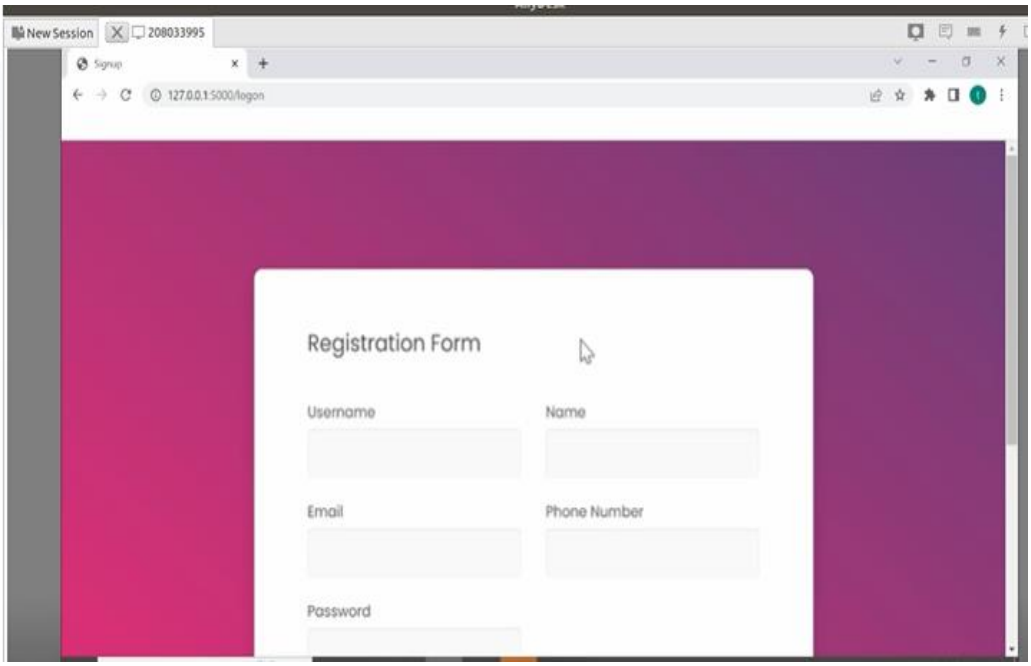

Figure 7.2: Registration page

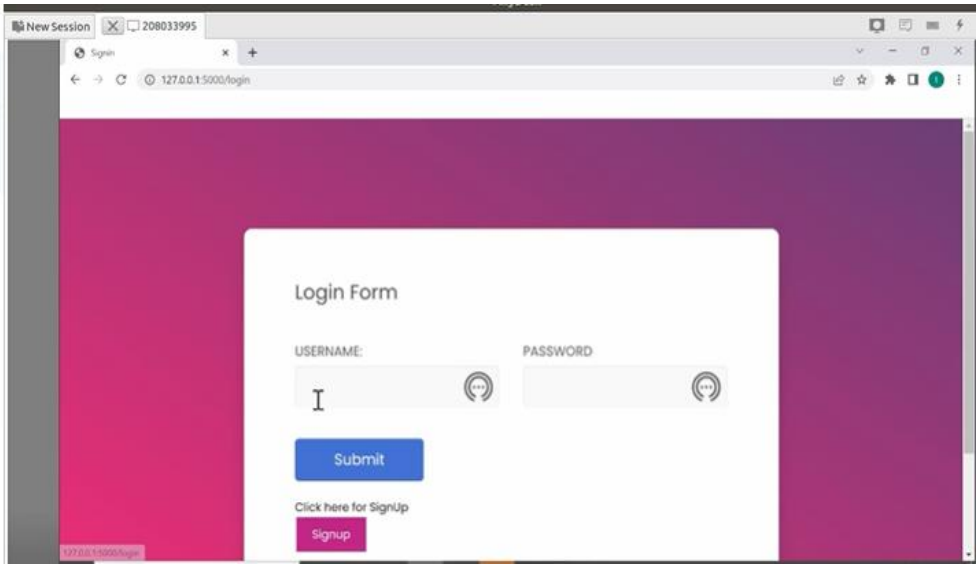

Figure 7.3: User Login

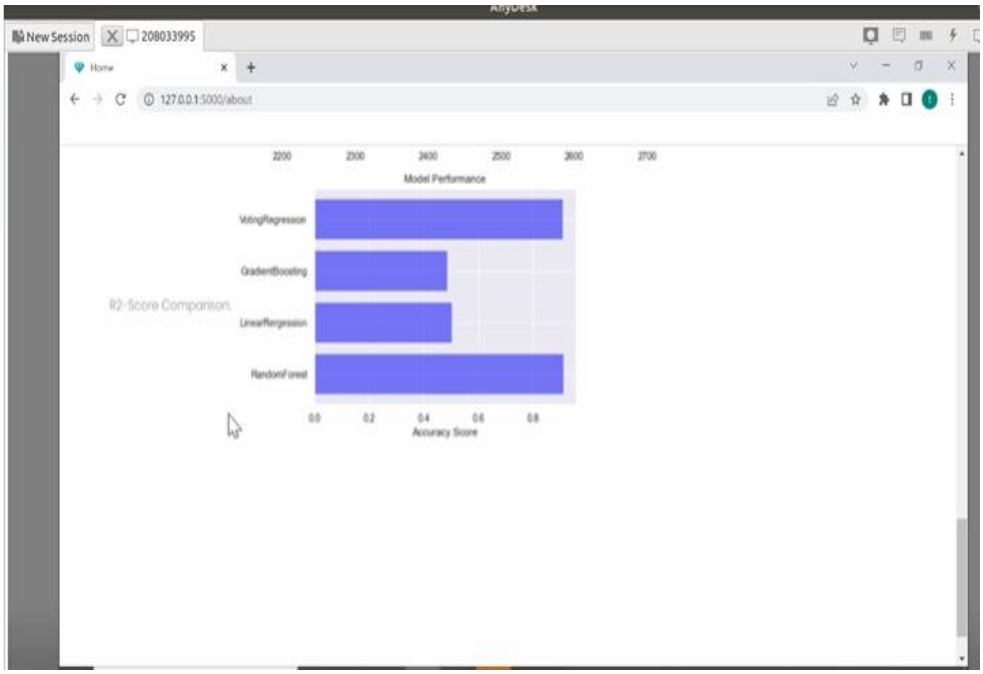

Figure 7.4: Accuracy Graph

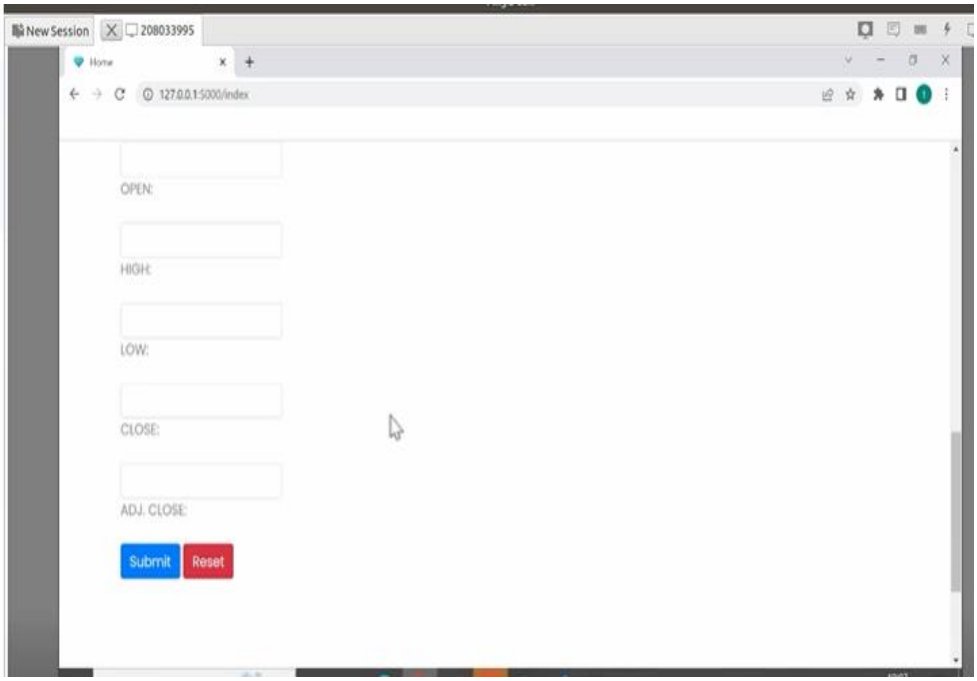

Figure 7.5: Main Page

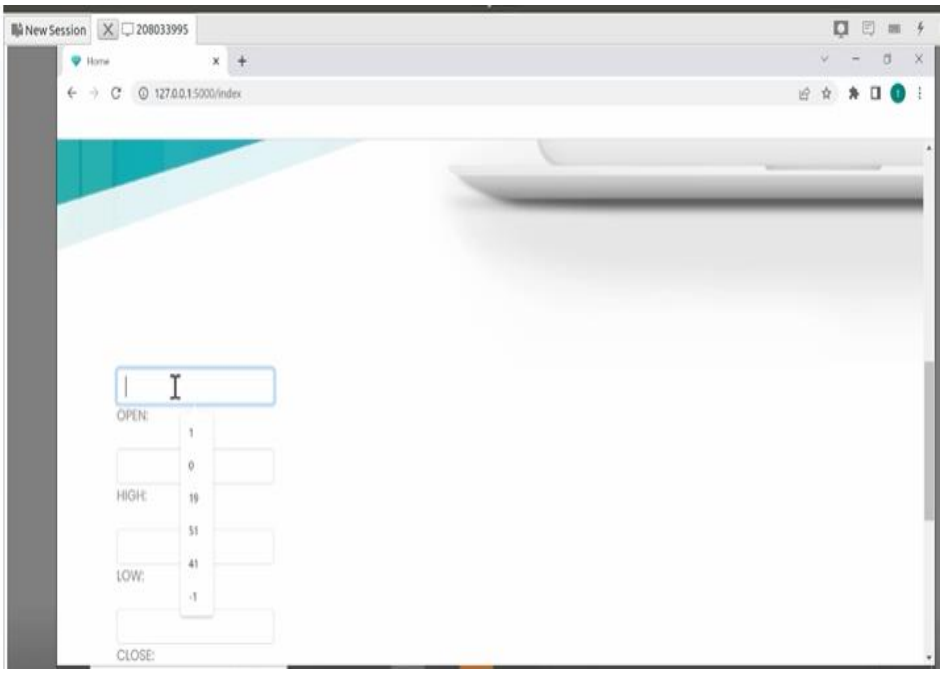

Figure 7.6: User Input

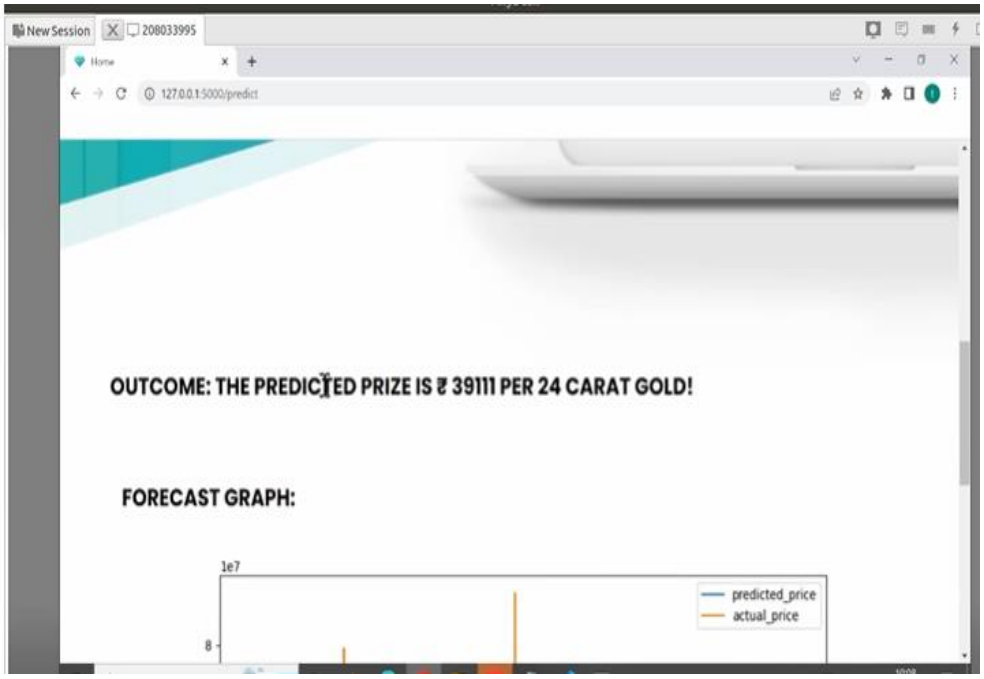

Figure 7.7: Prediction Result

8. CONCLUSION

8. CONCLUSION AND FEATURE SCOPE

This study was conducted to understand the relationship between gold price and selected factors influencing its price, namely stock market, crude oil price, rupee dollar exchange rate, inflation and interest rate. Monthly price data for the period January 2000 to December 2018 was used for the study. The data was further split into two periods, period I from January 2000 to October 2011 during which period the gold price exhibits a raising trend and period II from November 2011 to December 2018 where the gold price is showing a horizontal trend. Four machine learning algorithms, linear regression, random forest regression and gradient boosting regression and Voting classifier were used in analyzing these data. It is found that the correlation between the variables is strong during the period I and weak during period II. While these models show good fit with data during period I, the fitness is not good during the period II. Random forest regression is found to have better prediction accuracy for the entire period and gradient boosting regression 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. Further research with such data and different techniques may be conducted for better understanding of the performance of these techniques.

9. BIBILOGRAPHY

9. BIBILOGRAPHY

[1] W. Du and J. Schreger, "Local Currency Sovereign Risk," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2976788, Dec. 2013.

[2] J. Jagerson and S. W. Hansen, "All about investing in gold", McGraw-Hill Publishing, 2011.

[3] Z. Ismail, A. Yahya, and A. Shabri, "Forecasting gold prices using multiple linear regression method," Am. J. Appl. Sci., vol. 6, no. 8, p. 1509, 2009.

[4] C. Toraman, Ç. Basarir, and M. F. Bayramoglu, "Determination of factors affecting the price of gold: A study of MGARCH model," Bus. Econ. Res. J., vol. 2, no.4, p. 37, 2011.

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

[6] 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.

[7] 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.

[8] H. Naser, "Can Gold Investments Provide a Good Hedge Against Inflation? An Empirical Analysis," Int. J. Econ. Financ. Issues, vol. 7, no. 1, pp. 470–475, 2017.

[9] P. Khaemasunun, "Forecasting Thai gold prices," Available Http://www.Wbiconpro Com3-Pravit. Pdf Acess, vol. 2, 2014.

[10] S. M. Hammoudeh, Y. Yuan, M. McAleer, and M. A. Thompson, "Precious

metals–exchange rate volatility transmissions and hedging strategies," Int. Rev. Econ. Finance, vol. 19, no. 4, pp. 633–647, 2010.

[11] A. Han, K. K. Lai, S. Wang, and S. Xu, "An interval method for studying the relationship between the Australian dollar exchange rate and the gold price," J. Syst. Sci. Complex., vol. 25, no. 1, pp. 121–132, 2012.

[12] B. T. Ewing and F. Malik, "Volatility transmission between gold and oil futures under structural breaks," Int. Rev. Econ. Finance, vol. 25, pp. 113–121, 2013.

[13] D. Ghosh, E. J. Levin, P. Macmillan, and R. E. Wright, Gold as an inflation hedge?," Stud. Econ. Finance, vol. 22, no. 1, pp. 1–25, 2004.

[14] H. Mombeini and A. Yazdani-Chamzini, "Modeling gold price via artificial neural network," J. Econ. Bus. Manag., vol. 3, no. 7, pp. 699–703, 2015.

9.1 GITHUB LINK:

https://github.com/ChettiAbhinay/goldpricepredictionusingensemblebasedmachinelearning

10. PUBLICATION REPORT

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT

IJIEMR Transactions, online available on 17th Mar 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 04

10.48047/IJIEMR/V12/ISSUE 04/152

Title **Gold Price Prediction using Ensemble based Machine Learning Techniques**

Volume 12, ISSUE 04, Pages: 1187-1192

Paper Authors

Abhinay Chetti, K Anjali, Kommu Bindhu, Dr.Punyaban Patel

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

Gold Price Prediction using Ensemble based Machine Learning Techniques

[1]Abhinay Chetti, [2]K Anjali, [3]Kommu Bindhu [4] Dr.Punyaban Patel
[1,2,3] B. Tech Students, Department of Computer Science and Engineering,
CMR Technical Campus,Medchal, Hyderabad,Telangana, India,
[1]abhinayabhi0122@gmail.com, [2]katakamanjali3@gmail.com, [3]bindhukommu7755@gmail.com,
[4]Professor, Department of Computer Science and Engineering, CMR Technical Campus,
Medchal, Hyderabad, Telangana, India,
[4]punyabanpatel.cse@cmrtc.ac.in

ABSTRACT: This paper depends on research that was finished to become familiar with the connection between the cost of gold and a couple of significant variables, similar to the securities exchange, the cost of unrefined petroleum, the conversion scale of the rupee, expansion, and loan fees. Month to month cost measurements from January 2000 to December 2018 were utilized in the review. What's more, the information were isolated into two periods: period I, when the cost of gold rose from January 2000 to October 2011, and period II, when it tumbled from November 2011 to December 2018. These facts were broken down using four distinct ML calculations: linear regression, voting regression, gradient boosting regression, and random forest regression. There are significant areas of strength for a between the variables in period I and a powerless association in period II, it has been determined. During period I, these models fit the data well, but during period II, they don't. While gradient boosting regression is more accurate for each of the two time periods, random forest regression has a higher forecast accuracy over the entire time period.

Keywords *– Machine Learning, Regression, Prediction.*

1. INTRODUCTION

Investing and saving money are essential to everyone's existence. Ventures are the utilization of current assets determined to make a beneficial yield on them later on. In monetary terms, a venture is the securing of wares that are not used today yet will be utilized to create flourishing from now on. Speculation is the holding of a financial asset with the expectation that it will eventually generate income or that it will be sold at a higher price to gain access to the benefits that lie ahead. Quite possibly of the quickest developing economy on the planet is India, which has prompted more discretionary cashflow and various business amazing open doors. Buyers can choose from stocks, savings, commodities, real estate, and other forms of financial investment. Every one of them has one of a kind gamble and yield characteristics. Due to its rising value and wide range of applications, gold is another commodity that many investors consider to be an appealing business option. Investors' desire for gold as a defensive commodity is growing[1] as a result of their negative perceptions of the situation in established capital markets and foreign exchange markets. Gold is also referred to as "the asset of last resort," or the asset upon which buyers rely when industrialized global financial markets fail to deliver the desired profitability[2]. Because of this, buyers see gold as a way to hedge against market volatility.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

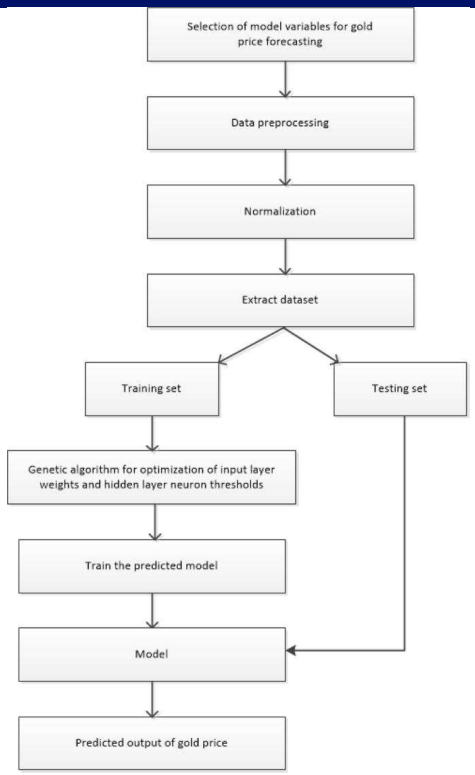

Fig.1: Example figure

As with any other product, the price of gold should be determined by supply and demand because it is a valuable commodity. However, this year's output has little effect on the stock's values because gold can be stored and accumulates over time. Gold is an item as well as a monetary instrument. Gold acts more like a drawn out instrument like stocks or ventures than an item. Various interconnected factors, including expansion rates, financial vacillations, and political agitation, decide the cost of gold [3].

2. LITERATURE REVIEW

Forecasting gold prices using multiple linear regression method:

Estimating is an administrative place that helps decide. The strategy for assessing in dubious planned conditions is one more name for it. Determining alludes to the assessment of time series or ceaseless information from a more extensive perspective. Gold was once utilized as money and is an important brilliant metal. It was broadcasted unlawful in the US quite a while ago, yet it is as of now being considered as an expected money. The interest for this ware is rising. The making of a projection model for assessing gold qualities utilizing financial factors like expansion, money related cost changes, etc was the target of this review. Financial backers are putting resources into gold in light of the devaluation of the US dollar since gold fills in as a monetary record stabilizer. It is fundamental to foster a model that both predicts the development of gold costs and mirrors the design and cadence of the gold market because of the developing interest for gold in Malaysia and different regions of the planet. The Multiple Linear Regression (MLR) model is the most effective strategy for dealing with gold's appreciation qualities. MLR research focuses on the connection between a single ward variable and at least one free factor, in this case, the gold cost. Gold costs will be anticipated utilizing the altered MLR model. An innocent model known as "figure 1" was utilized as a benchmark model to assess the model's prosperity. The price of gold is influenced by a number of factors, and "a hunch of specialists" have identified a number of monetary factors that affect gold qualities. factors like the future record from the Product Exploration Agency (CRB); the EUROUSD (the USD/Euro Exchange Rate); the Inflation Rate (INF); the Money Supply (M1); the New York Stock Exchange, or NYSE; the Standard and Poor's 500, or SPX; the Treasury Bill, or T-BILL; and that the USDX index was thought to have an impact on the value of the US dollar. To overview assumption precision, limit assessments for the MLR were performed including the Measurable Bundles for Sociology program (SPSS) with Mean Square Blunder (MSE) as the wellbeing capacity. Two renditions were then analyzed toward the end. The underlying model thought about all conceivable free factors. By making sense of 85.2% of the information distinctions in month to month gold qualities, the model gave off an impression of being helpful at guaging gold costs. The accompanying four autonomous variables were considered critical constantly model: CRB was first, EUROUSD was

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

first, INF was second, and M1 was second. The subsequent calculation performed better compared to the first when it came to estimate accuracy. With roughly 70% of the variety made sense of, the relapse results likewise give a method for assessing the general meaning of individual elements in the general estimate of gold cost.

Determination of factors affecting the price of gold: A study of MGARCH model:

The new ascent in gold rates has revived revenue, as have the impacts of the new monetary emergency. The distinguishing proof of the factors that impact gold rates is the essential target of this review. Month to month measurements from June 1992 to Walk 2010 are remembered for the review. The model considers the paces of oil, the US cash, expansion, and genuine loan costs. Gold costs and the US conversion scale have the most grounded negative connection, as per experimental discoveries. Second, there is a positive connection between's fuel costs and bullion.

The price of gold and the exchange rate:

The speculative examination of the primary trade rates and the costs of unfamiliar goods is the focus of this study. In the exploration segment, the contextual analysis of gold is examined using expectation mistake measurements. The world gold market is overpowered by the European money alliance, so the appreciation or depreciation of European financial norms basically influences the expense of gold in various money related principles. Different discoveries incorporate the way that since the Bretton Woods Global Financial Framework imploded, drifting trade rates among significant monetary forms have been a significant reason for cost insecurity.

Precious metals–exchange rate volatility transmissions and hedging strategies:

This study employs a complex framework to investigate the restrictive unpredictability, affiliation dependence, and relationship of the four most important metals—palladium, gold, and platinum—in light of current international affairs. The ramifications of the inexact discoveries for portfolio development and it are additionally analyzed to fence techniques. The four metals framework's discoveries show a great deal of short-and long haul dependence and relationship on news and past flimsiness. These outcomes stand apart much more when the government subsidizes rate and the conversion scale are considered. The instability of trade rates and significant wares are both straightforwardly impacted by financial arrangement. The outcomes are then used to show the best loads for a two-resource portfolio and the counterbalancing proportions for long property.

An interval method for studying the relationship between the Australian dollar exchange rate and the gold price:

This article proposes utilizing week after week, month to month, and quarterly information to research the connection between the gold cost and the conversion scale of the Australian dollar. To show how variables change, the span method utilizes middle example information. Mathematical methodologies are introduced as well as the ILS strategy's application to multi-model forecast. The ILS estimates precisely describe the connection between the swapping scale and the gold cost over the long and present moment, as confirmed by observational information. The distinction between the span and point strategies demonstrates that the disparity between the OLS and ILS gauges grows from week-to-week to quarterly data, with the least recurrence point information losing the most vacillation data.

3. METHODOLOGY

According to Lawrence[5], there were no significant correlations found between changes in macroeconomic variables like GDP and inflation. Additionally, he discovered that gold profits are less correlated with stock and credit benchmarks than profits from other metals. In any case, Baker and VanTassel [7] observed that the expense of gold is affected by arranged extension rates, and Sjaastad and Scacciavillani[6] observed that gold is a wellspring of huge worth against development. With respect to association between gold cost and expansion, Hanan Naser[8] accepts that previous examinations on the viability of gold as a shield against expansion are inconsistent in view of an investigation of writing.

Disadvantages:

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

1. However, gold purchases are risky given that the price of gold has recently fluctuated significantly.

The reason for this article is to research the connection between the cost of gold and different market and financial elements. Monetary controllers, theorists, store chiefs, and portfolio supervisors will all profit from having a superior comprehension of this association. Moreover, three ML calculations are utilized in this review to dissect these information: gradient boosting regression, random forest regression, and linear regression. We will actually want to decide the accuracy of these three methodologies under different circumstances by contrasting them.

Advantages:

1. improved precision in forecasting

2. We will be able to determine their precision in different situations by comparing these three methods.

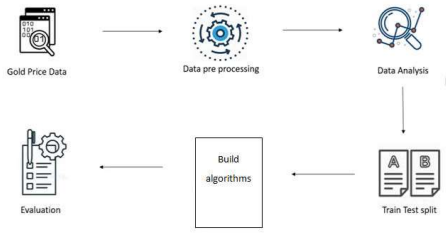

Fig.2: System architecture

MODULES:

- We created the modules that go with this project.
- Examining the information: Information will be entered into the framework using this module.
- Handling: We will review handling-relevant information using this module.
- Separation of data into training and testing: We will divide the information into train and test using this module.
- Making movies: Making the image: Linear regression, random regression, gradient-boosting regression, and a voting classifier (RF plus RF).
- Client register and login: Registration and authentication will take place as a result of using this module.
- User input: Forecast input will result from using this module.
- Prediction: The displayed final predicted value

4. IMPLEMENTATION

Linear Regression - Linear Regression is a procedure that utilizes a straight connection between a free factor and a reliant variable to make forecasts about what will occur from here on out. It is a factual strategy used in information science and ML for estimate examination.

Random Forest Regression - An Random Forest methodology is a coordinated ML procedure that is for the most part used in ML for Request and regression issues. We are aware that there are a lot of trees in a forest, and the more trees there are, the stronger the forest is.

Gradient Boosting Regression - A type of machine learning boosting is gradient boosting. It is predicated on the idea that when the best subsequent model is combined with models that came before it, the total forecast error is minimized. The crucial idea is to define the desired outcomes for the subsequent model to reduce error.

Voting Classifier (RF + RF): A voting classifier is an ML assessor that conjectures utilizing the consequences of different base models or assessors it trains. For every assessor yield, totaling models can be joined with casting a voting choices.

5. EXPERIMENTAL RESULTS

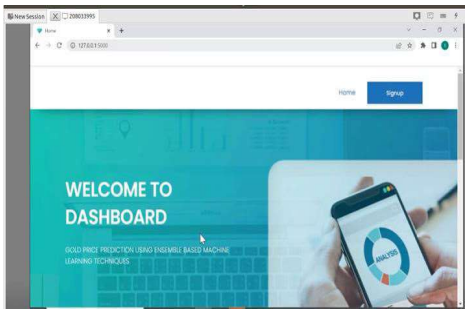

Fig.3: Home screen

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

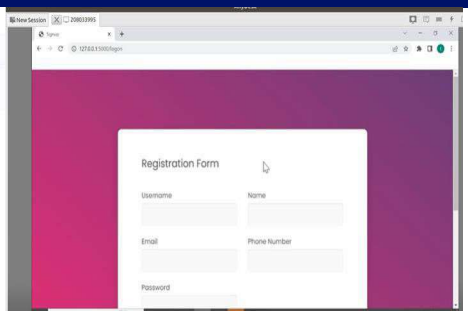

Fig.4: User registration

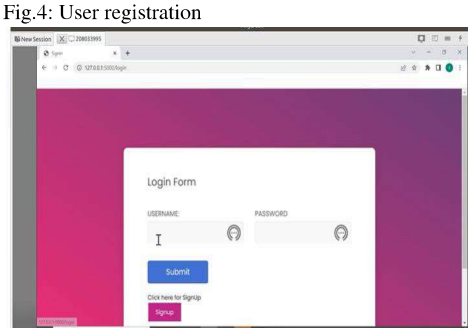

Fig.5: User login

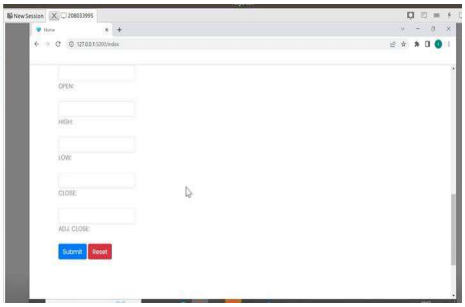

Fig.6: Main page


Fig.7: User input

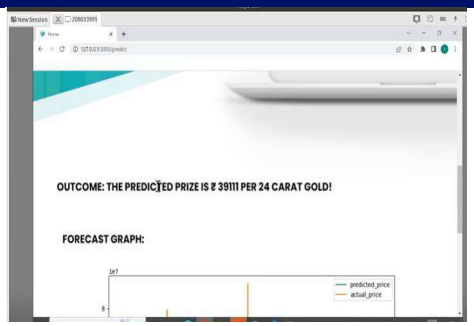

Fig.8: Prediction result

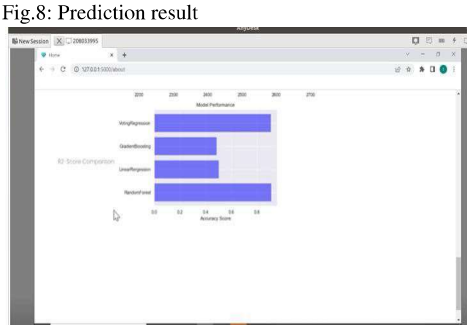
Fig.9: Accuracy graph

6. CONCLUSION

The objective of this study was to dive more deeply into the connection between the cost of gold and the securities exchange, the cost of oil, the rupee-dollar conversion standard, expansion, and loan fee. Month to month cost measurements from January 2000 to December 2018 were utilized in the review. In a similar vein, the data were separated into two distinct periods: The price of gold rose during period I, which ran from January 2000 to October 2011; The price of gold decreased during period II, which began in November 2011 and ended in December 2018. These facts were broken down using four distinct ML calculations: voting classifier, linear regression, gradient boosting regression, and random forest regression. It has been resolved that there is serious areas of strength for a between the variables in period I and a powerless association in period II. During period I, these models fit the information well, yet during period II, they don't. While gradient boosting regression is better for every one of the double cross time frames, random forest regression has better conjecture accuracy for the whole time. It has been

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

determined that these kinds of analyses benefit greatly from machine learning techniques, but the data's characteristics affect how precise they are. To get a better understanding of how these methods work, additional research using such data and various methods may be carried out.

REFERENCES

[1] W. Du and J. Schreger, "Local Currency Sovereign Risk," Social Science Research Network, Rochester, NY, SSRN Scholarly Paper ID 2976788, Dec. 2013.

[2] J. Jagerson and S. W. Hansen, "All about investing in gold", McGraw-Hill Publishing, 2011.

[3] Z. Ismail, A. Yahya, and A. Shabri, "Forecasting gold prices using multiple linear regression method," Am. J. Appl. Sci., vol. 6, no. 8, p. 1509, 2009.

[4] C. Toraman, Ç. Basarir, and M. F. Bayramoglu, "Determination of factors affecting the price of gold: A study of MGARCH model," Bus. Econ. Res. J., vol. 2, no.4, p. 37, 2011.

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

[6] 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.

[7] 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.

[8] H. Naser, "Can Gold Investments Provide a Good Hedge Against Inflation? An Empirical Analysis," Int. J. Econ. Financ. Issues, vol. 7, no. 1, pp. 470–475, 2017.

[9] P. Khaemasunun, "Forecasting Thai gold prices," Available Http://www Wbiconpro Com3-Pravit. Pdf Acess, vol. 2, 2014.

[10] S. M. Hammoudeh, Y. Yuan, M. McAleer, and M. A. Thompson, "Precious metals–exchange rate volatility transmissions and hedging strategies," Int. Rev. Econ. Finance, vol. 19, no. 4, pp. 633–647, 2010.

[11] A. Han, K. K. Lai, S. Wang, and S. Xu, "An interval method for studying the relationship between the Australian dollar exchange rate and the gold price," J. Syst. Sci. Complex., vol. 25, no. 1, pp. 121–132, 2012.

[12] B. T. Ewing and F. Malik, "Volatility transmission between gold and oil futures under structural breaks," Int. Rev. Econ. Finance, vol. 25, pp. 113–121, 2013.

[13] D. Ghosh, E. J. Levin, P. Macmillan, and R. E. Wright, Gold as an inflation hedge?," Stud. Econ. Finance, vol. 22, no. 1, pp. 1–25, 2004.

[14] H. Mombeini and A. Yazdani-Chamzini, "Modeling gold price via artificial neural network," J. Econ. Bus. Manag., vol. 3, no. 7, pp. 699–703, 2015.

Abhinay chetti is currently pursuing B. Tech final year in the stream of Computer Science and Engineering in CMR Technical Campus, Medchal, Hyderabad, Telangana, India.

K Anjali is currently pursuing B. Tech final year in the stream of Computer Science and Engineering in CMR Technical Campus, Medchal, Hyderabad, Telangana, India.

Kommu Bindhu is currently pursuing B. Tech final year in the stream of Computer Science and Engineering in CMR Technical Campus, Medchal, Hyderabad, Telangana,India.

Mr. Dr.Punyaban Patel is currently working as Professor in the department of Computer Science and Engineering, CMR Technical Campus, Medchal Hyderabad.

11. PUBLICATION CERTIFICATE

International Journal for Innovative Engineering and Management Research

IMPACT FACTOR 7.812

(ISSN 2456 - 5083)

ELSEVIER SSRN

CERTIFICATE

DOI: 10.48047/IJIEMR/V12/I04/152

This is to Certify that Prof./Dr./Mr./ Dr.Punyaban Patel From CMR Technical Campus, Hyderabad, Telangana, India.Participated in the International Journal of Engineering and Management Research.Presented a Paper Entitled"Gold Price Prediction using Ensemble based Machine Learning Techniques"In the Organizing Commitee of ROYAL SOCIETY FOR SCIENCE.

Published in International Journal of Innovation Engineering and Management Research (IJIEMR), Vol-12,Issue-04 Apr-2023

Hyderabad
2023

RAKESH ORUGANTI
Editor In Chief

+91 9989292561 Huda Techo Enclave, HiTech City Madhapur, Hyderabad, 500081 info.ijiemr@gmail.com

International Journal for Innovative Engineering and Management Research

IMPACT FACTOR 7.812

(ISSN 2456 - 5083)

ELSEVIER SSRN

CERTIFICATE

DOI: 10.48047/IJIEMR/V12/I04/152

This is to Certify that Prof./Dr./Mr./ K Anjali From CMR Technical Campus, Hyderabad, Telangana, India.Participated in the International Journal of Engineering and Management Research.Presented a Paper Entitled"Gold Price Prediction using Ensemble based Machine Learning Techniques"In the Organizing Commitee of ROYAL SOCIETY FOR SCIENCE.

Published in International Journal of Innovation Engineering and Management Research (IJIEMR), Vol-12,Issue-04 Apr-2023

Hyderabad
2023

RAKESH ORUGANTI
Editor In Chief

+91 9989292561 Huda Techo Enclave, HiTech City Madhapur, Hyderabad, 500081 info.ijiemr@gmail.com

IMPACT FACTOR 7.812

International Journal for Innovative Engineering and Management Research

(ISSN 2456 - 5083)

CERTIFICATE

ELSEVIER SSRN

DOI: 10.48047/IJIEMR/V12/I04/152

This is to Certify that Prof./Dr./Mr./ Kommu Bindhu From CMR Technical Campus, Hyderabad, Telangana, India.Participated in the International Journal of Engineering and Management Research.Presented a Paper Entitled"Gold Price Prediction using Ensemble based Machine Learning Techniques"In the Organizing Commitee of ROYAL SOCIETY FOR SCIENCE.

Published in International Journal of Innovation Engineering and Management Research (IJIEMR), Vol-12,Issue-04 Apr-2023

Hyderabad
2023

RAKESH ORUGANTI
Editor In Chief

+91 9989292561 Huda Techo Enclave, HiTech City Madhapur, Hyderabad, 500081 info.ijiemr@gmail.com

International Journal for Innovative Engineering and Management Research

IMPACT FACTOR 7.812

(ISSN 2456 - 5083)

ELSEVIER SSRN

CERTIFICATE

DOI: 10.48047/IJIEMR/V12/I04/152

This is to Certify that Prof./Dr./Mr./ Abhinay Chetti From CMR Technical Campus, Hyderabad, Telangana, India.Participated in the International Journal of Engineering and Management Research.Presented a Paper Entitled"Gold Price Prediction using Ensemble based Machine Learning Techniques"In the Organizing Commitee of ROYAL SOCIETY FOR SCIENCE.

Published in International Journal of Innovation Engineering and Management Research (IJIEMR), Vol-12,Issue-04 Apr-2023

Hyderabad
2023

RAKESH ORUGANTI
Editor In Chief

+91 9989292561 Huda Techo Enclave, HiTech City Madhapur, Hyderabad, 500081 info.ijiemr@gmail.com