Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

Prediction of Potential Gold Prices using Machine Learning Approach

Aruna S [1], Umamaheswari P [2], Sujipriya J [3], Dr.R.Umamaheswari [4]

[1]Associate Professor, Department of Master of Computer Applications,
Sona College of Technology, Salem-636005, Tamil Nadu, India

[2]Assistant Professor of Statistics, Department of Mathematics,
Sona College of Arts and Science, Salem-636005, Tamil Nadu, India

[3]Assistant Professor, Department of Master of Computer Applications,
Sona College of Technology, Salem-636005, Tamil Nadu, India

[4] Professor, Department of Computer Science and Engineering,
Gnanamani College of Technology, Namakkal.

Abstract—In comparison to other forms of payment, gold has traditionally been used to facilitate trading transactions around the world. Various states retained their gold deposits and strengthened them and were known as rich and progressive states. Currently, precious metals such as gold are kept by the central banks of all nations to ensure the redemption of international loans, as well as to manage inflation. Besides, it also represents the country's financial resilience. Various multi- national corporations and entities have also invested in gold deposits, in addition to government departments. Gold is often offered as gifts/souvenirs in cultural activities in Asian countries and at weddings, gold ornaments are presented in India, Pakistan, and other countries as Dowry. In addition to the demand and availability of goods on the market, gold prices are also heavily affected by the success of the world's leading economies. Centered on 21 market factors using machine learning techniques, we predicted potential gold prices. The findings demonstrate that we can quite accurately forecast daily gold prices. For investors and central banks to determine when to invest in this asset, our forecast models would be useful.

***Keywords*—**Gold rate prediction, ARIMA Model, linear regression, neural networks.

I. INTRODUCTION

Historically, in different parts of the world, including India, gold has been used as a type of currency [5]. Gold has also retained its importance in modern years and has been used as a way to measure a country's financial ability. Big investors were also drawn to this precious metal and spent massive quantities of it. Emerging world economies such as China, Russia, and India have lately become major buyers of gold, while the United States, South Africa, and Australia are among the major sellers of gold [8]. The price of gold is also influenced by traditional Chineseand Indian activities. More capital is spent on the purchase of this product at the time. Instead of alternative investment options, small investors often find this product for secure investment, which carries in- built investment risks. To set spot prices for gold, the internal financial conditions of the aforementioned countries play an important role. Government gold reserves are primarily dictated by their financial situations and interest rates, as they are indicators of their economy's strength. More economic activity is seen in the

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

US as US interest rates are lower, so capital inflows are seen in the gold sector. Similarly, when China's interest rates fell from 5.31 (2010) to 4.35 (2016), it actively purchased gold [8].

Global investors prefer to invest elsewhere, whether in countries or giant firms if they expect a major fall in gold prices. Some investors, in such a situation, turn to some other source of investing, such as US bonds or the stock exchange. Figure. 1 shows that while gold prices are poor, the New York Stock Exchange (NYSE) and S&P 500 appear to do well. The S&P 500 is an American stock exchange benchmark focused on the NYSE or NASDAQ market capitalizations of 500 major firms. This suggests that the movement of money from stock markets to the gold market has been noted. Any stakeholders, on the other hand, convert their gold reserves to USD, so the Euro USD index (Euro-USD exchange rate) continues to increase as gold prices fall. The valuation of the USD itself depends on different variables, including the interest rates set by the US government. The strength of the US economy is also demonstrated by the success of leading stock exchanges like NASDAQ and Dow Jones. Therefore, different phenomena are associated with gold prices and therefore influence the price.

The spot price is the prevailing selling price at which goods for immediate payment and distribution are bought or sold. That is distinct from the price of the future, which is the price at which the two sides agree on the future date of the deal. Based on supply and demand in the gold market, gold spot prices are determined twice a day. Fractional fluctuations in the price of gold will lead to tremendous gains or losses for both these holders and government banks. Increased and decreased forecasting of regular gold prices will help investors determine whether to purchase (or sell) the metal.

Researchers have carried out numerous studies to predict gold prices, many of them informative in their own right. In this report, we forecast gold prices using a) the most detailed set of characteristics than any of the previous research, which includes the performance measures of the Russian, Chinese, and Indian economies for the first time (as they are the largest buyer of gold) and the stock price of leading gold producing / trading firms, and b) apply various machine learning algorithms for forecasting. We also recognize which features most affect gold prices, some of which have not been seen before.

The remainder of the paper is structured as follows: the associated research that has been performed in this issue area is covered in Section II. We explain our data collection process in Section III, and the different attributes that we have used. In Section IV, findings are discussed. In Section V, we eventually conclude.

Figure.1 Effect of index prices on gold rates

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

II. RELATED WORK

In [1], the authors address the effect of the US dollar on the foreign market environment of crude oil and gold prices. They also examine the US dollar's impact over the 1996-2009 timeframe in terms of bond funds, stock markets, Federal Reserve interest rates, inflation, and economic decline. They take into account big developments around the world that could have influenced US dollar prices. Authors use the easiest approach in [2] to forecast potential gold prices since they do not take into account any characteristics that may affect gold rates directly or indirectly. Instead, they use only five features extracted from the gold prices themselves. The opening, close, highest, and lowest price of gold on a given day and the value of the asset that was sold on that day are these five characteristics. For the prediction of gold values, they use decision trees and help vector regression algorithms, but do not announce any findings. On the other hand, [6] make a very detailed analytical study of seventeen various time series modelling approaches to gold prices, and conclude that the best solution is the random walk approach. The shortcoming of the analysis is that only a few variables are used as input variables to the models, such as the price of other precious metals (palladium, platinum, etc.). They should not take the economic conditions of large economies or businesses processing gold into account.

In [1], the author uses text mining and artificial neural networks (ANN) to forecast the gold prices and compare their results with the autoregressive-moving average (ARMA) model. ARMA model is the most frequently used statistical model for analyzing time-series data. ARMA model consists of two parts, the first part AR, involves regressing the variable on its past values. The second part MA involves modeling the error term as a linear combination of error terms occurring simultaneously and at various times in the past. In [3] also, the author uses the ARMA model for predicting gold rates but uses monthly rates of gold for the past 124 months. They forecast actual gold prices and achieve an accuracy of 66.67%. In [10] also, the author uses the ARMA model but compares their results with ANN and show that ANN performs better than ARMA. For training, they used data from 1990 to 2006, while data from 2006 to 2008 was used. For performance measurement, Coefficient of Determination (R2), mean absolute error (MAE) and root mean square error was used for research. For sensitivity analysis, the Cosine Amplitude Methods (CAM) test was also performed to determine the relationship between relevant parameters.

The author uses the algorithm of extreme learning machines (ELM) in [4], a modification of ANN. They equate the effects of ELM without feedback, with backpropagation, radial base function, and ELMAN networks with the feed-forward neural network. They conclude that, with 93.82% accuracy, ELM performs the highest.

Gold, silver and crude oil prices are the factors considered by them. For the planning of their model, they also regard the Regular and Weak (S&P) 500 index and foreign exchange rate. In [5], the writers take economic variables such as inflation, currency values, the success of stock markets, etc. into account to forecast gold rates. They use multiple linear regression (MLR) models based on eight independent variables to predict gold prices. They argue that the Thomson Reuters Main Product (CRB) Index, EURUSD exchange rate, inflation rates, and money supply index (MI) are the most influential parameters. To remove associated error terms, the Praise-Winsten technique was used. They reached 96.92 percent precision using only these four attributes. Authors use the Logistic Regression (LR) model in [9] and have achieved 63.76 percent accuracy, 63.89 percent recall, and 61.92 percent accuracy using eight

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

years of results. LR outperforms the SVM, they conclude.

III. METHODOLOGY

The attributes considered by the studies mentioned in Section II are listed in Table II. The column called 'Proposed' lists the attributes used by us to create the models. Our attribute list is the most detailed and brings Russia, China, and India's success metrics into account for the first time since they are the main gold buyers. We do so because, among other countries, gold prices are continually changing due to the financial situations in those countries, such as the United States, the United Kingdom, China, and Russia [7]. Their financial power helps them to spend more in gold, and they sell their gold reserves to boost their currency while their economy gets fragile. Secondly, the stock values of the big gold trading firms are also taken into account.

A. Dataset

Data for this analysis was obtained from different sources from January 2005 to September 2019. Information was obtained for attributes such as Oil Price, NYSE, Standard and Poor's (S&P) 500 index, US Bond prices (10 years), Euro USD exchange rates. It has also gathered data from several central government banks and five major businesses that have spent vast quantities of gold. The price of precious metals will also be included in the measurement during this time. The online sources from which this data was collected are described in Table I. All these attributes are mentioned in Table II & III.

In machine learning approach, it is necessary to train the algorithm and to automate the process in order to make the predictions possible. In this proposed work, we have used PyCARET (Classification and Regression Training) package in Python programming software. The PyCAROT package consists of a set of functions that attempt to streamline the Process for creating predictive models. The package contains tools for data splitting, pre-processing, and feature selections.

The price of the gold we are trying to predict is in US dollars. On the dataset, a lot of cleaning and pre-processing was carried out. To complete the dataset, the question of missing values was treated acceptably.

Regularly, gold prices adjust and are also influenced by big world events. As seen in Figure-1., existing gold prices are far higher than a couple of years ago. It was planned to split the dataset in a sequential manner instead of random sampling, bearing in mind the large price gap. The most recent 25% data is then used as the test range and the earliest 75% data is used for training.

Thus, the training set constitutes the first 2295 records, while the evaluation set comprises the last 770 rows.The latest historical evidence would be more representative of the upcoming trend due to large variations in gold prices over the years. We further split the training set into four versions, thus. The first edition includes all documents from 0 percent to 75 percent, the second version consists of 15 percent to 75 percent records, the third version consists of 30 percent to 75 percent records, and 45 percent to 75 percent of the entire data in the last version.

B. Correlation Analysis

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

To decide which of the twenty-two qualities we obtain are strongly correlated to the price of gold, correlation research was carried out. Fig. 3 indicates the results of the study of the correlation. It offers a few useful observations. The success of the US (or any other major) economy or the price of other precious metals is not the attribute that has the greatest correlation with gold prices, but it is the stock price of Silver Wheaton Corporation (SLW), the world's largest streaming business for precious metals. The tenth and eleventh most associated characteristics include other major gold miners, Eldorado Gold Corporation, and Compania de Minas Buenaventura. This is the first research to estimate the price of gold using the prices of the main gold producers (shown in Table II).

The values of precious metals (such as silver) and output indices of major world economies such as the US and UK are the most associated characteristics, led by SLW, as predicted. A surprise is Russia's seventh interest rate, the first to be included in any gold price forecasting report. China's interest rate, on the other hand, does not have a significant impact on the price of gold.

C. Machine Learning Models

Two models of ML, namely neural networks and linear regression, are used. Neural networks are a class of models inspired by biological neural networks that are used to simulate functions that can depend on a broad number of inputs, also known as artificial neural networks (ANN). They consist of one or more hidden layers of neurons, in addition to the input and output layers that try to learn non-linear decision boundaries that distinguish different data groups. It can also be used in our case to forecast continually valued characteristics such as gold values. Figure.4 represents an ANN sample.

Linear regression (LR) is an approach used in statistics to model the relationship between one or more independent variables (attributes) and a dependent (class variable) relationship. For predicting constant value properties, linear regression may be used. We use the LR and ANN implementations supported by the CAROT package. Using the RMSE (Root Mean Square Error) measure of efficiency, both models are optimized.

Table I. Sources of Data Collection

S.No	Data	Source
1	US Bond Rates	www.treasury.gov
2	EuroUSD Index	www.finance.yahoo.com
3	NYSE Index	www.finance.yahoo.com
4	EGO Index	www.finance.google.com
5	SLW Index	www.finance.google.com
6	AU Index	www.finance.google.com
7	ABX	www.finance.google.com
8	BVN	www.finance.google.com
9	Interest Rate China	www.quandl.com
10	Interest Rate USA	www.treasury.gov.
11	Interest Rate UK	www.global-rates.com
12	Interest Rate Russia	www.global-rates.com
13	Gold Spot Rates	www.gold.org

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

14	Silver Spot Rates	www.perthmint.com.au
15	Platinum Spot Rates	www.perthmint.com.au
16	Palladium Spot Rates	www.perthmint.com.au
17	Rhodium Spot Rates	www.markets.businessinsider.com
18	Oil Spot Prices	www.investing.com
19	S&P500 Index	www.finance.yahoo.com

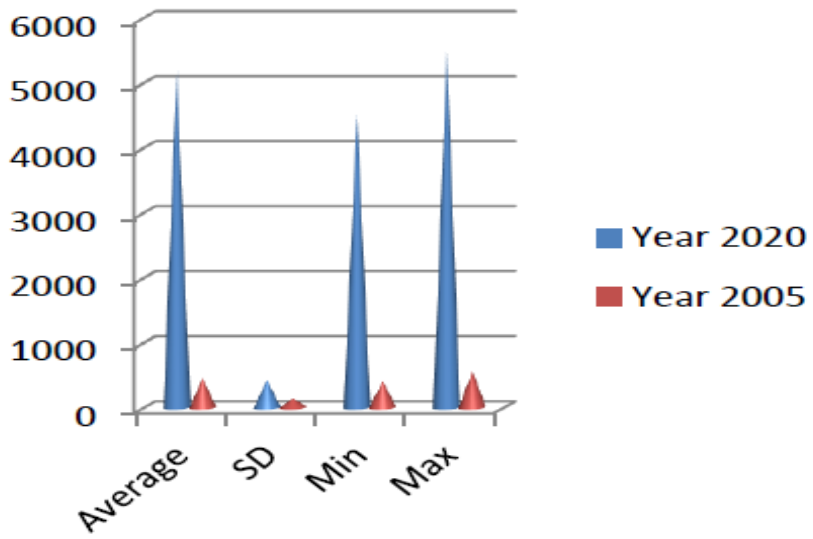


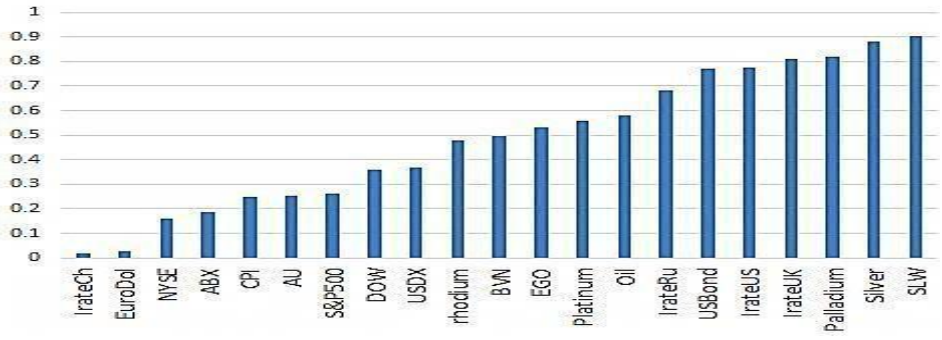
Figure. 2 The contrast in prices of gold

Table II. Attributes used for predicting gold rates

Variable	[1]	[4]	[5]	[6]	[9]	[10]	Proposed
Time Period	2000-2014	2003	1972-2013	2005-2013	1990-2008	2005-2012	2020-2025
Oil Spot Price	✓	✓				✓	✓
Oil Future Price	✓				✓		
Gold Spot Price	✓	✓	✓	✓	✓	✓	✓
Gold Future Price					✓		
S&P 500 Index	✓	✓	✓		✓		✓
Silver Price	✓	✓		✓		✓	✓
Silver Future Price					✓		
Copper Future Price					✓		
Platinum Price	✓			✓			✓
US Dollar Index	✓	✓	✓		✓	✓	✓
Base Interest Rate	✓						
Consumer Price Index (CPI)	✓						✓
Inflation	✓					✓	
Gold Production						✓	
US Bond (10 Years)	✓						✓
US Bond (5 Years)							

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

Dow Jones Index	✓				✓		✓
Commodity Research Bureau			✓				
(CRB)							
EUROUSD Rate			✓		✓		✓
Money Supply (M1)			✓				
NYSE Index			✓				✓
Treasury Bill			✓				
Palladium Price				✓			✓
Rhodium Price				✓			✓
Interest Rate (US)							✓
Interest Rate (UK)							✓
Interest Rate (China)							✓
Interest Rate (Russia)							✓
Eldorado Gold Corp. (EGO)							✓
Silver Wheaton Corp. (SLW)							✓
AngloGold Ashanti Limited							✓
(AU)							
Barrick Gold Corp. (ABX)							✓
Campania de Minas							✓
Buenaventura (BVN)							
NASDAQ					✓	✓	
FTSE					✓	✓	
Hang Seng Index					✓		
Nikkei Index					✓		
GBP-USD					✓		
JPY-USD					✓		
CNY-USD					✓		
Shanghai Composite					✓		
KOSPI Composite					✓		
Ibo Vespa					✓	✓	
DAX					✓	✓	
Euronext 100					✓		



Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

Figure.3 Correlation analysis.

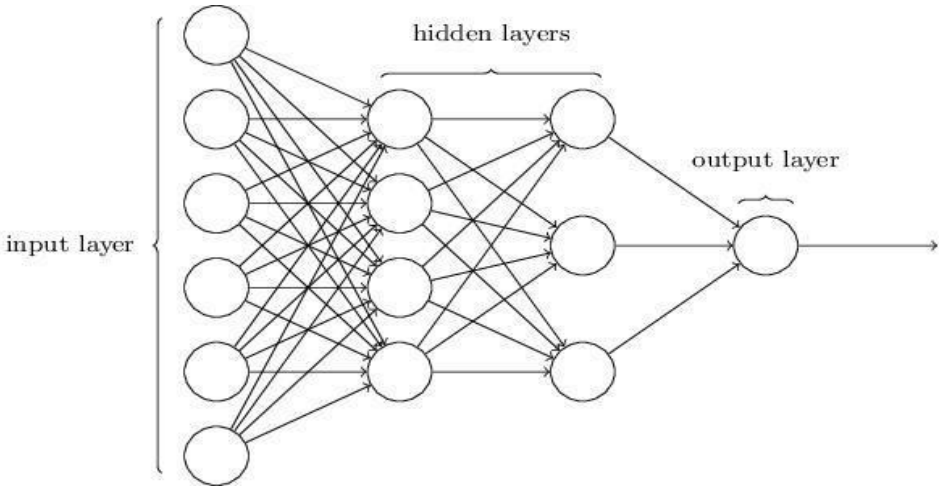


Figure.4 Artificial Neural Network

IV. RESULTS

ANN has different parameters for tuning. Most of them were experimented with and two parameters were found, namely the number of layers and the learning rate, to be the ones with the greatest effect on their results.

Consequently, Fig. 5(a)-(d) shall signify the consequence of the application of ANN to the test set, although the values of these two parameters differ from the training set. The four figures refer to the four (Section III) variations of the training set. Likewise, Fig. 6 indicates the output of LR when the ridge parameter varies. The findings are very promising. Fig. 5(d) reveals that the root mean square error is as low as 19 when using as little as 920 days of data (i.e. 45-75 percent) for preparation. Even though the average gold prices in the test results are over (Fig. 2), this is an exceptionally low mistake. Although LR's output is lower than ANN's, the difference is not important LR has the advantage of a preparation time that is shorter than ANN.

Usually, when greater training data is used, the efficiency improves for ML algorithms, although it is important to note that both classifiers perform better when smaller training data (i.e. 45-75% training set version) is used. The theory behind this phenomenon is that the broader training data (i.e. 0-75% training data version) includes records of gold rates back in 2005, although in the last few years the gold rates have changed dramatically. Using just recent history is useful. The preparation data of 45-75percent consists of just a little over two and a half years of data before the data in the test collection.

Fig. 5(a)-(d) also gives some insight into how ANN functions. When smaller training sets are used, ANN with two layers of neurons performs better, while ANN with five layers of neurons performs best when a wider training set is used. This suggests that ANN, with five layers for a smaller training set, thus over fitted the 45-75 percent training set with its relative efficiency, and thus increased its relative performance with the rise in the training set scale. As for ANN with two layers, with the rise in size of the training package, it was reduced under fitting. Whereas, as it has the most stable output, ANN with three layer suits the data

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

better.

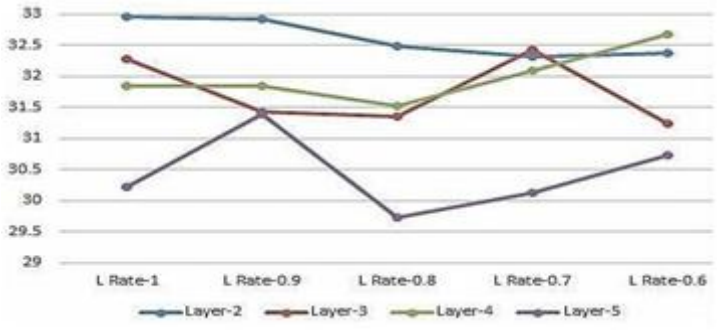


Figure. 5(a)

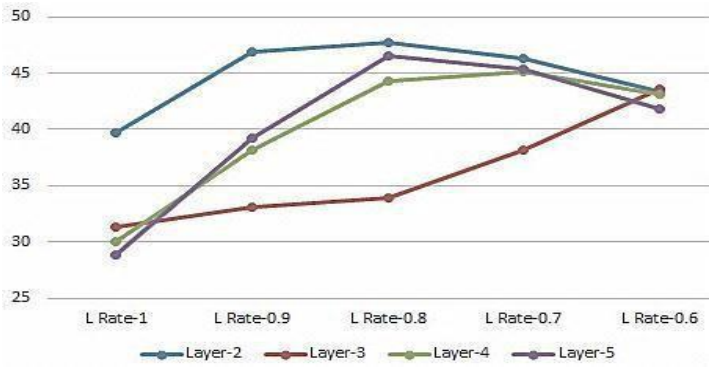


Figure. 5(b)

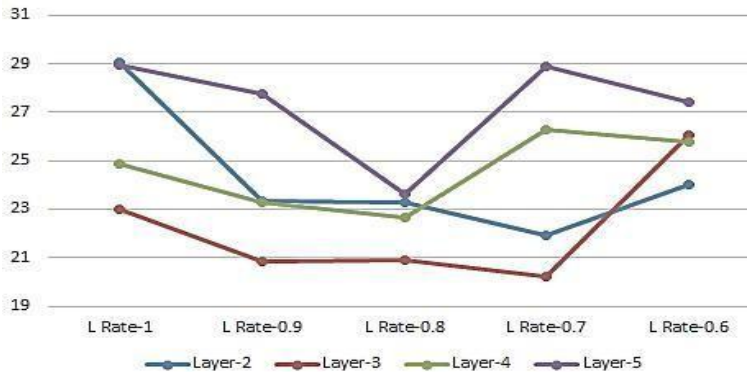


Figure. 5(c)

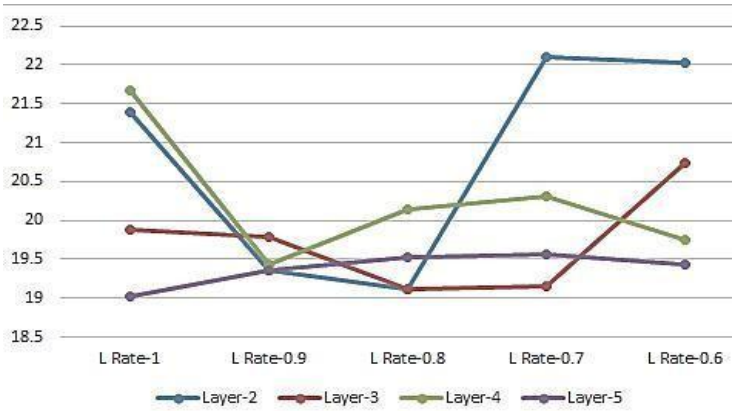


Figure. 5(d)

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

Figure. 5. (a) ANN error chart on 0-75% training dataset; (b) ANN error chart on 15-75% training dataset; (c) ANN error chart on 30 to 75% training dataset;(d) ANN error chart on 45 to 75% training dataset.

Pseudocode for Gold Price Prediction using Python

```
#Importing Libraries
import pandas as pd
from datetime import datetime import matplotlib.pyplot as plt
from googletrends import googletrends

ticker_details = pd.read_excel("tickerlist.xlsx") ticker = ticker_details['Ticker'].to_list()
names = ticker_details['Description'].to_list()


#Preparing Date Range
end_date= datetime.strftime(datetime.today(),'%Y-%m-
%d')
start_date = "2019-01-01" date_range =
pd.bdate_range(start=start_date,end=end_date) values = pd.DataFrame({ 'Date': date_range})
values['Date']= pd.to_datetime(values['Date'])
     values = values.drop(labels='Date1',axis=1)

#Extracting Data from Google Trendsand Adding them to Values table using date as key
for i in ticker:
raw_data = googletrends(i) raw_data =
raw_data.get_historical_price_data(start_date, end_date, "daily")
   df = pd.DataFrame(raw_data[i]['prices'])[['formatted_date','adj close']]
df.columns = ['Date1',i]
df['Date1']= pd.to_datetime(df['Date1']) values =
values.merge(df,how='left',left_on='Date',right_on='Date 1')


#Filling the NaN values in the data set values = values.fillna(method="ffill",axis=0) values =
values.fillna(method="bfill",axis=0)

# Co-ercing numeric type to all columns except Date
cols=values.columns.drop('Date') values[cols] =
values[cols].apply(pd.to_numeric,errors='coerce').round( decimals=1)
imp = ['Gold','Silver', 'Crude Oil', 'S&P500','MSCI EM ETF']

# Calculating Short term -Historical Returns change_days = [1,3,5,14,21]data =
pd.DataFrame(data=values['Date'])
for i in change_days:
   x= values[cols].pct_change(periods=i).add_suffix("-T- "+str(i))
data=pd.concat(objs=(data,x),axis=1) x=[]

# Calculating Long term Historical Returns
change_days = [60,90,180,250]for i in change_days:
   x= values[imp].pct_change(periods=i).add_suffix("-T- "+str(i))
data=pd.concat(objs=(data,x),axis=1) x=[]
```

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

The proposed pseudocode also addresses the calculation of moving averages of gold, merging average values to the feature space and once the data is prepared, the model is loaded (ie., executed) for making prediction on the new data and gives the historic prices, projected return and forecasted prices in 3 weeks each date in the dataset.

V. CONCLUSION

In history, gold has been one of the most significant resources. To maintain the world's existing economic system, the preservation of gold reserves by central banks is vital. A large sum of capital is still spent in gold by several big businesses and investors. Predicting the rate of gold, while not straightforward, would help investors and central banks to determine when to sell and buy it, thus maximizing their income. We used machine learning algorithms to very accurately estimate the gold rates in this analysis. Our analysis is perhaps the most detailed to date, taking numerous economic metrics from different countries and businesses into account. It is the first time that major gold trading / producing companies ' market value, and Russia's interest rates, have been effectively used as an indicator for gold rate forecasts. On the opposite, we illustrate that a big company's market valuation has a greater impact on gold prices than the US economy. In the future, by using ensemble learning and deep learning, we expect to boost our performance.

REFERENCES

[1] M. C. &. C.-C. C. Hsin-Hung Chen, "The Integration of Artificial Neural Networks and Text Mining to Forecast Gold Futures Prices," Communications in Statistics - Simulation and Computation, pp. 1532- 414, 2016.

[2] D. G. Navin, "Big Data Analytics for Gold Price Forecasting Based on Decision Tree Algorithm and Support Vector," International Journal ofScience and Research, pp. 2026-2030, 2015.

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

[4] M. S. S. S. S. Kumar Chandar, "Forecasting Gold Prices Based on Extreme Learning Machine," International Journal of ComputersCommunications & Control, pp. 372-380, 2016.

[5] A. Y. A. S. Z. Ismail, "Forecasting Gold Prices Using Multiple Linear Regression Method," American Journal of Applied Sciences, pp. 1509-1514, 2009.

[6] Hassani, H., Silva, E. S., Gupta, R., & Segnon, M. K. (2015). Forecasting the price of gold. Applied Economics, 47(39), 4141-4152.

[7] I. S. G. a. A. Mody, "International Gold Price Movements, 1972- 1982," Economic and Political Weekly, vol. Vol. 17, no. No. 46/47, pp. 1861-1870, 1982.

[8] H. Long, "http://money.cnn.com," cnn.com, 10 february 2016.

[9] P. V. M. V. P. G. M. P. Sima P Patel, "Gold Market Analyzer using Selection based Algorithm," International Journal of AdvancedEngineering Research and Science (IJAERS), vol. 3, no. 4, pp. 55-102,2016.

[10] A. Y.-C. Hossein Mombeini, "Modelling Gold Price via Artificial Neural Network," Journal of Economics, Business and Management, vol. 3, no. 7, pp. 699-703, 2015.

[11] Usmani, Mehak, et al. "Stock market prediction using machine learning techniques." 2016 3rd international conference on computer and information sciences (ICCOINS). IEEE, 2016.

[12] Mallqui, Dennys CA, and Ricardo AS Fernandes. "Predicting the direction, maximum,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 1385 - 1396
Received 05 March 2021; Accepted 01 April 2021.

minimum and closing prices of daily Bitcoin exchange rate using machine learning techniques." Applied Soft Computing 75 (2019): 596-606.

[13] ur Sami, I., and Khurum Nazir Junejo. "Predicting Future Gold Rates using Machine Learning Approach." International Journal of Advanced Computer Science and Applications 8.12 (2017): 92-99.

[14] Chen, Wei, et al. "Machine learning model for Bitcoin exchange rate prediction using economic and technology determinants." International Journal of Forecasting (2020)