



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IV **Month of publication:** April 2022

DOI: <https://doi.org/10.22214/ijraset.2022.41623>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Gold Price Prediction System

Abhimanyu Wagh¹, Shreyas Shetty², Adrian Soman³, Prof. Deepali Maste⁴

^{1, 2, 3}B.E. Student, ⁴Professor, Dept. of Information Technology, Atharva College of Engineering, Mumbai, India

Abstract: *This paper aims to promote the use of the LSTM, random forest Regression & Linear Regression algorithm to predict stock prices of India and to compare their accuracy. These are machine learning algorithms used for historical real-time gold prices. The Historical Gold data we made with reference of www.goldpriceindia.com. The full source code of the project was written via Python. It is thought that the LSTM model are more compatible than the Linear Regression and Random Forest model for the Gold price Prediction forecasting model. Gold is a valuable metal and has been historically owned and traded as an asset /commodity. The price of Gold is often a derivative of the investor's sentiment and perception of other asset classes (real estate, equities, commodities, futures and cash equivalents) as Gold has very little fundamentals of its own. Our project is aimed at studying the relationship between gold price, selected economies and various market variables to try and accurately predict the future price of gold using Machine Learning algorithms.*

Keywords: *Machine Learning LSTM, Random Forest, Linear Regression, Gold , Prediction, Time Series, Historical Data, Python , TensorFlow.*

I. 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 the 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 [4]. 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. 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. Investor's 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 [2]. 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. 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. This raising value of gold coupled with the volatilities and fall in prices of other markets like capital markets and real estate markets has attracted more and more investors towards gold as an attractive investment. But, of late price of gold is also witnessing high volatility and investments in gold are turning to be riskier. There is a fear as to whether these high prices are sustainable and when the prices would reverse. Even though there are a number of studies analysing the correlation between the price of gold and some economic variables. It is still considered that a study to reveal the influence and impact of various macro-economic factors on the price of gold in the present situation will be helpful in determining the dynamic effects of these relationships [5].

II. LITERATURE REVIEW

There are many studies dealing with the price of gold in the literature. Although various different variables are used in these studies, it is observed that gold prices are regressed against USA dollar and stock return in general[4]. The relationship between other macroeconomic variable and gold prices has also been studied by many researchers. The relationship between gold price and prices of other commodities especially crude oil has also been extensively studied. But the results from these studies are found to be contradicting. Some of the studies on the factors influencing gold price and various techniques used for studying these relationships are discussed in the following sections. 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 Van Tassel [7] also have found that the price of gold depends on the future inflation rate.

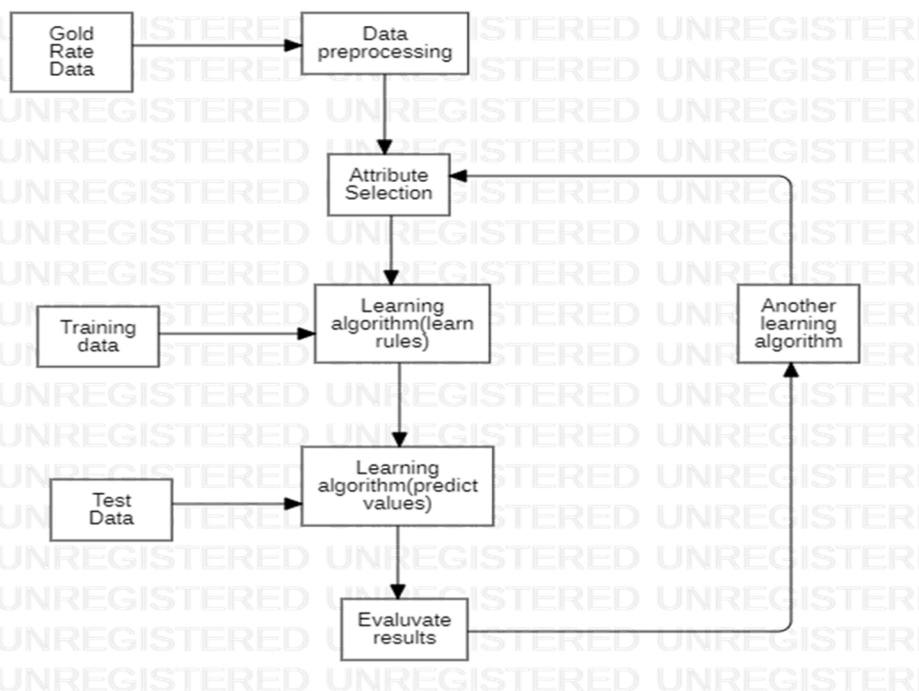
With respect to the relationship between gold price and inflation, based on the review of literature Hanan Naser[8] is of the opinion that historical studies with regards to the effectiveness of gold as a hedge against inflation are contradicting Ismail[3] have forecasted gold prices based on multiple economic factors such as commodity research bureau future index, USD/Euro foreign exchange rate, inflation rate, money supply, New York Stock Exchange Index; Standard and Poor 500 index, Treasury bill and USD index. The study finds that Commodity Research Bureau future index, USD/Euro foreign exchange rate, Inflation rate and money supply have a significant impact on gold price. Khaemusunun [9] has examined the impact of currencies of selected countries, Oil Prices and Interest Rate on the gold price. Hammoudeh [10] conclude that there is an interdependent exist between the volatility of gold price and the exchange rate. Ai, et al.[11] report empirical evidence that the exchange rate relates to the gold price both in the long-run and short-run. Ewing and Malik [12] find evidence of volatility transmission between gold and oil future prices. Ghosh et al. [13] have concluded that gold prices are related with US Inflation level, interest rates and dollar exchange rate. They have also reported a long run relationship between gold prices and US Consumer Price Index as a result of the cointegration analysis. From the review of related literature, it can be concluded that the relationship between gold price and various factors considered to influence it are contradicting. In studying volatility in gold price and the relationship with the factors considered to influence it, researchers have used a variety of techniques. Hossein and Abdolreza[14]have predicted the gold price by using artificial neural networks (ANN) and ARIMA model. Khaemusunun, (2009) predicts the Thai gold price by using Multiple Regression and ARIMA model. Toraman[4] has reported that various studies have been conducted using multivariate regression models to test the sensitivity of gold prices among various variables. In this regard Ismail et al.[3] have used multiple linear regression (MLR) model for forecasting the gold prices and are of the opinion that MLR model appeared to be useful for predicting the gold price. From the review of literature, it can be seen that multiple linear regression is widely used technique for understanding relationship among such variables.

III. DATA AND METHODOLOGY

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

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

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



Linear regression models with more than one independent variable are called multiple linear models. A representation of multiple linear regressions is where, Y is dependent variable and $X_1, X_2 \dots$ are independent variables are as seen below.

$$Y = a + b_1 * X_1 + b_2 * X_2 + \dots + b_p * X_p$$

As such, linear regression was developed in the field of statistics and is studied as a model for understanding the relationship between input and output numerical variables, but has been borrowed by machine learning. It is both a statistical algorithm and a machine learning algorithm now.

A Random Forest is an ensemble technique capable of performing both regression and classification tasks with the use of multiple decision trees and a technique called Bootstrap Aggregation, commonly known as bagging. The basic idea behind this is to combine multiple decision trees in determining the final output rather than relying on individual decision trees. The Random Forest uses bootstrapping on Decision Trees to reduce the variance while maintaining the low bias that is resulted from a Decision Tree model. A Random Forest algorithm has the following advantages when compared to most of the other algorithms - The overfitting problem will never come when we use the random forest algorithm in any classification problem. The same random forest algorithm can be used for both classification and regression task. And, the random forest algorithm can be used for feature engineering for identifying the most important features out of the available features from the training dataset.

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

A block has many basic and a few complex components that make it smarter as compared to the standard neuron. It consists of many gates that coordinate relative input functions with output functions. Whenever a block receives an input, a gate is triggered which takes decision about whether or not to pass the block forward for further processing.

The standard LSTM block, in its simplest form, consists of an input gate, an output gate, a cell and a forget gate.

Cell: It is used to remember the values over arbitrary time intervals.

Input Gate: It decides which information to keep in the cell.

Output Gate: It is used to decide which part of cell state should be given as an output.

Forget Gate: It is used to decide which information to throw away from the cell.

In the Line Redistribution model, the calculation line calculation is used to combine a set of input data values (x) into a predicted output data set of input values (y). Both the input and output variables and values are considered integers. The unique number given by the Line Rotation equation is represented using the Greek capital letter Beta (B) and is commonly known as a coefficient. In addition to this, another coefficient is added to give the line additional degrees of freedom. This additional term is often referred to as the bias coefficient. Typically, the bias coefficient is calculated or otherwise measured by finding the distance of our mathematical points from the most relevant line. This can be displayed as a straight line at right angles to the vertex and calculated using the line bias. Statistically, a line tangent is used to measure its proximity to the relative linear Regression.

A problem model model in Linear Regression will be provided as follows:

$$y = B_0 + B_t * x_t + E_t$$

This same line is also called a plane or plane when we are dealing with more than one input. This is often the case with high-volume data. The Linear Regression model is therefore represented by the mathematical and introverted values measured by the specific coefficients. However, before using this line number, we are faced with a number of issues. These issues often increase the complexity of the model which makes accurate estimates difficult. This complexity is often discussed in terms of the number of dependent and independent factors.

IV. RESULTS AND DISCUSSION

The Daily gold price values are taken from period 2010-2022. Monthly Gold Price values are taken from Period 2015-2022. Six Months Gold Price values are taken from Period 2010-2022. Yearly Gold Price values are taken from Period 2010-2022. Actual vs Predicted prices Bar Graph Plotted below for daily Prices during the period 2010-2022.

A. Linear Regression

The bar graph for actual predicted for linear regression is taken and the pic is taken.

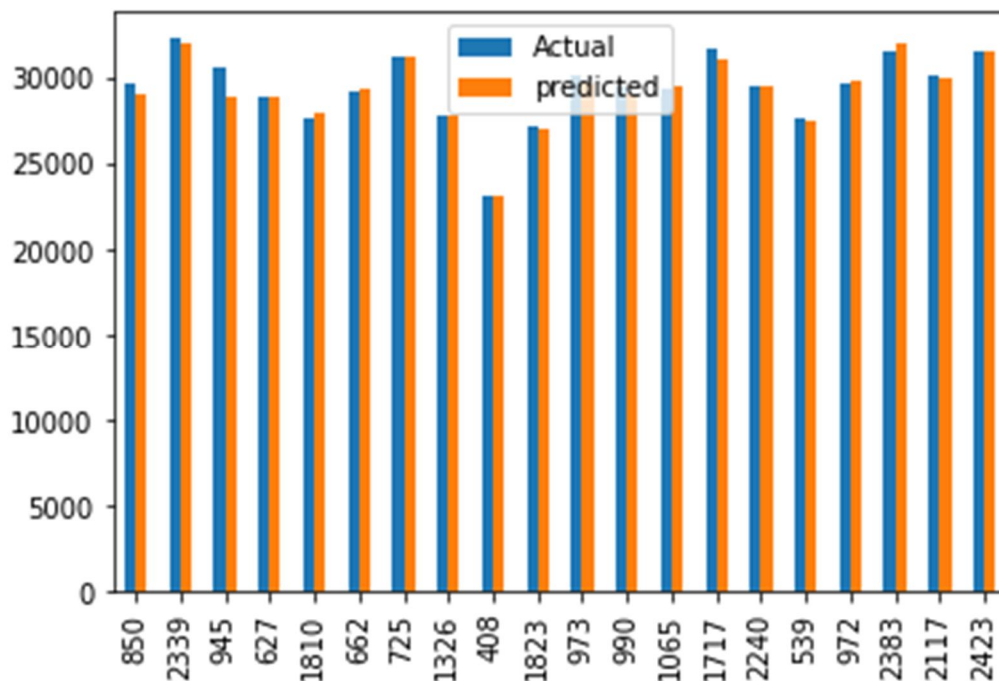


Fig.4.1. Bar graph of linear regression

B. Random forest Regression

The bar graph for actual predicted for random forest regression is taken and the pic is taken.

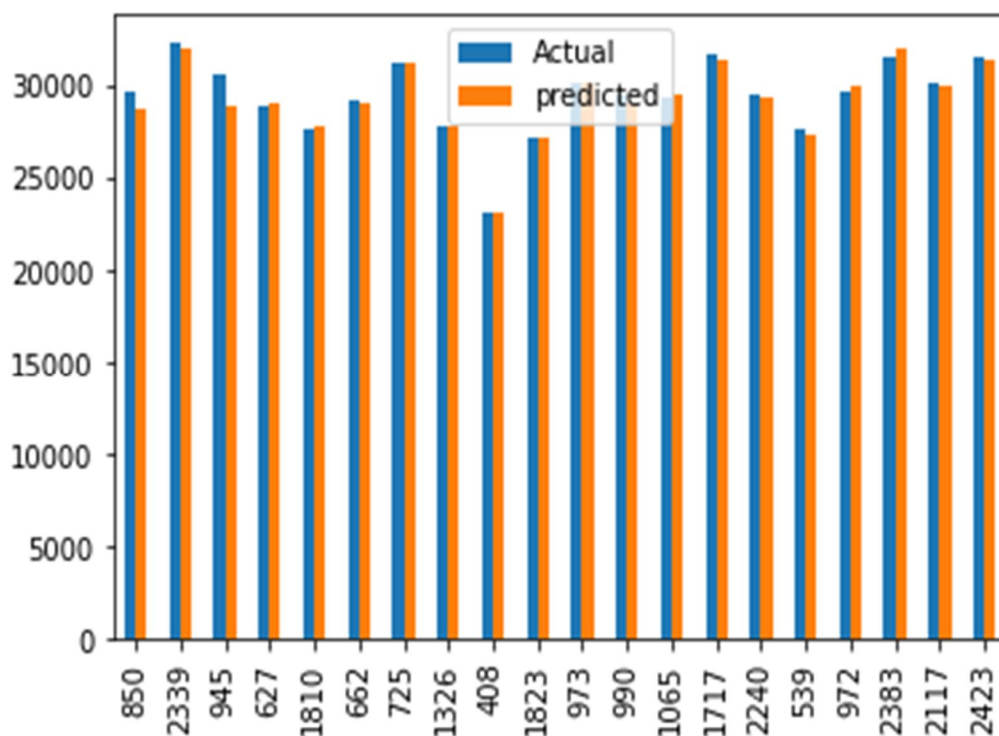


Fig.4.3. Bar graph of Random Forest regression

C. Actual vs Predicted Graph

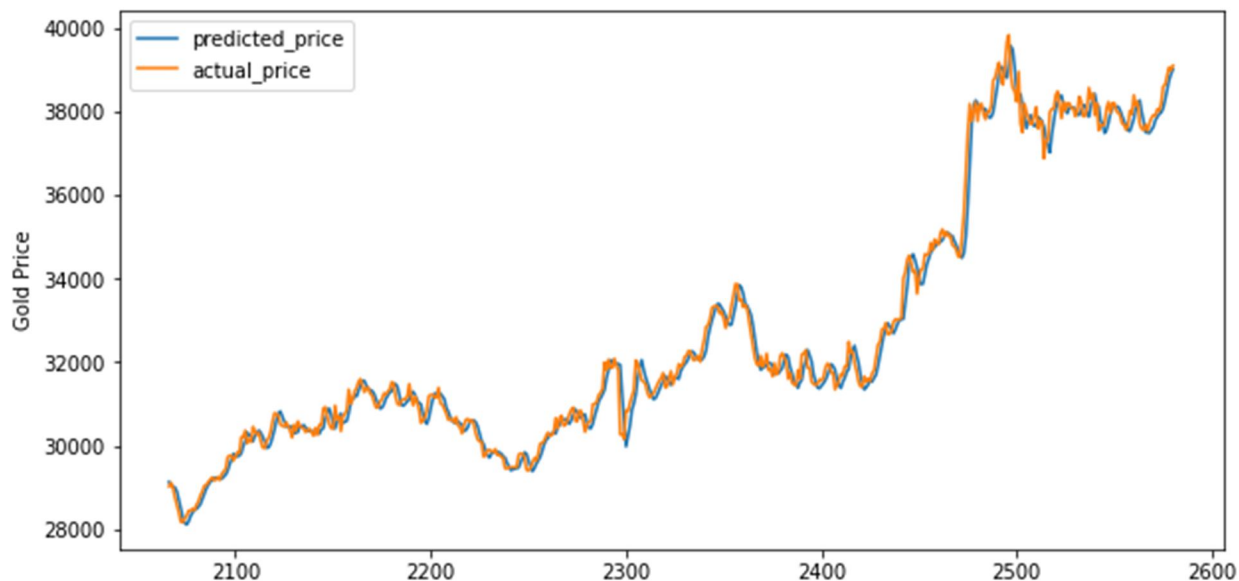


Fig 4.3 Actual vs. Predicted Graph for linear regression

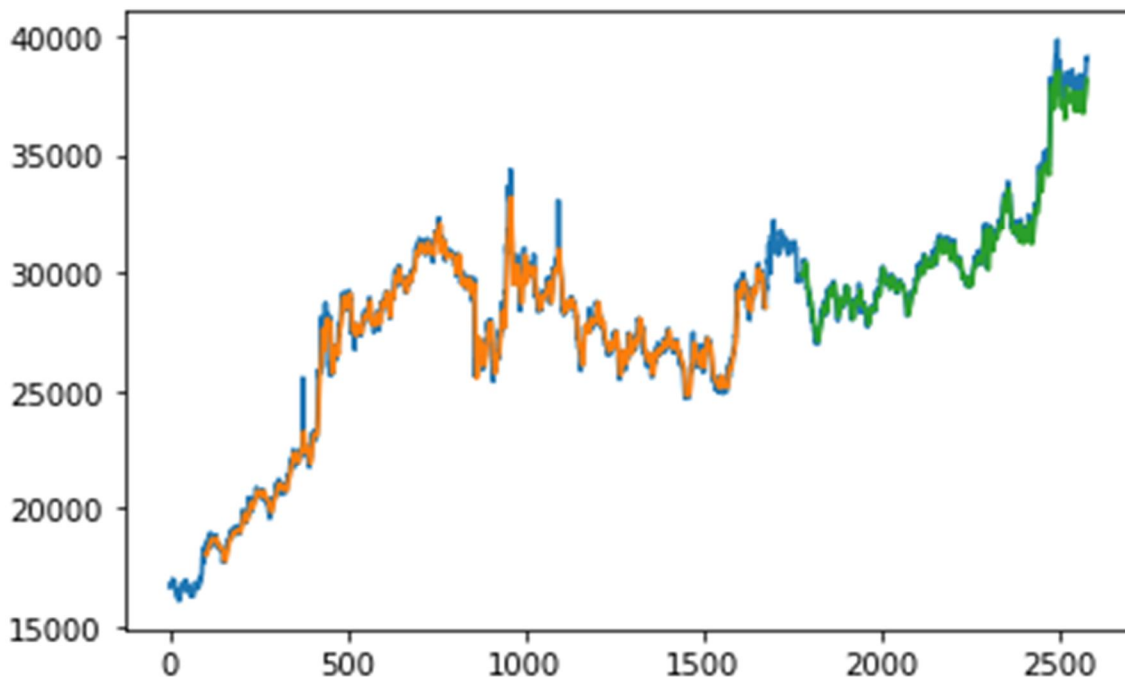


Fig 4.4 Actual vs. Predicted Graph for LSTM Model

Comparing the prediction accuracy of the three models LSTM Model gives Best Results and shows great accuracy throughout the period. Based on the results from these analyses, it may be inferred that there has been a change in the trend in the gold price during the period considered for this study. In such situations where there is change in the trend of the dependent variable and no significant changes in the trend of independent variables, the accuracy of various methods may differ. Hence the model used should depend on the relationship between the variables used in the study.

V. CONCLUSION

This study was conducted to understand the relationship between gold price and selected factors influencing its price. The Daily gold price values are taken from period 2010-2022. Monthly Gold Price values are taken from Period 2015-2022. Six Months Gold Price values are taken from Period 2010-2022. Yearly Gold Price values are taken from Period 2010-2022. Three machine learning algorithms, linear regression, random forest regression and LSTM Model were used in analyzing these data. LSTM Model is found to have better prediction accuracy for the entire period is found to give better accuracy for the two period taken separately. It is concluded that machine learning algorithms are very useful in such analysis, but the characteristics of the data influences their accuracy. Further research with such data and different techniques may be conducted for better understanding of the performance of these techniques.

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. Khaemasunon, "Forecasting Thai gold prices," Available [Http://www.Wbiconpro.Com3-Pravit.Pdf](http://www.Wbiconpro.Com3-Pravit.Pdf) Access, 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.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)