iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 Issue: IX Month of publication: September 2022

DOI: https://doi.org/10.22214/ijraset.2022.46555

www.ijraset.com

Call: ☎08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

Time Series Analysis and Forecasting of Gold Price using ARIMA and LSTM Model

Dhruvi Sarvaiya[1], Disha Ramchandani[2]
[12]Department of Computer Engineering, Thadomal Shahani Engineering College, Mumbai, India

Abstract: Gold is an age-old method of investing money. It is tangible and can be passed on from one generation to another. It is the one form of investment that many people consider very safe to employ, to keep their money safe and easily multiplicative. Investment guides and consultants usually read charts to predict the future price of this commodity. In this research paper, we are making use of Machine Learning models to predict the price of gold based on past prices. The dataset consists of the opening, closing, highest and lowest prices of gold through 7 years, daily. This is a comparative study that highlights the best model, between a classical statistical model ARIMA and a recurrent neural network model LSTM

Keywords: Time series Forecasting, ARIMA, LSTM, Gold price, RMSE.

I. INTRODUCTION

Investing your money, correctly and carefully; is a very effective way to try and grow your wealth. If done with caution, it can yield majorly positive results, which might help you outpace inflation. In the past few decades, the investment of money has become an immensely popularised idea. There is a variety of ways to invest your money. Primary or traditional sources of investment include stocks, mutual funds, and bonds, these are heavily volatile and unpredictable as they are determined due to the global economy and public sentiment. Alternative investment options include precious metals, real estate, and other similar commodities.

When talking specifically about gold, the value is stable and hence can be a safe investment option. Speaking about the economic crisis in the year 2008, while a wide range of financial tools failed to give good returns; gold maintained its performance. This makes it a reliable investment option as the price of the metal is influenced by the dollar exchange rate,

Inflation, or monetary policy, to name a few. [1] Gold is viewed as a means by which people might maintain and transmit their riches from one generation to the next. People have cherished the special qualities of precious metals from the beginning of time [2] Although gold prices can considerably vary due to supply and demand reasons in the near term, they have historically held their worth.

With the use of machine learning (ML), which is a form of artificial intelligence (AI), software programs can predict outcomes more accurately without having to be explicitly instructed. In order to forecast new output values, machine learning algorithms use historical data as input. In business, finance, supply chain management, production, and inventory planning, time series forecasting is one of the most widely used data science methodologies. Another crucial field of machine learning (ML) is time series forecasting, which may be viewed as a supervised learning issue. It can be subjected to ML techniques including regression, neural networks, support vector machines, random forests, and XG Boost. For practical researchers in Economics and Finance, machine learning techniques have emerged as a crucial tool for estimation, model selection, and forecasting. Making accurate and dependable projections is crucial in the Big Data era due to the accessibility of enormous data sets. [3]

In this paper, we aim at comparing two-time series forecasting models based on their accuracy in predicting future gold prices. Firstly, we are employing the ARIMA model, which is an Autoregressive model with integrated moving averages. The second model is the LSTM model, which stands for long short-term memory and is a part of neural networks. ARIMA model, though certainly faster and more cost-effective to develop for the purpose of fitting moderately high order autoregression models with linear predictions, comes with its drawbacks. It requires the estimation of many parameters which can lead to inaccurate and dissatisfactory results. [4]On the other hand, LSTM models work on the principle of recurrent neural networks. While we need to tune some essential hyperparameters, the data is manipulated based on the previous and sequential data. This leads to more accurate and reliable results. [5]

In addition to this concise introduction, the paper has been divided into five more sections. The section following this one contains the literature review that we have employed for our research. Section three talks about the dataset that we have made use of to test the accuracy of the two models we wish to discuss. The fourth section describes, in-depth, the models that are being compared on the bases of their accuracy and the fifth section throws light on the results that we obtained. The last section of this paper summarizes a conclusion and briefly mentions the future scope that it holds.

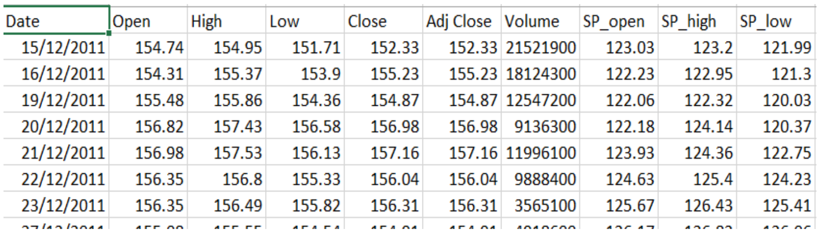
II. LITERATURE REVIEW

1) *A Gold price prediction using ARIMA model by Parag Saharia; Research Scholar and Sangita Kalita; Associate professor; Department of Statistics; Cotton University & Ranjita Goswami; Assistant Professor and Pranab Das; Assistant Professor; Department of Statistics; Mangaldai college:* In this paper, the authors have made use of various ARIMA models of permutations of p,d,q values to conclude that ARIMA model of order (1,1,1) yields the least amount of error. [6]
2) *Gold price forecasting using LSTM, Bi-LSTM, and GRU by Mustafa Yurtsever:* This paper has made use of three multivariate versions of the models to predict the price of gold based on the economic metrics such as crude oil, stock market index, currency exchange rate, interest rates, and consumer price index. [7]
3) *Factors affecting Gold Prices: A Case study of India by Prerana Baber, Ruturaj Baber & Dr. George Thomas:* This analytical essay explores the various factors that have contributed to the consistently rising prices of gold in India between 2002 and 2012, including the global business environment, the political climate, market conditions, its introduction to the commodity market, consumer purchasing patterns, and inflation. [8]
4) *A study of forecasts in Financial Time Series using Machine Learning methods by Mowniesh Asokan:* This article highlights that deep learning-based algorithms and techniques have immense potential in the fields of economics and finance. In addition to that, several other predictions can be made using machine learning as well. The authors have fitted a total of nine models- ranging from ML models to NN models- to the same data from the S&P 500, SSE, and FTSE 100 index and evaluated their performance accuracy. [9]

III. DATASET

As mentioned previously, our paper focuses on the comparison between two time series forecasting based on how accurately they can predict the price of gold. For this purpose, we required a dataset that would provide data on gold prices. The dataset we chose to employ for our models was picked from Kaggle. It contains daily data entries of gold prices from January 2011 to December 2018, in US dollars. Initially, the dataset was comprised of 1718 rows and 81 columns. The values of these columns were the opening price, closing price, lowest price, highest price, and the adjacent closing price of the day for multiple indexes- the S&P 500 index, and Dow Index, to name a few.

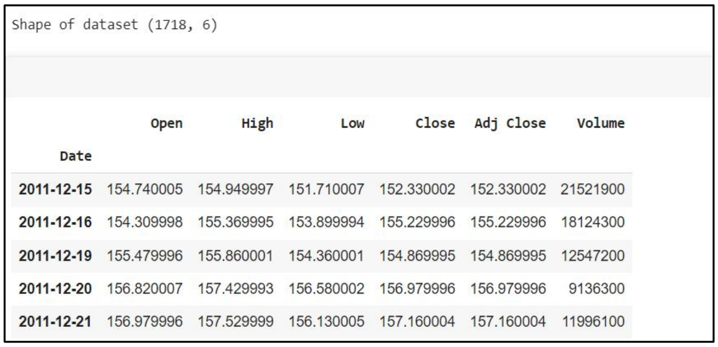
TABLE I



Date	Open	High	Low	Close	Adj Close	Volume	SP_open	SP_high	SP_low
15/12/2011	154.74	154.95	151.71	152.33	152.33	21521900	123.03	123.2	121.99
16/12/2011	154.31	155.37	153.9	155.23	155.23	18124300	122.23	122.95	121.3
19/12/2011	155.48	155.86	154.36	154.87	154.87	12547200	122.06	122.32	120.03
20/12/2011	156.82	157.43	156.58	156.98	156.98	9136300	122.18	124.14	120.37
21/12/2011	156.98	157.53	156.13	157.16	157.16	11996100	123.93	124.36	122.75
22/12/2011	156.35	156.8	155.33	156.04	156.04	9888400	124.63	125.4	124.23
23/12/2011	156.35	156.49	155.82	156.31	156.31	3565100	125.67	126.43	125.41

We decided to retain only the standard index values while eliminating the rest. After the elimination, we were left with 1718 rows and 7 columns (including the date column) which served as our final dataset. To achieve this, we deleted the columns of the other indices and created a data frame of our required values.

TABLE 2



```
Shape of dataset (1718, 6)
```

Date	Open	High	Low	Close	Adj Close	Volume
2011-12-15	154.740005	154.949997	151.710007	152.330002	152.330002	21521900
2011-12-16	154.309998	155.369995	153.899994	155.229996	155.229996	18124300
2011-12-19	155.479996	155.860001	154.360001	154.869995	154.869995	12547200
2011-12-20	156.820007	157.429993	156.580002	156.979996	156.979996	9136300
2011-12-21	156.979996	157.529999	156.130005	157.160004	157.160004	11996100

IV. METHODOLOGY

A. ARIMA

For the purpose of determining how accurate classical statistical models are in the field of time series forecasting, we decided to employ the ARIMA model. ARIMA is a typical autoregression model which also includes the application of moving averages to increase the accuracy. This model is said to work best with a non-seasonal or stationary dataset. A stationary data must have no trend, constant amplitude variations around its mean, and consistent ups-and-lows, which means that statistically speaking, its short-term random time patterns remain the same. The latter requires that its power spectrum, or more precisely, its autocorrelations—correlations with its own prior departures from the mean—remain constant across time. [10]

An "ARIMA(p,d,q)" model is a nonseasonal ARIMA model, where:

1) p is the number of autoregressive terms,
2) d is the number of nonseasonal variations required for stationarity,
3) and q is the number of lags forecast errors.

In terms of y, the general forecasting equation is:

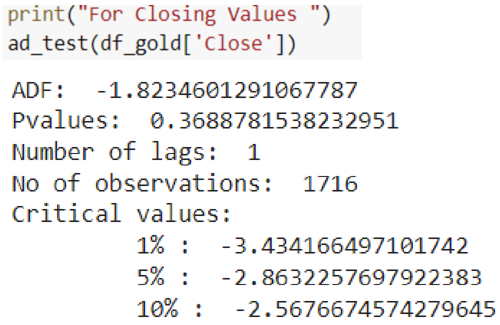
$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}$$

Where θ represents the moving average parameters.

For the purpose of our research, we used the ARIMA model of the order (1,1,2) as was deemed fit by the ADFuller test. ADFuller stands for Augmented Dickey-Fuller Test. Let us first define the Dickey-Fuller test before moving on to the ADF test. The following model equation's null hypothesis that α=1 is tested using a Dickey-Fuller test, a unit root test. The first lag on Y's coefficient is called alpha. Alpha=1 is the null hypothesis (H0). [11] The ADFuller test includes an "augmented," as the name suggests, version of the Dickey-Fuller test and can be used to include a high-order regressive process in the model.

The following steps were taken to measure the predicted values

a) We imported the gold price dataset which ranges from January 2011 to December 2018, in US dollars.
b) We pre-processed the data and checked for seasonality in our dataset.
c) Next, we performed the ADFuller test to determine which model will work best for our needs.

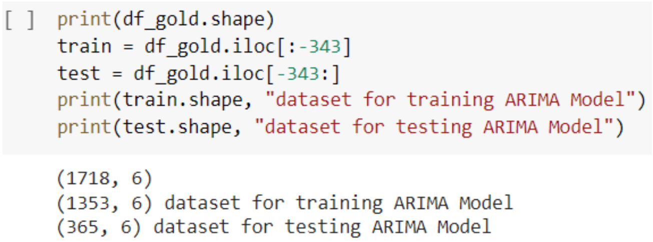


```
print("For Closing Values ")
ad_test(df_gold['Close'])
```

```
ADF:  -1.8234601291067787
Pvalues:  0.3688781538232951
Number of lags:  1
No of observations:  1716
Critical values:
         1% :  -3.434166497101742
         5% :  -2.8632257697922383
         10% :  -2.5676674574279645
```

Fig. 1: ADFULLER TEST

d) We split the dataset into training and testing which is 80 percent of the dataset will be used to train the model and 20 percent to test it.



```
[ ]  print(df_gold.shape)
     train = df_gold.iloc[:-343]
     test = df_gold.iloc[-343:]
     print(train.shape, "dataset for training ARIMA Model")
     print(test.shape, "dataset for testing ARIMA Model")
```

```
(1718, 6)
(1353, 6) dataset for training ARIMA Model
(365, 6) dataset for testing ARIMA Model
```

Fig. 2: SPLITTING THE DATA

e) We trained the model using ARIMA model of order (1,1,2)

f) Lastly, we plotted the graph and checked the performance metric which was RMSE (root mean squared error) which came out to be 4.04316.

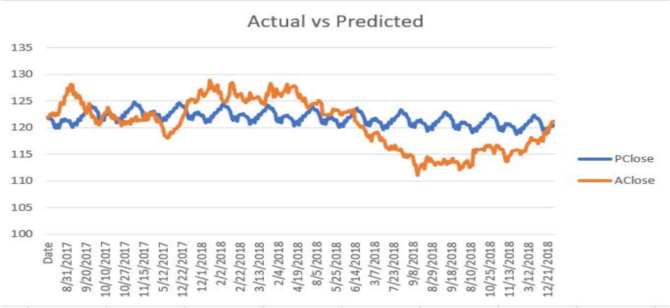


Fig. 3: COMPARING ACTUAL AND PREDICTED VALUES.

B. LSTM

Long short-term memory networks, or LSTMs, are employed in deep learning. Many recurrent neural networks (RNNs) can learn long-term dependencies, particularly in tasks involving sequence prediction. Aside from singular data points like photos, LSTM has feedback connections, making it capable of processing the complete sequence of data. This has uses in machine translation and speech recognition, among others. A unique version of RNN called LSTM exhibits exceptional performance on a wide range of issues. Long Short-Term Memory (LSTM) is a Recurrent Neural Network (RNN) architecture that has been demonstrated to outperform conventional RNNs on a variety of temporal processing tasks. [12] Numerous applications of neural networks have been made to model and forecast the dynamics of complex systems. There are many different types of networks accessible, but the quality of the modelling is greatly influenced by how well the network architecture fits the task at hand. [13]

LSTM in time series analysis-Demand forecasting is difficult in the current environment, and getting the data necessary for precise large-scale forecasting can be difficult. Time series forecasting models can forecast future values based on prior, sequential data by utilizing LSTM. This improves demand forecasters' accuracy, which helps the business make better decisions.

A memory cell that can keep its state over time and nonlinear gating units that control information flow into and out of the cell make up the core of the LSTM architecture. Many current studies take advantage of the LSTM architecture's numerous advancements since it was first developed. [14]

Convolutional layers excel at extracting relevant information from time-series data and learning the internal representation of the data, while LSTM networks excel at spotting both short- and long-term dependencies. [15]

Long Short-Term Memory networks – usually just called “LSTMs” – are a special kind of RNN, capable of learning long-term dependencies All recurrent neural networks have the form of a chain of repeating modules of neural network. In standard RNNs, this repeating module will have a very simple structure, such as a single tanh layer. LSTMs also have this chain-like structure, but the repeating module has a different structure. Instead of having a single neural network layer, there are four, interacting in a very special way. The repeating module in an LSTM contains four interacting layers.

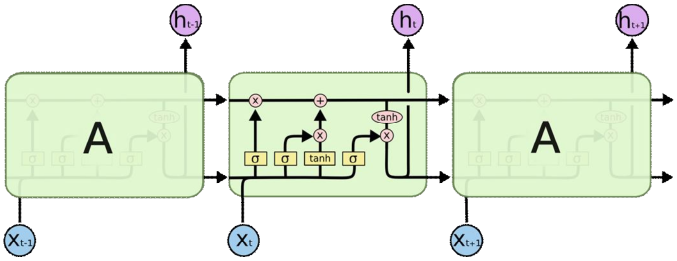


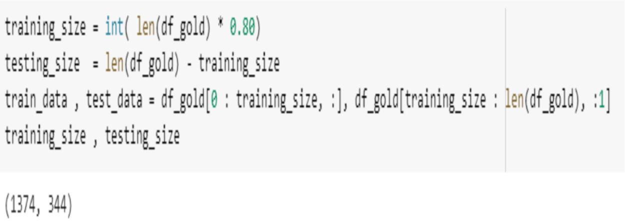
Fig. 4: LSTM NETWORKS

The following steps were taken to measure the predicted values

1) We imported the gold price dataset from Kaggle which ranges from January 2011 to December 2018, in US dollars.

2) We pre-processed the data and checked for seasonality in our dataset

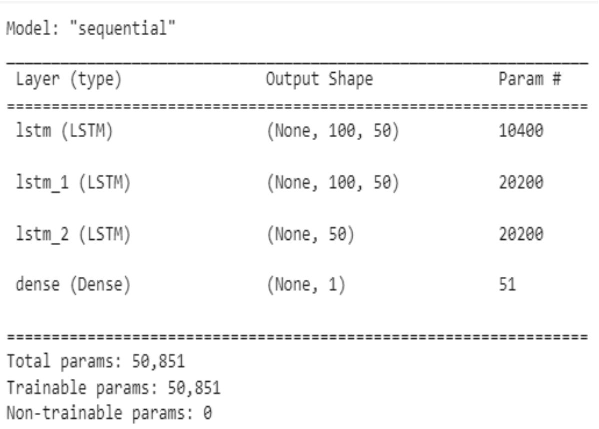
3) Next, we performed feature scaling using a MinMax scaler to transform and set a range for our dataset. The range is from zero to one.
4) We split the dataset into training and testing which is 80 percent of the dataset will be used to train the model and 20 percent to test it.



```
training_size = int( len(df_gold) * 0.80)
testing_size  = len(df_gold) - training_size
train_data , test_data = df_gold[0 : training_size, :], df_gold[training_size : len(df_gold), :1]
training_size , testing_size

(1374, 344)
```

Fig. 5: SPLITTING THE DATA

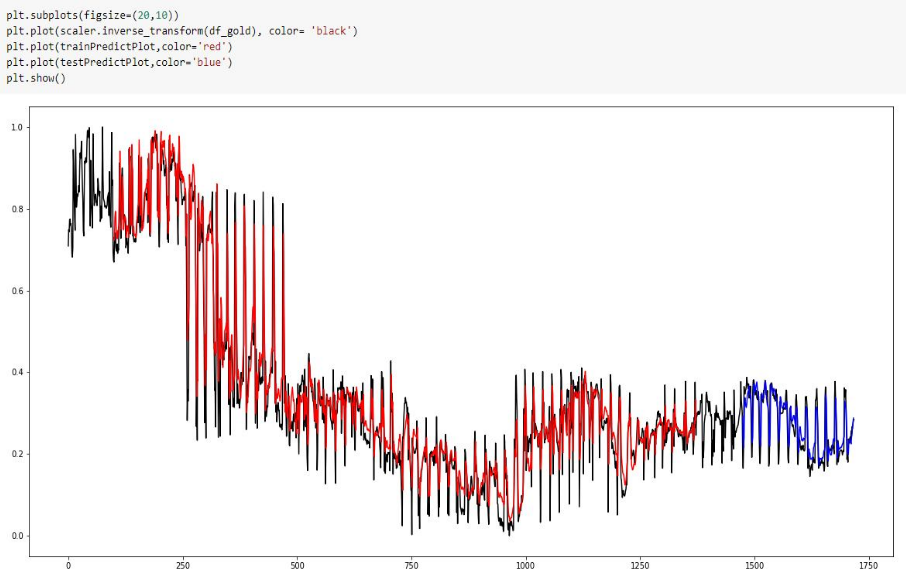
5) We printed the LSTM model summary



```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm (LSTM)                 (None, 100, 50)           10400

 lstm_1 (LSTM)               (None, 100, 50)           20200

 lstm_2 (LSTM)               (None, 50)                20200

 dense (Dense)               (None, 1)                 51

=================================================================
Total params: 50,851
Trainable params: 50,851
Non-trainable params: 0
```

Fig. 6: LSTM MODEL SUMMARY

6) Lastly, we plotted the graph and checked the performance metric which was RMSE (root mean squared error) which came out to be 0.038.



```
plt.subplots(figsize=(20,10))
plt.plot(scaler.inverse_transform(df_gold), color= 'black')
plt.plot(trainPredictPlot,color='red')
plt.plot(testPredictPlot,color='blue')
plt.show()
```

Fig. 7: COMPARING ACTUAL AND PREDICTED VALUES

V. RESULTS

To check the accuracy of our models, we employed the RMSE (Root Mean Squared Error) metric. It is one of the methods most frequently used to assess accuracy. It illustrates the Euclidean distance between measured true values and forecasts.

Calculate the residual (difference between prediction and truth) for each data point, along with its norm, mean, and square root in order to determine the RMSE. [16]
The formula for RMSE is as follows

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

Fig. 8: RMSE FORMULA

TABLE 3

FORECASTING MODEL	ERROR MEASURES	ERROR VALUE
ARIMA	RMSE	4.043
LSTM	RMSE	0.038

VI. CONCLUSION AND FUTURE SCOPE

The main aim of this research paper was to compare two types of time series forecasting models to understand which one would work better for future predictions. After finishing our work, we can conclude that LSTM is a more accurate model for gold price prediction as compared to classical statistical models like ARIMA. The project can be taken further by implementing deep learning to improve the accuracy of the prediction. We can also employ this model on a website which might be used to predict the future price of gold, helping people decide whether it is a good decision to invest or not.

REFERENCES

[1] Y. Xiaohui, "The prediction of gold price using ARIMA model," In 2nd International Conference on Social Science, Public Health and Education , vol. volume 196, pp. 273-276, 2019.

[2] T. Daltorio, "8 Good Reasons To Own Gold," [Online]. Available: https://www.investopedia.com/articles/basics/08/reasons-to-own-gold.asp.

[3] R. P. Masini, M. C. Marcelo and M. F. Eduardo, ""Machine learning advances for time series forecasting," Journal of Economic Surveys, 2021.

[4] N. Paul, "ARIMA model building and the time series analysis approach to forecasting," Journal of forecasting, vol. 2, pp. 23-35, 1983.

[5] "The Value of LSTM in Time Series Forecasting," [Online]. Available: https://www.predicthq.com/events/lstm-time-series-forecasting.

[6] P. Saharia,, . S. Kalita, R. Goswami and P. Das, "GOLD PRICE FORECASTING IN INDIA USING ARIMA (AUTOREGRESSIVE INTEGRATED MOVING AVERAGE) MODEL".

[7] M. Yurtsever, "Gold price forecasting using LSTM, Bi-LSTM and GRU," European Journal of Science and Technology, 2021.

[8] P. Baber, R. Baber and T. G, "Factors affecting Gold Prices: A Case study of India," in National Conference on Evolving Paradigms in Manufacturing and Service Sectors, 2013.

[9] M. Asokan, "A study of forecasts in Financial Time Series using Machine Learning methods," Division of Statistics and Machine Learning, 2022.

[10] "Introduction to ARIMA: nonseasonal models," [Online]. Available: https://people.duke.edu/~rnau/411arim.htm#pdq.

[11] . S Prabhakaran, "Augmented Dickey Fuller Test (ADF Test) – Must Read Guide," Machinelearningplus, 2 November 2019. [Online]. Available: https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/.

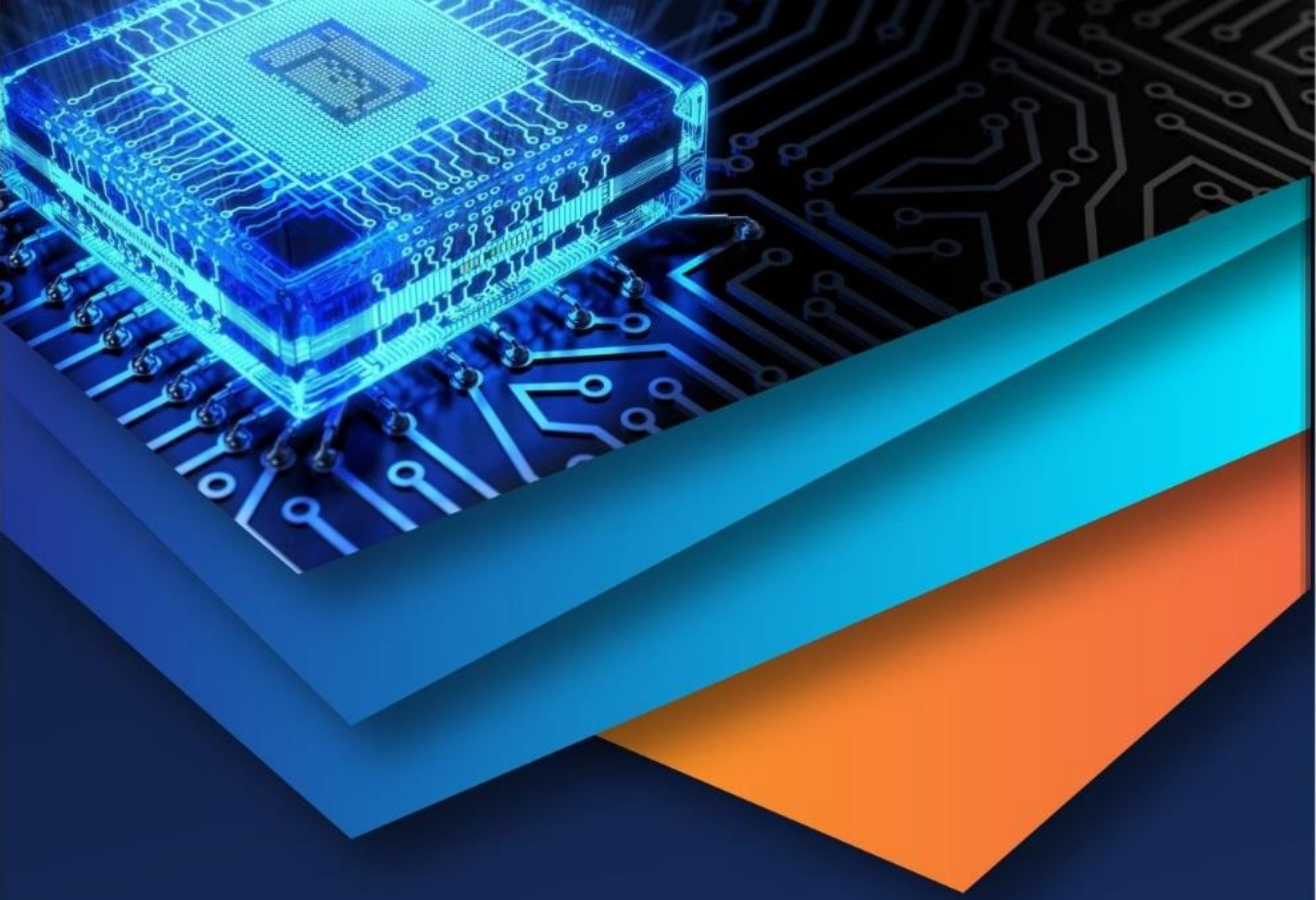
[12] F. A. Gers, . E. Douglas and J. Schmidhuber, "Applying LSTM to time series predictable through time-window approaches," In Neural Nets WIRN Vietri-01, vol. 01, pp. 193-200, 2002.

[13] B. Lindemann, T. Müller, H. Vietz, . N. Jazdi and M. Weyrich, "A survey on long short-term memory networks for time series prediction," in Procedia CIRP 99 , 2021.

[14] G. Klaus, R. K. Srivastava, J. Koutník, B. R. Steunebrink and J. Schmidhuber, "LSTM: A search space odyssey," IEEE transactions on neural networks and learning systems 28, vol. 10, pp. 2222-2232., 2016.

[15] I. E. Livieris, E. Pintelas and P. Pintelas, "A CNN–LSTM model for gold price time-series forecasting.," Neural computing and applications 32, vol. 23, pp. 17351-17360., 2020.

[16] "Root Mean Square Error (RMSE)," C3.ai, [Online]. Available: https://c3.ai/glossary/data-science/root-mean-square-error-rmse/.



doi crossref

10.22214/IJRASET

INDEX COPERNICUS

45.98

ISRA JIF

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)