iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** III **Month of publication:** March 2020

DOI:

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: 7.429
Volume 8 Issue III Mar 2020- Available at www.ijraset.com

Predicting NASDAQ and NSE Stocks using Machine Learning Algorithms: ARIMA, LSTM & Linear Regression

Prof. Amit Narote[1], Kaushik Arvind Jadhav[2], Jay Dharmendra Barot[3], Shubham Santosh Sawant[4]
[1]*Assistant Professor,* [2, 3, 4] *Students, Department of Information Technology, Xavier Institute of Engineering, Mumbai-400016, India*

Abstract*: This paper aims to propose the use of ARIMA, LSTM & Linear Regression algorithm to predict NASDAQ (American) and NSE (Indian) stock prices as well as to compare their respective accuracies. These Machine Learning algorithms were applied to the historical stock data of the past 2 years as well as real time stock prices. The stock data for NASDAQ stocks was fetched from the Yahoo Finance API and that for NSE stocks was fetched from Alpha Vantage API. The complete source code of the project was written in Python. It was inferred that ARIMA and LSTM models are more consistent than Linear Regression model for forecasting NASDAQ (American Company) stocks. Whereas, for NSE (Indian Company) stocks, LSTM and Linear Regression prove to be more efficient than ARIMA.

***Keywords*: Machine Learning, ARIMA, LSTM, Linear Regression, Stock Market, Prediction, Stock Exchange, Trading, Time Series, Historical Data, Python**

I. INTRODUCTION

Stock prices are very fluctuating in nature. They vary based on various factors such as previous stock prices, present scanrio of the market, financial news, rival companies etc. It is important to have accurate prediction of future trends in stock prices for smart investing decisions [1] [2]. However, the fluctuating nature of the stock prices makes it difficult to have a precise estimation. Stock Market Prediction is an attempt to determine the future price of a company stock [3]. Historical stock prices for NASDAQ stocks were fetched from the Yahoo Finance API and those for NSE stocks were fetched from the Alpha Vantage API. This data was pre-processed and then passed into Machine Learning models. Finally, results of each of the models were visualized.

II. LITERATURE REVIEW

A. Machine Learning Techniques and Use of Event Information for Stock Market Prediction

Paul D. Yoo, Maria H. Kim and Tony Jan compared and evaluated some of the existing ML techniques used for stock market prediction. After comparing simple regression, multivariate regression, Neural Networks, Support Vector Machines and Case Based Reasoning models they concluded that Neural Networks offer ability to predict market directions more accurately as compared to other techniques. Support Vector Machines and Case Based Reasoning are also popular for stock market prediction. In addition, they found that incorporating event information with prediction model plays a very important role for more accurate prediction. The web provides the latest and latent event information about stock market which is required to yield higher prediction accuracy and to make prediction in a short time frame [1].

B. Stock Price Prediction Using Financial News Articles

M.I. Yasef Kaya and M. Elef Karshgil analysed the correlation between the contents of financial news articles and the stock prices. News articles were labeled positive or negative depending on their effect on stock market. Instead of using single word as features, they used word couples as features. A word couple consisted of a combination of a noun and a verb. SVM classifier was trained with labeled articles to predict the stock prices. [2].

C. Forecasting of Stock Market Indices Using Artificial Neural Network

Dr. Jay Joshi, Nisarg A Joshi in his work, used artificial neural network (ANN) to predict the stock prices in reputed indexes of Bombay Stock Exchange (BSE) Sensitive Index (Sensex). They conducted experiments and case studies to compare the performance of neural network with random walk and linear autoregressive models. They reported that neural network outperforms linear autoregressive and random walk models by all performance measures in both in-sample and out-of-sample forecasting of daily BSE Sensex returns.The model forecasted the desired target with an average accuracy of 82% [3].

D. Stock Price Prediction Using ARIMA Model

Ayodele A. Adebiyi, Aderemi O. Adewumi and Charles K. Ayo used ARIMA model to predict the stock price on the data obtained from New York Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE). They made use of a data set consisting of four features: open, low, close and high price. In their work, they have taken the closing price as the target feature to be predicted. The reason behind this is that the Closing price is the most relevant price at the end of the day. They have demonstrated that there is no relation between the autocorrelation functions (ACFs) and partial autocorrelation functions (PACFs) using Q-statistics and Correlation plots. Moreover, for non-stationary data, it was made stationary with the help of differencing techniques. It was concluded towards the end of the research that ARIMA model is very useful for short-term prediction [4].

III.METHODOLOGY

A. Auto Regressive Integrated Moving Average (ARIMA)

The full form of ARIMA is Auto Regressive Integrated Moving Average. There are two types of ARIMA models that may be used in forecasting: seasonal ARIMA and Non-seasonal ARIMA. In our case, a non-seasonal ARIMA model has been used due to the nature of stock data. ARIMA is actually an instance of models is based on its own past values, such that this relation can be used to predict future values. ARIMA model takes in three main parameters, defined as follows:

p = Number of periods to lag. For example, when p=4, we make use of the previous four time laps of our data in autoregressive calculation. p enables us to adjust the fitting line of the time series.

d = In ARIMA, we fit and convert the relative time series into standard time series by means of differencing. We use d to specify the count of differencing computations.

q = q is used to denote the error component lag. Error component is a portion of the historical data which cannot be justified by usual variation of values

Autoregressive component: An independent AR model relies upon a mixture of the historical values. This dependency is similar to that observed in classical linear regression such that the count of Auto Regressive components has a direct dependency to the count of previous periods.

We make use of the Auto Regressive component when:

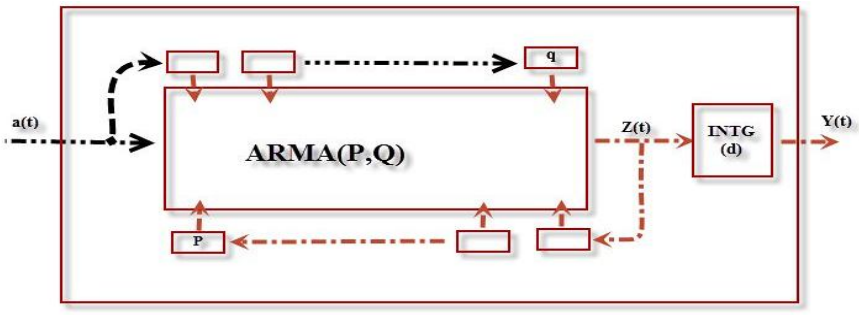
1) ACF graphs portray slope decreasing towards zero
2) A positive at lag-1 is portrayed by the ACF plot of the time series
3) The graph of PACF values suddenly drops down to null

Moving Averages: Moving Averages are random jumps in the data which results in more than two periods which may or may not be successive. These hops are used to specify the calculated error and define what object would the MA component lag for. A completely MA model would justify and clarify these jumps similar to the exponential smoothing method.

We make use of the Moving Averages component when:

a) A significant drop is observed in ACF just after few lags
b) The model shows a Lag which is negative
c) The slope is tend to go downwards increasingly for PACF

Integrated component: The integrated component is triggered only when the historical or real time series data is not stationary or seasonal. The number of times the time series needs to be differenced and computed in order to make it stationary forms i-term of the generic integrated component.



ARIMA(P,d,Q)

Fig 3.1: ARIMA model [5]

B. Long Short Term Memory (LSTM)

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.

1) *Cell:* It is used to remember the values over arbitrary time intervals.
2) *Input Gate:* It decides which information to keep in the cell.
3) *Output Gate:* It is used to decide which part of cell state should be given as an output.
4) *Forget Gate:* It is used to decide which information to throw away from the cell.

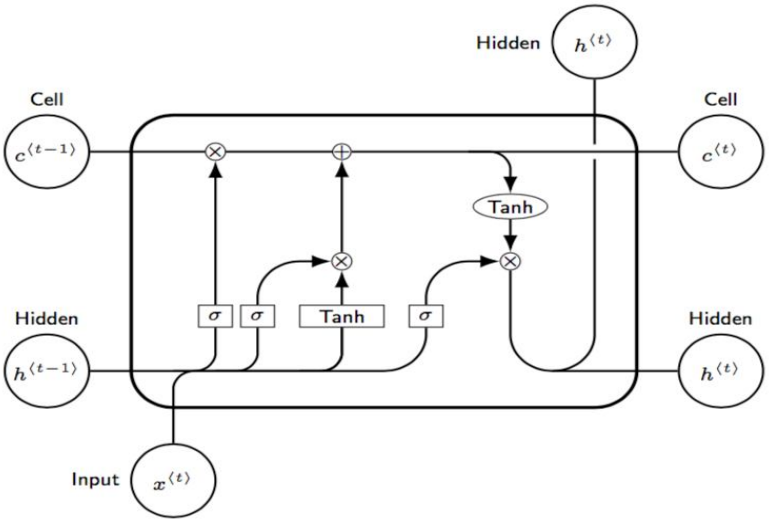


Fig 3.2: LSTM model [6]

C. Linear Regression

In Linear Regression model, the simulation of equation linearity is used to combine a input data set of values (x) to the predicted output data set of input values (y). Both the input and output variables and values are treated as integers. The variable integer assigned by the equation of Linear Regression is represented using the capital Greek letter Beta (B) and is most commonly known as the coefficient. In addition to this, another coefficient is added to give the line an extra degrees of freedom. This extra term is commonly known as the bias coefficient. Often, the bias coefficient is calculated or otherwise estimated by finding the distance of our equation points from the best fit line. This may be represented as a straight line at right angles to the vertex and calculated using slope of the line. Mathematically, the tangent of the line is used to estimate its proximity to the relative equation of Linear Regression

The equation of a problem model in Linear Regression would be given as follows:

y = B0 + Bt*xt+Et

This same line is also called a plane or a hyper-plane when we are dealing with more than one inputs. This is often the case with higher dimensional data. The model of Linear Regression is therefore, represented in the form of the equation and introverted and estimated values used for specific coefficients. However, before using this linear equation, we are faced with several issues. These issues often increase the complexity of the model making precise estimation difficult. This complexity is usually discussed in terms of the number of dependent and independent variables.

The influence of the input variable on the model is effectively hampered when a particular coefficient becomes zero. Therefore, due to null values, the accuracy is reduced for the prediction made from the model (0 * x = 0). When we analyse regularization methods which are capable of modifying learning algorithm to reduce the complexity of models by emphasizing the importance on the absolute size of the coefficients, driving some to zero, this specific case becomes relevant.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue III Mar 2020- Available at www.ijraset.com

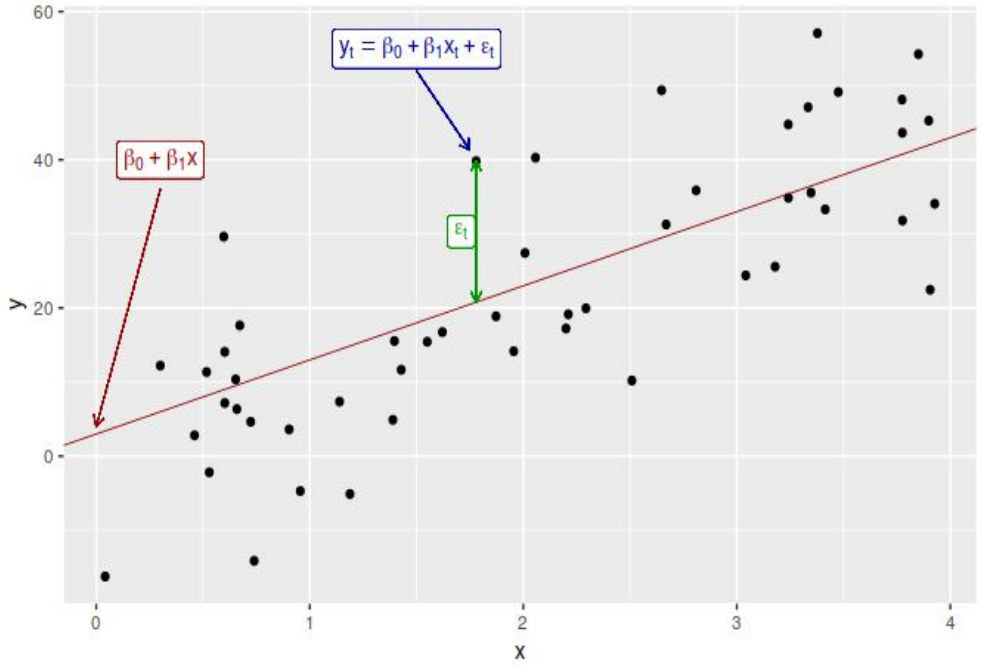


Fig 3.3: Linear Regression [7]

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Fetching and Visualising NASDAQ Data

The NASDAQ (American Company) stock data for the past 2 years along with real time prices is fetched from the Yahoo Finance API and visualized in python.

Fig 4.1 Historic Stock Data for NASDAQ (AAPL) stock

```
##################################################################################
Today's AAPL Stock Data:
          Date        Open        High         Low       Close   Adj Close     Volume
504  2020-02-28  257.26001  278.410004  256.369995  273.359985  273.359985  106627500
##################################################################################
```

Fig 4.2 Real Time Stock Data for NASDAQ (AAPL) stock

B. ARIMA forecast for NASDAQ stock

ARIMA model was applied to the test set data (20% of the entire dataset). The predicted values were compared against the actual values and the results were visualised in python.

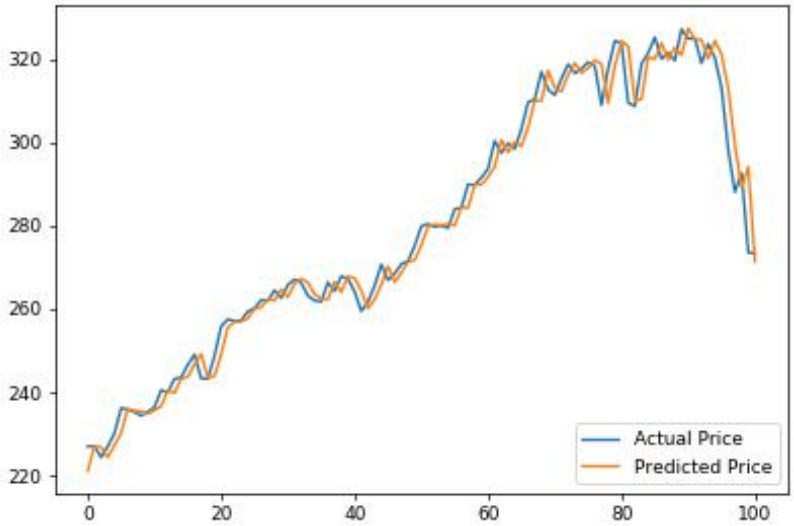


Fig 4.3: ARIMA forecast for NASDAQ (AAPL) stock

```
################################################################################
Tomorrow's AAPL  Closing Price Prediction by ARIMA: 294.1296773532678
ARIMA RMSE: 4.652448775151771
################################################################################
```

Fig 4.4: ARIMA prediction and Root Mean Squared Error (RMSE) for NASDAQ (AAPL) stock

C. LSTM forecast for NASDAQ stock

LSTM model was applied to the test set data (20% of the entire dataset). The predicted values were compared against the actual values and the results were visualised in python.

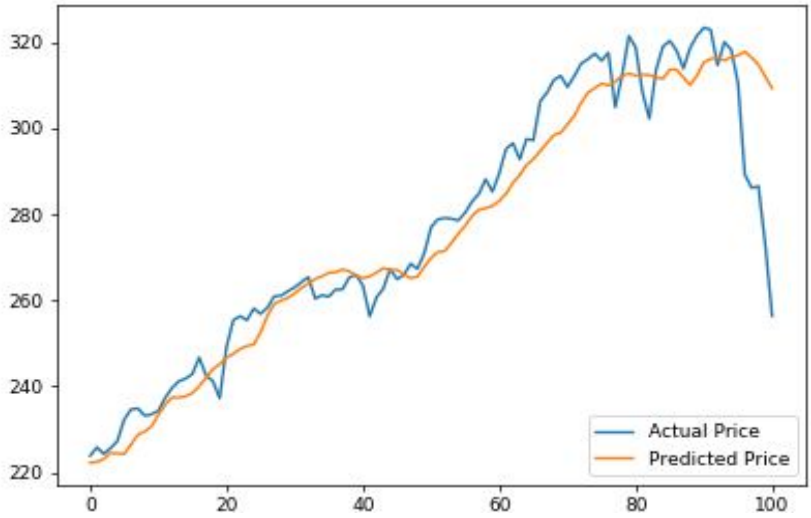


Fig 4.5: LSTM forecast for NASDAQ (AAPL) stock

```
################################################################################
Tomorrow's  AAPL  Closing Price Prediction by LSTM:  300.1012
LSTM RMSE: 10.040912573271779
################################################################################
```

Fig 4.6: LSTM prediction and Root Mean Squared Error (RMSE) for NASDAQ (AAPL) stock

D. Linear Regression forecast for NASDAQ stock

Linear Regression model was applied to the test set data (20% of the entire dataset). The predicted values were compared against the actual values and the results were visualised in python.

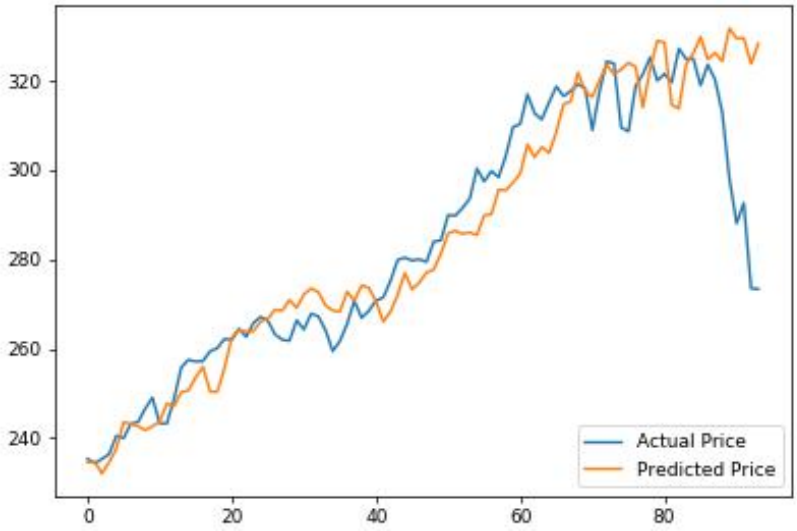


Fig 4.7: Linear Regression forecast for NASDAQ (AAPL) stock

```
##########################################################################################
Tomorrow's  AAPL  Closing Price Prediction by Linear Regression:  325.0728229041051
Linear Regression RMSE: 12.014273414277266
##########################################################################################
```

Fig 4.8: Linear Regression prediction and Root Mean Squared Error (RMSE) for NASDAQ (AAPL) stock

E. Fetching and Visualising NSE Data

The NSE (Indian Company) stock data for the past 2 years along with real time prices is fetched from the Alpha Vantage API and visualized in python.

Fig 4.9 Historic Stock Data for NSE (HDFCBANK) stock

```
##########################################################################################
Today's HDFCBANK Stock Data:
           Date    Open    High     Low   Close  Adj Close      Volume
502  2020-02-28  1175.5  1185.0  1170.1  1177.65    1177.65  12155940.0
##########################################################################################
```

Fig 4.10 Real Time Stock Data for NSE (HDFCBANK) stock

F. ARIMA forecast for NSE stock

ARIMA model was applied to the test set data (20% of the entire dataset). The predicted values were compared against the actual values and the results were visualised in python.

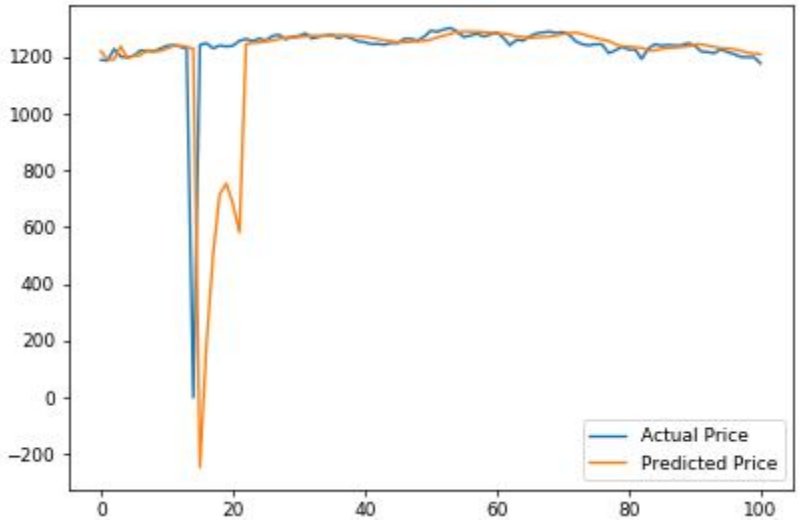


Fig 4.11: ARIMA forecast for NSE (HDFCBANK) stock

```
##########################################################################################
Tomorrow's HDFCBANK  Closing Price Prediction by ARIMA: 1212.1630564293232
ARIMA RMSE: 257.1649710815921
##########################################################################################
```

Fig 4.12: ARIMA prediction and Root Mean Squared Error (RMSE) for NSE (HDFCBANK) stock

G. LSTM forecast for NSE stock

LSTM model was applied to the test set data (20% of the entire dataset). The predicted values were compared against the actual values and the results were visualised in python.

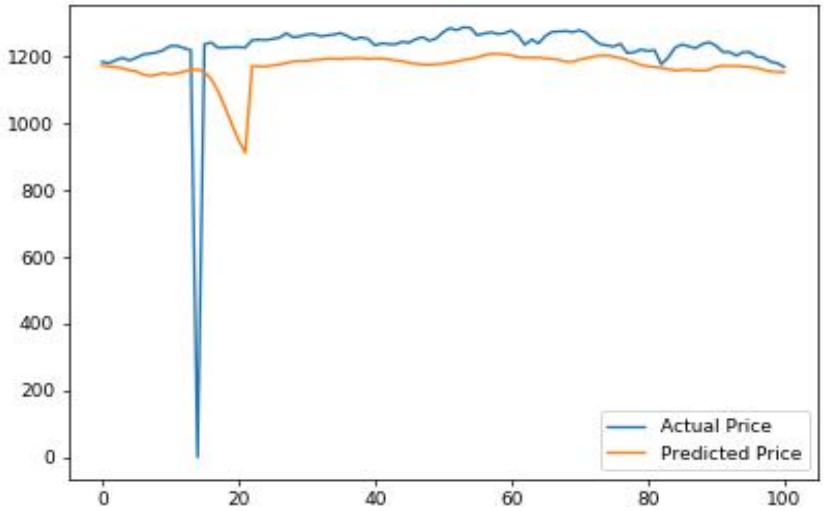


Fig 4.13: LSTM forecast for NSE (HDFCBANK) stock

```
##########################################################################################
Tomorrow's  HDFCBANK  Closing Price Prediction by LSTM:  1146.6365
LSTM RMSE: 141.62551994527215
##########################################################################################
```

Fig 4.14: LSTM prediction and Root Mean Squared Error (RMSE) for NSE (HDFCBANK) stock

H. Linear Regression forecast for NSE stock

Linear Regression model was applied to the test set data (20% of the entire dataset). The predicted values were compared against the actual values and the results were visualised in python.

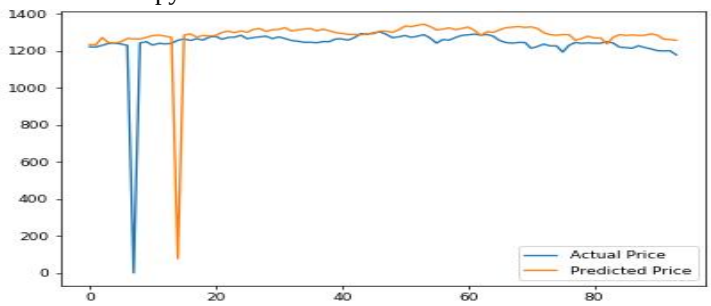


Fig 4.15: Linear Regression forecast for NSE (HDFCBANK) stock

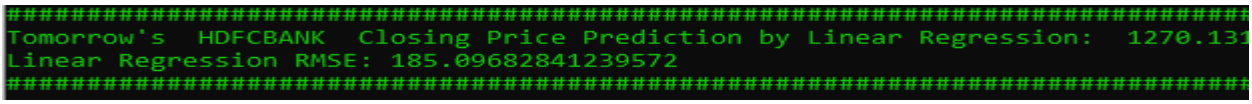


Fig 4.16: Linear Regression prediction and Root Mean Squared Error (RMSE) for NSE (HDFCBANK) stock

I. Comparison of Model Performance of implemented models

The Root Mean Squared Error (RMSE) of ARIMA, LSTM and Linear Regression models for NASDAQ and NSE stocks are tabulated and compared below. It is seen that ARIMA and LSTM models have a lower error rate than Linear Regression model for forecasting NASDAQ (American Company) stocks. Whereas, for NSE (Indian Company) stocks, LSTM and Linear Regression have a lower error rate than ARIMA.

[Table 4-1: Comparison of Model Performance]

RMSE	ARIMA	LSTM	Linear Regression
NASDAQ (American) stocks	4.65	10.04	12.01
NSE (Indian) stocks	257.16	141.62	185.09

V. CONCLUSION

The proposed algorithms worked very well with the stock market data of NASDAQ and NSE stocks. From the plots and tables presented above, it is seen that though the model predictions somewhat deviate from the actual values, they give a good estimation of future trends in stock prices. This estimation helps in getting valuable insights about the stocks, thereby aiding smart investment decisions. It is observed that ARIMA and LSTM models are more consistent than Linear Regression model for forecasting NASDAQ (American Company) stocks. Whereas, for NSE (Indian Company) stocks, LSTM and Linear Regression prove to be more efficient than ARIMA. This further supports the argument that different models and algorithms react differently to stocks belonging to different indexes. So, one should select models and algorithms depending upon the scale and indexes of their respective stocks.

REFERENCES

[1] P. D. Yoo, M. H. Kim and T. Jan, "Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation," International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), Vienna, 2005, pp. 835-841.

[2] M. İ. Y. Kaya and M. E. Karsligil, "Stock price prediction using financial news articles," 2010 2nd IEEE International Conference on Information and Financial Engineering, Chongqing, 2010, pp. 478-482.

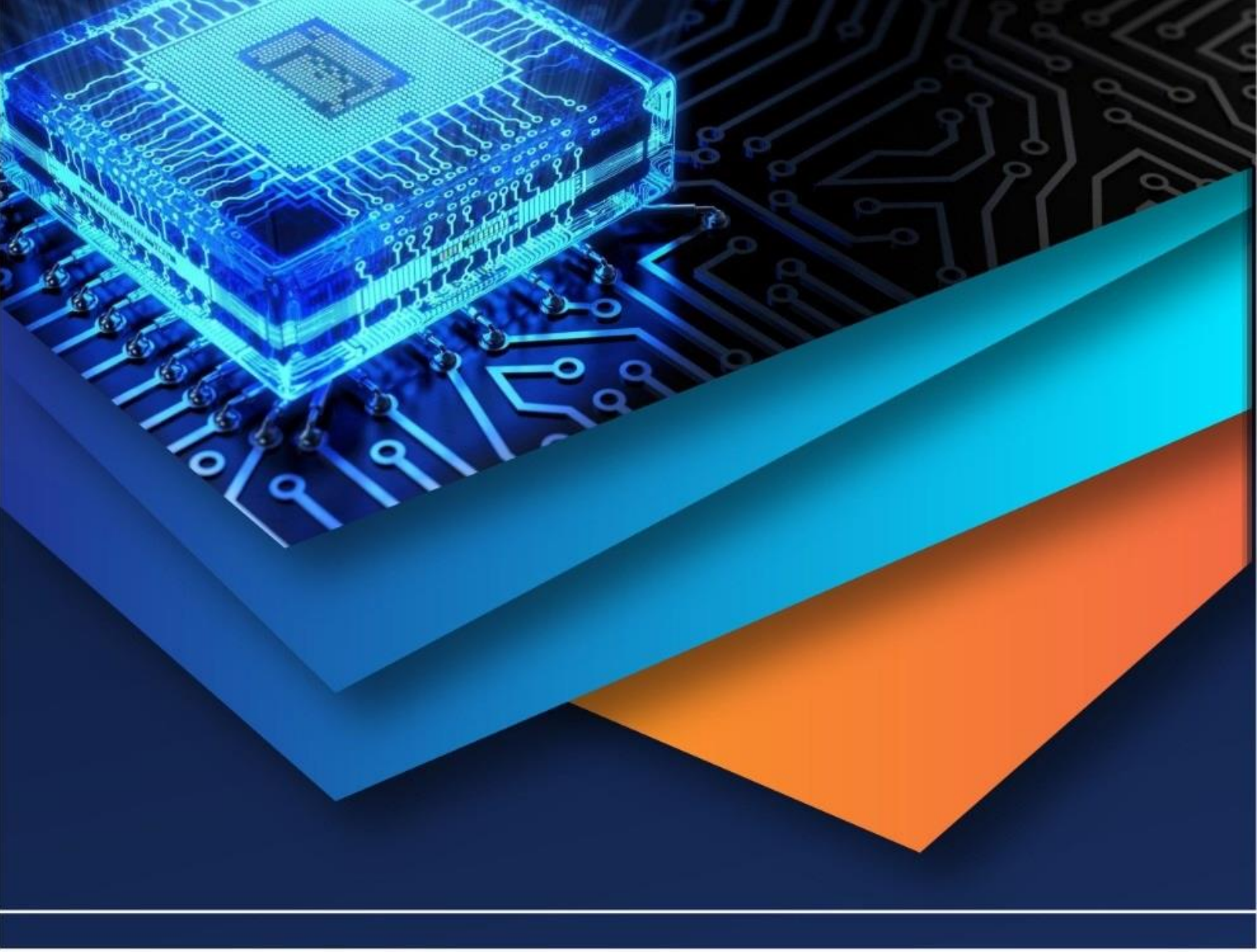
[3] Hedayati, Amin & Moghaddam, Moein & Esfandyari, Morteza. (2016). Stock market index prediction using artificial neural network:. Journal of Economics, Finance and Administrative Science. 10.1016/j.jefas.2016.07.002.

[4] Ayodele A. Adebiyi., Aderemi O. Adewumi, "Stock Price Prediction Using the ARIMA Model", IJSST, Volume-15, Issue-4. [Online]. Available :https://ijssst.info/Vol-15/No-4/data/4923a105.pdf

[5] Chen, Peiyuan. (2020). STOCHASTIC MODELING AND ANALYSIS OF POWER SYSTEM WITH RENEWABLE GENERATION, Research Gate Publication

[6] Angle Qian (2018), Structure of LSTM RNNs, Stack Exchange [Online]. Available: https://ai.stackexchange.com/questions/6961/structure-of-lstm-rnns

[7] Rob J Hyndman and George Athanasopoulos, Forecasting: Principles and Practice, OTexts, Kindle Edition. [Online]. Available: https://otexts.com/fpp2/



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY