iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35822

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 9 Issue VI Jun 2021- Available at www.ijraset.com

Machine Learning Model for Stock Market Prediction

Dr. S. T. Patil[1], Madhura Ashok Mane[2]
[1]*Professor, Vishwakarma Institute of Technology, Pune*
[2]*M. Tech. Student, Vishwakarma Institute of Technology, Pune*

Abstract: ***In recent time's stock market predictions is gaining more attention, maybe due to the fact that if the trend of the market is successfully predicted, the investors may be better guided. A stock exchange is a system where you can buy and sell stocks. By stock we mean the share in the ownership of the company. Companies buy stocks to get the money they need to grow. Whereas people buy the stocks, also called as securities as investment or ways of possibly earning money. A stock Market Prediction model will help people to predict particular company's stock price before they want to invest. This system will help people to invest wisely.***

Keywords: Stock Market Prediction, Machine Learning, Deep learning, ANN, RNN, LSTM, Sentiment analysis, Predictive analysis

I. INTRODUCTION

Now a days, there are many new ways of investments where in people can put their money to earn more and more money. One of the popular ways for investment is into the Stock Market. Stock Market is basically a system to buy and sell stocks, and stocks are nothing but the share in the ownership of the company. Companies sell these stocks to make their organization grow and people buy these stocks as a way of earning money with the investment. Various companies belong to each stock market. The companies sell these stocks to people and people then use these stocks to trade the stocks among themselves. The price of these stocks may rise or fall throughout the day. Every day the stock market is opened. Many factors affects the price of the stock. If the company is successful, the price of the stock will usually go up.

Basically, an ability to predict direction of stock price accurately is crucial for market dealers or investors to maximize their profits. Machine learning techniques have been successfully shown to generate high forecasting accuracy of stock market future by applying various algorithms over the data. Stock market plays an important role in the economic performance. Prediction of stock market is a challenging task because of its randomness in nature. Researchers have made several attempts to predict financial market values using various techniques. Stock price prediction is a classic and important problem. The motivated idea is that, if we know all information about today's stock trading (of all specific traders), the price is predictable. Thus, if we can obtain just a partial information, we can expect to improve the current prediction lot.

II. LITERATURE SURVEY

Predicting stock prices is an uncertain task which is modelled using machine learning to predict the return on stocks. There are a lot of methods and tools used for the purpose of stock market prediction. The stock market is considered to be very dynamic and complex in nature. An accurate prediction of future prices may lead to a higher yield of profit for investors through stock investments. As per the predictions, investors will be able to pick the stocks that may give a higher return.

Over the years, various machine learning techniques have been used in stock market prediction, but with the increased amount of data and expectation of more accurate prediction, the deep learning models are being used nowadays which have proven their advantage over traditional machine learning methods in terms of accuracy and speed of prediction.

1) *Linear Regression:* Linear regression is a linear approach to modeling the relationship between a dependent variable and one or more independent variables. The way we are going to use linear regression here is that we will fit a linear regression model to the previous N values, and use this model to predict the value on the current day.
2) *CNN:* CNN was widely used in the field of image recognition because of its powerful pattern recognition ability; its use was also extended to the field of economic prediction. Similar to the traditional neural network, CNN is composed of multiple neurons connected by a hierarchical structure, and the weights and bias between layers can be trained. CNN is different from the network structure of a fully connected network such as deep brief network (DBN), Sparse Auto encoder (SAE), back propagation (BP), as the CNN can share the weight among the neurons in each layer of the network. Hence, the model

significantly reduces the weight of the network and avoids falling into dimensional disaster and local minimization. If the characteristics of the stock market at a specific time point are regarded as a feature graph, CNN has the potential to extract the characteristics of the stock market at the corresponding period from these feature graphs. Therefore, CNN can be used to build a timing-selection model and can ultimately be used to complete the construction of the timing-selection strategy.

3) *ANN:* Most research with machine learning forecasting has focused on Artificial Neural Networks (ANN) [4]. ANNs have a series of interconnected nodes that simulate individual neurons, and are organized into different layers based on function (input layer, processing layer, output layer, etc.). The ANN assigns weights to connections, and the output is calculated based on the inputs and the weights. As the machine trains, it notices patterns in the training data and reassigns the weights.
4) *SVM:* Recent research in the field has used another technique known as Support Vector Machines in addition to or as an alternative to ANNs. Whereas ANNs are models that try to minimize classification error within the training data, SVMs may make classification errors within training data in order to minimize overall error across test data. A major advantage of SVMs is that it finds a global optimum, whereas neural networks may only find a local optimum.
5) *RNN:* RNN belongs to the neural network, and it is good at modelling and processing sequential data. The specific expression is that the RNN is able to memorize the previous state, and the previous state can be used in the current state calculation. The different hidden layers are non-independent, and the input of the current hidden layer includes not only the output of the input layer but also the output of the previously hidden layer. For this reason, RNN has a good performance in dealing with sequential data. The advantage of RNN is that it considers the context of data in the process of training, which is very suitable for the scenario of stocks and Forex because the fluctuation at a particular time often contains some connection to the previous trend.
6) *LSTM:* The LSTM model is one of the variants of the RNN. Its core contribution is to introduce the design of self-loop to generate the path of a gradient which could continuously flow for an extended period. The weight of the self-loop is also updated in each iteration, which solves the gradient vanishing problem that is easily generated when the RNN model updates the weights. The modelling of a time series is essentially a process of nonlinear parameter fitting. The LSTM model performs well to reveal the correlation of a nonlinear time-series in the delay state space and to realize the purpose of stock prediction. The stock or Forex trend prediction model based on LSTM obtained the corresponding data characteristics from the stock or Forex history data.

III. PROPOSED SYSTEM

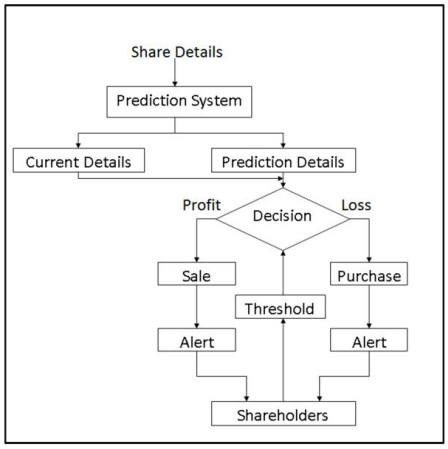


Fig 1. Stock Market Prediction System Flow

Long Short-Term Memory (LSTM) is one of many types of Recurrent Neural Network RNN, it's also capable of catching data from past stages and use it for future predictions. In general, an Artificial Neural Network (ANN) consists of three layers:

1) Input layer,
2) Hidden layers,
3) Output layer.

In a NN that only contains one hidden layer the number of nodes in the input layer always depend on the dimension of the data, the nodes of the input layer connect to the hidden layer via links called 'synapses'. The relation between every two nodes from (input to the hidden layer), has a coefficient called weight, which is the decision maker for signals. The process of learning is naturally a continues adjustment of weights, after completing the process of learning, the Artificial NN will have optimal weights for each synapses. The hidden layer nodes apply a sigmoid or tangent hyperbolic (tanh) function on the sum of weights coming from the input layer which is called the activation function, this transformation will generate values, with a minimized error rate between the train and test data using the SoftMax function. The values obtained after this transformation constitute the output layer of our NN, these value may not be the best output, in this case a back propagation process will be applied to target the optimal value of error, the back propagation process connect the output layer to the hidden layer, sending a signal conforming the best weight with the optimal error for the number of epochs decided. This process will be repeated trying to improve our predictions and minimize the prediction error. After completing this process, the model will be trained. The classes of NN that predict future value base on passed sequence of observations is called Recurrent Neural Network (RNN) this type of NN make use of earlier stages to learn of data and forecast futures trends. The earlier stages of data should be remembered to predict and guess future values, in this case the hidden layer act like a stock for the past information from the sequential data. The term recurrent is used to describe the process of using elements of earlier sequences to forecast future data. RNN can't store long time memory, so the use of the Long Short-Term Memory (LSTM) based on "memory line" proved to be very useful in forecasting cases with long time data. In a LSTM the memorization of earlier stages can be performed trough gates with along memory line incorporated. The Figure 4 describe the composition of LSTM nodes.

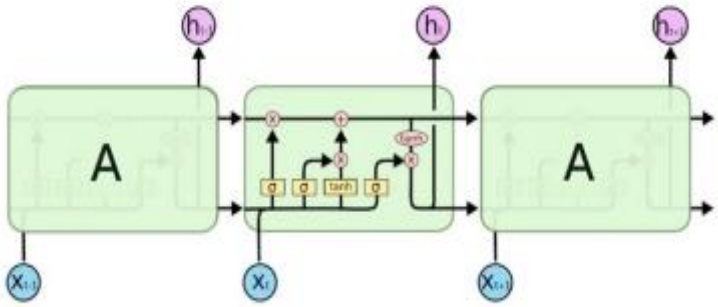


Fig 2. Internal structure of LSTM

A. Implementation Details

Web-scraping will be performed for data collection. For web-scraping we used Beautiful Soup. The latest data is got from Yahoo finance and then stored it in CSV file. Then this csv file (data) is loaded using Pandas.

The obtained data contained five features:

1) *Date:* Date of stock price.
2) *Opening Price:* When trading begins each day this is opening price of stock.
3) *High:* The highest price at which the stock was traded during a period(day).
4) *Low:* The Lowest price at which the stock was traded during a period(day).
5) *Volume:* How much of a given financial asset has traded in a period of time.
6) *Close Interest:* The last price at which a particular stock traded for the trading session.

Stock market information is available from key sources: Tiingo API, Yahoo and Google Finance. These websites give APIs from which stock dataset can be obtained from various companies by simply specifying parameters.

The data is processed into a format suitable to use with prediction model by performing the following steps:

Transformation oftime-series data into input-output components for supervised learning.

Data cleaning is performed.

Normalization (Min-max scalar): The min-max scalar form of normalization uses the mean and standard deviation to box all the data into a range lying between a certain min and max value. For most purposes, the range is set between 0 and 1. At other times, other ranges may be applied but the 0 to 1 range remains the default.

$$Xsc=X-Xmin/Xmax-Xmin.$$

LSTM: The ability of memorizing sequence of data makes the LSTM a special kind of RNNs. Every LSTM node most be consisting of a set of cells responsible of storing passed data streams, the upper line in each cell links the models as transport line handing over data from the past to the present ones, the independency of cells helps the model dispose filter of add values of a cell to another. In the end the sigmoidal neural network layer composing the gates drive the cell to an optimal value by disposing or letting data pass through. Each sigmoid layer has a binary value (0 or 1) with 0 “let nothing pass through”; and 1 “let everything pass through.” The goal here is to control the state of each cell, the gates are controlled as follow:

Forget Gate outputs a number between 0 and 1, where 1 illustration “completely keep this”; whereas, 0 indicates “completely ignore this.”

Memory Gate chooses which new data will be stored in the cell. First, a sigmoid layer “input door layer” chooses which values will be changed. Next, a tanh layer makes a vector of new candidate values that could be added to the state.

Output Gate decides what will be the output of each cell. The output value will be based on the cell state along with the filtered and freshest added data.

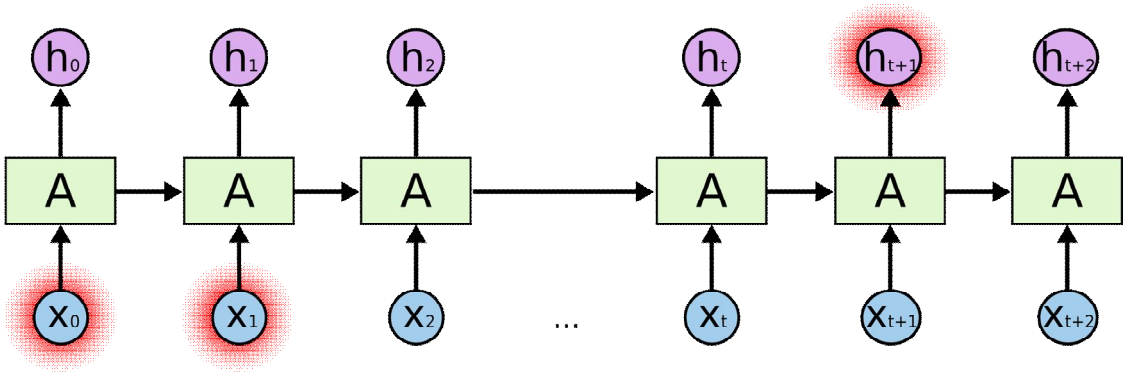


Fig 3. Layers in LSTM

The main advantage of LSTM is its ability to read intermediate context. Each unit remembers details for a long or short period without explicitly utilizing the activation function within the recurring components. An important fact is that any cell state is repeated only with the release of the forget gate, which varies between 0 and 1. That is to say, the gateway for forgetting in the LSTM cell is responsible for both the hardware and the function of the cell state activation. Thus, the data from the previous cell can pass through the unchanged cell instead of explicitly increasing or decreasing in each step or layer, and the instruments can convert to their appropriate values over a limited time. This allows LSTM to solve a perishable gradient problem - because the amount stored in the memory cell is not converted in a recurring manner, the gradient does not end when trained to distribute backwards.

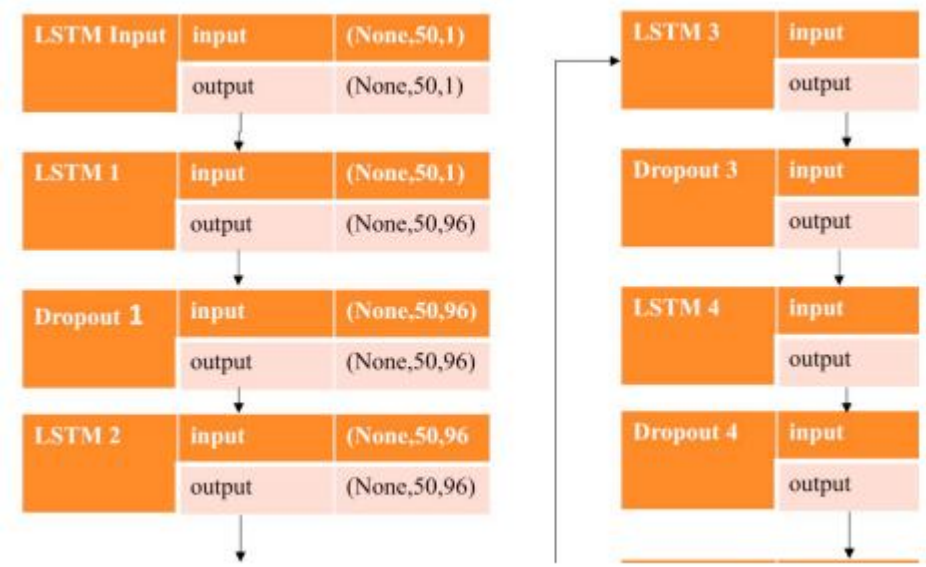


Fig 4. LSTM Model Structure

IV. RESULT

The basic stock prediction model is able to predict stock prices for coming 30 days depending on the historical data. Data parameters considered for testing are previous date, opening price and closing price. 70%of the data was given for training and 30% is used for testing. Root mean square error was used to calculate the accuracy of the model. Following figure shows the raw data used for prediction.

	High	Low	Open	Close	Volume	Adj Close
Jun 7, 2021 5:30 AM	898.8000	835.6500	837.4000	878.6000	33684310	878.6000
Jun 8, 2021 5:30 AM	887.8000	859.1500	883.4000	874.2000	16422574	874.2000
Jun 9, 2021 5:30 AM	901	831.3000	879.2500	852.6000	31225441	852.6000
Jun 10, 2021 5:30 AM	863.9500	843.3000	858.5500	846.7500	11744190	846.7500
Jun 11, 2021 5:30 AM	856.4000	820.7000	852	839.7500	17460610	839.7500
Jun 14, 2021 5:30 AM	783.7000	681.1000	755.8000	768.4500	89771631	768.4500
Jun 15, 2021 5:30 AM	802.9000	755	802.9000	761.8500	45408426	761.8500
Jun 16, 2021 5:30 AM	755	697.8000	755	706.7000	29440056	706.7000
Jun 17, 2021 5:30 AM	704.6500	638.3500	683	646.9000	70812142	646.9000
Jun 18, 2021 5:30 AM	704.6000	638.1000	643.2000	694.3500	76238380	694.3500

Fig 5. Raw data

Figure 5 shows raw data (only for last 10 days data frame) used for prediction for stock name ADANIPORTS. Figure 6 below shows Close Price VS Date Interactive chart for analysis:

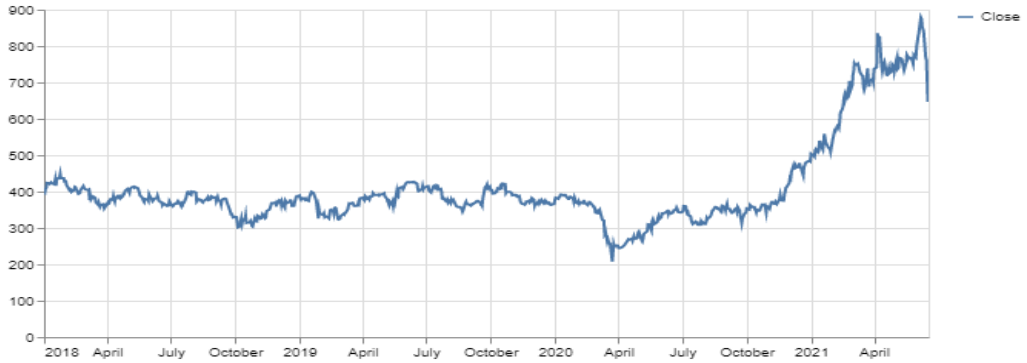


Fig 6. Close Price VS Date Interactive chart

Figure 7 below shows curve of prediction.

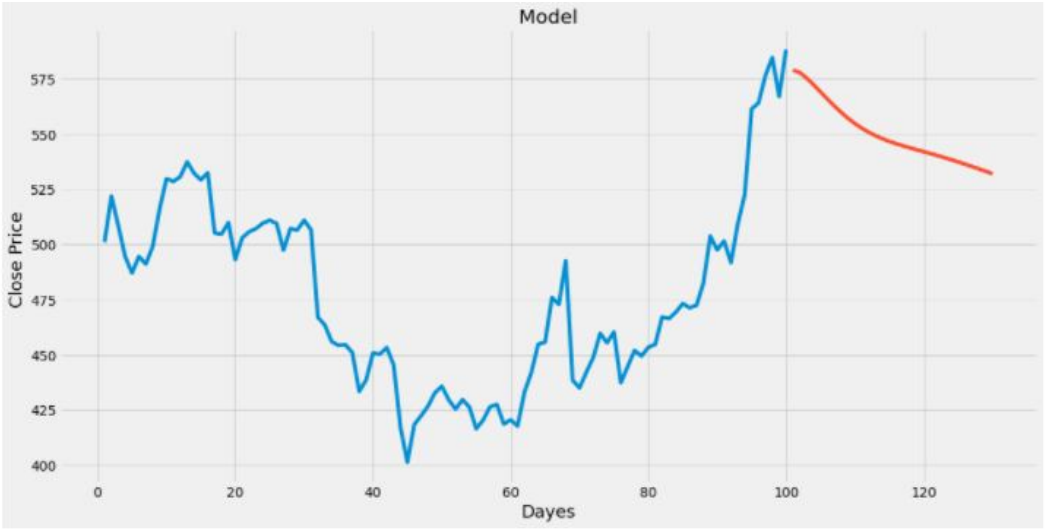


Fig 7. Curve of prediction

Figure 8 below shows graph for trained data set, actual and predicted values.

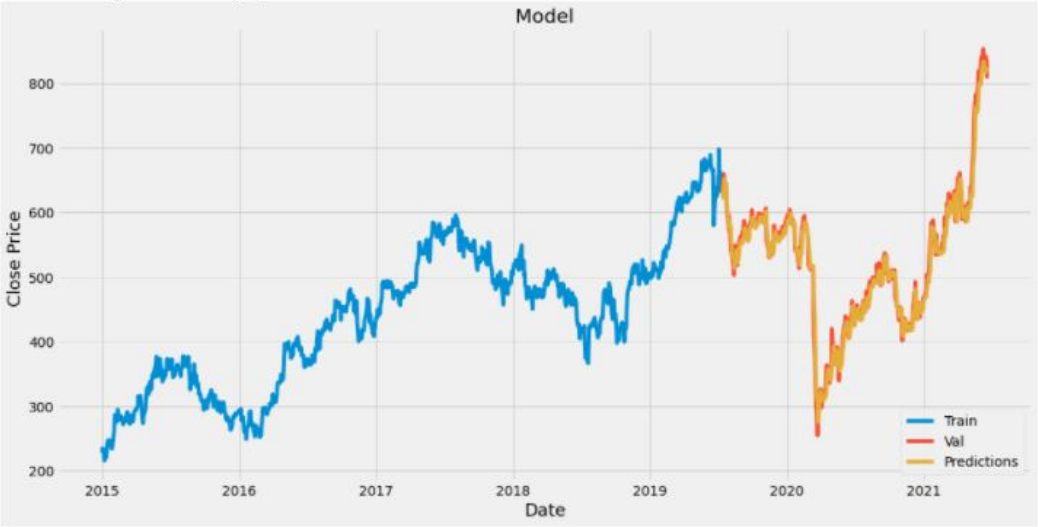


Fig 8. Actual graph of predicted values

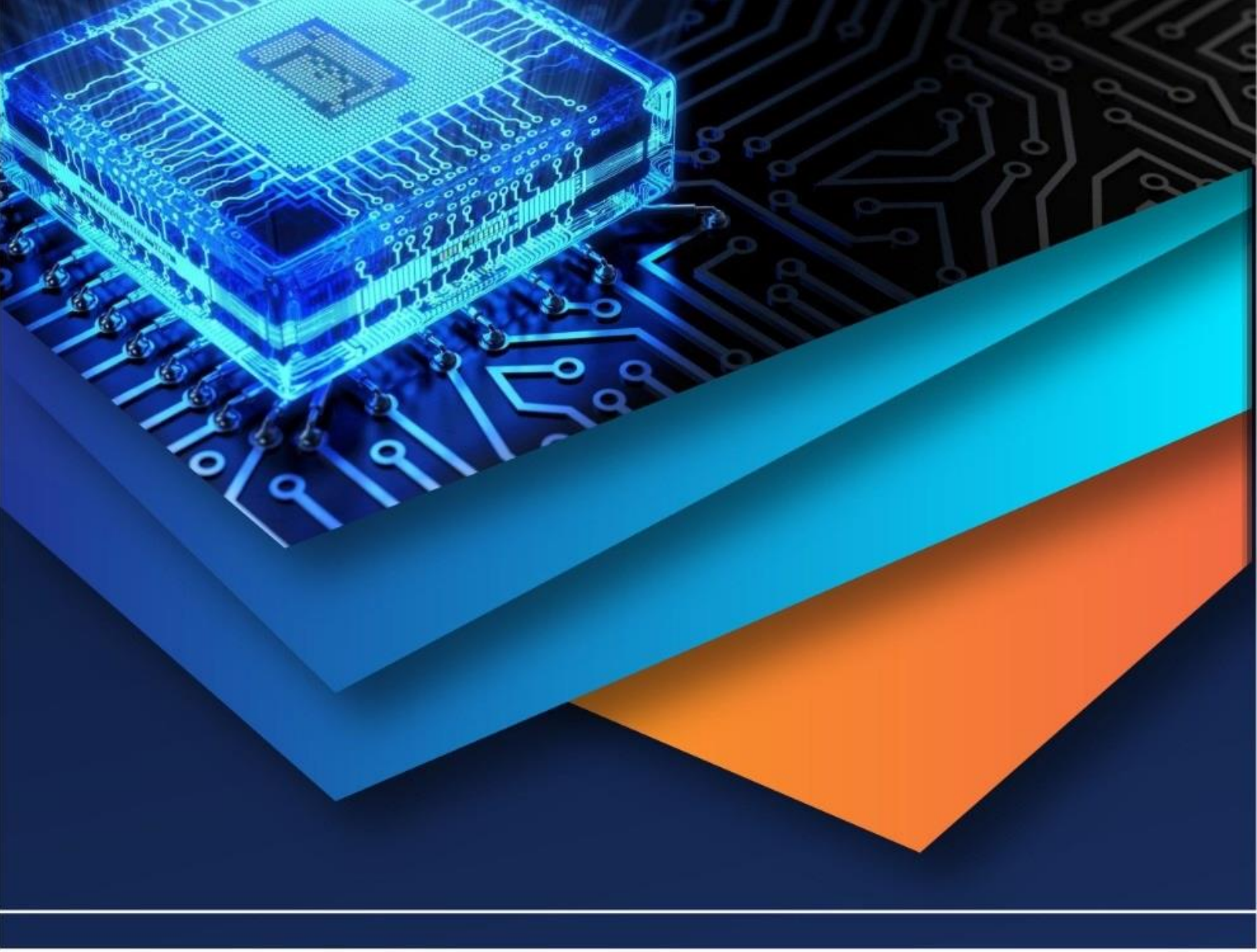
Accuracy - This performance measure is most intuitive and can be calculated as ration between correctly predicted observations to total number of observations. The more the accuracy the better the model has performed. In this case the accuracy is about 85%.
Root Mean Square Error (RMSE) – It is the standard deviation of the residuals (prediction errors). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the line of best fit. Root mean square error is commonly used in climatology, forecasting, and regression analysis to verify experimental results.

V. CONCLUSION

Stock Market Prediction model is built with RNN along with LSTM algorithm to forecast future values to get some promising predictions. To increase the accuracy of the model, hyper parameter tuning is performed. To evaluate the accuracy of the model confusion matrix is used.

REFERENCES

[1] V Kranthi Sai Reddy | Stock Market Prediction Using Machine Learning | International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 10 | Oct 2018 www.irjet.net p-ISSN: 2395-0072
[2] Adil MOGHAR ,Mhamed HAMICHE | Stock Market Prediction Using LSTM Recurrent Neural Network | International Workshop on Statistical Methods and Artificial Intelligence (IWSMAI 2020) April 6-9, 2020, Warsaw, Poland
[3] Zineb Lanbouri , Said Achcha | Stock Market prediction on High frequency data using Long-Short Term Memory | International Workshop on Artificial Intelligence & Internet of Things (A2IoT) August 9-12, 2020, Leuven, Belgium
[4] Anita Yadava, C K Jhaa, Aditi Sharan| Optimizing LSTM for time series prediction in Indian stock market | International Conference on Computational Intelligence and Data Science (ICCIDS 2019)
[5] Ishita Parmar | Stock Market Prediction Using Machine Learning | 2018 First International Conference on Secure Cyber Computing and Communication(ICSCCC)
[6] Sahaj Singh Maini | Stock Market Prediction using Data Mining Techniques | Proceedings of the International Conference IEEE Xplore Compliant - Part Number:CFP17M19-ART, ISBN:978-1-5386-1959-9
[7] Vivek Rajput , SarikaBobde Stock Market Prediction Using Hybrid Approach | International Conference on Computing, Communication and Automation (ICCCA2016) SBN:978-1-5090-1666-2/16/$31.00 ©2016 IEEE 82
[8] Radu Iacomin | Stock Market Prediction | 2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, Cheile Gradistei, Romania
[9] Yunus YETISl, Halid KAPLAN | Stock Market Prediction by Using Artificial Neural Network | WAC 2014 1569925719
[10] Chang Sim Vui, Gan Kim Soon | A Review of Stock Market Prediction with Artificial Neural Network (ANN) | 2013 IEEE International Conference on Control System, Computing and Engineering, 29 Nov. - 1 Dec. 2013, Penang, Malaysia



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY