iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43976

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 VI June 2022- Available at www.ijraset.com

Stock Market Prediction Using LSTM Technique

Drashti Talati[1], Dr. Miral Patel[2], Prof. Bhargesh Patel[3]
[1]*ME Student,* [2]*Associate Professor,* [3]*Assistant Professor, IT Department, G.H. Patel College of Engineering & Technology, CVM University, Vallabh Vidhyanagar*

Abstract: ***One of the most intricate machine learning problems is the share value prediction. Stock market prediction is an activity in which investors need fast and accurate information to make effective decisions. Moreover, the behavior of stock prices is uncertain and hard to predict. For these reasons, stock price prediction is an important process and a challenging one. This leads to the research of finding the most effective prediction model that generates the most accurate prediction with the lowest error percentage. Prices of stocks are depicted by time series data and neural networks are trained to learn the patterns from trends in the existing data. This system employed algorithm using LSTM to improve the accuracy of stock price prediction.***
Keywords: KNN, SVM, LSTM, NB, Machine Learning, Deep Learning, Finance, Stock Price Prediction

I. INTRODUCTION

The stock market is known for its volatility, randomness, and unpredictability. It is a chaotic place with an unbelievably huge continuously changing stream of data which makes predicting and acting on those predictions to make a profit very hard. It is actually one of the most challenging tasks in times series forecasting.

This research's main goal is to study and apply deep learning techniques to the stock market in order to predict stock behavior and thus act on those predictions to avoid investment risk and generate profit. The goal is to be achieved by using transfer learning in order to take advantage of pre-built neural networks models. Predictions are then tested against actual historical stock price data. This research will be a helpful tool that aims to help beginner traders make better decisions.

In order to do so, many tools will be used to accurately reach the objectives of this research. Deep learning Studio (software) is a great starting point especially for beginners in the field as it helps to easily create different neural network models to see which works best in the case of times series forecasting. As for the model and languages to be used, after a thorough research the programming language to be used for implementation will be Python, this is due to its flexibility and the availability of pre-built models and open source particularly useful libraries that can help us with our goal and maybe even enhance results.

In addition, this paper will cover a simple example of the most fitting model (the one that yields the best results) in the case of time series forecasting which is certainly the LSTM model that stands for Long Short Term Memory. Compared to a conventional deep neural network, its effectiveness is due to the addition of a crucial component in time series predictions, the memory component.

II. MOTIVATION

Stock market prediction is basically defined as trying to determine the stock value and offer a robust idea for the people to know and predict the market and the stock prices. It is generally presented using the quarterly financial ratio using the dataset. Thus, relying on a single dataset may not be sufficient for the prediction and can give a result which is inaccurate. Hence, we are contemplating towards the study of machine learning with various datasets integration to predict the market and the stock trends (1).

The problem with estimating the stock price will remain a problem if a better stock market prediction algorithm is not proposed. Predicting how the stock market will perform is quite difficult. The movement in the stock market is usually determined by the sentiments of thousands of investors. Stock market prediction, calls for an ability to predict the effect of recent events on the investors. These events can be political events like a statement by a political leader, a piece of news on scam etc. It can also be an international event like sharp movements in currencies and commodity etc. All these events affect the corporate earnings, which in turn affects the sentiment of investors. It is beyond the scope of almost all investors to correctly and consistently predict these hyper parameters. All these factors make stock price prediction very difficult. Once the right data is collected, it then can be used to train a machine and to generate a predictive result (1).

III. RELATED WORKS

Stock market prediction is basically defined as trying to determine the stock value and offer a robust idea for the people to know and predict the market and the stock prices.

©IJRASET: All Rights are Reserved

Mehar Vijh et al. (2019) proposed Artificial Neural Network and Random Forest techniques that have been utilized for predicting the next day closing price for five companies belonging to different sectors of operation. The financial data: Open, high, low and close prices of stock are used for creating new variables which are used as inputs to the model. This models are evaluated using standard strategic indicators: RMSE and MAPE. The low values of the two indicators show that the model are efficient in predicting stock closing price. The comparative analysis based on RMSE, MAPE and MBE values clearly indicate that ANN gives better prediction of stock prices as compared to RF (2).

V Kranthi Sai Reddy (2018) propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies. SVM algorithm works on the large dataset value which is collected from different global financial markets. Also, SVM does not give a problem of over fitting (3).

Wasiat Khan et. al. (2020) use algorithms on social media and financial news data to discover the impact of this data on stock market prediction accuracy for ten subsequent days. For improving performance and quality of predictions, feature selection and spam tweets reduction are performed on the data sets. Moreover, we perform experiments to find such stock markets that are difficult to predict and those that are more influenced by social media and financial news. We compare results of different algorithms to find a consistent classifier. Finally, for achieving maximum prediction accuracy, deep learning is used and some classifiers are ensembled. Our experimental results show that highest prediction accuracies of 80.53% and 75.16% are achieved using social media and financial news, respectively (4).

Jingyi Shen et. al. (2020) conducted comprehensive evaluations on frequently used machine learning models and conclude that our proposed solution outperforms due to the comprehensive feature engineering that we built. The system achieves overall high accuracy for stock market trend prediction. With the detailed design and evaluation of prediction term lengths, feature engineering, and data pre-processing methods, this work contributes to the stock analysis research community both in the financial and technical domains (5).

M Umer Ghania et. al. (2019) use Machine Learning Algorithm specially focus on Linear Regression (LR), Three month Moving Average(3MMA), Exponential Smoothing (ES) and Time Series Forecasting using MS Excel as best statistical tool for graph and tabular representation of prediction results. We obtained data from Yahoo Finance for Amazon (AMZN) stock, AAPL stock and GOOGLE stock after implementation LR we successfully predicted stock market trend for next month and also measured accuracy according to measurements (6).

Nusrat Rouf et. al. (2021) explains the systematics of machine learning-based approaches for stock market prediction based on the deployment of a generic framework. Findings from the last decade (2011–2021) were critically analyzed, having been retrieved from online digital libraries and databases like ACM digital library and Scopus. Furthermore, an extensive comparative analysis was carried out to identify the direction of significance. The study would be helpful for emerging researchers to understand the basics and advancements of this emerging area, and thus carry-on further research in promising directions (7).

A M Pranav et. al. (2021) use open-source libraries and pre-existing methods to create machine learning models in a WebApp to forecast future stock prices for exchange, in order to help make this volatile kind of commerce a little more predictable. To avoid the conventional method and avoid getting the outcome completely based on numbers, a system to incorporate text-based machine learning model and pattern recognition is implemented. The objective is to create a platform for small and amateur traders, from the existing stock prediction models considering variables like news articles, stock volume, previous close, etc. to predict the future stock market values (8).

Subhadra Kompella et. al. (2019) implemented a Random Forest approach to predict stock market prices. Random Forests are very effectively implemented in forecasting stock prices, returns, and stock modeling. We outline the design of the Random Forest with its salient features and customizable parameters. We focus on a certain group of parameters with a relatively significant impact on the share price of a company. With the help of sentiment analysis, we found the polarity score of the new article and that helped in forecasting accurate result. Although share market can never be predicted with hundred per-cent accuracy due to its vague domain, this paper aims at proving the efficiency of Random forest for forecasting the stock prices. Variance score of Random forest is better than that of logistic regression. Mean absolute score of Random forest is better than that of logistic regression. Mean squared score of Random forest is better than that of logistic regression. Mean squared log error score of Random forest is better than that of logistic regression. In all, it can be concluded that the random forest algorithm is much efficient compared to logistic regression for the stock market prediction based on sentiment analysis (9).

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

K. Hiba Sadia et. al. (2019) examines the use of the prediction system in real-world settings and issues associated with the accuracy of the overall values given. The paper also presents a machine-learning model to predict the longevity of stock in a competitive market. The successful prediction of the stock will be a great asset for the stock market institutions and will provide real-life solutions to the problems that stock investors face. By measuring the accuracy of the different algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data is the random forest algorithm (1).

Faisal Momin et. al. (2019) examines the use of the prediction system in real-world settings and issues associated with the accuracy of the overall values given. The back propagation gives output as final predicted rate comes. The proposed system can get the output of prediction list of stock price and graph of prediction table like that user can view the final predicted result. The successful prediction of the stock will be a great asset for the stock market institutions and will provide real-life solutions to the problems that stock investors face (10).

Troy J. Strader et. al. (2020) identify directions for future machine learning stock market prediction research based upon a review of current literature. A systematic literature review methodology is used to identify relevant peer-reviewed journal articles from the past twenty years and categorize studies that have similar methods and contexts. Four categories emerge: artificial neural network studies, support vector machine studies, studies using genetic algorithms combined with other techniques, and studies using hybrid or other artificial intelligence approaches. Studies in each category are reviewed to identify common findings, unique findings, limitations, and areas that need further investigation. The final section provides overall conclusions and directions for future research (11).

Yang Li · Yi Pan (2021) proposes a novel deep learning approach to predict future stock movement. The model employs a blending ensemble learning method to combine two recurrent neural networks, followed by a fully connected neural network. In our research, we use the S&P 500 Index as our test case. Our experiments show that our blending ensemble deep learning model outperforms the best existing prediction model substantially using the same dataset, reducing the mean-squared error from 438.94 to 186.32, a 57.55% reduction, increasing precision rate by 40%, recall by 50%, F1-score by 44.78%, and movement direction accuracy by 33.34%, respectively. The purpose of this work is to explain our design philosophy and show that ensemble deep learning technologies can truly predict future stock price trends more effectively and can better assist investors in making the right investment decision than other traditional methods (12).

Mehtabhorn Obthong a et. al. (2020) reviewed and compared the state-of the-art of ML algorithms and techniques that have been used in finance, especially the stock price prediction. The number of ML algorithms and techniques has been discussed in terms of types of input, purposes, advantages, and disadvantages. For stock price prediction, some of ML algorithms and techniques have been popularly selected as to their characteristics, accuracy and error acquired (13).

Aparna Nayak et. al. (2016) build two models: one for daily prediction and the other one is for monthly prediction. Supervised machine learning algorithms are used to build the models. As part of the daily prediction model, historical prices are combined with sentiments. Up to 70% of accuracy is observed using supervised machine learning algorithms on daily prediction model. Monthly prediction model tries to evaluate whether there is any similarity between any two months trend. Evaluation proves that trend of one month is least correlated with the trend of another month (14).

Dharmaraja Selvamuthu et. al. (2019) apply most common techniques used in the forecasting of financial time series are Support Vector Machine (SVM), Support Vector Regression (SVR) and Back Propagation Neural Network (BPNN). All three algorithms provide an accuracy of 99.9% using tick data. The accuracy over 15-min dataset drops to 96.2%, 97.0% and 98.9% for LM, SCG and Bayesian Regularization respectively which is significantly poor in comparison with that of results obtained using tick data (15).

Xiao-Yang Liu et. al. (2020) have presented FinRL library that is a DRL library designed specifically for automated stock trading with an effort for educational and demonstrative purpose. FinRL is characterized by its extendability, more-than-basicmarket environment and extensive performance evaluation tools also for quantitative investors and strategy builders. Customization is easily accessible on all layers, from market simulator, trading agents' learning algorithms up towards profitable strategies (16).

Fuli Feng et. al. (2019) contribute a new deep learning solution, named Relational Stock Ranking (RSR), for stock prediction. Our RSR method advances existing solutions in two major aspects: 1) tailoring the deep learning models for stock ranking, and 2) capturing the stock relations in a time-sensitive manner. The key novelty of our work is the proposal of a new component in neural network modeling, named Temporal Graph Convolution, which jointly models the temporal evolution and relation network of stocks. To validate our method, we perform back-testing on the historical data of two stock markets, NYSE and NASDAQ. Extensive experiments demonstrate the superiority of our RSR method. It outperforms state-of-the-art stock prediction solutions achieving an average return ratio of 98% and 71% on NYSE and NASDAQ, respectively (17).

Raehyun Kima et. al. (2019) propose a hierarchical attention network for stock prediction (HATS) which uses relational data for stock market prediction. Our HATS method selectively aggregates information on different relation types and adds the information to the representations of each company. Specifically, node representations are initialized with features extracted from a feature extraction module. HATS is used as a relational modeling module with initialized node representations. Our method is used for predicting not only individual stock prices but also market index movements, which is similar to the graph classification task. The experimental results show that performance can change depending on the relational data used. HATS which can automatically select information outperformed all the existing methods (18).

Lior Sidi (2021) evaluate the models on seven S&P stocks from various industries over five years period. The prediction model we trained on similar stocks had significantly better results with 0.55 mean accuracy, and 19.782 profit compare to the state of the art model with an accuracy of 0.52 and profit of 6.6 (19).

Ya Gao et. al. (2021) design a new model for optimizing stock forecasting. We incorporate a range of technical indicators, including investor sentiment indicators and financial data, and perform dimension reduction on the many influencing factors of the retrieved stock price using depth learning LASSO and PCA approaches. In addition, a comparison of the performances of LSTM and GRU for stock market forecasting under various parameters was performed. Our experiments show that (1) both LSTM and GRU models can predict stock prices efficiently, not one better than the other, and (2) for the two different dimension reduction methods, both the two neural models using LASSO reflect better prediction ability than the models using PCA (20).

Amin Hedayati Moghaddama et. al. (2016) investigate the ability of artificial neural network (ANN) in forecasting the daily NASDAQ stock exchange rate. Several feed forward ANNs that were trained by the back propagation algorithm have been assessed. The methodology used in this study considered the short-term historical stock prices as well as the day of week as inputs (21).

Khalid Alkhatib et. al. (2013) applied k-nearest neighbor algorithm and non-linear regression approach in order to predict stock prices for a sample of six major companies listed on the Jordanian stock exchange to assist investors, management, decision makers, and users in making correct and informed investments decisions. According to the results, the kNN algorithm is robust with small error ratio; consequently the results were rational and also reasonable. In addition, depending on the actual stock prices data; the prediction results were close and almost parallel to actual stock prices (22).

Akash Patel et. al. focus on preprocessing of datasets. Second, after processing the datasets earlier, we will review the use of major AI technique for that data and productive results. In addition, the proposed system evaluates the application of the forecast system to the real-world scenario and the problems associated with the accuracy of the total values provided. The high accuracy and profitability was achieved when results of all algorithms are combined and considered all factors affecting the stock prices. Successful valuation prediction of share price can become a big asset for stock market firms and provide real life solutions to the difficulties faced by stock market individual investors have (23).

Nitin N. Sakhare et. al. (2020) predict the rise in stock market and downfall of stock market through various web sites using classification and prediction algorithms using Sentiment Analysis (24).

Kien Wei Siah et. al. proposes to study the potential of using both behavioral and technical features in stock price prediction models based on traditional classifiers and popular neural networks. We believe that behavioral data may offer insights into financial market dynamics in addition to that captured by technical analysis. An improved price forecasting model can yield enormous rewards in stock market trading (25).

IV. METHODOLOGY

A. Long Short Term Memory model (LSTM)

LSTM, which stands for Long Short Term Memory, is a type of neural network which is particularly useful in the case of time series forecasting. According to an article by Srivastava on LSTM's and essentials of deep learning, an LSTM network is the most effective solution to time series analysis and thus stock market prediction. With the recent breakthroughs that have been happening in data science, it is found that for almost all of these sequence prediction problems, long short Term Memory networks have been observed as the most effective solution. LSTMs have an edge over conventional feed-forward neural networks and Recurrent Neural Networks in many ways. This is because of their property of selectively remembering patterns for long durations of time.

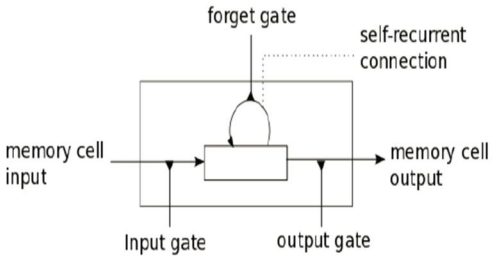


Figure 1. Simplified Look at LSTM Cell

In the case of a basic neural network, in order to add a new information, it transforms the existing information completely by applying a sigmoid function. Because of this, the entire information is modified as a whole. Thus, there is no consideration for 'important' information and 'not so important' information. LSTMs on the other hand, make small modifications to the information by multiplications and additions. With LSTMs, the information flows through a mechanism known as cell states. This way, LSTMs can selectively remember or forget things.
The following figure, represents a more detailed view at the internal architecture of an LSTM network:

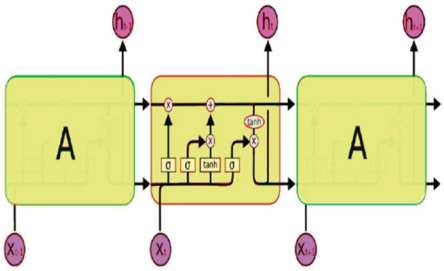


Figure 2. LSTM Cell Architecture

A typical LSTM network is comprised of different memory blocks called cells. There are two states that are being transferred to the next cell; the cell state and the hidden state. The memory blocks are responsible for remembering things, and manipulations to this memory is done through three major mechanisms called gates:

B. Forget Gate

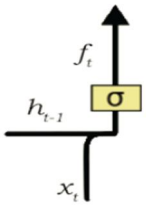


Figure 3. Forget Gate, Internal Architecture

A forget gate is responsible for removing information from the cell state. The information that is no longer required for the LSTM to understand things or the information that is of less importance is removed. This gate takes in two inputs; h_t-1 and x_t.
h_t-1 is the hidden state from the previous cell or the output of the previous cell and x_t is the input at that particular time step. The given inputs are multiplied by the weight matrices and a bias is added. Following this, the sigmoid function is applied to this value. The sigmoid function outputs a vector, with values ranging from 0 to 1, corresponding to each number in the cell state. Basically, the sigmoid function is responsible for deciding which values to keep and which to discard. If a '0' is output for a particular value in the cell state, it means that the forget gate wants the cell state to forget that piece of information completely. Similarly, a '1' means that the forget gate wants to remember that entire piece of information. This vector output from the sigmoid function is multiplied to the cell state.

C. Input Gate

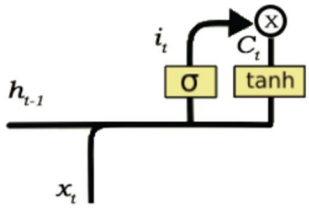


Figure 4. Input Gate, Internal Architecture

The input gate is responsible for the addition of information to the cell state. This addition of information is basically three-step process as seen from the diagram above.

1) Regulating what values need to be added to the cell state by involving a sigmoid function. This is basically very similar to the forget gate and acts as a filter for all the information from h_t-1 and x_t.

2) Creating a vector containing all possible values that can be added (as perceived from h_t-1 and x_t) to the cell state. This is done using the tanh function, which outputs values from -1 to +1.

3) Multiplying the value of the regulatory filter (the sigmoid gate) to the created vector (the tanh function) and then adding this useful information to the cell state via addition operation.

Once this three-step process is done with, we ensure that only that information is added to the cell state that is important and is not redundant.

D. Output Gate

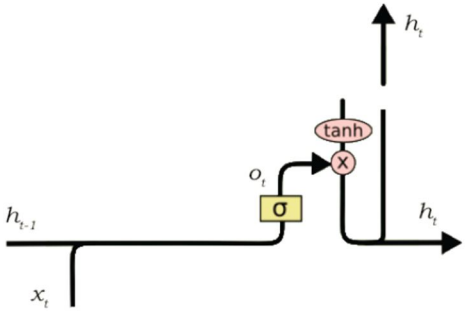


Figure 5. Output Gate, Internal Architecture

The output gate is responsible for selecting useful information from the current cell state and showing it out as an output. The functioning of an output gate can again be broken down to three steps:

1. Creating a vector after applying tanh function to the cell state, thereby scaling the values to the range -1 to +1.
2. Making a filter using the values of h_t-1 and x_t, such that it can regulate the values that need to be output from the vector created above. This filter again employs a sigmoid function.
3. Multiplying the value of this regulatory filter to the vector created in step 1, and sending it out as an output and also to the hidden state of the next cell.

E. Our Dataset

The data contains records about the stock price of IT Leading company such a
Infosys:
https://finance.yahoo.com/quote/INFY.NS/history/
Microsoft:
https://finance.yahoo.com/quote/MSFT/history/
TCS:
https://finance.yahoo.com/quote/TCS.NS/history/

The dataset also contains a date-wise price of stock with open, close, high, and low prices along with volume traded as well as turnover on that day.

Here I use adjusted close value for prediction. The Adjusted Close Value is the final output value that will be forecasted using the Machine Learning model.

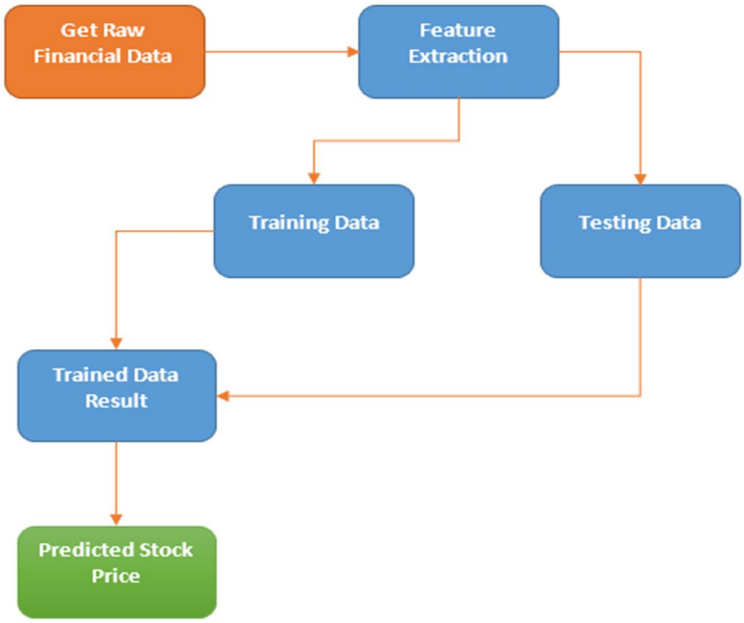
At last compare adj close's true values and adj close's predicted values using the LSTM Machine learning model.

V. PROPOSED WORK

A. *Proposed Algorithm is give below*

Step 1: Importing the Libraries
Step 2: Getting to Visualising the Stock Market Prediction Data
Step 3: Check for Null Values by printing the DataFrame Shape
Step 4: Setting the Target Variable and Selecting the Features
Step 5: Creating a Training Set and a Test Set for Stock Market Prediction
Step 6: Building the LSTM Model for Stock Market Prediction
Step 7: Training the Stock Market Prediction Model
Step 8: LSTM Prediction
Step 9: Comparing Predicted vs True Adjusted Close Value – LSTM

B. *System Flow Diagram*

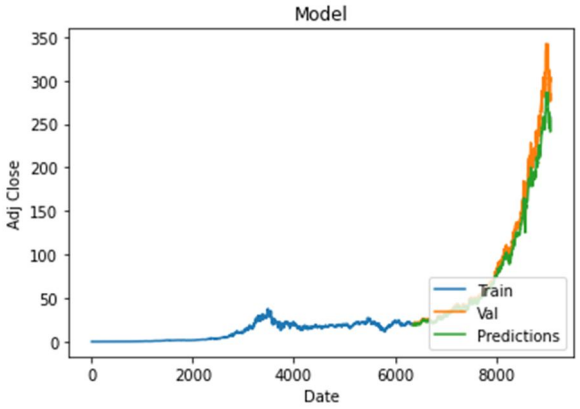


C. *Performance Parameter*

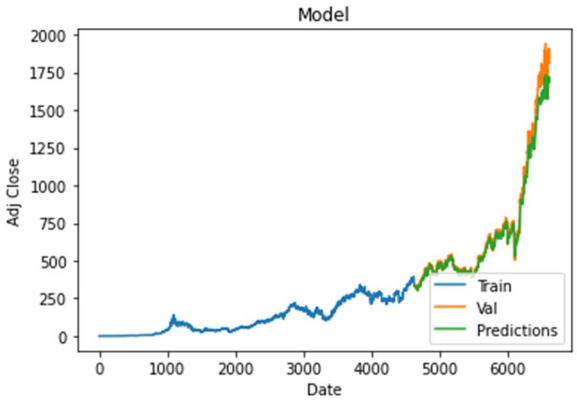
Parameters	Explanation
Absolute Error	ABS Error=*Fx=abs(predicted-actual)*
Square Error	Square error=(ABS ERRR)^2
%age Error	% Error=ABS Error/Actual Close price
Mean absolute deviation (MAD)	Average absolute error
Mean of square error (MSE)	Average square error
Mean absolute %age error (MAPE)	Average percentage error

VI. EXPERIMENTAL RESULTS

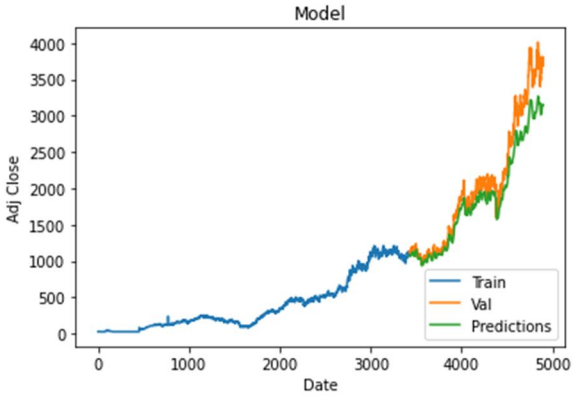
A. *Prediction based on Microsoft Stock Data*



B. *Prediction based on Infosys Stock Data*



C. *Prediction based on TCS Stock Data*



D. *Performance Parameters based on Different Dataset*

Dataset	MAD	MSE	MAPE
Infosys (1996 to 2022)	35.46	1538.08	6.12
Microsoft (1986 to 2022)	10.03	390.56	5.82
TCS (2002 TO 2022)	86.47	16095.97	3.52

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

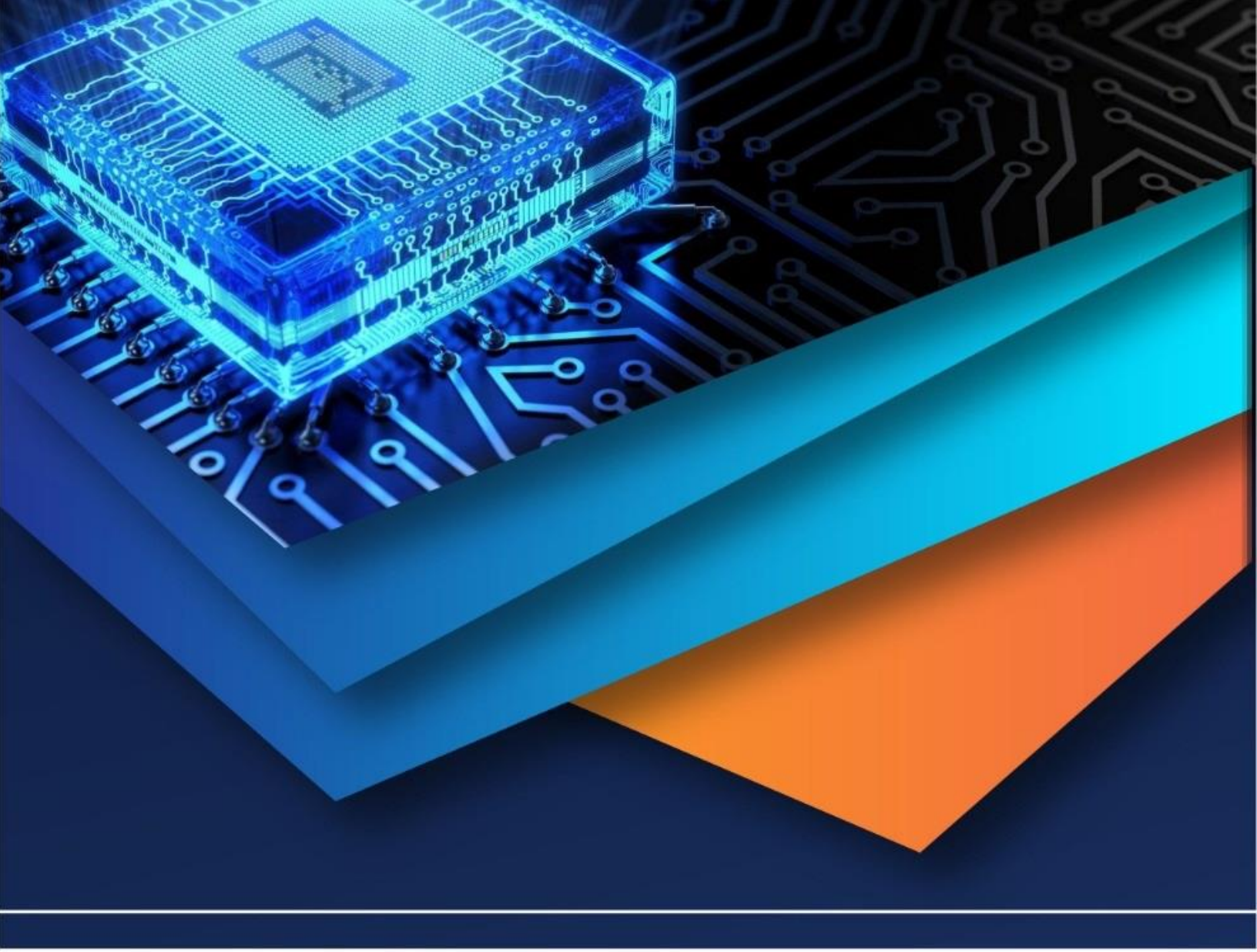
VII. CONCLUSION

With the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data. The Opening Value of the stock, the Highest and Lowest values of that stock on the same days, as well as the Closing Value at the end of the day, are all indicated for each date. Predicting the stock market was a time-consuming and laborious procedure a few years or even a decade ago. However, with the application of machine learning using LSTM for stock market forecasts, the procedure has become much simpler. Machine learning not only saves time and resources but also outperforms people in terms of performance. It will always prefer to use a trained computer algorithm since it will advise you based only on facts, numbers, and data and will not factor in emotions or prejudice.

REFERENCES

[1] Stock Market Prediction Using Machine Learning Algorithms. K. Hiba Sadia, Aditya Sharma, Adarrsh Paul, SarmisthaPadhi, Saurav Sanyal. 2019, International Journal of Engineering and Advanced Technology (IJEAT), p. 7.
[2] Stock Closing Price Prediction using Machine Learning Techniques. Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal, Arun Kumar. 2019, International Conference on Computational Intelligence and Data Science (ICCIDS 2019), p. 8.
[3] Stock Market Prediction Using Machine Learning. Reddy, V Kranthi Sai. 2018, International Research Journal of Engineering and Technology (IRJET), p. 5.
[4] Stock market prediction using machine learning classifiers and social media, news. Wasiat Khan, Mustansar Ali Ghazanfar, Muhammad Awais Azam, Amin Karami, Khaled H Alyoubi, Ahmed S Alfakeeh. 2020, Journal of Ambient Intelligence and Humanized Computing, p. 24.
[5] Short-term stock market price trend prediction using a comprehensive deep learning system. Shafiq, Jingyi Shen and M. Omair. 2020, Journal of Big Data, Springer, p. 33.
[6] Stock Market Prediction Using Machine Learning(ML) Algorithms. M Umer Ghania, M Awaisa and Muhammad Muzammula. 2019, ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, p. 20.
[7] Stock Market Prediction Using Machine Learning Techniques:A Decade Survey on Methodologies, Recent Developments, and Future Directions. Nusrat Rouf, Majid Bashir Malik, Tasleem Arif, Sparsh Sharma, Saurabh Singh, Satyabrata Aich. 2021, Electronics 2021,MDPI, p. 25.
[8] StockClue: Stock Prediction using Machine Learning. A M Pranav, Sujooda S, Jerin Babu, Amal Chandran, Anoop S. 2021, International Journal of Engineering Research & Technology (IJERT), p. 4.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)