IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

PREDICTIVE MODELING OF REAL TIME STOCK DATA

[1]Manavi Mishra, [2]Mr. TusharChaudhari, [3]Manjushree Patil, [4]GeetanjaliRaut

[2]Assistant Professor, [134]Student

[1234]Department of Electronics and Telecommunication,

St. John College of Engineering and Management, Palghar, Maharashtra, India

Abstract: ***Stock price prediction is one of the major problems faced by many investors. The volatile nature of the stock market increases the risk of investing our money in fear of losing it. Fluctuating stock prices affects the investor's belief and thus there is a need to predict the future stock price to make more informed and accurate investment decisions. Also, one of the most quickly expanding research areas in today's world is Sentiment Analysis. Social Media sites like Twitter have become an origin of a wide variety of information. This is because the people of the modern world prefer blogs or social media pages to post their ideas or opinions about a particular product, service, or organization.***

This paper aims to propose a prediction system using the Machine Learning algorithms; ARIMA, LSTM & Linear Regression and makes use of a popular social media platform, Twitter to implement a system that gets insights about the sentiments from the fetched Tweets to predict stock prices. These Machine Learning algorithms are applied to the historical stock data of the past 2 years. Parameters considered are open, close, low, high, and volume.

Index Terms – **Stock price, Investment decisions, Tweets, Social Media.**

I. INTRODUCTION

The stock market prediction has been a crucial issue in the field of finance, engineering, and mathematics due to its potential financial return. As a massive amount of capital is traded through the stock market, the stock market is seen as a peak investment outlet. In addition, stock market forecasting brings with it the challenge of proving whether the financial market is predictable or not. Since there has been no consensus on the validity of the Efficient Market Hypothesis (EMH) which states the market is efficient and there is no space for prediction, researchers have aimed for proving the forecast ability of the financial market. With the development of faster computers and vast information over the Internet, stock markets have become more accessible to either well-thought-out investors or the general public. As the Internet provides a primary source of event information that significantly impacts stock markets, the techniques to extract and use the information to support decision-making have become a critical task. To predict the stock market precisely, different prediction algorithms and models have been proposed by many researchers in both academics and industry.

II. LITERATURE REVIEW

The aim of [1] is to classify different machine learning algorithms by presenting the concept of economic derivatives just like the "no arbitrage" principle and people of the predictive model, like stochastic process theory and efficient market hypothesis (EMH). News reports were analyzed and also the prediction was made on their basis. Information was collected daily and predictions were made on LogisticRegression model. [2] predicts the market performance of the Karachi stock market (KSE) using different machine learning techniques. Various procedures were discussed like Single Layer Perceptron, Multi-Layer Perceptron, Radial Basis Function, and Support Vector Machines. [3] proposed the survey of a widely known efficient regression approach to predict the stock exchange price from the stock exchange database. Various regression techniques like polynomial regression, rectilinear regression, RBF regression, and sigmoid regression were described. In [4], the prediction of the exchange trends was done using machine learning prototypes just like the Random Forest model and SVM. The past data, further as sentimental data, were analyzed as they players an important role in market performance. [5] predicted the price of the stocks of the stock market of Thailand (SET). Different Procedures to predict the stock were the Multi-Layer Perceptron model, Support Vector Machine model, and Partial Least Square Classifier. [6] aims to review the Stock market prediction using multiple Traditional, Machine learning, and Deep learning algorithms. Together with these algorithms, the survey has also focused on different datasets used for Stock market prediction, the characteristics of those datasets were selected as input parameters, and also the evaluation metrics used for comparing the results of predictions. [7] proposes a novel profound deep neural network DPLSTM for stock value prediction. It uses news stories as controlled data and joins distinctive news sources through the differential security system. [8] introduced an examination of AI-supported calculations to assess the stock costs later on to break down market conduct. [9] examined the expectation of future stock costs utilizing the assumption esteems for each stock. It concentrates on two different methodologies, Word2vec and N-gram, for deriving insights from sentiments in tweets. Operation of sentiment analysis in incorporation to machine learning algorithms to tweets extracted from Twitter and analyzing the correlation between the stock market movement of

company and sentiments in tweets. The authors of [10] used the ARIMA model to anticipate the stock cost on the information got from New York Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE). They have used data sets of four features: open, low, close, and high cost. They have accepted the end cost as the objective viewpoint to be anticipated. In [11], the proposed model is based on the study of historical data and technical data. It develops LS-SVM with a PSO algorithm for daily stock prediction. Levenberg-Marquardt's (LM) algorithm is employed as a benchmark for comparison with LS-SVM and LS-SVM-PSO models. [12] proposed a comparative study of nine machine learning models, namely Random Forest, Decision Tree, eXtreme Gradient Boosting, Adaptive Boosting, K-Nearest Neighbors, Support Vector Classifier, Naïve Bayes, Artificial Neural Network, and LR. Along with this Long Short-Term Memory and Recurrent Neural Networks, Deep learning techniques were additionally studied. In [13], the authors have used four types of deep learning architectures which are Recurrent Neural Networks, Multilayer Perceptron, Convolutional Neural networks, and Long ShortTerm Memory. All these architectures were used for predicting the future value of the stock of a company from the historical dataset obtained. In [14] various stock market prediction techniques such as classification, regression, ensemble, evolutionary, deep learning, hybrid, and ABS have been compared. From the results, Random Forest has the highest accuracy of 80.8% for the Zagreb Stock Exchange whereas deep learning has the least MAPE value(0.0002) for the Chinese securities index 10(CSI 300).

III. PROPOSED METHODOLOGY

In the proposed design we worked with the trading parameters such as Open, Close, High, Low, Adj. Close and Total Traded Volume. The past 2 years historical data was collected from Yahoo Finance API.

1. Linear Regression:

Linear Regression is a supervised machine learning algorithm. It is used to predict a correlation amongst more than one independent variable.

$Y0 = b0 + b1*x1$

Where, Y = Dependent variable

X= Independent variable

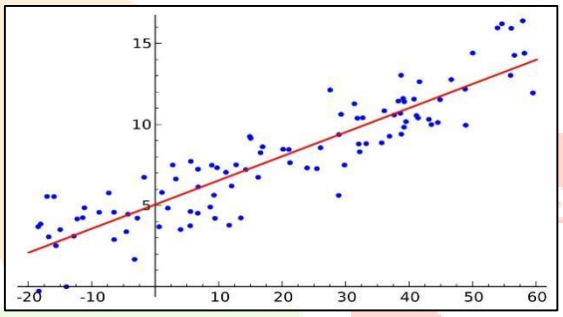

Figure 1: Linear Regression
(https://en.wikipedia.org/wiki/Linear_regression)

2. ARIMA:

ARIMA is an abbreviation that stands for AutoRegressive Integrated Moving Average. This abbreviation is illustrative, capturing the key aspects of the model itself. Briefly, they are:

- AR: Autoregression. A prototype that uses the dependent relationship between an observation and some number of lagged observations.
- I: Integrated. The use of differencing raw observations (e.g. subtracting an observation from observation at the previous time step) makes the time series stationary.
- MA: Moving Average. A model that uses the dependency between an observation and a residual error from a moving average model is applied to lagged observations.

Each of these components is clearly specified in the model as a variable. A regular notation is used of ARIMA(p,d,q).

The ARIMA model parameters are defined as follows:

- p: The number of delayed observations included in the model, also called the lag order.
- d: The number of times that the new observations are differenced, also called the degree of difference.
- q: The proportion of the moving average window, also called the order of moving average.

3. LSTM:

The Long Short-Term Memory network, or LSTM network, is a recurrent neural network that is trained using Back propagation Through Time and conquers the disappearing gradient problem. As such, it can be used to create large recurrent networks that in turn can be used to address complex order problems in machine learning and attain state-of-the-art results. Instead of neurons, LSTM networks have memory blocks that are attached through layers.

A block has elements that make it shrewd than a classical neuron and a memory for recent orders. A block carries gates that manage the block's state and output. A block runs upon an input sequence and each gate within a block uses the sigmoid enabled units to control whether they are triggered or not, making the change of state and addition of information circulating through the block conditional.

There are three types of gates inside a unit:

- Forget Gate: conditionally decides what data to throw away from the block.
- Input Gate: conditionally decides which values from the input to update the recollection state.
- Output Gate: conditionally decides what to output based on input and the recollection of the block.

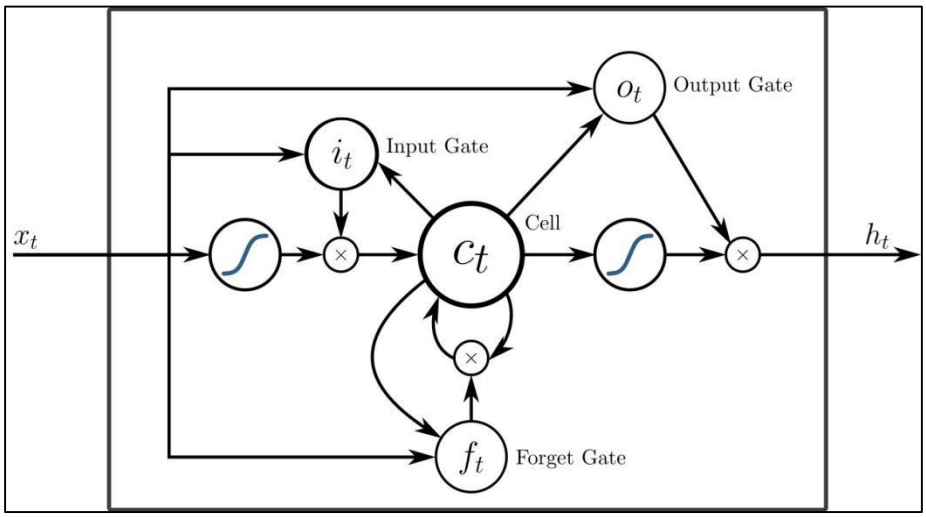

Figure 2: Long Short Term Memory
(https://en.wikipedia.org/wiki/Long_short-term_memory)

IV. BLOCK DIAGRAM

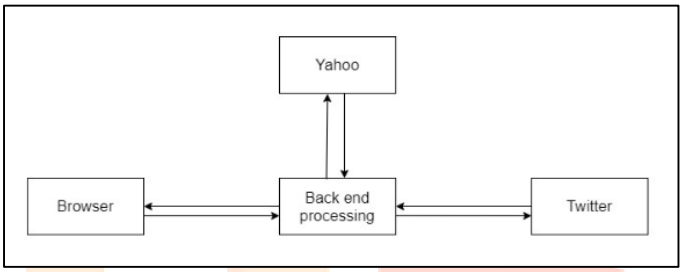

Figure 3: System Block Diagram

Figure 3, illustrates the Block Diagram, which represents the design and operation of the proposed system. Whenever the user visits the prediction webpage in his browser, the user enters the ticker code of the company of which he wants to get stock quote prediction. The Browser sends the user's input as a request to the server which handles Back end Processing. The Back end processing unit, further sends the request to Yahoo and Twitter API's to fetch the historical stock prices and the current stock news and sentiments respectively. After the processing is complete, the output in the form of predicted Quote Information and Graphs will be displayed in a web page on the Browser.

V. FLOW CHART

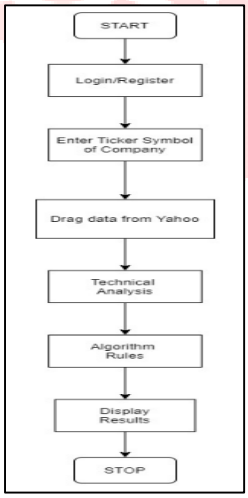

Figure 4: Historical Methodology

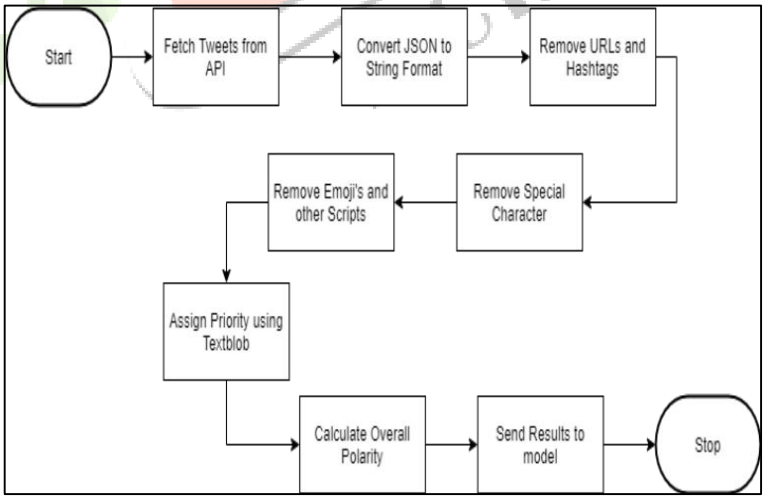

Figure 5: Sentimental Methodology

VI. RESULT and DISCUSSION

Result observation made for two companies namely Wipro Limited (WIPRO.NS) and Coal India Limited (COALINDIA.NS) are listed below. The prediction were made on 16/06/2021 and the next 7 days forecast values were predicted.

a. Predictions

1. Wipro Limited (WIPRO.NS)

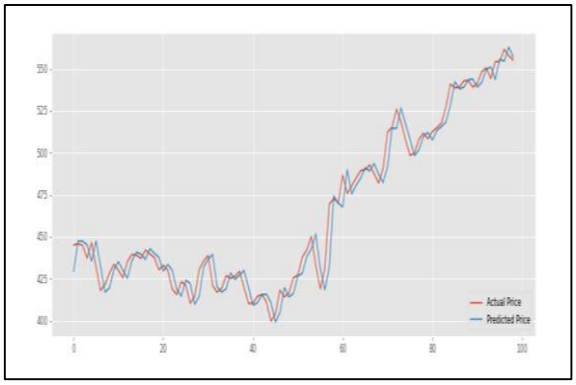

Figure 5: ARIMA of WIPRO.NS

Figure 6: LR of WIPRO.NS

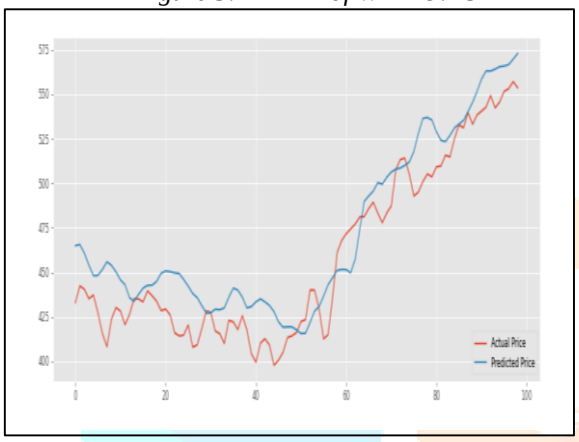

Figure 7: LSTM of WIPRO.NS

Figure 8: Trends of WIPRO.NS

Table 1: Prediction of WIPRO.NS

CompanyTicker code	Date	Predicted	Actual	Difference
WIPRO.NS	17-06-2021	561.01	558.70	2.31
	18-06-2021	554.12	549.80	4.32
	21-06-2021	564.95	542.15	22.80
	22-06-2021	564.99	556.55	8.44
	23-06-2021	572.86	540.15	32.71
	24-06-2021	568.88	549.05	19.83
	25-06-2021	566.02	547.50	18.52

2. Coal India Limited (COALINDIA.NS)

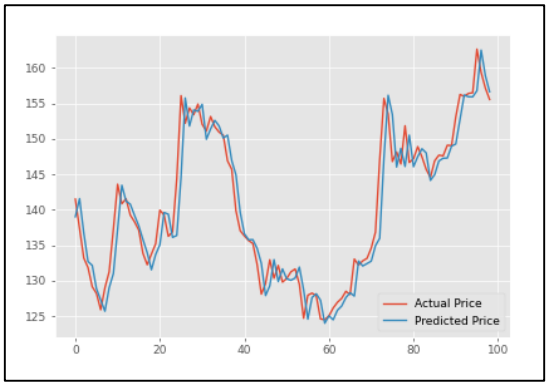

Figure 9: ARIMA of COALINDIA.NS

Figure 10: LR of COALINDIA.NS

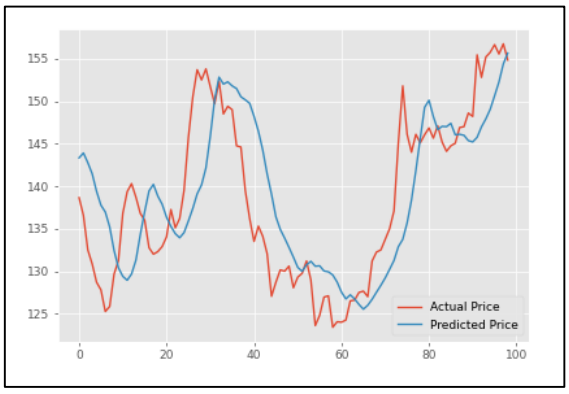

Figure 11: LSTM of COALINDIA.NS

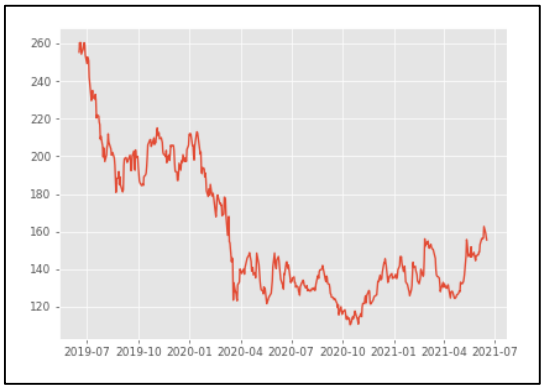

Figure 12: Trends of COALINDIA.NS

Table 2: Prediction of COALINDIA.NS

CompanyName	Date	Predicted	Actual	Difference
COALINDIA.NS	17-06-2021	155.62	152.15	3.47
	18-06-2021	156.00	146.40	9.60
	21-06-2021	156.09	147.50	8.59
	22-06-2021	161.96	148.70	13.26
	23-06-2021	158.82	148.45	10.37
	24-06-2021	156.72	146.40	10.32
	25-06-2021	155.19	148.75	6.44

b. System Feasibilities

The system is feasible technically, operationally and also economically. Every technology required to create this system is accessible. It is too flexible and can be expanded further guaranteeing ease of use, accuracy, data security and reliability. Operational Feasibility means to find out if a proposed system could satisfy users' objectives which could be fitted into the present system operation. Our solution is operationally possible to implement. Economically, this system is completely feasible because it requires no extra financial investment and with respect to time, it can be completely set up in few minutes.

VII. CONCLUSION

The objective of adding news feeds, Twitter data to increase the prediction accuracy along with historical stock prices did give better results compared to only historical stock prices. Along with this, using LSTM and CNN layers helped in mapping the events better. Though the root mean squared error value of our models were not substantially different from that of AR and ARIMA, but they had a higher classification accuracy. The addition of Twitter data to the model had a significant effect on the prediction decision not on the values. This could be because of a smaller dataset. Collecting more streaming tweets with paid APIs and incorporating them in the system might help in achieving better accuracy.

VIII. FUTURE SCOPE

The future scope would involve incorporating more stocks from different domains and finding out the correlation among them to predict the trend for an entire sector. Also, finding out how much does one sector depends on another and incorporating that into the model. Making use of Twitter data in a much better way incorporating with a large corpus. For the stocks whose tweets were not available, paid API's can be used to get financial news. Other alternatives to fetch news could also be incorporated. Also looking out for some more features which can help in getting even closer to the actual values. Also, the system could be made more customizable by including various features. The system code could be optimized more to reduce the execution time and obtain even faster results.

REFERENCES

[1] Misra, Meghna, et al. “Stock Market Prediction Using Machine Learning Algorithms: A Classification Study.” 2018 International Conference on Recent Innovations in Electrical, Electronics and Communication Engineering, ICRIEECE 2018, IEEE, 2018, pp. 2475–78.

[2] Sirimevan, Naadun, et al. “Stock Market Prediction Using Machine Learning Techniques.” 2019 International Conference on Advancements in Computing, ICAC 2019, no. 4, 2019, pp. 192–97.

[3] Sharma, Ashish, et al. “Survey of Stock Market Prediction Using Machine Learning Approach.” Proceedings of the International Conference on Electronics, Communication and Aerospace Technology, ICECA 2017, vol. 2017-Janua, 2017, pp. 506–09.

[4] Maini, Sahaj Singh, and K. Govinda. “Stock Market Prediction Using Data Mining Techniques.” Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017, no. Iciss, IEEE, 2018, pp. 654–61.

[5] Vijh, Mehar, et al. “Stock Closing Price Prediction Using Machine Learning Techniques.” Procedia Computer Science, vol. 167, 2020, pp. 599–606.

[6] Sable, Rachna, et al. “Empirical Study on Stock Market Prediction Using Machine Learning.” 2019 6th IEEE International Conference on Advances in Computing, Communication and Control, ICAC3 2019, IEEE, 2019, pp. 1–5.

[7] Li, Xinyi, et al. “DP-LSTM: Differential PrivacyInspired LSTM for Stock Prediction Using Financial News.” ArXiv, no. NeurIPS, 2019, pp. 1–9.

[8] Bhardwaj, Neha, and Akil Ansari. “Prediction of Stock Market Using Machine Learning Algorithms.” International Research Journal of Engineering and Technology, vol. 5994, no. May, 2019, pp. 5994–6005.

[9] Jadhav, Rupawari, and Wakode M. S. “Survey: Sentiment Analysis of Twitter Data for Stock Market Prediction.” Ijarcce, vol. 6, no. 3, 2017, pp. 558–62.

[10] Adebiyi, Ayodele A., et al. “Stock Price Prediction Using the ARIMA Model.” Proceedings - UKSimAMSS 16th International Conference on Computer Modelling and Simulation, UKSim2014, 2014, pp. 106–12.

[11] Hegazy, Osman, et al. "A Machine Learning Model for Stock Market Prediction" International Journal of Computer Science and Telecommunications, Volume 4, Issue 12. [Online]. Available: http://arxiv.org/abs/1402.7351

[12] Parmar, Ishita, et al. “Stock Market Prediction Using Machine Learning.” ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications, IEEE, 2018, pp. 574–76.

[13] Nabipour, Mojtaba, et al. “Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; A Comparative Analysis.” IEEE Access, vol. 8, 2020, pp. 150199–212.

[14] Singh, Sukhman, et al. “Stock Market Forecasting Using Machine Learning: Today and Tomorrow.” 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies, ICICICT 2019, 2019, pp. 738–45

[15] Narote, Prof Amit, et al. Predicting NASDAQ and NSE Stocks Using Machine Learning Algorithms : ARIMA , LSTM &. no. Iii, 2020, pp. 39–46.

[16] Hiransha, M., et al. “NSE Stock Market Prediction Using Deep-Learning Models.” Procedia Computer Science, vol. 132, no. Iccids, Elsevier B.V., 2018, pp. 1351–62.