IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

STOCK PREDICTION USING MACHINE LEARNING TECHNIQUES.

1.Ms.N.VivekaPriya

2.Dr.S.Geetha

ABSTRACT

The main objective of this paper is to find the best model to predict the value of the stock market. During the process of considering various techniques and variables that must be taken into account, it is found out that techniques like random forest, support vector machine were not exploited fully. In, this paper it is about to present and review a more feasible method to predict the stock movement with higher accuracy. The first thing that have been taken into account is the dataset of the stock market prices from previous year. The dataset was pre-processed and tuned up for real analysis. Hence, this paper will also focus on data preprocessing of the raw dataset. Secondly, after preprocessing the data will be reviewed to use the random forest, support vector machine on the dataset and the outcomes it generates. In addition, the proposed paper 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.

Keywords: Machine Learning, Data Pre-processing, Data Training, Dataset, Stock, Data Storing.

I-INTRODUCTION

I.I-INTRODUCTION TO STOCK MARKET PREDICTION

Recently, a lot of interesting work has been done in the area of applying Machine Learning Algorithms for analyzing price patterns and predicting stock prices and index changes. Most stock traders nowadays depend on Intelligent Trading Systems which help them in predicting prices based on various situations and conditions, thereby helping them in making instantaneous investment decisions. Stock Prices are considered to be very dynamic and susceptible to quick changes because of the underlying nature of the financial domain and in part because of the mix of known parameters (Previous Year, Opening Price, Closing Price,.) and unknown factors. An intelligent trader would predict the stock price and buy a stock before the price rises, or sell it before its value declines. Though it is very hard to replace the expertise that an experienced trader has gained, an accurate prediction algorithm can directly result into high profits for investment firms,

indicating a direct relationship between the accuracy of the prediction algorithm and the profit made from using the algorithm.

SCOPE OF THE PAPER.

This paper discusses the Machine Learning techniques which have been applied for stock trading to predict the rise and fall of stock prices before the actual event of an increase or decrease in the stock price occurs. In particular the paper discusses the application of Support Vector Machines, Linear Regression, Prediction using Decision Stumps, Expert Weighting and Online Learning in detail along with the benefits and pitfalls of each method. The paper introduces the parameters and variables that can be used in order to recognize the patterns in stock prices which can be helpful in the future prediction of stocks and how Boosting can be combined with other learning algorithms to improve the accuracy of such prediction systems.

II-LITERATURE REVIEW

AN EMPIRICAL ANALYSIS ON BETWEEN STOCK PREDICTION DEVELOPMENT

Development of the stock markets is important for economic development and growth. This research is an effective attempt to examine the significance of India's relation between economic growth and stock market development. The study covers the period from 1990-2018. The ADF Unit Root test findings indicate all variables of order one to be combined. Results of Cointegration reveal the existence of variables in a long-term relation. The findings of long-term estimations showed a strong and important relation between stock market and GDP. Short-term dynamics of the VECM revealed that the stock market had a strong short-term effect on India's economic development. Impulse Response Function (IRF) reveals how GDP reacts to a stock market impact.

RELEVANCE OF EFFICIENT MARKET HYPOTHESIS: A STUDY OF PRESENT SCENARIO IN INDIA

The Indian stock market has witnessed several peaks and troughs in the span of last few years. It has now become a serious concern among the economists, traders and policy developers to deal with such volatility. The stock market volatility impacts over the economic stability of the country. So it is necessary to understand the movement of stock market. Several researches in this field deals with different opinion amongst which random walk theory and efficient market theory is of great apprehension. In the world of speculation, is the market efficient enough to reflect the price movement? This paper is an attempt to test the market efficiency theoretically and empirically with respect to the publicly or privately available information and simultaneously effort has been made whether

INVESTMENT STRATEGY FOR INTERNATIONAL INVESTORS IN ASIAN EMERGING MARKETS

Incorporating the extra risky stocks in portfolios can lead to realisation of extra returns. Due to financial liberalisation, there are huge opportunities of investment in Asian stock markets as these markets have gone through a considerable expansion. In large number of empirical studies of risk return analysis, it is observed that economic stability and good perspectives have been key assets for the development of emerging markets. Thus, emerging capital markets are becoming increasingly important for institutional as well as individual investors. Liberalisation of financial systems in these emerging Asian markets has attracted domestic and foreign institutional investors to diversify their funds across the markets and reduce their portfolio risk

III-PROPOSED SYSTEM

In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. The recent trend in stock market prediction technologies is the use of machine learning which makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of Regression and LSTM based Machine learning to predict stock values. Factors considered are open, close, low, high and volume.

In this proposed system, we focus on predicting the stock values using machine learning algorithms like Random Forest and Support Vector Machines. It proposes the system "Stock market price prediction" have predicted the stock market price using the random forest algorithm. In this proposed system, it were able to train the machine from the various data points from the past to make a future prediction. The proposed system has taken data from the previous year stocks to train the model. There are majorly used two machine-learning libraries to solve the problem. The first one was numpy, which was used to clean and manipulate the data, and getting it into a form ready for analysis. The other was scikit, which was used for real analysis and prediction. The data set we used was from the previous year's stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data. The basic approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data. The system uses the python panda's library for data processing which combined different datasets into a data frame. The tuned-up data frame allowed us to prepare the data for feature extraction. The data frame features were date and the closing price for a particular day. In this method it used all these features to train the machine on random forest model and predicted the object variable, which is the price for a given day. Also it quantified the accuracy by using the predictions for the test set and the actual values. The proposed system touches different areas of research including data pre-processing, random forest, and so on.

This proposed system in work at year first 14 days because all company working year begging. So, previews data available in day end closing values or opening values compare the two values and any one values using upon time.

IV-SYSTEM ARCHITECTURE

The first step is the conversion of this raw data into processed data. This is done using feature extraction, since in the raw data collected there are multiple attributes but only a few of those attributes are useful for the purpose of prediction. So the first step is feature extraction, where the key attributes are extracted from the whole list of attributes available in the raw dataset. Feature extraction starts from an initial state of measured data and builds derived values or features. These features are intended to be informative and non-redundant, facilitating the subsequent learning and generalization steps. Feature extraction is a dimensionality reduction process, where the initial set of raw variables is diminished to progressively reasonable features for ease of management, while still precisely and totally depicting the first informational collection. The feature extraction process is followed by a classification process wherein the data that was obtained after feature extraction is split into two different and distinct segments. Classification is the issue of recognizing to which set of categories a new observation belongs. The training data set is used to train the model whereas the test data is used to predict the accuracy of the model. The splitting is done in a way that training data maintain a higher proportion than the test data. The random forest algorithm utilizes a collection of random decision trees to analyze the data. In layman terms, from the total number of decision trees in the forest, a cluster of the decision trees look for specific attributes in the data. This is known as data splitting. In this case, since the end goal of this proposed system is to predict the price of the stock by analyzing its historical data

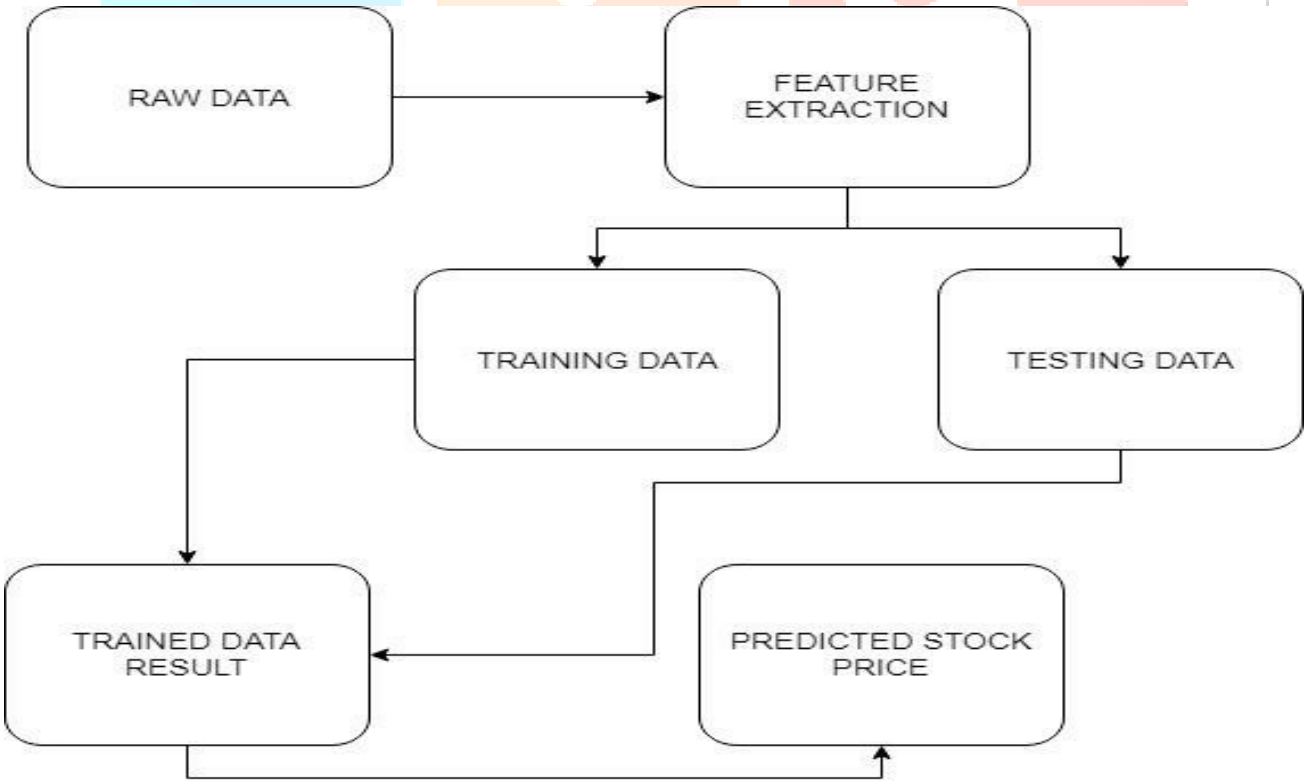


Fig.4.1. System Architecture

I. DATA COLLECTION

Data collection is a very basic module and the initial step towards the project. It generally deals with the collection of the right dataset. The dataset that is to be used in the market prediction has to be used to be filtered based on various aspects. Data collection also complements to enhance the dataset by adding more data that are external.The data mainly consists of the previous year stock prices. Initially,it will be analyzing the Kaggle dataset and according to the accuracy, we will be using the model with the data to analyze the predictions accurately.

II. PRE-PROCESSING

Data pre-processing is a part of data mining, which involves transforming raw data into a more coherent format. Raw data is usually, inconsistent or incomplete and usually contains many errors. The data pre-processing involves checking out for missing values, looking for categorical values, splitting the data-set into training and test set and finally do a feature scaling to limit the range of variables so that they can be compared on common environs.

III. TRAINING THE MACHINE

Training the machine is similar to feeding the data to the algorithm to touch up the test data. The training sets are used to tune and fit the models. The test sets are untouched, as a model should not be judged based on unseen data. The training of the model includes cross-validation where this system get a well-grounded approximate performance of the model using the training data. Tuning models are meant to specifically tune the

IV. DATA SCORING

The process of applying a predictive model to a set of data is referred to as scoring the data. The technique used to process the dataset is the Random Forest Algorithm. Random forest involves an ensemble method, which is usually used, for classification and as well as regression. Based on the learning models, the system achieves interesting results. The last module thus describes how the result of the model can help to predict the probability of a stock to rise and sink based on certain parameters. It also shows the vulnerabilities of a particular stock or entity. The user authentication system control is implemented to make sure that only the authorized entities are accessing the results.

V-MACHINE LEARNING TECHNIQUES FOR STOCK PREDICTION

I-STOCK PREDICTION IN DETAIL

In practice, there are two Stock Prediction Methodologies: Fundamental Analysis: Performed by the Fundamental Analysts, this method is concerned more with the company rather than the actual stock. The analysts make their decisions based on the past performance of the company, the earnings forecast etc. Technical Analysis: Performed by the Technical Analysts, this method deals with the determination of the stock price based on the past patterns of the stock (using time-series analysis.) When applying Machine Learning to Stock Data, we are more interested in doing a Technical Analysis to see if the algorithm can accurately learn the underlying patterns in the stock time series. This said, Machine Learning can also play a major role in evaluating and forecasting the performance of the company and other similar parameters helpful

in Fundamental Analysis. In fact, the most successful automated stock prediction and recommendation systems use some sort of a hybrid analysis model involving both Fundamental and Technical Analysis.

The Efficient Market Hypothesis (EMH)

The EMH hypothesizes that the future stock price is completely unpredictable given the past trading history of the stock. There are 3 types of EMH's: strong, semi-strong, and weak form. In the weak EMH, any information acquired from examining the stock's history is immediately reflected in the price of the stock.

The Random Walk Hypothesis

The Random Walk Hypothesis claims that stock prices do not depend on past stock prices, so patterns cannot be exploited since trends to not exist. With the advent of more powerful computing infrastructure (hardware and software) trading companies now build very efficient algorithmic trading systems that can exploit the underlying pricing patterns when a huge amount of data-points are made available to them. Clearly with huge datasets available on hand, Machine Learning Techniques can seriously challenge the EMH.

Functions of time-series analysis

We now take a brief look at the Functions of time-series analysis that are normally used in the technical analysis of stock prices:

Functions of time-series analysis can be any of the following:

Calculate Average (CA): The average of the past n values till today. Exponential Calculate Average (EMA): Gives more data to the most recent values while not discarding the older observation entirely.

Rate of Change (ROC): The ratio of the current price to the price n quotes earlier. n is generally 1 to 260 days. Relative Strength Index (RSI): Measures the relative size of recent upward trends against the size of downward trends within the specified time interval (usually 1 – 260 days). For this Project, the EMA was considered as the primary indicator because of its ability to handle an almost infinite amount of past data, a trait that is very valuable in time series prediction (It is worth noting that the application of other indicators might result in better prediction accuracies for the stocks under consideration).

EMA (n) = EMA (n-1) + Closing stock * (Opening stock (t) - EMA (n-1))

Find the number of day and time-series analysis is calculate in previews 7 days closing data an find in average in close. Put in open values in 8 columns. The overall output in display in Excel(xlsx) format. All data store in local OS path.

II CLASSIFICATION IS AN INSTANCE OF SUPERVISED LEARNING WHERE A SET IS ANALYZED AND CATEGORIZED BASED ON A COMMON ATTRIBUTE.

From the values or the data are given, classification draws some conclusion from the observed value. If more than one input is given then classification will try to predict one or more outcomes for the same. A few classifiers that are used here for the stock market prediction includes the random forest classifier, SVM classifier. Random Forest Classifier Random Forest classifier is a type of ensemble classifier and also a supervised algorithm. It basically creates a set of decision trees, that yields some result. The basic approach of random class classifier is to take the decision aggregate of random subset decision tress and yield a final class or result based on the votes of the random subset of decision trees. Parameters The parameters included in the random forest classifier are estimators which is total number of decision trees, and other hyper parameters like oobscore to determine the generalization accuracy of the random forest, max_features which includes the number of features for best-split. min_weight_fraction_leaf is the minimum weighted fraction of the sum total of weights of all the input samples required to be at a leaf node. Samples have equal weight when sample weight is not provided. SVM classifier SVM classifier is a type of discriminative classifier. The SVM uses supervised learning i.e. a labeled training data. The output are hyperplanes which categorizes the new dataset. They are supervised learning models that uses associated learning algorithm for classification and as well as regression. Parameters The tuning parameters of SVM classifier are kernel parameter, gamma parameter and regularization parameter.

· Kernels can be categorized as linear and polynomial kernels calculates the prediction line. In linear kernels prediction for a new input is calculated by the dot product between the input and the support vector.

· C parameter is known as the regularization parameter; it determines whether the accuracy of model is increases or decreases. The default value of c=10.Lower regularization value leads to misclassification.

· Gamma parameter measures the influence of a single training on the model. Low values signify far from the plausible margin and high values signifies closeness from the plausible margin.

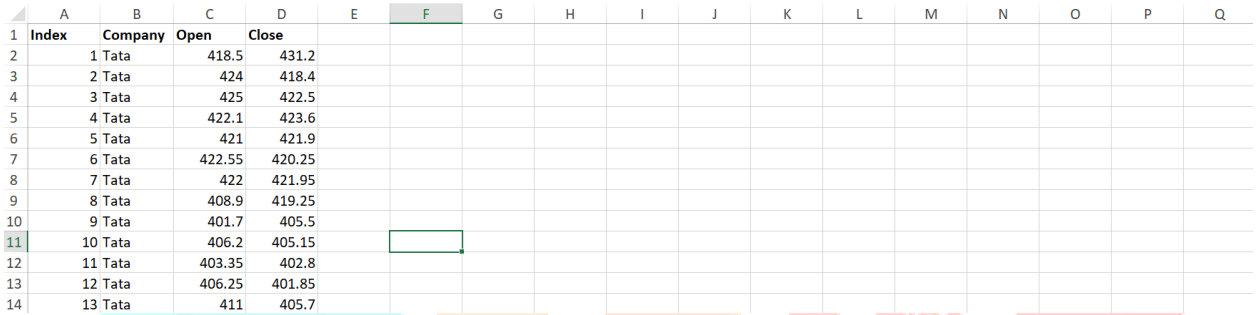
III-RANDOM FOREST ALGORITHM RANDOM FOREST ALGORITHM IS BEING USED FOR THE STOCK MARKET PREDICTION.

Since it has been termed as one of the easiest to use and flexible machine learning algorithm, it gives good accuracy in the prediction. This is usually used in the classification tasks. Because of the high volatility in the stock market, the task of predicting is quite challenging. In stock market prediction we are using random forest classifier which has the same hyperparameters as of a decision tree. The decision tool has a model similar to that of a tree. It takes the decision based on possible consequences, which includes variables like event outcome, resource cost, and utility. The random forest algorithm represents an algorithm where it randomly selects different observations and features to build several decision tree and then takes the aggregate of the several decision trees outcomes. The data is split into partitions based on the questions on a label or an attribute. The data set we used was from the previous year stock markets collected from the public database available online, 80 % of data was used to train the machine and the rest 20 % to test the data. The basic

approach of the supervised learning model is to learn the patterns and relationships in the data from the training set and then reproduce them for the test data.

VI-EXPERIMENTAL RESULTS

The xlxs file contains the raw data based on which we it go to publish the findings. There are eleven columns or eleven attributes that describe the rise and fall in stock prices. Some of these attributes are (1) HIGH, which describes the highest value the stock had in previous year. (2) LOW, is quite the contrary to HIGH and resembles the lowest value the stock had in previous year (3) OPENP is the value of the stock at the very beginning of the trading day, and (4) CLOSEP stands for the price at which the stock is valued before the trading day closes. There are other attributes such as YCP, LTP, TRADE, VOLUME and VALUE, but the above mentioned four play a very crucial role in our findings.

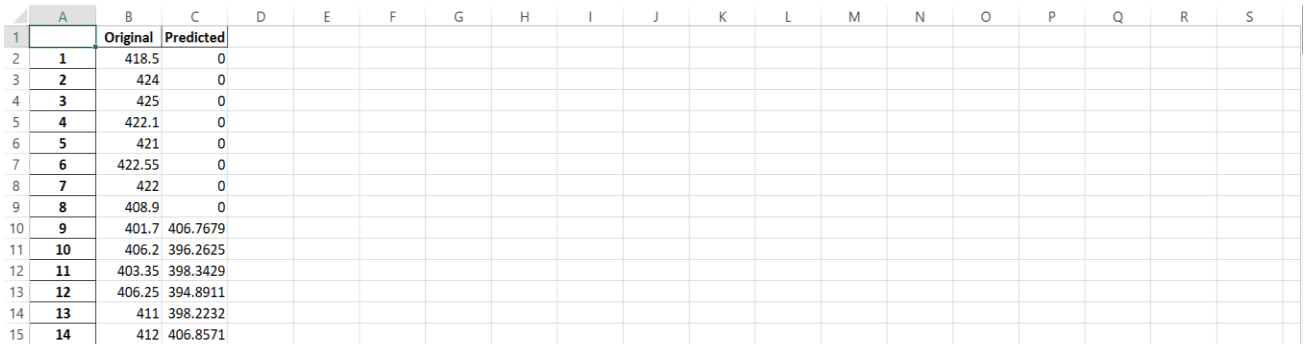


	Index	Company	Open	Close
1	1	Tata	418.5	431.2
2	2	Tata	424	418.4
3	3	Tata	425	422.5
4	4	Tata	422.1	423.6
5	5	Tata	421	421.9
6	6	Tata	422.55	420.25
7	7	Tata	422	421.95
8	8	Tata	408.9	419.25
9	9	Tata	401.7	405.5
10	10	Tata	406.2	405.15
11	11	Tata	403.35	402.8
12	12	Tata	406.25	401.85
13	13	Tata	411	405.7

Fig.6.1. Raw data

This is a pictorial representation of the data present in our xlxs file. This particular file contains 418.5 such records. There are more than 14 days available in the dataset and some of the records do not have relevant information that can help us train the machine, so the logical step is to process the raw data. Thus, it obtains a more refined dataset which can now be used to train the machine.

Then analyses in open or close values and time-series algorithm apply in calculate in 14 values and sum of values. After raw data move on summation of X and V.



		Original	Predicted
2	1	418.5	0
3	2	424	0
4	3	425	0
5	4	422.1	0
6	5	421	0
7	6	422.55	0
8	7	422	0
9	8	408.9	0
10	9	401.7	406.7679
11	10	406.2	396.2625
12	11	403.35	398.3429
13	12	406.25	394.8911
14	13	411	398.2232
15	14	412	406.8571

Fig.6.2. Output Prediction

The time-series algorithm work at in raw data flow on open values. This is the result of using the head(). Since the system is using the pandas library to analyses the data, it returns the first five rows. Here five is the default value of

the number of rows it returns unless stated otherwise. The trading code in the processed data set is not relevant so the system uses the strip() to remove it and replace all of the trading codes with a value „GP"

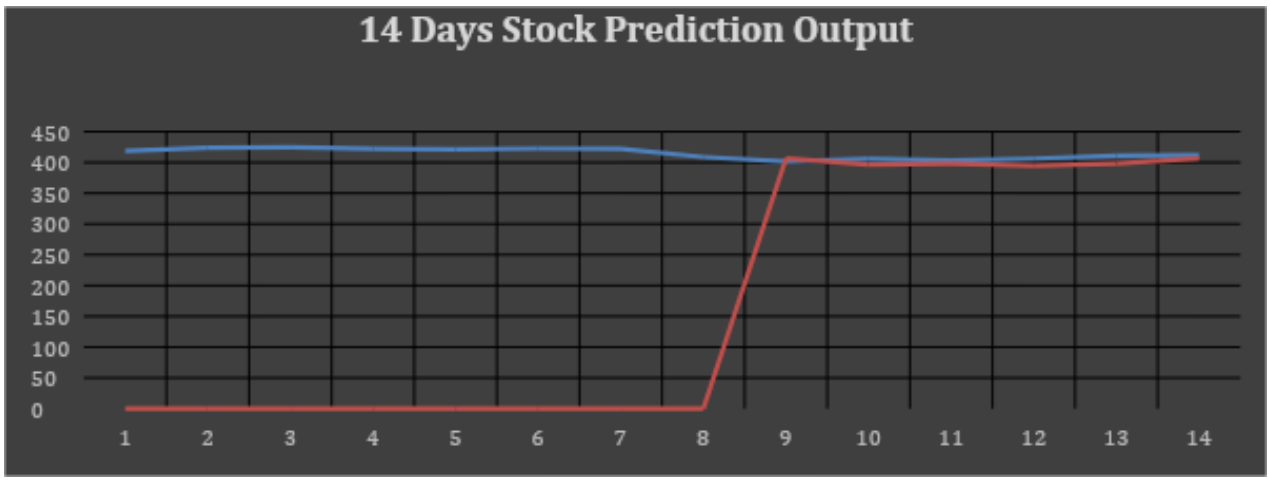


Fig.6.3. 14 Days Stock Prediction Output

This is the result of using the head(). Since the system uses the pandas library to analyses the data, it returns the first five rows. Here five is the default value of the number of rows it returns unless stated otherwise. The trading code in the processed data set is not relevant so it use the strip() to remove it and replace all of the trading codes with a value "Open".

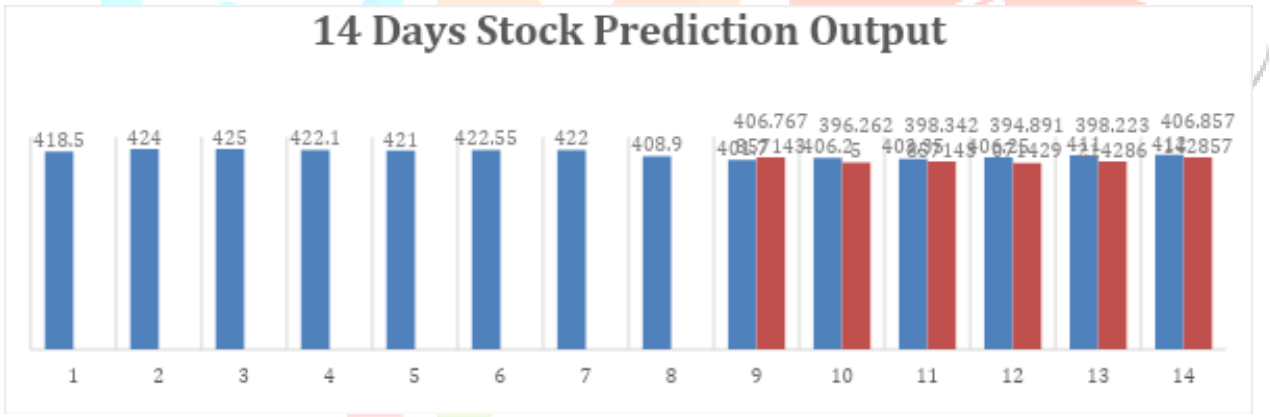


Fig.6.4. 14 Days Stock Prediction Output

This is a time series plot generated from using the "matplotlib.pyplot" library. The plot is of the attribute's "CLOSE" vs "DATE". This is to show the trend of closing price of stock as time varies over a span of two years. The figure provided below is the candle stick plot, which was generated using the library "malignance". The candle stick plot was generated suing the attributes 'DATE', 'OPEN', 'HIGH', 'LOW','CLOSE'. The above two figures are histograms plotted between „CLOSEP" and „OPENP" and the attributes „HIGH" and „LOW". This is done because it is believed today's closing price and opening price along with the high and lowest price of the stock during last year will affect the price of the stock at a later date. Based on such reasoning it is devised a logic "if today's CLOSEP is greater than yesterday's CLOSEP then it is assigned the value 1 to DEX or else it assigns the value -1 to DEX. Based on such the whole data set is processed and upon using the head () we get a glimpse of the data obtained thus far. The next step entailed the setting of feature and target variable, along with the setting of train size. Using the sklearn libraries we import SVC classifier and fit it with the training

data. After training the model with the data and running the test data through the trained model the confusion matrix obtained is shown below.

VII-CONCLUSION

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. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

VIII-FUTURE ENHANCEMENT

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee. The use of traditional algorithms and data mining techniques can also help predict the corporation's performance structure as a whole.

IX-REFERENCES

1. Ashish Sharma, Dinesh Bhuriya, Upendra Singh. "Survey of Stock Market Prediction Using Machine Learning Approach", ICECA 2022.

2. Loke.K.S. "Impact Of Financial Ratios And Technical Analysis On Stock Price Prediction Using Random Forests", IEEE, 2021.

3. Xi Zhang1, Siyu Qu1, Jieyun Huang1, Binxing Fang1, Philip Yu2, "Stock Market Prediction via Multi-Source Multiple Instance Learning." IEEE 2021.

4. VivekKanade, BhausahebDevikar, SayaliPhadatare, PranaliMunde, ShubhangiSonone. "Stock Market Prediction: Using Historical Data Analysis", IJARCSSE 2022.

5. SachinSampatPatil, Prof. Kailash Patidar, Asst. Prof. Megha Jain, "A Survey on Stock Market Prediction Using SVM", IJCTET 2021.

6. https://www.cs.princeton.edu/sites/default/files/uploads/Saahil_magde. pdf

7. Hakob GRIGORYAN, "A Stock Market Prediction Method Based on Support Vector Machines (SVM) and Independent Component Analysis (ICA)", DSJ 2021.

8. RautSushrut Deepak, ShindeIshaUday, Dr. D. Malathi, "Machine Learning Approach In Stock Market 9. Prediction", IJPAM 2022.

10. Pei-Yuan Zhou, Keith C.C. Chan, Member, IEEE, and Carol XiaojuanOu, "Corporate Communication Network and Stock Price Movements: Insights From Data Mining", IEEE 2021

11. Atsalakis GS, Valavanis KP. Forecasting stock market short-term trends using a neuro-fuzzy based methodology. Expert Syst Appl. 2009;36(7):10696–707.

12. Ayo CK. Stock price prediction using the ARIMA model. In: 2022UKSim-AMSS 16th international conference on computer modelling and simulation. 2014. https://doi.org/10.1109/UKSim.2014.67.

13. Brownlee J. Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python. Machine Learning Mastery. 2021. https://machinelearningmastery.com/time-series-prediction-lstm-recurrent-neural-networkspython-keras/

14. Eapen J, Bein D, Verma A. Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction. In: 2019 IEEE 9th annual computing and communication workshop and conference (CCWC). 2022. pp. 264–70. https://doi.org/10.1109/CCWC.2019.8666592.

15. Fischer T, Krauss C. Deep learning with long short-term memory networks for fnancial market predictions. Eur J Oper Res. 2018;270(2):654–69.

16.1 https://doi.org/10.1016/j.ejor.2017.11.054. 6. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classifcation using support vector machines. Mach Learn 2002;46:389–422.

17. Hafezi R, Shahrabi J, Hadavandi E. A bat-neural network multi-agent system (BNNMAS) for stock price prediction: case study of DAX stock price. Appl Soft Comput J. 2015;29:196–210. https://doi.org/10.1016/j.asoc.2014.12.028.

18. Halko N, Martinsson PG, Tropp JA. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. SIAM Rev. 2001;53(2):217–88.

19. Hassan MR, Nath B. Stock market forecasting using Hidden Markov Model: a new approach. In: Proceedings—5th international conference on intelligent systems design and applications 2022, ISDA'05. 2022. pp. 192–6. https://doi. org/10.1109/ISDA.2005.85.

20. Hochreiter S, Schmidhuber J. Long short-term memory. J Neural Comput. 1997;9(8):1735–80. https://doi. org/10.1162/neco.1997.9.8.1735. 11. Hsu CM. A hybrid procedure with feature selection for resolving stock/futures price forecasting problems. Neural Comput Appl. 2013;22(3–4):651–71. https://doi.org/10.1007/s00521-011-0721-4.

21. Lei L. Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. Appl Soft Comput J. 2018;62:923–32. https://doi.org/10.1016/j.asoc.2020.09.029.

22. Lin X, Yang Z, Song Y. Expert systems with applications short-term stock price prediction based on echo statenetworks. Expert Syst Appl. 2020;36(3):7313–7. https://doi.org/10.1016/j.eswa.2008.09.049.

23. Liu G, Wang X. A new metric for individual stock trend prediction. Eng Appl Artif Intell. 2019;82(March):1–12. https:// doi.org/10.1016/j.engappai.2019.03.019.

24. Liu S, Zhang C, Ma J. CNN-LSTM neural network model for quantitative strategy analysis in stock markets. 2020;1:198–206. https://doi.org/10.1007/978-3-319-70096-0.

25. Long W, Lu Z, Cui L. Deep learning-based feature engineering for stock price movement prediction. Knowl Based Syst. 2020;164:163–73. https://doi.org/10.1016/j.knosys.2020.10.034.

26. Malkiel BG, Fama EF. Efcient capital markets: a review of theory and empirical work. J Finance. 2022;25(2):383–417.

27. McNally S, Roche J, Caton S. Predicting the price of bitcoin using machine learning. In: Proceedings—26th Euromicro international conference on parallel, distributed, and network-based processing, PDP 2020. pp. 339–43. https:// doi.org/10.1109/PDP2018.2020.00060.

28. Nagar A, Hahsler M. News sentiment analysis using R to predict stock market trends. 2018. http://past.rinfnance .com/agenda/2012/talk/Nagar+Hahsler.pdf. Accessed 20 July 2019.

29. Nekoeiqachkanloo H, Ghojogh B, Pasand AS, Crowley M. Artifcial counselor system for stock investment. 2019. ArXiv Preprint arXiv:1903.00955.

30. Ni LP, Ni ZW, Gao YZ. Stock trend prediction based on fractal feature selection and support vector machine. Expert Syst Appl. 2020;38(5):5569–76. https://doi.org/10.1016/j.eswa.2022.10.079.

31. Pang X, Zhou Y, Wang P, Lin W, Chang V. An innovative neural network approach for stock market prediction. J Supercomput. 2018. https://doi.org/10.1007/s11227-017-2228-y.

32. Pimenta A, Nametala CAL, Guimarães FG, Carrano EG. An automated investing method for stock market based on multiobjective genetic programming. Comput Econ. 2022;52(1):125–44. https://doi.org/10.1007/s1061 4-017-9665-9.

33. Piramuthu S. Evaluating feature selection methods for learning in data mining applications. Eur J Oper Res. 2022;156(2):483–94. https://doi.org/10.1016/S0377-2217(02)00911-6.

34. Qiu M, Song Y. Predicting the direction of stock market index movement using an optimized artifcial neural network model. PLoS ONE. 2022;11(5):e0155133.

35.Scikit-learn.Scikit-learnMin-MaxScaler.2019.https://scikit-learn.org/stable/modules/generated/sklearn.preprocess ing.MinMaxScaler.html. Retrieved 26 July 2020.

36. Shen J. Thesis, “Short-term stock market price trend prediction using a customized deep learning system”, supervised by M. Omair Shafq, Carleton University. 2019.

37. Shen J, Shafq MO. Deep learning convolutional neural networks with dropout—a parallel approach. ICMLA. 2018;2018:572–7.

38. Shen J, Shafq MO. Learning mobile application usage—a deep learning approach. ICMLA. 2019;2019:287–92.

39. Shih D. A study of early warning system in volume burst risk assessment of stock with Big Data platform. In: 2019 IEEE 4th international conference on cloud computing and big data analysis (ICCCBDA). 2019. pp. 244–8.

40. Sirignano J, Cont R. Universal features of price formation in fnancial markets: perspectives from deep learning. Ssrn. 2018. https://doi.org/10.2139/ssrn.3141294.

41. Thakur M, Kumar D. A hybrid fnancial trading support system using multi-category classifers and random forest. Appl Soft Comput J. 2018;67:337–49. https://doi.org/10.1016/j.asoc.2018.03.006.

42. Tsai CF, Hsiao YC. Combining multiple feature selection methods for stock prediction: union, intersection, and multiintersection approaches. Decis Support Syst. 2020;50(1):258–69. https://doi.org/10.1016/j.dss.2020.08.028.

43. Tushare API. 2018. https://github.com/waditu/tushare. Accessed 1 July 2020.

44. Wang X, Lin W. Stock market prediction using neural networks: does trading volume help in short-term prediction?. n.d.

45. Weng B, Lu L, Wang X, Megahed FM, Martinez W. Predicting short-term stock prices using ensemble methods and online data sources. Expert Syst Appl. 2020;112:258–73. https://doi.org/10.1016/j.eswa.2018.06.016.

46. Zhang S. Architectural complexity measures of recurrent neural networks, (NIPS). 2019. pp. 1–9.