STOCK MARKET PREDICTION USING MACHINE LEARNING

SUBMITTED BY

KANHAIYA SUKHLAL KHANSARE
M.SC.IT (Part-2) Student of University Department Of Information Technology , Mumbai university

&

ABHIJEET SHIVAJI DABHOLE
M.SC.IT (Part-2) Student of University Department Of Information Technology , Mumbai university

Under the Guidance of
Dr. Srivaramangai Ramanujam

For the fulfilment of M.Sc in Information Technology
Part II Semester III

Technical Writing and Entrepreneurship Development
Practical Examination 2020-2021

(First Part of Documentation of Project**)**

Acknowledgement

I express my sincere thanks to University Department of Information Technology (UDIT), University of Mumbai for giving me the opportunity to pursue my M.Sc.(Information Technology)also express my thanks to teaching and non-teaching staff of UDIT.

I would like to thank my guide Mrs. Srivaramangai Ramanujam for her encouragement and guidance, which helped me in completing the project. Finally, I would like to thank my classmates and friends who have helped me in completing the project successfully.

Mr. KANHAIYA SUKHLAL KHANSARE;
Mr. ABHIJEET SHIVAJI DABHOLE

INDEX

CHAPTER 1:
INTRODUCTION

The Covid pandemic has brought millions of new investors into the Indian stock markets, ushering in a massive buying spree. These newbie investors, have ventured into the market. 6.3 Million Demat Account Opened Between April-September as Covid Drives New Entrants in Stock Markets. The prediction process of stock values is always a challenging problem. A correct prediction of stocks can lead to huge profits for the Investers.

Frequently, it is brought out that prediction is chaotic rather than random, which means it can be predicted by carefully analyzing the history of respective stock market. Machine learning is an efficient way to represent such processes. It predicts a market value close to the tangible value, thereby increasing the accuracy. Introduction of machine learning to the area of stock prediction has appealed to many researches because of its efficient and accurate measurements.

The dated market hypothesis believes that it is impossible to predict stock values and that stocks behave randomly, but recent technical analyses show that most stocks values are reflected in previous records; therefore the movement trends are vital to predict values effectively. In our system we will collect the past 3 years of company data and analysis that and using linear regression we will predict the accurate or nearest price of share. It will help investor to make profit.

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.

Linear regression is one of the most commonly used predictive modelling techniques.It is represented by an equation $Y = a + bX + e$, where a is the intercept, b is the slope of the line and e is the error term. This equation can be used to predict the value of a target variable based on given predictor variable(s).

Exchanging the stocks on money markets is one of the significant speculation exercises. Already, scientists developed different stock examination system that could empower them to envision the bearings of stock esteem development. Predicting and foreseeing of significant worth future cost, in perspective of the present cash related information and news, is of colossal use to the financial pros. Financial masters need to know whether some stock will get higher or lower over particular time-period.

Chapter 2: Problem Statement

The Stock Market prediction task is interesting as well as divides researchers and academics into two groups those who believe that we can devise mechanisms to predict the market and those who believe that the market is efficient and whenever new information comes up the market absorbs it by correcting itself, thus there is no space for prediction

In order to predict the stock prices in future markets, we have analyzed papers and has given an overview on how these algorithms give precise and accurate future predictions. In this paper, we used several algorithms from which we observed that not all the algorithms implemented can predict data we need. There has been a basic requirement for computerized and automized ways to deal with powerful and proficient usage of huge measure of money related information to help organizations and people in vital arranging and decision making on investments.

Chapter 3: Description of the Technical Domain and Application Domain

The technical domain of this project is Finance because this system is used in financial sector. The system helps to make profit in financial sector. It will boost the financial sector.
We use machine learning and algorithms of machine learning to for this system that's why the application domain of the project is Machine learning.

Chapter 4: Literature Survey

1) M.Suresh Babu et al., 2012, this paper investigates the significant clustering calculations: K-Means, Hierarchical grouping calculation and turn around K-Means and look at the execution of these noteworthy clustering calculations on part of effectively class savvy group building capacity of calculation. The proposed strategy comprises of three stages. To start with, they change over each money related report into an element vector and utilize the various levelled agglomerative grouping strategy to isolate the changed over element vectors into bunches. They consider both subjective and quantitative highlights in monetary reports. Second, they join the upsides of two grouping techniques to propose a compelling clustering strategy. Third, picking a fitting number of parts in HAC can limit the bunches produced and in this way enhance the nature of the grouping created by the K-means clustering.

2) Mahajan Shubhrata D et al., 2016, this paper is to anticipate future stock value utilizing forecast idea. In that Parse Records at that point figure anticipated esteem and send to client. Also, consequently perform activities like buy and deal shares utilizing Automation idea. For that utilization Naïve Bayes Algorithm. There is Real time Access by Download log shapes hurray back site and Store in dataset. The investigations uncover a high capability of Naïve Bayes Algorithm in foreseeing the arrival on interest in the offer market.

3) Xiao Ding et al., 2015, recommended that a deep learning technique for occasion driven securities exchange expectation. Initially, occasions are removed from news message, and spoke to as thick vectors, prepared utilizing a International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072 © 2019, IRJET | Impact Factor value: 7.211 | ISO 9001:2008 Certified Journal | Page 5995 novel neural tensor system. Second, a deep convolutional neural system is utilized to demonstrate both here and now and long haul impacts of occasions on stock value developments. They exhibited that deep learning is helpful for occasion driven stock value development forecast by proposing a novel neural tensor system for learning occasion embeddings, and utilizing a deep convolutional neural system to demonstrate the joined impact of long haul occasions and here and now occasions on stock value developments.

4) Adebiyi Ayodele et al., 2012, proposed investigation work to enhance the exactness of every day stock value forecast of securities exchange records utilizing artificial neural networks. The examination utilized three-layer, multilayer perceptron models (a feedforward neural system demonstrate) prepared with backpropagation calculation. This paper displays that hybridized approach can possibly upgrade the nature of basic leadership of financial specialists in money markets by offering more precise stock forecast contrasted with existing specialized investigation based approach.

5) Ayodele A. Adebiyi et al., 2012 , presents comprehensive |technique of building stock price predictive model using the ARIMA model. Published stock data obtained from New You are able to Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE) {are being used} with stock price predictive model developed. Outcomes or Benefits obtained revealed that the ARIMA model has a strong potential for immediate prediction and can remain competitive favorably with existing {processes tactics for stock price prediction.

6) Peihao Li, Chaoqun Jing, et al., 2015 , through this paper they proposed a modelling process and present the estimate SSE (Shanghai Stock Exchange) Composite Index to see the model's

estimation performance, which testifies to be feasible and effective. The forecast condition is essentially a direct condition that alludes to past estimations of unique time series and past estimations of the blunders. As should be obvious in the last expectation, the pattern can be accurately anticipated by the model implying that both the model and free factors are accurately chosen.

7) C.Narendra Babu and B.Eswara Reddy et al., 2014, proposed a direct half and half model utilizing ARIMA and GRACH is created which protects the information pattern and renders great forecast exactness. Appropriately, the given TSD is deteriorated into two distinctive arrangement utilizing a basic moving normal (MA) filter. One of them is displayed utilizing ARIMA and the other is demonstrated utilizing GARCH relevantly. The forecasts got from both the models are then melded to acquire the final demonstrate expectations. Indian Stock market information is considered keeping in mind the end goal to assess the exactness of the proposed display. The execution of this model is contrasted and conventional models, which uncovers that for multi-venture ahead expectation, the proposed display beats the others as far as both forecast precision and safeguarding information drift.

8) Abdulsalam Sulaiman Olaniyi et al., 2011, proposed in this paper, the serial development of stock costs over some stretch of time extricated from the everyday official rundown of Stock Exchange, are utilized as a part of building a database and estimations of factors were separated from the database to anticipate the future estimations of different factors using time arrangement information that utilized moving normal technique. They exhibited regression investigation as an information mining strategy and created instrument for abusing particularly time arrangement information in money related organization. An expectation framework has been constructed that utilizations information mining method to create intermittently estimates about securities exchange costs and ready to utilize regression investigation as an information mining system to portray the patterns of stock exchange costs and foresee the future securities exchange costs.

9) Jatinder N.D. Gupta et al., 2000, gives a review to the tasks research reader of the fundamental neural network methods, and in addition their verifiable and ebb and flow use in business. Neural networks and information mining are not enchantment answers for issues, in spite of the message indicated by merchants of programming items. They explored neural network strategies in business from the viewpoint of the tasks researcher. The three fundamental neural network ways to deal with taking care of business issues have been presented: multi-layered feedforward neural networks, Hopfield neural networks, and self-sorting out neural networks. Every one of these methods discovers regular similarity with more customary factual and tasks research strategies, and these analogies have been talked about.

10) Dinesh Bhuria et al., 2017, surveyed about stock market prediction using regression techniques and proposed productive regression way to deal with foresee the stock market cost from stock market information based. In future the consequences of multiple regression approach could be enhanced utilizing more number of factors. This examination ponder is to help the stock merchants and speculators for putting cash in the stock market. The expectation International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072 © 2019, IRJET | Impact Factor value: 7.211 | ISO 9001:2008 Certified Journal | Page 5996 plays an imperative part in stock market business which is exceptionally confused and testing process due to dynamic nature of the stock market.

11) Md Jan Nordin et al., 2012 , looks at the hypothesis and routine with regards to regression methods for expectation of stock value slant by utilizing a changed informational collection in ordinal information format. In this examination, all information in numerical esteems are changed into ordinal or counted qualities to frame the dataset. Regression based classifiers from WEKA are then utilized as prescient examination to test the ordinal information. The results were thought about and assessed.

12) Bhagyashree Nigade et al., 2017 , surveyed that the stock market does not have an effective programming where the best possible proposals of accessible stocks and the correct speculation investigation are displayed in an effective way. The paper clarifies the advancement and execution of a stock value prediction application utilizing machine learning algorithm and protest situated approach of programming framework improvement. The algorithm was utilized as a part of preparing an arrangement of market information gathered for the time of one thousand, two hundred and three days. And a prediction framework has been manufactured that utilizations information mining procedure to deliver intermittently gauges about stock market costs. The utilization of back engendering in neural system empowers us to decrease blunders and enhance exactness of the framework.

13) Hemangi Karchalkar et al., 2017, explained a stock value prediction technique in this undertaking and for this reason regression algorithm and question arranged approach of programming advancement is used. The winning strategies demonstrate a pattern on future development of stocks and not the conceivable cost for any stock later on. It is in this manner desirable over have an instrument that does point a bearing towards value development, as well as demonstrates the doubtlessly cost of the stock itself

14) Luckyson Khaidem et al., 2016, proposed a novel method to limit the danger of interest in stock advertise by foreseeing the profits of a stock utilizing a class of intense machine learning calculations known as ensemble learning. They have utilized four administered learning calculations, i.e “Logistic Regression, Gaussian Discriminant Analysis, Quadratic Discriminant Analysis, and SVM”.

15) T. Manojlović et al., 2015, utilized the ‘Random-forest’ calculation to construct the model used to anticipate 5- days-ahead and 10-daysahead bearings of the CROBEX record andchose stocks. Their outcomes demonstrate that random forests can be effectively used for building predictive models for anticipating the course of securities exchange patterns.

16) Mohammad Bolandraftar et al., 2014, attempted to create three models and looked at their exhibitions in anticipating the heading of development in every day Tehran Stock Exchange (TSE) file. The models depend on three order procedures, Decision Tree, Random Forest and Naïve Bayesian Classifier. What's more, reasoned that principal investigation assumes less vital part than specialized examination during the time spent basic leadership of brokers and partners.

17) Khalid Alkhatib et al., 2013, applied “K-nearest Neighbour” algorithm and indirect relapse approach so as to anticipate stock costs for an example of six noteworthy organizations recorded on the Jordanian stock exchange to help financial specialists, administration, chiefs, and clients in making right and educated ventures choices. As indicated by the outcomes, the k-

NN algorithm is vigorous with little mistake proportion; subsequently the outcomes were sound and furthermore sensible.

18) San- hing Liul et al., 2010, build up a fundamental anticipating model in view of KNN and BP Neural Network. They attempted it in anticipating the stock cost of China stock and test comes about demonstrate that the normal blunders happening in KNN-ANN algorithm are littler than those in KNN algorithm, showing that the anticipating model in light of KNN-ANN algorithm can do better in the stock forecast.

19) Tian Ye et al., 2017 , showed a stock estimating model in light of wavelet examination and ARIMA-SVR. The stock cost is disintegrated into remade part and blunder part by wavelet deterioration and wavelet remaking. At that point, the ARIMA show and the SVR display are utilized to gauge the recreated part and the blunder part separately, and the last forecast outcomes are joined to get the last expectation comes about. The day by day shutting cost of Shanghai Pudong Development Bank from January 5, 2015 to January 29, 2016, was chosen as the trial information, with a sum of 250, of which the initial 220 were the preparation set and the last 30 were the test set to do recreation try. The International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 05 | May 2019 www.irjet.net p-ISSN: 2395-0072 © 2019, IRJET | Impact Factor value: 7.211 | ISO 9001:2008 Certified Journal | Page 5997 test comes about demonstrate that contrasted and the single determining model, the proposed display is a powerful technique for anticipating stock value, which extraordinarily enhances the precision of estimating.

20) Robert P. Schumaker et al., 2008, discovered that two speculations have significantly affected statistical surveying: Efficient Market Hypothesis (EMH) and Random Walk Theory. In EMH, the cost of a security is an impression of finish showcase data. At whatever point a change in money related viewpoint happens, the market will in a flash alter the security cost to mirror the new data (Fama, 1964). EMH contained three unique levels of data sharing: the powerless frame, the semi-solid and the solid shape. Inside powerless EMH, just recorded information is implanted inside the present cost. The semi- solid frame goes more remote by joining chronicled and current open data into its costs. The solid shape incorporates verifiable and current open data and additionally private data. From these three structures, it was trusted that business sectors acted proficiently and that immediate value amendments would forestall expectation models.

Chapter 5: Methodology

Stock market prediction seems a complex problem because there are many factors that have yet to be addressed and it doesn't seem statistical at first. But by proper use of machine learning techniques, one can relate previous data to the current data and train the machine to learn from it and make appropriate assumptions.

Machine learning as such has many models but this paper focuses on two most important of them and made the predictions using them. Fundamental analysis , Technical analysis , Machine learning and Data sources for market prediction.
Linear regression is a supervised learning algorithm to predict the outcome of a continuous variable

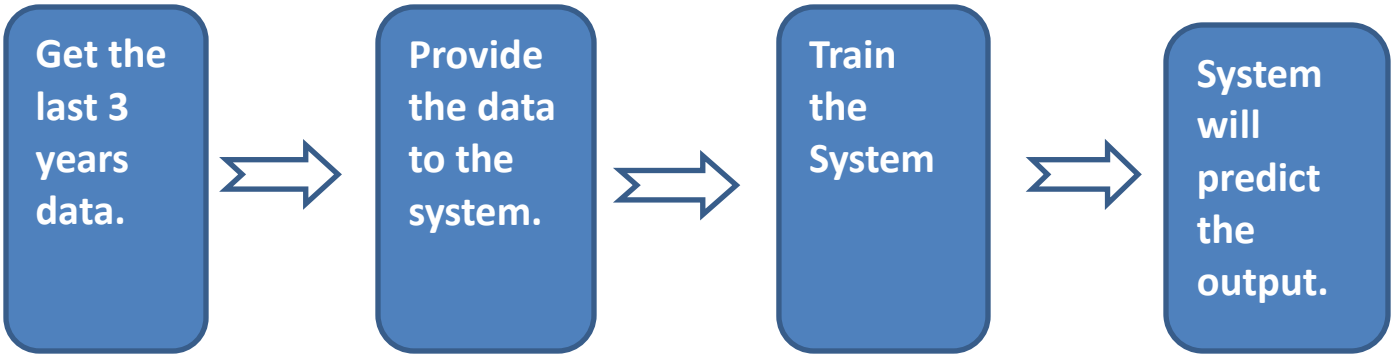
- **Hardware and Software to be used:**

Hardware:

1. Intel i3 4th generation processor
2. RAM 3GB DDR3
3. 50 GB hard disk space.

Software:

a. Window OS.
b. Python
c. Chrome
d. jupyter notebook
e. Python library



❖ **Step**

1. Get the last 3 years data.
2. Use Python programming language and jupyter notebook framework
3. Use pandas pandas , numpy and matplotlib library
4. Linear regression code execute
5. Predication graph output

Chapter 6: REFERENCES

1. https://www.researchgate.net/publication/331345883_Stock_Market_Prediction_U ng_Machine_Learning
2. https://towardsdatascience.com/predicting-future-stock-market-trends-with-python-machine-learning-2bf3f1633b3c
3. M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.
4. K. Raza, "Prediction of Stock Market performance by using machine learning techniques," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1
5. H. Gunduz, Z. Cataltepe and Y. Yaslan, "Stock market direction prediction using deep neural networks," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.
6. M. Billah, S. Waheed and A. Hanifa, "Stock market prediction using an improved training algorithm of neural network," 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4.
7. H. L. Siew and M. J. Nordin, "Regression techniques for the prediction of stock price trend," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.
8. C. N. Babu and B. E. Reddy, “Selected Indian stock predictions using a hybrid ARIMA-GARCH model,” 2014 Int. Conf. Adv. Electron. Comput. Commun. ICAECC 2014, 2015.
9. M. D. Shubhrata, D. Kaveri, T. Pranit, and S. Bhavana, “Stock Market Prediction and Analysis Using Naïve Bayes,” Int. J. Recent Innov. Trends Comput. Commun., vol. 4, no.11, pp. 121–124, 2016.
10. X. Ding, Y. Zhang, T. Liu, and J. Duan, “Deep Learning for Event-Driven Stock Prediction,” no. Ijcai, pp. 2327–2333, 2015.
11. K. A. Smith and J. N. D. Gupta, “Neural networks in business: Techniques and applications for the operations researcher,” Comput. Oper. Res., vol. 27, no. 11–12, pp.1023–1044, 2000.
12. M. S. Babu, N. Geethanjali, and P. B. Satyanarayana, “Clustering Approach to Stock Market Prediction,” vol. 1291, pp. 1281– B. Sivakumar and K. Srilatha, “A novel method to segment blood vessels and optic disc in the fundus retinal images,” Res. J. Pharm. Biol. Chem. Sci., vol. 7, no. 3, pp. 365–373, 2016.
13. M. P. Mali, H. Karchalkar, A. Jain, A. Singh, and V. Kumar, “Open Price Prediction of Stock Market using Regression Analysis,” Ijarcce, vol. 6, no. 5, pp. 418–421, 2017.
14. Yunus Yetis; Halid Kaplan; Mo Jamshidi (2014), “Stock Market Prediction by Using Artificial Neural Network”, 2014 World Au- tomation Congress (WAC), pages: 718 - 722, DOI: 10.1109/ WAC. 2014.6936118.
15. Cakra, Y. E., & Distiawan Trisedya, B. (2015). Stock price prediction using linear regression based on sentiment analysis. 2015 International Conference on

Advanced Computer Science and Information Systems (ICACSIS). doi:10.1109/icacsis.2015.7415179
16. Kumar, K.; Haider, M.T.U. Blended computation of machine learning with the recurrent neural network for
17. intra-day stock market movement prediction using a multi-level classifier. Int. J. Comput. Appl. 2019, 1–17.
18. Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining - Practical Machine Learning Tools and Techniques (3rd Ed). Burlington: Morgan Kaufmann, pp. 191-192
19. Mehak Usmani; Syed Hasan Adil; Kamran Raza; Syed Saad Azhar Ali “Stock Market Prediction Using Machine Learning Tech- niques” 2016 3rd International Conference on Computer and Infor- mation Sciences (ICCOINS) Pages: 322 - 327, DOI: 10.1109/ICCOINS.2016.7783235
20. Murphy, K.P. Machine Learning: A Probabilistic Perspective; MIT Press: Cambridge, MA, USA, 2012.
21. Parmar, I., Agarwal, N., Saxena, S., Arora, R., Gupta, S., Dhiman, H., & Chouhan, L. (2018). Stock Market Prediction Using Machine Learning. 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC). doi:10.1109/icsccc.2018.8703332
22. M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.
23. K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions," INTERACT-2010, Chennai, 2010, pp. 59-63.
24. Mankar, T., Hotchandani, T., Madhwani, M., Chidrawar, A., & Lifna, C. . (2018). Stock Market Prediction based on Social Sentiments using Machine Learning. 2018 International Conference on Smart City and Emerging Technology (ICSCET). doi:10.1109/icscet.2018.8537242
25. Khan, W., Ghazanfar, M. A., Azam, M. A., Karami, A., Alyoubi, K. H., & Alfakeeh, A. S. (2020). Stock market prediction using machine learning classifiers and social media, news. Journal of Ambient Intelligence and Humanized Computing. doi:10.1007/s12652-020-01839-w
26. Li, Lei, Yabin Wu, Yihang Ou, Qi Li, Yanquan Zhou, and Daoxin Chen. (2017) “Research on machine learning algorithms and feature extraction
27. for time series.” IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC): 1-5.
28. Tae Kyun Lee et al. “Global stock market investment strategies based on financial network indicators using machine learning tech niques”, Expert Systems with Applications, 117(2019):228-242.