

Step by Step Implementation of Machine Learning Algorithms for Stock Price Prediction

Priyanka Gupta

Ph.D. Research Scholar, School of Data Science and Forecasting, Devi Ahilya Vishwavidyalaya, Indore (MP)
Assistant Professor, Dept. of IT, D. Y. Patil College of Engineering, Pune (MH)
indian.priyanka@gmail.com

Abstract: *The Stock market prediction is a very important aspect in the financial sector. Stock market investigation empowers financial investors to distinguish the inherent worth of a security even prior to putting resources into it. All securities exchange tips are formed after intensive examination by specialists. Stock experts attempt to discover action of an instrument/area/market in future. By utilizing prediction analysis of stock, financial investors and brokers show up at value purchasing and selling choices. Prediction research analysis of stock prior to making a speculation is an unquestionable requirement. It is exclusively after a careful exploration that you can make a few suppositions into the worth and future exhibition of a stock price. Regardless of whether you are following stock exchanging tips, it is ideal to do some examination, just to guarantee that you are making a trade that is relied upon to get you the most extreme returns. In this paper, various techniques of machine learning is used to predict share price in the stock market. The successful prediction of the stock market will have a very positive impact on the stock market institutions and the investors also.*

Keywords: Machine Learning, Linear Regression, K-Nearest Neighbors, Naïve Bays, Decision Tree.

I. INTRODUCTION

Stock market consists of various buyers and sellers of stock. Prediction available market is a potential and a helpful space of analysis for business call makers. The stock trading is an organic process, advanced and volatile method. Stock prediction is characterized by information intensity, noise, uncertainty and hidden relationships. Prediction of stock techniques obtainable is incredibly complicated and important analysis topic. it's difficult once the data available is noisy and not stationary. it's vital as a result of it yield important results for decision markers. Stock trading may be a location where corporations invest high capital and do their shares trading. The predicting cost is crucial issue for investors for creating cash. Researchers have tried that it's attainable to predict stock costs. It put together helps investors to from commercialism or buying choices to induce higher profits.[1]. There are various algorithms used in data science for stock market value used in data science for stock market value prediction as shown in Figure 1.



Figure 1: Machine Learning Algorithms

A. Linear Regression

One of the simplest and most well-liked supervised Machine Learning algorithms is linear regression. It is an applied math approach that's used for prophetic analysis. Linear regression shows the linear relationship between dependent and independent variables as shown in figure 2. which implies however, the worth of the variable quantity is varied in keeping with the worth of the experimental variable. Linear regression could be an easy technique and quite straightforward to interpret, however downside is that regression algorithms model over fits to the date and month worth rather than taking under consideration the previous values from the purpose of prediction, the Figure 2. Linear Regression can contemplate the worth from a similar date a month past, or a similar date/month a year or past. is the assumption of linear relationship between dependent and independent variables. [2]

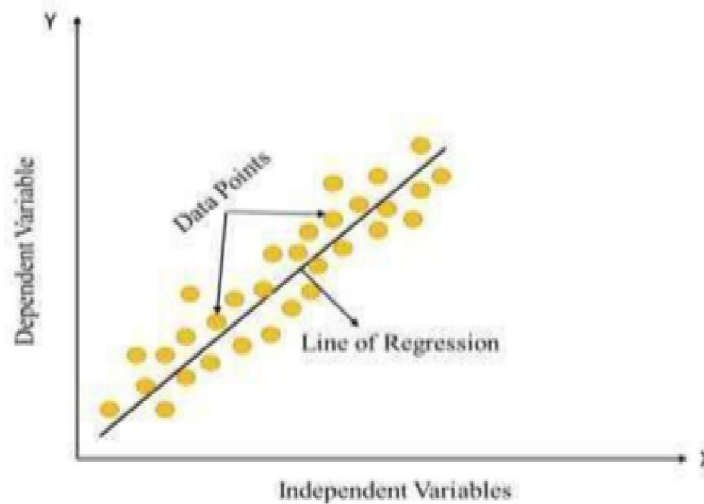


Figure 2: Linear Regression

B. K-Nearest Neighbors (K-NN)

KNN is that the simplest machine learning algorithm that is predicated on supervised learning techniques as shown in figure [3],[4] Assumption for KNN is that the similarity between the new data/cases and offered cases and keep the new case into the class that is analogous to the accessible classes.

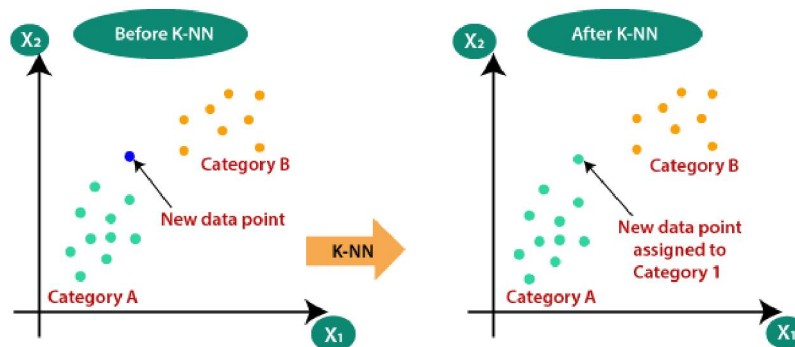


Figure 3. Before KNN

Figure 4. After KNN

The K-NN formula is used for each Regression and Classification however primarily it's used for the Classification issues. it's a non-parametric formula. it's a lazy learner formula because it doesn't acquire from the coaching set at once. for big coaching information it's simpler. to see the worth of K is also complicated your time. it's advantageous if we have a tendency to choose the options properly then it offers superb results. Second is that the K-Nearest neighbor Classifier works well on basic recognition.

The most vital drawback is that the existence of immaterial parameters is incredibly sensitive to them [5]. In stock prediction analysis there is no definite variable to predict future cost of stock. People opinion or emotion about particular stock also affects the price of stock. To categorize tweet to calculate sentiment worth relating to specific stock, Naive Bayes and Random Forest algorithm area unit used. The results of sentiment analysis area unit want to predict the corporate stock value. The linear regression technique used to build the prediction model. [6]

C. Naïve Bayes

It is a supervised learning algorithm that is predicated on mathematician theorem. This algorithm is employed for solving classification issues. it's in the main used for text classification that consists of high dimension training information sets. it's a {quick} machine learning algorithm which may offer quick predictions. It is a probabilistic classifier. it's used for binary furthermore as multiclass classification. The main disadvantage of Naive mathematician is that the assumption of freelance predictor options. Naïve Bayes assumes that each one prognosticators area unit freelance. In world it's seldom happening. [7]

D. Decision Tree

Decision Tree algorithm is supervised learning algorithm. This algorithm accustomed solve each regression and classification issues. Chiefly it's accustomed solve classification issues. It is a tree structured classifier, during which internal nodes represent the options of a dataset, branches represent the choice rules and every leaf node represents the end result. There are a unit 2 nodes within the call tree, Decision node and also the leaf node as shown in figure 5. Decision nodes area unit used for creating a choice and it's multiple branches. Leaf nodes don't contain from now on branches and area unit the output of these choices. the choices area unit taken on the premise of options of the given dataset. It's a graphical illustration for all potential solutions to a tangle based on given conditions. To create a tree Classification and Regression Tree (CART) algorithm is used. It is straightforward to grasp. information cleansing demand is a smaller amount as compared to different algorithms. The decision tree could contain many layers, which can build it advanced. [8]

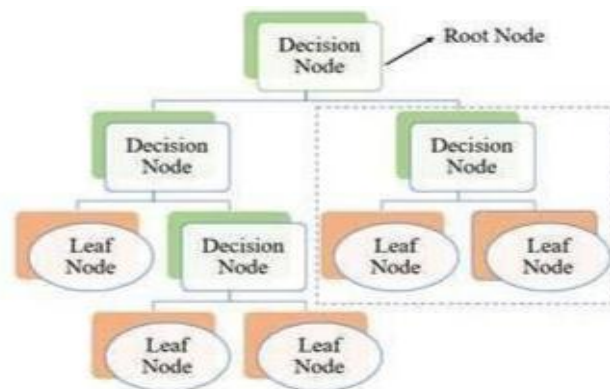


Figure 5. Decision Tree

E. Support Vector Machine (SVM)

Support Vector Machines area unit the foremost standard supervised Learning algorithms. it's used for Classification still as Regression complications. SVM algorithm is to form the simplest line or call boundary which will segregate n-dimensional area into categories so we will simply place the new datum within the correct class. Hyperplane is that the best decision boundary. It works well in generalization. It conjointly reduces process burden. SVM selects the acute points/vectors that facilitate in forming the hyper plane as shown in figure 6. These extreme points area unit referred to as support vectors, and thence the formula is termed as Support Vector Machine. it's simple to tackle the problem of call rules and error frequency. For nonlinearly separable training data, it is not easy to decide optimal parameters. It does not provide good transparency. It is difficult to understand. [9]

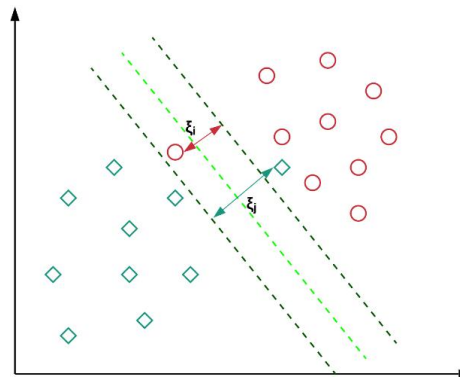


Figure 6. Support Vector Machine

II. LITERATURE REVIEW

One of the most daunting areas of computer science is study of stock market movements and emotions. The author proposed a method in [10] for automatically predicting the stock price using the neural network in this paper. The author demonstrates that the forecast of stock prices from historical prices and sentiments is substantially linked to the real stock price of a specific stock price. In [11] article, the author suggested a new methodology that incorporates the qualitative method and the quantitative method. Data from social media, including text mining and sentiment analysis, was obtained in qualitative analysis. Historical data of individual stock was used in quantitative analysis to forecast market movement. For the latest news headlines, the author also measures sentiment value. The stock prediction value, which supports the opening price and closing price, was calculated by the author using the machine learning module. Fuzzy rules are set to forecast specific market movement on the basis of sentiment value and stock prediction value. This model has suggested the beneficial stock for investment successfully. There are two common methods in this strategy to forecast stock market prices. Technical theories are one of these, and fundamental or intrinsic value analysis is the second. The method proposed in [12] is based on the concept of technological theory. The forecast model can be useful for historical data to obtain potential patterns when and when new data varies in the valuation of the business stock market. Technical analysis and a semi-strong type of successful study. In the proposed work, the market hypothesis is followed to construct the prediction model. This approach uses historical data and social media data to construct a model that forecasts the movement of stock patterns. Two models are constructed, using both models administered by the algorithm of machine learning [13]. The first model is a regular prediction model and the second model is predetermined weekly. The future pattern for the next day is projected by the Regular Forecast model. The monthly forecast model reflects historical data only and forecasts the pattern for the next month. It is very common among investors to forecast the stock market value. Investors want to know what the return on their investment. Stock values are historically forecast by technical analysts and brokers who use historical prices, trends and fundamental trends the stock price forecast for a day is now becoming more complicated as stock prices depend on the political climate, the country's economic situation and natural disasters, etc. The author uses the linear regression method in this approach to estimate the stock value. The technique of linear regression is used because it is very straightforward and is usually appropriate. If we try to estimate the value of a variable based on the value of another variable, linear regression is a statistical model. The Moore and Penrose technique are used by the author in [14] to estimate the regression equation coefficients.

By using Java, the author implements this method. The author collected stock details from the New York Stock Exchange. In [15] two algorithms were used by the author: Least-squares support-vector machines (LS—SVM) are the least-square variants of support vector (SVM) machines and particle swarm machines Optimization (PSO). LS-SVM is an algorithm supervised by machine learning that can be used for the problems of classification or regression.

III. CONCLUSION

Technical research identifies with the investigation of past stock costs to foresee the pattern of costs in future. It shows you the bearing of development of the offer costs. With the assistance of specialized exploration, you can recognize

whether there will be sharp ascent or fall in the cost of offer. It is not reliant on ongoing information or occasions which have just been joined in the price of the stock. As the stock costs are reliant on financial backer sentiment which continues changing as indicated by news and occasions, technical research underlines the utilization of Stop-misfortunes. It will save financial investors from enduring a major misfortune in future Specialized exploration gives significant outcomes just for stocks which are high popular and exchanged enormous volumes. In this paper, we summarized various machine learning algorithms which are relevant to stock prediction.

REFERENCES

- [1]. Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & operations research*, 32(10), 2513-2522.
- [2]. Gururaj, V., Shriya, V. R., & Ashwini, K. (2019). Stock market prediction using linear regression and support vector machines. *Int J Appl Eng Res*, 14(8), 1931-1934.
- [3]. Khan, W., Ghazanfar, M. A., Asam, M., Iqbal, A., Ahmad, S., & Khan, J. A. (2016). Predicting Trend In Stock Market Exchange Using Machine Learning Classifiers *Science International*, 28(2).
- [4]. Y. E. Cakra and B. Distawan Trisedya, "Stock price prediction using linear regression based on sentiment analysis," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Depok, Indonesia, 2015, pp. 147-154, doi: 10.1109/ICACSIS.2015.7415179.
- [5]. Kamble, R. A. (2017, June). Short and long term stock trend prediction using decision tree. *International Conference on Intelligent Computing and Control Systems (ICICCS)* 1371-1375, IEEE.
- [6]. Nair, B. B., Dharini, N. M., & Mohandas, V. P. (2010, October). A stock market trend Prediction system using a hybrid decision tree-neuro-fuzzy system. In *2010 international conference on advances in recent technologies in communication and computing* (pp. 381-385). IEEE.
- [7]. Ashish Pathak, Nisha P Shetty, Indian Stock Market Prediction using Machine Learning and Sentiment Analysis, In *Computational Intelligence in Data Mining* (pp. 595-603). Springer, Singapore.
- [8]. Vivek John George, Darshan M. S, Sneha Pricilla, Arun S, Ch. Vanipriya, India Stock Market Predictor System, *International Journal of Advanced Computational Engineering and Networking*, ISSN: 2320-2106
- [9]. Aparna Nayak, M. M. Manohara Pai and Radhika M. Pai, Prediction Models for Indian Stock Market, Twelfth International Multi-Conference on Information Processing-2016 (IMCIP- 2016).
- [10]. Dr. P. K. Sahoo, Mr. Krishna charlapally, Stock Price Prediction Using Regression Analysis, *International Journal of Scientific & Engineering Research*, Volume 6, Issue 3, March-2015 ISSN 2229-5518.
- [11]. Osman Hegazy, Omar S. Soliman and Mustafa Abdul Salam, A Machine Learning Model for Stock Market Prediction, *International Journal of Computer Science and Telecommunications* [Volume 4, Issue 12, December 2013].
- [12]. Sirimevan, N.; Mamalgaha, I.G.U.H.; Jayasekara, C.; Mayuran, Y.S.; Jayawardena, C. Stock Market Prediction Using Machine Learning Techniques. In *Proceedings of the IEEE 2019 International Conference on Advancements in Computing (ICAC)*, Malabe, Sri Lanka, 5–7 December 2019; Volume 1, pp. 192–197.
- [13]. Strader, T.J.; Rozycki, J.J.; Root, T.H.; Huang, Y.H.J. Machine Learning Stock Market Prediction Studies: Review and Research Directions. *J. Int. Technol. Inf. Manag.* 2020, 28, 63–83.
- [14]. Nair, B.B.; Dharini, N.M.; Mohandas, V. A Stock Market Trend Prediction System Using a Hybrid Decision Tree-Neuro-Fuzzy System. In *Proceedings of the 2010 International Conference on Advances in Recent Technologies in Communication and Computing*, Kottayam, India, 16–17 October 2010; pp. 381–385.
- [15]. Gururaj, V.; Shriya, V.R.; Ashwini, K. Stock market prediction using linear regression and support vector machines. *Int. J. Appl. Eng. Res.* 2019, 14, 1931–1934