Stock Market Prediction Approach: An Analysis

Mayank Joshi
M.Tech Student, Department Of Computer Science and
Engineering
Chandigarh Group of College
Ghauran Mohali, India
mayanksandeepjoshi@gmail.com

Gaurav Goel
Assistant professor, Department Of Computer Science and
Engineering
Chandigarh Group of College
Ghauran Mohali, India
gaurav.goel9@gmail.com

Abstract—**Data mining is a method that can glean meaningful information from a wealth of data. The data mining method known as prediction analysis uses current data to forecast potential future outcomes. Pre-processing the image, to extract the features, and classify the image are just a few of the operations used in the prediction analysis methodology. The stock market forecast is the foundation of this review essay. This study reviews a number of classification-based stock market prediction systems. Python is used to put the stock market forecasting concepts into practice.**

Keywords— Stock market, machine learning, feature extraction

I. Introduction

Data mining is often referred to as "better data use." Since every human being must deal with ever-increasing volume of data, DM or KDD appears to have an impact on everyone. As a result, it is acknowledged as one of the primary research fields. In an ideal world, we would aim to provide methods for "better using any form of data for any purpose." We contend, however, that this objective is still too difficult [1]. More and more historical data has been electronically stored during last few years, and the maximization in thisvolume can be seen in the coming years. Different fund managers are unable to fully realize the significance of this plethora of data, though.Artificial intelligence and data mining methods namelyDTs, the rough set method, and ANNs is recently used in such field [2]. The term "data mining" describes the process of obtaining knowledge from huge data sets or warehouses. Discovering concept or class descriptions, associations and associations, categorization, forecasting, trend analysis, etc. are some of its functions. Data classification is performed easily using several techniques, andDT is a significant method. Every scenario is represented visually and the routes that could lead to it.By offering shares to the general public and institutions, public corporations raise money. An investor gains a portion of the company's ownership by purchasing shares. Stocks of publicly traded companies are traded on a stock market. Buyers and sellers exchange stocks, creating transaction data. Prices fluctuate according to stock supply and demand [3]. The stock market where publicly traded corporations are listed collects all trading data. Data on stock trading is highly time-variant, nonlinear, and volatile. Many pieces of information that are concealed in the enormous amount of data that the stock exchange has collected are time- and labor-intensive for humans to uncover without strong technologies.So it is quite difficult to anticipate the stock's future price. Due to changes in stock market supply and demand, stock prices fluctuate. Investors purchase or sell stocks in publicly traded firms based on historical data and up-to-date information. After examining the fundamentals and technical data of stocks, financiers, stock analysts, etc.

forecast demand and supply of stocks [4]. Investors must sift through a great volume of data in order to find secret patterns, that is an extremely time-consuming and exhausting endeavor. Typically, the analysts evaluate the stocks on the basis of 3 crucial factors: technical analysis, basis analysis, and external environmental factors. Severalconcepts used to perform stock market forecasting, some of these concepts are described as:

A. Efficient Market Hypothesis (EMH)

In 1965, this concept was proposed. According to EMH, a security's price will represent all available market data. Investors will purchase a security as soon as there is any indication that it is undervalued and thus presents a profit opportunity, and its price will climb to its fair value. Weak, semi-strong, and strong financial market efficiency were the three types that were introduced [5]. Since past prices cannot reliably anticipate future ones in weak-form efficiency, the technical analysis was invalid. The price will quickly respond to newly revealed knowledge that is made publicly available in semi-strong-form efficiency; neither fundamental analysis nor technical analysis techniques will consistently generate excess returns.

B. Behavioral finance

Market inefficiency is explained by behavioral finance, including why it may occur. Its purpose is to investigate how psychological, social, cognitive, and emotional aspects affect markets and security. The investor may be psychologically impacted by past trends and the current market environment, and their response will determine the direction of the market [6].Somesignificant approaches make the deployment of methods to predict the stock market which are defined as:

1) Time Series Forecasting (TSF): Analyzing TS data and making predictions about the future is the process of time series forecasting. It models and explains TDS of data points using statistical techniques. It is a method that is frequently used to predict commercial factors.

2) Back Propagation Neural Network (BPNN): A computational model called a neural network (NN) is capable of pattern recognition and estimation (TSF) (Classification). It is resistant to inaccurate and noisy data, and it can also learn and adjust to its surroundings. When an algorithm is indefinable or an exhaustive search is impractical, neural networks can be used. Due to the consideration of the first priori, NN is taken into account. The neural network used in this study is a multi-layer perceptron (MLP), and it learns using the back propagation (BP) algorithm.

The input is considered to predict a particular output while classifying the data. An algorithm is implemented to analyze a training set in which the numerous attributes and the expectedresults are included. It is called an objective or a predictive attribute [7]. This system requires an association

among attributes for predicting the outcomes. After that, a novel dataset is generated and it is referred asprediction set. In this, a set of attributes having similarity with the prior one, is included. The predictive feature is not present and not known. The analysis of the input assists the algorithm in achieving forecasting. The accuracy parameter is utilized to compute the efficacy of the algorithm.

a) Decision tree (D-Tree): The classifier decision tree produces results in the form of a tree structure that resembles a flowchart. Although the D-Tree result is very interpretable, it must be given as categorical data [8]. The "J48" Decision Tree algorithm is typically adopted as it can categorize the trend of the future stock market.

b) Naïve bayes: It is a probabilistic algorithm planned on the basis of Bayes theorem; naive Bayes makes the erroneous assumption that each pair of characteristics is not related to other.

c) k-Nearest neighbors (KNN): KNN is implemented for predicting the class of an object using the k training samples that are closest to it in the feature space. Based on the majority vote of its neighbors, an object is classified; in this study, k = 15, and the object is given the class which is appropriate with its k closest neighbors.

II. LITERATURE REVIEW

Ze Zhang, et.al (2017)suggested an approach using the ENNin order to forecast the stock market's opening price. A local RNN with a context layer which is used for remembering the past situations, the Elman neural network is frequently used to solve time series problems. This research uses a self-adapting version Particle Swarm Optimization algorithm for optimizing the weights and thresholds of the network [9] in light of the constrained Elman network. In order to demonstrate that the suggested an approachis accurate and stablein comparison with the CNN, this study tests the model using various stock prices and performs its comparison with BPNN and ENN.

Nonita Sharma, et.al (2017)highlighted the use of previous data to forecast future stock market index values. The historical data, extracted from 2 indices: the CNX Nifty and S&P BSE Sensex, spanning a period of ten years, applies to compute the presented approach [10]. The proposed model's prediction performance is contrasted against SVR. Each of the prediction models uses a different set of technical indicators as inputs. The predicted variable is the stock price's closing value. The proposed method works better than Support Vector Regression, according to the results, and can be successfully used to create predictive systems in predicting stock values.

MustainBillah, et.al (2016)suggested an enhanced LM artificial neural network method [12]. On the earlier data of DSE containingdifferent prices, and total share traded, an improved LM algorithm of NN is effective for forecasting the closing stock price at lower memory and time at the end of the day. The results of BSE indicated the suggested method as a superior computing tool inpredicting closing stock price. We'll apply this theory to forecasting network traffic in the future.

R. YaminiNivetha, et.al (2017)built a prediction model by analyzing numerous prediction methods. This algorithm is planned on the basis of daily and monthly predictions to anticipate the market price for the following day [13]. This model predicts the market's opening price for the following day. MLR, SVM and ANN algorithmsare compared against one another. The finest forecasting algorithm for sentiment analysis predicts the stock price. The outcome clearly shows that DL approach outperforms the other algorithms. The hidden layer neuron is employed for learning with each prediction in the deep learning algorithm. As a result, the output layer neuron generates the optimal results. The best prediction algorithm is an artificial neural network.

Muhammad Waqar, et.al (2017)used PCA in conjunction with LR to study the issue of the higher dimensional rate of the stock exchange to anticipate the market patterns [15]. Three stock exchanges are the subjects of experiments on a high-dimensional spectrum. Using principal component analysis, the efficacy of the LRalgorithm is evaluated. Generally, the results of the tests demonstrate that PCA can enhance ML performance in general, but only when relative association between input attributes is studied and consideration is given while selecting PCs. The presented approach is assessed with regard to RMSE evaluation measure.

III. TOOL AND TECHNOLOGIES

Python is a higher quality, object-oriented programming language in which dynamic semantics are included. The pairing inherent data frameworks of this language is done with dynamic typing and binding to develop an application and for using glue language forattaching the priorelements. Python is straightforward and easy for learning syntax based readability. This lowers the expenses formaintaining the program. Python's supporting modules and packages promote the modularity of programs and code reuse. The large standard library and Python interpreter can be accessed without any cost on all popular platforms .Python is a very popular choice among programmers due to the productivity boost it offers. There is no compilation stage, therefore a quick edit-test-debug cycle is used instead. Debugging Python is also a breeze. Neither a defect notimproper input leads to create a segmentation fault. Additionally, in case, the interpreter finds a mistake, an exclusion is issued. The interpreter focuses on printing a stack trace in case of failure of the program in handling the exception. A source level debugger examines diverse variables and tests the arbitrary terms. The quickest way of debugging a program is to insert some print statements to the source code.

IV. PERFORMANCE ANALYSIS PARAMETERS

Following are the various parameters for the performance Analysis

A. Accuracy

The ratio of exactly identified points with the total points, whose multiplication is done with 100, is the definition of accuracy.

$$Accuracy = \frac{Number\ of\ points\ correctly\ classified}{Total\ Number\ of\ points} \times 100$$

Authors Names	Year	Description	Outcomes

Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song, Yan Zhu	2017	Suggested an approach using the ENNin order to forecast the stock market's opening price.	In order to demonstrate that the suggested an approachis accurate and stablein comparison with the CNN, this study tests the model using various stock prices and performs its comparison with BPNN and ENN.
Nonita Sharma, AkankshaJuneja	2017	Highlighted the use of previous data to forecast future stock market index values.	The suggested method works better than Support Vector Regression, according to the results, and can be successfully used to create predictive system in predicting stock values.
Yaojun Wang	2016	Employed the technology for mining SM (social media) to quantitatively evaluate market segments and estimate the short-term trajectory of the stock price.	The findings of the experiment demonstrate that the presentedalgorithm is capable of forecasting more accurately when social media mining is supplemented with other data.
MustainBillah, SajjadWaheed, Abu Hanifa	2016	It was suggested to enhance the Levenberg Marquardt (LM) ANN algorithm.	The results of BSE indicated the suggested method as a superior computing tool in predicting closing stock price.
R. YaminiNivetha, Dr. C. Dhaya	2017	MLR, SVM and ANN algorithmsare compared against one another.	As a result, the output layer neuron generates the best results. The best prediction algorithm is an ANN.
Harun Ercan	2017	This work intends to forecast the value of OMXBBGIbased on artificial neural networks.	In particular, characteristics are used to anticipate BSM values using Artificial Neural Network provide the research its distinctiveness.
Muhammad Waqar, Hassan Dawood, Muhammad Bilal Shahnawaz, Mustansar Ali Ghazanfar, Ping Guo	2017	By combining PCA with linear regression, the issue of the stock exchange's higher dimensional rate was addressed in order to forecast market patterns.	The classification model is assessed using the RMSE assessment measure.

B. Precision

The proportion of pertinent examples among the recovered instances in recognizing the pattern, extracting the information, and classifying the data is known as precision.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

C. Recall

It is the portion of applicable examples over all relevant instances that have been retrieved.

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

V. REVIEW METHODOLOGY

The related work of stock market prediction techniques is generated through the vast amount of data provided viaSEs. For this, the enclosureprinciple is deployed on the basis of methodologies. Figures 2(a) and 2(b) display the explanation of the chosen article type and work year. The accompanying graph demonstrates that journals cover 51%, followed by conferences at 40% and book chapters at 9% of the study. Additionally, the graph that follows shows an analysis of stock market forecasting research per year.
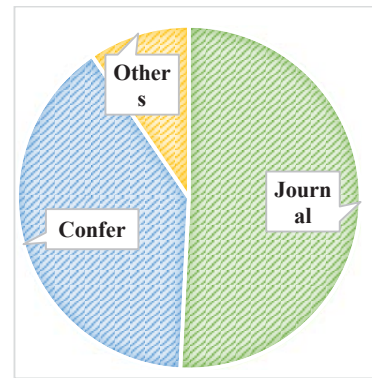

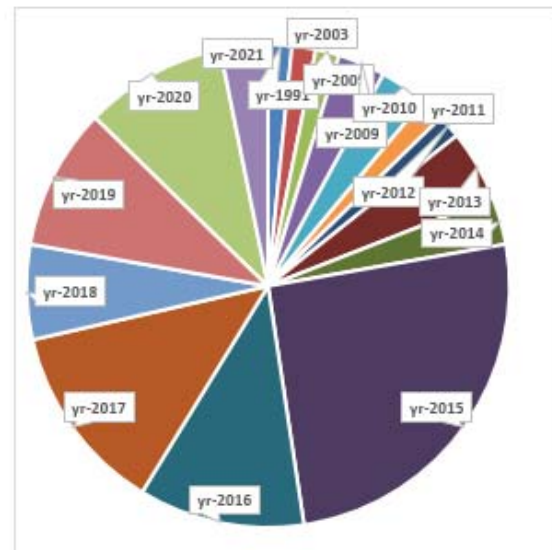Fig. 1. Included research works- Type of Paper

Fig. 2. Year-wise included research

CONCLUSION

This study comes to the conclusion that because the dataset is so complicated, stock market forecasting poses the most challenge to PA. The approach of predicting the stock market includes various processes, such as to extract the attributes and classify the data. A classifier serves as the foundation for the algorithms of predicting the stock market. A hybrid classification approach with excellent stock market forecast accuracy will be developed in the future.

REFERENCES

[1] S. S. A. Ali, M. Moinuddin, K. Raza, & S. H. Adil, "An adaptive learning rate for RBFNN using time-domain feedback analysis," The Scientific World Journal, 2014.

[2] S. Asur, & B. A. Huberman, "Predicting the future with social media," International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010.

[3] C. D. R. Aurangzeb, "Factors Affecting Performance of Stock Market: Evidence from South Asian Countries," International journal of academic research in business and social sciences, vol. 2, no. 9, 2012.

[4] S. Shen, H. Jiang, & T. Zhang, "Stock market forecasting using machine learning algorithms, 2012.

[5] I. Ali, K. U. Rehman, A. K. Yilmaz, M. A. Khan, & H. Afzal, "Causal relationship between macro-economic indicators and stock exchange prices in Pakistan," African Journal of Business Management, vol. 4. No. 3, pp. 312, 2010.

[6] J. Bollen, H. Mao, & X. Zeng, "Twitter mood predicts the stock market. Journal of Computational Science," vol. 2, no. 1, pp. 1-8, 2011.

[7] Ze Zhang, Yongjun Shen, Guidong Zhang, Yongqiang Song, Yan Zhu, "Short-term Prediction for Opening Price of Stock Market Based on Self-adapting Variant PSO-Elman Neural Network", 2017, IEEE

[8] Nonita Sharma, AkankshaJuneja, "Combining of Random Forest Estimates using LSboost for Stock Market Index Prediction", 2017 2nd International Conference for Convergence in Technology (I2CT)

[9] Yaojun Wang, Yaoqing Wang, "Using Social Media Mining Technology to Assist in Price Prediction of Stock Market", 2016 IEEE International Conference on Big Data Analysis (ICBDA)

[10] MustainBillah, SajjadWaheed, Abu Hanifa, "Stock Market Prediction Using an Improved Training Algorithm of Neural Network", 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) 8-10 December 2016

[11] R. YaminiNivetha, Dr. C. Dhaya, "Developing a Prediction Model for Stock Analysis", 2017 International Conference on Technical Advancements in Computers and Communications

[12] Harun Ercan, "Baltic Stock Market Prediction by Using NARX", 2017, 12th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)

[13] Muhammad Waqar, Hassan Dawood, Muhammad Bilal Shahnawaz, Mustansar Ali Ghazanfar, Ping Guo, "Prediction of Stock Market by Principal Component Analysis", 2017 13th International Conference on Computational Intelligence and Security

[14] Jintao Liu, Hongfei Lin, Liang Yang, Bo Xu, Dongzhen Wen, "Multi-Element Hierarchical Attention Capsule Network for Stock Prediction", 2020, IEEE Access

[15] FarnoushRonaghi, Mohammad Salimibeni, FarnooshNaderkhani, ArashMohammadi, "ND-SMPF: A Noisy Deep Neural Network Fusion Framework for Stock Price Movement Prediction", 2020, IEEE 23rd International Conference on Information Fusion (FUSION)

[16] Yuye Liu, Chen Cao, Weixin Huang, Shi Hao, "A Deep Neural Network Based Model For Stock Market Prediction", 2021, IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)

[17] Goel, G., & Tiwari, R. (2022). Resource Scheduling in Fog Environment Using Optimization Algorithms for 6G Networks. International Journal of Software Science and Computational Intelligence (IJSSCI), 14(1), 1-24.

[18] Goel, G., Tiwari, R., Rishiwal, V., & Upadhyay, S. (2018, December). Data preservation by hash algorithm for matrix multiplication over venomous cloud. In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 210-214). IEEE.

[19] Goel, G., & Tiwari, R. (2022). Task management in IoT-Fog-Cloud environment employing static scheduling Techniques. ENP Engineering Science Journal, 2(1), 13-20.

[20] Goel, G., Tiwari, R., Anand, A., & Kumar, S. (2022). Workflow Scheduling Using Optimization Algorithm in Fog Computing. In International Conference on Innovative Computing and Communications (pp. 379-390). Springer, Singapore.

[21] Komal, K., Goel, G., & Kaur, M. (2021, December). Hybrid Scheduling Strategy in Cloud Computing based on Optimization Algorithms. In 2021 2nd International Conference on Computational Methods in Science & Technology (ICCMST) (pp. 51-56). IEEE.

[22] Goel, G., Tiwari, R., Koundal, D., & Upadhyay, S. (2021). Analysis ofResource Scheduling algorithms for optimization in IoT-Fog-Cloud System.

[23] Gill, J. S., Singh, R., Palta, P., Sharma, T., & Goel, G. (2016). CBIR of trademark images in different color spaces using XYZ and HSI. J. Netw. Commun. Emerg Technol.(JNCET), 6(5).

[24] Singh, R., Gill, J. S., Sharma, T., & Goel, G. (2016). Product Aspect Ranking, Sentiment Analysis and Classification of Product Reviews on E-Commerce Sites-A Review. Journal of Network Communications and Emerging Technologies, 6(5).

[25] Gaurav Goel, Tejpal Sharma, Jagbir Singh, Upinderpal Singh, Gagan Singla. Secure Outsourced Computation of Geometric Progression on Vicious Cloud. Journal of Artificial Intelligence Research & Advances. 2019; 6(2): 88–91p.