

A Machine Learning Model for Stock Market Prediction

Pranav Bhat

*Electronics and Telecommunication Department, Maharashtra Institute of Technology, Pune.
Savitribai Phule Pune University*

Abstract- Stock price forecasting is a popular and important topic in financial and academic studies. Share Market prediction is a complicated task since there are no significant rules to estimate or predict the price of share in the share market. Many methods like technical analysis, fundamental analysis, time series analysis and statistical analysis, etc. are all used to attempt to predict the price in the share market none of these methods are proved as a consistently acceptable prediction tool. The new proposed algorithm is tested for its effectiveness in implementation in forecasting stock prices, returns, and stock modeling. This project is for Indian users as the prediction is done on the listed companies of India Stock Exchange. We outline the design of the proposed model with its salient features and customizable parameters. We select a certain group of parameters with relatively significant impact on the share price of a company. With the help of statistical analysis, the relation between the selected factors and share price is formulated which can help in forecasting accurate results. Although, share market can never be predicted, due to its vague domain, this project aims at applying CNN Algorithm in forecasting the stock prices.

I. INTRODUCTION

Trading in shares is big business in many economies. Based on the information on their websites, Stockbrokers do not seem to have any intelligent tool that can help them advise clients on which stocks are suitable for any buy or sale trade. These websites provide information that points to use of fundamental, technical and time series analysis methods. These prevalent methods show a trend on future movement and not the likely trade price for any stock in future. It is therefore desirable to have a tool that does not just point a direction of price movement, but also indicates the most likely price value of the stock itself. An ANN model that is well tuned with the appropriate parameters can be used to develop such a predictive tool.

In recent times stock market predictions is gaining more attention, maybe due to the fact that if the trend of the market is successfully predicted the investors may be better guided. The profit gained by investing and trading in the stock market greatly depends on the predictability. If there is a system that can consistently predict the direction of the dynamic stock market will enable the Users of the system to make informed decisions. More over the predicted trends of the market will help the regulators of the market in taking corrective measures.

Stock value forecast is a work of art and imperative issue. With a fruitful model for stock forecast, we can pick up

knowledge about market conduct after some time, spotting patterns that would somehow not have been taken note. With the inexorably computational intensity of the PC, AI will be an effective strategy to take care of this issue. In any case, the open stock dataset is unreasonably restricted for some, AI calculations to work with, while requesting more highlights may cost a huge number of dollars consistently. In this paper, we will present a structure in which we coordinate client forecasts into the present AI calculation utilizing open verifiable information to improve our outcomes. The inspired thought is that, in the event that we know all data about the present stock exchanging (of every particular dealer), the cost is unsurprising. Consequently, on the off chance that we can get simply halfway data, we can hope to improve the present forecast parcel. With the development of the Internet, informal communities, and online social communications, getting day by day client expectations is practical job. In this manner, our inspiration is to structure an open administration joining chronicled information and clients expectations to make a more grounded model that will profit everybody.

II. LITERATURE SURVEY

A lot of studies provide strong evidence that traditional predictive regression models face significant challenges in out-of sample predictability tests due to model uncertainty and parameter instability. Recent studies introduce particular strategies that overcome these problems. Support Vector Machine (SVM) is a relatively new learning algorithm that has the desirable characteristics of the control of the decision function, the use of the kernel method, and the sparsely of the solution. In this paper, we present a theoretical and empirical framework to apply the Support Vector Machines strategy to predict the stock market. Firstly, four company-specific and six macroeconomic factors that may influence the stock trend are selected for further stock multivariate analysis. Secondly, Support Vector Machine is used in analyzing the relationship of these factors and predicting the stock performance. Our results suggest that SVM is a powerful predictive tool for stock predictions in the financial market. [1]

The prediction of the stock market can generate an actual financial loss or gain, so it is practically important to enhance the predictability of models. Consequently, many studies have been trying to model and predict financial time series, using statistical or soft computational skills that are capable of examining the complex and chaotic financial market. In recent years, deep learning techniques have been actively applied based on their excellent achievements in various classification problems. In this study, we constructed a stock price prediction model based on RNN using LSTM units, which is one of the typical methodologies of deep learning. We

integrated GA and LSTM network to consider the temporal properties of the stock market, and utilized the customized architectural factors of a model. The LSTM network used in this study is composed with two hidden layers, which is a deep architecture for expressing nonlinear and complex features of the stock market more effectively. GA was employed to search the optimal or near-optimal value for the size of the time window and number of LSTM units in an LSTM network. [2]

Sentiment analysis has proven to be effective while analyzing people's attitudes by examining large social data. In this, a novel approach is designed to extract people's opinions on specific topics by relying on social media content. 70% of dataset is used for training whereas remaining dataset is used during testing. It has been determined that the proposed classifier, classifies the positive, negative and neutral sentiments with an accuracy of 98.32% From the experiment, it has been analyzed that the detection accuracy of sentiments has been increased by 8.99 % from the existing work. The main advantage of this work is that a stop word panel is added in to the GUI, so that a user can add or remove the stop words as per the need. [3]

Hegazy, Soliman and Salam [5] implemented a machine learning model to predict stock market price. The algorithm integrated Particle swarm optimization (PSO) and least square support vector machine (LS-SVM). The PSO algorithm selected best free parameters combination from the study of stocks historical data and technical indicators for LS-SVM to avoid over-fitting and local minima problems. The proposed model was applied and evaluated using thirteen benchmark financials datasets and compared with artificial neural network with Levenberg-Marquardt (LM) algorithm. The obtained results revealed that the proposed model has better prediction accuracy and the potential of PSO algorithm in optimizing LS-SVM. [4]

Owing to feature sparse problem and unbalance problem between the number of positive and negative within one day, in order to get the correct corresponding relationship between sentiment polarity and the stock trend, this work proposed a prediction model with an inverse bias algorithm. In order to label the tweets to sentiment related group, this work also proposed an extension wordlist based semisupervised Naïve Bayes classification algorithm. The experimental results show that the hybrid feature selection method improves the accuracy of the stock trend prediction, doing sentiment analysis on example sets is feasible and it can improve the prediction accuracy likewise. The accuracy of linear SVM classification model based on two SVM feature selection hybrid method is improved from 84.06% to 86.95% when using 10-fold-cross validation. In addition, inverse phrase also improves the accuracy from 86.95% to 90.33% by using linear SVM. Different learning algorithms were compared in the experiments. It is found that linear SVM performed the best when dealing with text example sets. For future work, a more

complicated bias calculation will be taken into consideration. [5]

A system is called offline when system takes images as inputs and tries to predict characters from given images by applying machine learning methods. We have worked on offline character recognition problem in this project. Many machine learning methods have been proposed over the years for solving this problem. In this paper we implemented 6 most popular machine learning methods to solve offline handwriting character recognition problem and compare the performance results to decide which method gives best accuracy results under pre-defined conditions. We have selected 92255 images from NIST Special 19 Database and used them as input images during the training phase of the selected machine learning methods. These methods are SVM, Decision Tree, Bag of Trees, Artificial Neural Networks (ANN), Deep learning network with autoencoders and Convolutional Neural Networks (CNN). We implemented all of these methods and compare the performance of the results according to accuracy metric. The results obtained from the comparison is going to help in deciding which ML method should be used to solve Offline Handwriting Character Recognition problem. [6]

It is seen that, regression architectures are capable of capturing dynamics and are able to make predictions. We trained the model using the data of stock and was able to predict stock price of stock. This shows that, the proposed system is capable of identifying some inter relation with in the data. Also, it is evident from the results that, SVR, RFR & DTR model is capable of identifying the changes in trends. For the proposed methodology DTR is identified as the best model. It uses the information given at a particular instant for prediction. Even though the other two models are used in many other time dependent data analysis, it is not out performing the DTR model in this case. This is due to the sudden changes that occurs in stock markets. The changes occurring in the stock market may not always be in a regular pattern or may not always follow the same cycle. Based on the companies and the sectors, the existence of the trends and the period of their existence will differ. The analysis of these type of trends and cycles will give more profit for the investors. To analyze such information we must use networks like DTR as they rely on the current information. [7]

Based on the results obtained, we conclude that the two companies under consideration have less correlation. The stock value change does not depend on the stock exchange index. It is dependent on the sentiments of social media. The prediction using machine learning algorithms do not give accurate results since the correlation between them is less. Results are not accurate as the dependency is less than 50% for all variables. But the graph trends between TCS and Infosys show similar variation except at some points where it was opposite. Combination of natural language processing techniques for analysis and summarization of text can help in

handling such cases. Apart from the parameters which are considered in this paper there can be other parameters which can affect the stock shares such as Inflation, Deflation, International currency and gold rates and International economic policies, etc. Other techniques that can be used are Momentum, Mean Reversion and Martingales. [8]

This paper presents a survey of different techniques such as machine learning techniques, hidden Markov model, ARIMA model and also deep learning techniques. It is observed that selection of the right parameters for the dataset used for prediction plays important role good prediction accuracy. Various machine learning models as well as hybrid and ensemble model give higher rate of accuracy. To get even better accuracy fundamental analysis can be used which uses sentiment analysis and feature selection along with machine learning and deep learning techniques. [9]

Stock price prediction is an important issue in the financial world, as it contributes to the development of effective strategies for stock exchange transactions. In this paper, we propose a generic framework employing Long Short-Term Memory (LSTM) and convolutional neural network (CNN) for adversarial training to forecast high-frequency stock market. This model takes the publicly available index provided by trading software as input to avoid complex financial theory research and difficult technical analysis, which provides the convenience for the ordinary trader of nonfinancial specialty. Our study simulates the trading mode of the actual trader and uses the method of rolling partition training set and testing set to analyze the effect of the model update cycle on the prediction performance. Extensive experiments show that our proposed approach can effectively improve stock price direction prediction accuracy and reduce forecast error. [10]

III. PROPOSED SYSTEM

Diverse securities exchange parameters are utilized for break down financial exchange expectation are as per the following:

- SET list: The SET Index is a composite financial exchange list which is determined from the costs of every single basic stock on the principle leading body of the Stock Exchange, with the exception of stocks that have been suspended for over one year.
- Basic analysis: Fundamental examination of stocks is a method which is useful in settling on speculation choices. Its fundamental significance lies in deciding the inherent estimation of a security. It would then be able to be contrasted with the current stock cost and decided whether the stock is exaggerated or underestimated.
- Shutting estimation of the index: "Closing cost" for the most part alludes to the last cost at which a stock trades amid an ordinary exchanging session. For some U.S. markets, ordinary exchanging sessions keep running from 9:30 a.m. to 4:00 p.m.
- Moving normal hybrid inputs: The most fundamental kind of hybrid is the point at which the cost of an advantage moves from one side of a moving normal and closes on the

other. Value hybrids are utilized by merchants to distinguish moves in force and can be utilized as an essential passage or leave methodology.

- Stock offer esteem: An offer cost is the cost of a solitary offer of various saleable loads of an organization, subordinate or other money related resource. In layman's terms, the stock cost is the most astounding sum somebody is eager to pay for the stock, or the least sum that it very well may be purchased for.

CNN is a numerical model that has been propelled by the creature sensory system comprising of neurons and the manner in which data is gone from all aspects of human body to the mind. Data as certain information esteems is sustained into the system (interconnection between neurons). Based on explicit capacity utilized at each layer and the info esteem the yield can be assessed. It has three kinds of layers:-

- 1) The input layer gets the qualities on which the calculation must be finished. These are the distinctive estimations of the tuples in the dataset.
- 2) In the concealed layer the calculations are done as the qualities are gone through each dimension. The quantity of shrouded layers may fluctuate in various models and applications.
- 3) In the yield layer we get the estimation of the parameter after have been prepared and processed by specific activation work as indicated by the application for which it is structured. It might be a numeric, double or a straight out esteem.

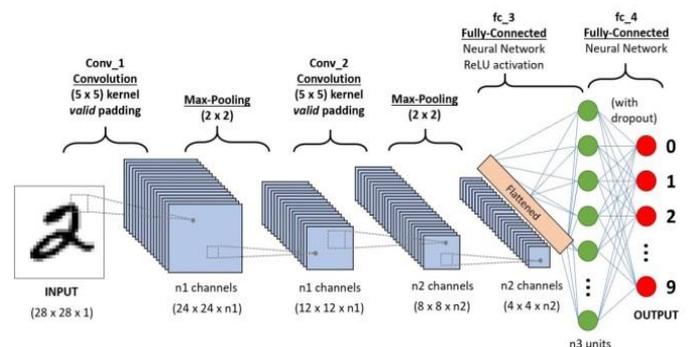


Fig.1: proposed system

CNN has discovered its application in grouping, mechanical technology, relapse, time arrangement assessment and some more. Learning is finished by computing the mean square blunder for each consequent perception and a model is picked that has the least mistake and high prescient power.

IV. CONCLUSION

Anticipating the securities exchange cost is exceptionally well known among financial specialists as speculators need to know the arrival that they will get for their ventures. Generally the specialized experts and intermediaries used to foresee the stock costs dependent on chronicled costs, volumes, value designs and the essential patterns. Today the stock value

expectation has turned out to be mind boggling than before as stock costs are influenced because of organization's money related status as well as due to socio practical state of the nation, political environment and cataclysmic events and so on. The arrival from the offer market is constantly unsure and equivocalness in nature thus conventional procedures won't give precise expectation. A ton explore has been made around there and progressed insightful procedures going from unadulterated numerical models and master frameworks to neural systems have likewise been proposed by numerous budgetary exchanging frameworks for stock value expectation. In this paper we are going to utilize CNN (Convolution Neural Network) for anticipating stocks an incentive for following day.

V. REFERENCES

- [1]. International Journal of Current Trends in Engineering & Technology Volume: 02, Issue: 01 (JAN-FAB, 2016) 18 Stock Market Prediction Using Support Vector Machine Mr. Sachin Sampat Patil, Prof. Kailash Patidar, Assistant Prof. Megha Jain
- [2]. Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction Hyejung Chung and Kyung-shik Shin Sustainability 2018, 10, 3765; doi:10.3390/su10103765
- [3]. An optimized CNN based robust sentiment analysis system on big social data using text polarity feature Komalpreet Kaur, Chitender Kaur, Tarandeep Kaur Bhatia International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-6, April 2019
- [4]. Hegazy, Soliman & Salam. (2013). A Machine Learning Model for Stock Market Prediction. Retrieved June 22, 2018
- [5]. Li, J., & Meesad, P. (2016). *Combining Sentiment Analysis with Socialization Bias in Social Networks for Stock Market Trend Prediction*. *International Journal of Computational Intelligence and Applications*, 15(01), 1650003. doi:10.1142/s1469026816500036
- [6]. Performance Comparison of Machine Learning Methods for Solving Handwriting Character Recognition Problem Ş.G.KIVANÇI, A.E. BAKTIRI and B.ŞENİ International Conference on Advanced Technologies, Computer Engineering and Science (ICATCES'18), May 11-13, 2018 Safranbolu, Turkey
- [7]. JSRD - International Journal for Scientific Research & Development| Vol. 6, Issue 07, 2018 | ISSN (online): 2321-0613 All rights reserved by www.ijrsrd.com 408 Stock Market Prediction using RFR, DTR & SVR Ravikant1 Suman Kumar Swarnkar2L. P. Bhaiya
- [8]. International Journal of Computer Sciences and Engineering Open Access Research Paper Vol.-6, Issue-5, May 2018 E-ISSN: 2347-2693 Stock Market Analysis and Prediction using Hadoop and Machine Learning Piyush Jain, Kaustubh Bhat, HarshalKesharwani3 , Pritesh Bhate, Khushboo P Khurana
- [9]. 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 4842 Stock Market Forecasting Techniques: A Survey Rashmi Sutkatti1, Dr. D. A. Torse
- [10]. Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets, [Xingyu Zhou](#),¹ [Zhisong Pan](#),¹ [Guyu Hu](#),¹ [Siqi Tang](#),¹ and [Chen g Zhao](#) Mathematical Problems in Engineering Volume 2018, Article ID 4907423, 11