

An Efficient Approach on Big Data for Stock Prediction with the Aid of Optimal Machine Learning Approach

T. P Sameerapallavi^{1*}, B. Manjula²

^{1,2}Kakatiya University, Hanamkonda, Warangal, Telangana State, India

*Corresponding Author: thallapally.sameerapallavi@gmail.com

DOI: <https://doi.org/10.26438/ijcse/v7i2.482486> | Available online at: www.ijcseonline.org

Accepted: 20/Feb/2018, Published: 28/Feb/2019

Abstract- The research on stock market is considered as an important issue from recent years. The investment in stock market is performed based on prediction and analysis. The current market economy has numerous variables which need to be considered before doing a transaction in stock market. So, the analysis of the variables manually is a tough task. In order to predict the variables in the stock market and analyse the affecting factors machine learning approach is best suited. The machine learning can provide prediction of different aspects such as index value, higher stock price, exchange rate etc. Different machine learning approaches like naive Bayes classifier, support vector machine, Artificial neural networks are reviewed which helps in stock price prediction and the market prediction. stock market prediction helps the investors and traders make better and quick decisions and ensure profits. Furthermore, advantages and limitations are discussed based on the prediction accuracy and performance.

Keywords: Stock market, Prediction, Big data, Machine learning

I. INTRODUCTION

Due to exponential growth of data, big data analytics and data science has become the emerging areas to organize the raw data. The data analysis and processing offer different kinds of strength towards effective organization and predict the generated data in different formats [10]. Big data analytics helps in analysing both the structured and unstructured data found in many applications or devices such as sensors, mobile communications, geospatial data etc [5]. The importance of the big data has increased over years as it helps in providing flexible management of information assets in organizations thus improving the supply chain performance [2]. One of the important areas of research of big data is stock market prediction. By the aid of the social media and technical advancements, the popularity of stock market prediction has been increased.

In the stock market prediction, the market prices are based on the information present and in the act of modern intelligence arrives into the display the systems shows the unbalanced state. So, the given information is not possible to think towards the specific forthcoming demand as concerns effective sale.

Random walk theory: within the theory that effective stock price changes and holds independent so that the trend of stock price remains not more necessary regarding forecasting towards forthcoming [7]. The prices as concerns effective share market can be modelled using two

approaches such as technical and fundamental. In technical approach the numerical study regarding share prices is determined whereas in original study the public agency analysis is well chosen [1].

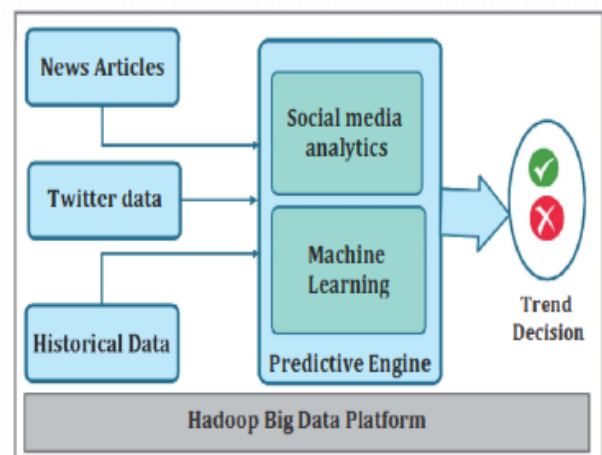


Figure 1. Framework for stock forecast

Expert system has being considered as one of the primary approach considering effective forecasting in regard to commodities exchange. Spectacular inputs for effective data are considered within the stock market itself. Public agency analysis have placed also powerful than ever and the data which is developed in public agency obtain enormous along with unstructured and for this technical guidance remains absolutely necessary towards development of data (Figure

1). So, the tremendous knowledge access has been considered to execute forecast prototype [4]. Powerful forecast prototypes remain correlated moreover calculated by applying expert system approaches such as neural net, boosted tree and regression vector. Expert system has three types such as administered study, unrestrictedly study and augmentation study. Other techniques include artificial neural network, support vector machines, radial basis function etc. The different techniques of machine learning used for the stock prediction are reviewed below.

This paper mainly discussed on the different machine learning approaches on big data for stock price prediction as follows section I contain the Introduction of big data and important research areas of stock market prediction, section II is related to the literature review mainly focused on several techniques proposed by different authors regarding stock market prediction using different machine learning techniques, section III discussed about the big data stock prediction with an aid of many approaches and process involved in stock prediction, section IV contains the different machine learning techniques applied for stock price prediction and results are evaluated, section V concludes that among different machine learning techniques the support vector machine is the most used technique for stock prediction.

II. LITERATURE REVIEW

There are several techniques proposed by different authors regarding stock market prediction using machine learning techniques. These techniques are explored and discussed in the succeeding sections. The efficient regression approach to predict the stock market price from the big data of stock market [14]. Regression approach is a statistical process for estimating the relationships among the variables which includes techniques for modelling and analysing the variables. The different types of regression techniques such as polynomial regression, sigmoid regression, RBF regression are explained. This research study proposed by the author aims in helping the stock brokers for investing the money in the stock market. The prediction also plays an important role in the stock market. Further added, the results of the different regression models could be improved using more number of variables. A machine learning approach for the business intelligence applications such as prediction of the stock price [9]. The support vector machine is applied to perform classification on complex inputs of graph nodes. The structural support vector machine is used for learning a prediction model for a complex input with resulting edges of nodes. The resulting model was applied to the problem of stock price prediction. The experimental results show that the support vector machine is having 78% efficiency higher (Table 1) and has minimum graph cuts. Thus, the practical approach of stock prediction is performed.

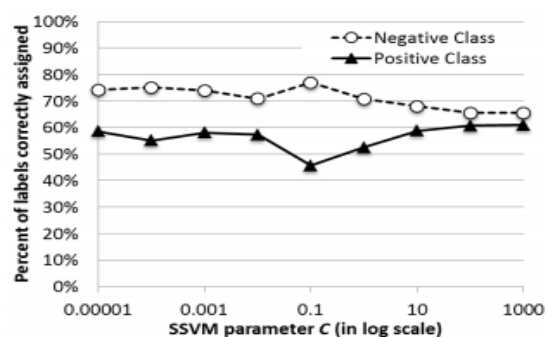


Figure 2. Cross validation of SSVM using two-fold test

The task of predicting the stock market index using future values [13]. The experimental analysis is made by considering the Bombay stock exchange using the historical data from 10 years. The study compares the four prediction models such as artificial neural network, support vector machine, random forest, naive Bayes. The paper focused on two approaches involving support vector regression is used to predict the stock movements in the first stage and the second stage of the fusion approach involves the artificial neural network, Random forest, and SVR which results in SVR-ANN, SVR-RF prediction model. The results are analysed using the prediction performance of the hybrid models and is compared against the single stage algorithms. The results show that two stage hybrid models perform better than the single stage prediction models (Figure 2). The performance improvement is more when ANN and RF are hybridized with SVR. The advantage of two stage prediction models over single stage prediction models is more evident as the predictions are made for more number of days in prior. The design and architecture of the trading signal mining platform which employs machine learning to make the prediction of the stock price based on the data sources [10]. The comparisons are made between the extreme machine learning and the existing algorithms which includes support vector machine, back propagation neural network. The results of the analysis show that the Radial basis function extreme learning approach (RBF-ELM) and radial basis function support vector machine (RBF-SVM) achieve higher prediction accuracy and faster prediction speed than back propagation neural network. The speed of prediction in extreme machine learning approach is greater than the support vector machine approach. A new feature selection algorithm called as Normalized Relative Discriminant Criterion Algorithm (NRDC) and also new feature weighting algorithm called as term frequency inverse document frequency (TF-IDF) to improve the stock market prediction accuracy and used for financial news documents [15]. The fast learning model called as extreme machine learning is employed and according to make ELM suitable for multiple data sources, ELM is extended to the kernel-based ELM to improve the prediction speed. The comparison is made

between the two approaches such as hybrid proposed kernel based extreme learning approach (KL-ELM) including feature selecting and support vector machine with back propagation neural network. The experimental results show that K based ELM achieve with better performance in both prediction accuracy and prediction speed. But another type of kernel based extreme learning approach is used to predict the stock price performance i.e., Multiple kernel learning based extreme learning machine performs much better than compared to other methods [16].

Table 1. Algorithms with performance accuracy

S. No	Algorithm	Efficiency
1	SVM	78.00%
2	Random forest	95.08%
4	ELM-RBF	76.48%
5	ELM-SVM	77.70%
6	RBFNN	96.38%
7	K-ELM	96.38%
8	MKL-ELM	97.90%

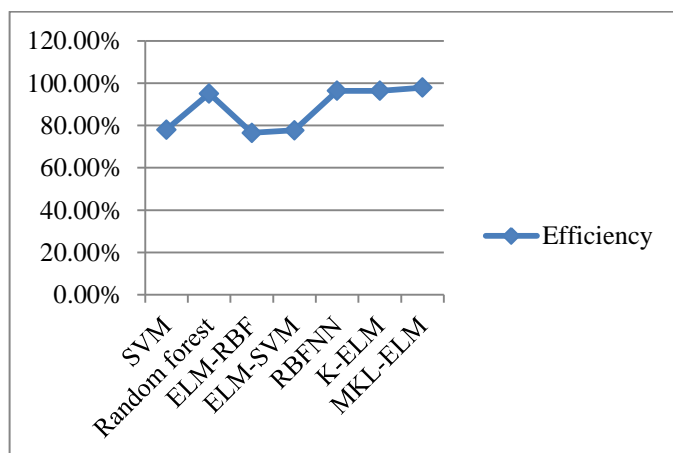


Figure 3. Graphs showing the efficiency in different algorithms

III. BIG DATA ON STOCK MARKET PREDICTION

The big data stock prediction is usually performed by many approaches such as social media analytics and machine learning techniques. Let us consider the two approaches and the process involved in stock prediction.

III. I Public agency analysis as effective forecast: The prediction involves three stages such as capture, analyse and present. In the stock market, the social media plays a significant role in predicting the price of the stock. By predicting the sentiment of the news article and social media, determines the rise or fall of stock price [12]. The data is collected and is analysed for four weeks and

compared. The social media prediction model comprises five modules such as data gathering, data preparation, sentiment analysis, aggregation and visualisation (Figure 4).

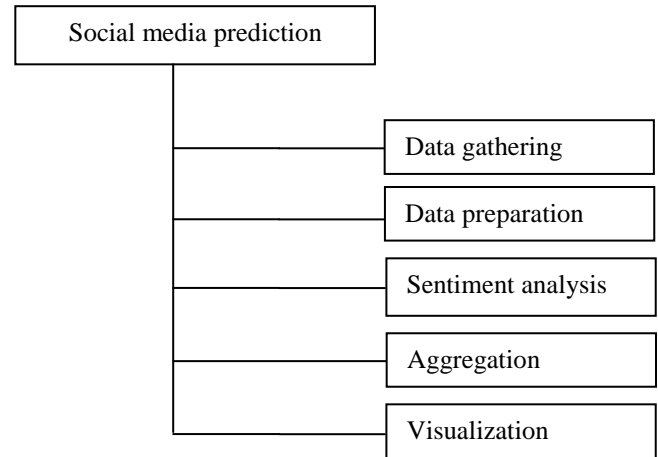


Figure 4. Social media prediction model

Data gathering: The articles of news displayed by different companies are collected. Tweets of the companies using search API. In real time analytics using the big data when the data is generated it can be later used by HDFS. This can be applied for real time as well.

Data preparation: The news and tweets collected are prepared for the analysis through May steps such as removing URL, lemmatization process etc.

Sentiment analysis: The data whichever is collected from the above step is further dumped into the HDFS for the analysis.

Aggregation: The analysis performed in the above step is aggregated to provide the sentiment to the company.

Visualization: The results obtained are plotted using the graph [1].

III. II Stock market prediction using the machine learning approach:

Prediction of the stock market is considered as a complicated system. Forecasting the stock prediction is characterised by the noise, data intensity, uncertainty etc. So, for this purpose, a well-suited machine learning approach needs to be employed. The data can be collected according to the requirement. For example, if the data is received from a finance company and later the received information is considered for training dataset. The machine learning techniques which are usually applied are regression method, neural networks; support vector machines [1].

IV. STOCK PREDICTION MODEL USING MACHINE LEARNING APPROACH

The process of prediction using optimal machine learning approach is reviewed. The various types of procedure and their process vary with different authors. Machine learning allows computers to find the hidden data without being programmed.

Support Vector Machine (SVM): The SVM is used for the statistical analysis and the application of SVM is used for the stock classification. The data used in the study is from banks and the annual reports. The four variables are considered by the author [8] such as net revenue, net income, earnings ratio of stock and earnings per share and included macroeconomic variables which influence the stock market. The training sample was based on random selection of few companies from different years. The test sample is created. The SVM non-linear classification is applied for the attributes. The model which is trained by the SVM shows good effect with the accuracy rate of 96.15 %. There might be minor fluctuations or error in the training data. To predict a stock, the classification model is necessary and is tested under a test sample. Thus, the companies' stock market is predicted using statistics.

Naive Bayes Classifier: The prediction of stock can be performed in real time and dummy as well. The naïve Bayes algorithm, the classification technique [11] generates the Bayesian networks for the defined dataset based on the theorem proposed by Bayes. The main advantages of the naïve Bayes algorithm are, it is straightforward to build and useful for large datasets. The essential steps in the algorithm are,

The given dataset is always converted to the frequency table and helps in calculating the probabilities

The posterior probability is found by the naïve Bayesian equation.

The predictable outcome is the class with the highest posterior probability.

The stock market prediction of future price and finding the share market value are performed by this algorithm and the high potential of Naïve Bayes Algorithm in predicting the return on investment in the share market. The yield on the investment is always unpredictable, so this approach is used for prediction which saves the user's time.

Artificial Neural Networks (ANN): As a prediction of stock markets is considered as one of the crucial issues in finance, now a day's artificial neural networks are applied to the financial problems such as the bankruptcy prediction, stock exchange index prediction. The forecasting of the stock exchange market index value in the financial sector is defined [3]. To illustrate the objective, ANN is used in the Istanbul stock exchange index values. The training is performed by the ANN where 6 ANN models were applied

to the system using a software package. The relative percentage error is used to measure the accuracy of prediction. The mean relative errors calculated for all models verified that ANN outperforms the other models. The prediction was accurate in ANN than with the other models. Thus, the different machine learning techniques are applied to the stock market prediction, and the results are evaluated accordingly.

V. CONCLUSION

This paper presents a review of stock market prediction in big data and machine learning techniques. The stock market prediction using different machine learning approaches is performed along with the advantages and limitations. Some of the methods reviewed are support vector machine, naïve Bayes classifier and artificial neural networks. The data collected in stock market exchange is analysed and classified using the training sets. The different variables in the stock market are considered and predict the accuracy of the stock. The techniques have shown improvisation when compared to the existing methods. The support vector machines are the most used technique for stock prediction. As a future work, an improvement over the current algorithms for enhancing the stock prediction without any errors using new machine learning and data mining approaches are performed.

REFERENCES

- [1] G. V. Attigeri, M. M. Manohara Pai, M. Radhika Pai, N. Aparna, "Stock market prediction: A big data approach". In: TENCON 2015-2015 IEEE Region 10 Conference Macao, pp. 1-4 November 2015.
- [2] M. Birasnav, R. Mittal, S. Loughlin, "Linking leadership behaviours and information exchange to improve supply chain performance: A conceptual model", Global Journal of Flexible Systems Management, Vol. 16, issue, 2, pp. 205-217, 2015.
- [3] Birgul Egeli, O. Meltem, B. Bertan, "Stock market prediction using artificial neural networks", Decision Support Systems, Vol. 22, pp.171-185, 2003.
- [4] J. Bollen, H. Mao, X. Zeng, "Twitter mood predicts the stock market", Journal of Computational Science, Vol. 2, issue, 1, pp.1-8, 2011.
- [5] S. Chauhan, N. Agarwal, A.K. Kar, "Addressing big data challenges in smart cities: A systematic literature review". Info, Vol. 18, issue, 4, pp. 73-90, 2016.
- [6] P. Grover, A.K. Kar, "Big data analytics: a review on theoretical contributions and tools used in literature" Global Journal of Flexible Systems Management, Vol. 18 issue, 3, pp. 203-229, 2017.
- [7] Hell Strom, T.A. Random, "Walk through the Stock Market, Licentiate Thesis" Department of Computer Science, Umea University, Sweden, 1998.
- [8] Z. Hu, J. Zhu, K. Tse, "Stocks market prediction using support vector machine". In: 6th International Conference on Information Management, Innovation Management and Industrial Engineering, 23-24 November, 2013.
- [9] Leung, C. K. S., MacKinnon, R. K., & Wang, Y. (2014). A machine learning approach for stock price prediction. In: IDEAS'14

Proceedings of the 18th International Database Engineering & Applications Symposium, ACM, July 7-9 2014, Porto, Portugal.

- [10] Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F. ... & Deng, X. (2014). Empirical analysis: stock market prediction via extreme learning machine. *Neural Computing and Applications*, **27(1)**, 67-78.
- [11] Mahajan, S. D., Deshmukh, K. V., Thite, P. R., Samel, B. Y., & Chate, P. J. (2016). Stock Market Prediction and Analysis Using Naïve Bayes. *International Journal on Recent and Innovation Trends in Computing and Communication*, **4(11)**, 121-124.
- [12] O'Connor, A. J. (2013). The power of popularity: An empirical study of the relationship between social media fan counts and brand company stock prices. *Social Science Computer Review*, **31(2)**, 229-235.
- [13] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2014). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, **42(1)**, 259-268.
- [14] Sharma, A., Bhuriya, D., & Singh, U. (2017). Survey of stock market prediction using machine learning approach. In: *International Conference on Electronics, Communication and Aerospace Technology*, 20-22 April 2017, Indore, Madhya Pradesh.
- [15] Wang, F., Zhang, Y., Xiao, H., Kaung, L., & Lai, Y. (2015). Enhancing Stock Price Prediction with a Hybrid Approach Based Extreme Learning Machine. In: *International Conference on Data Mining Workshops, Atlantic City, New jersey*, 14-17 November 2015.
- [16] Feng wang, Zhujong Zhao, Xiaodong Li, Fei yu, Hao Zhang (2014). Stock Volatility prediction using multi-kernal learning machine. In: *International Joint Conference on Neural Networks (IJCNN)*, July 6-11, Beijing, China, IEEE, 2014.

Authors Profile

Mrs T. P. Sameera Pallavi completed mtech in computer science from SR Engineering College 2013 .

She is currently pursuing PhD from Department of Computer Science, Kakatiya University, Warangal. She has published 3 Research Papers. Her Research work focuses on Big Data Analytics.



Mrs B. Manjula has presently working as Assistant Professor in Department of Computer Science, Kakatiya University, Warangal since February, 2008. She has published more than 20 Research Papers in Reputed International Journals including New York, USA, Hong Kong, U.S. and Conferences including Springer, Elsevier. Her main Research work focuses on Data mining, Big Data, Network Security. She has 11 years of Teaching Experience.

