SURVEY REPORT ON VARIOUS STOCK PREDICTION MODEL BASED ON MACHINE LEARNING ALGORITHMS

[1]Abhay Mohan Gupta, [2]Shikha Agrawal, [3]Rajeev Pandey
[123]Department of Computer Science and Engineering,
[1] UIT- RGPV, Bhopal-402033, India

Abstract : With the evolution of Machine learning in the field of computer science and engineering, its application in different domain also grew like in Medical, Marketing, Telecommunication, Finance, etc. The reason for the growth of popularity of machine learning in these domains is the high accuracy prediction. This is why stock return prediction became very popular in this field over past few years. Apart from this, various studies also claimed that the Investor's sentiment also important for stock return prediction and using their sentiments with the stock market data result in high accuracy and precision. The only problem with that is to choose perfect model for the prediction. As different model return different accuracy depending upon the data. So in order to determine the best predictive model, in this paper various models for stock prediction and User-generated textual sentiments analysis are studied so as to extract the detailed version of best models for stock market prediction.

Keywords **- Stock Prediction, Machine Learning, Support Vector Machine, Piecewise linear representation, Latent Dirichlet Allocation, Bayesian Model**

I. INTRODUCTION

Stock market is one of the best places to get good returns. It is one of the core and critical component of capital markets, which plays important part in supporting and promoting the economic development. Predicting the stock market helps to get good returns, but it's not easy task. Prediction of stock market is one of the foremost challenges that are faced by investors for better gain. In a book[14] by Burton Malkiel which is published in 1973 "A Random Walk Down Wall Street" claimed that the stock price can be predicted by looking at price history. After that various paper and journals are being published on stock market prediction and forecasting. Researchers used various models for stock prediction like Neural network, Time series analysis method, Support vector Machine, etc. Out of which Support Vector machine is considered to be the best model for stock market.

II. MACHINE LEARNING

Originally coined by Arthur Samuel in 1959, Machine Learning(ML) is one of the most important subfield of artificial intelligence in the field of computer science. It provides computer an ability to "learn by itself". That is, it gives computer, capability to learn from previous data without being programmed explicitly. Evolved from pattern recognition and learning theory, it has provided us capability like image recognition in self-driving cars, practical speech recognition in virtual assistants or in humanoid robot, recommendation system for movies and ecommerce platform, effective web search, and a vastly improved understanding of the human genome. It also provides prediction ability like for weather forecasting, cancer prediction in medical diagnosis, stock return in financial market, etc.

III. LITERATURE SURVEY

In this paper, researchers [Pei-Chann et.al 2008] uses a new approach in which Piecewise Linear Representation (PLR) is used to disintegrate historical data into different parts. The result of this data is the turning points which can be trough (low) or can be peak (high) of the previous data. This result is used as an input data for the Neural Network which is used for training the model. After this, testing data can trigger the model whenever any buying or selling point is observed by Neural network. The researchers further developed an intelligent PLR (IPLR) model to iteratively enhance the threshold value of the model. Thus, it further increases the efficiency of the model.[12]

In the paper, According to the researcher [Ming-Chih Lin et al 2011], there are two features of financial reports which are important for better stock prediction. First ones are the quantitative and other are qualitative features. Considering only one(quantitative and qualitative) feature will not provide complete information. Hence in this paper, author extracted both features of the these reports and assign a weights to justify both the features together. They then use HRK(Hierarchical agglomerative and Recursive K-means clustering)in which they apply HAC to do the initial clustering and then repeatedly used K-means to do further clustering so as to get purity of cluster. The reason for using this method is that it provide noise free data and avoids the over fitting problem.[11]

This paper proposed a different method to forecast stock returns in which researchers [JiangJiao Duan et. al 2013] mine sentiments and opinions from messages posted on Web forums. The Parameters for the model are estimated in such way that they

are best suited with the training data by using regular Baum-Welch algorithm and Genetic Algorithm(GA).A Bayesian model is used to established relation between the stock market returns and the combination of sentiments and opinions.[4]

In this paper researchers[Xiaotian Jin et.al 2014] proposes a new and better stock prediction method which is based on the social network and regression model(SVM regression Model, Least Square regression, Ridge Regression). In this paper dataset from NASDAQ stock market and from Twitter is used for verification. Experimental result shows that the proposed method accurately predict stock price.[3]

In this paper, Support vector machine along with Piecewise linear representation (PLR) is used for predicting the stock price. According to the researchers, SVM has great capability of generalization which prevent it from overfitting. In the mean time, PLR produces a junction from adjacent segment in which the junction is the inflection point of stock price.[5]

In this paper, researchers compared two different stock prediction methodology based on new articles. First they build up a framework which take the textual documents(news) as an input and generates the predicted price movements as output. Secondly, they construct the summaries of the news articles and use them as an input for framework. Then, they compared the performance of the prediction based on those two different information sets. The reason that they summarize the news is that this helps them to extracts the major information from the article and filters the noises that might distract the focus of readers.[10]

In this paper, stock return prediction method is defined by using sentiment analysis using financial microblogs. Researchers [Bo Zhao et. al 2016] analyze the microblogs contexts to find the financial sentiments of microblogs and then combine these sentiments and the training data to predict the stock market movements. In this model, Support vector machine is used along with the LDA (Latent Dirichlet Allocation) for microblogs filtering.[2]

In this paper, On day closing of Karachi Stock Exchange (KSE) is used for predicting the stock market return by using different learning models. In the paper, prediction model uses various attributes as an input and predicts stock market as Negative and Positive. The attributes that are used in the model includes rates of Oil, Gold, Silver, Foreign Exchange (FEX), Interest, NEWS and social media feed. Compared with various machine learning algorithms, the maximum accuracy was resulted by SVM. The algorithm produced 100% accuracy on training set but 60% on the test set with the average accuracy of 80%. [6]

In this paper, instead of using the news, social media or webforum's data, researchers [Zhihua Jiang et. al 2016] uses search indexes(Baidu Search Index, most popular search engine of China). Baidu Index is used for predicting the fluctuations in stock market. In the paper, a relation between the searched keyword and volatility of stock is found and a logistic regression model is built upon this for prediction.[7]

This paper studied the analysis of the announcement during the suspension period. It establish a prediction model based on content of announcement through NLP and try to predict the trend of next trade day. The researchers [Zhihua Jiang et. al 2016] focus on the announcement of the stock price during the suspension period (time when stock market stop trading). After that they select the appropriate classification algorithm for classifying the announcement as up and down. After properly classifying the announcements, they are used with stock data for prediction. Overall in this paper feasibility of using text classification to predict the trend of stock are analysed.[8]

In this paper, a new hybrid approach is used for the prediction of stock market in which sentiment are collected from different sources such as from news articles or from some social media platform like Twitter along with prediction using clustering algorithm. This is because it helps researchers to construct a model for predicting stock market movement. This model is based upon the opinion mining and clustering. In this paper, National Stock Exchange(NSE) market data is used. According to the researchers [Vivek Rajput et.al 2016], this methodology will result two output set. One is from sentiment analysis while another one from clustering based prediction with respect to some specialized parameters of stock exchange.[9]

In this paper researcher integrates sentiment analysis using the method based upon support vector machine. Since the stock data is nonlinear and dynamic, hence Support vector machine(SVM) is used in the prediction of the stock price as the SVM solution is unique and globally optimal. Along with the model researchers [Rui et. al 2018] employs sentimental analysis to create list of sentiments and then mix it with the stock's data. Furthermore, researchers [Rui et. al 2018] incorporate the day-of-a-week effect into cogitation which state that the average stock market return on Mondays is lower respective of the other days of the week and this effect helps to construct a more reliable and pragmatic sentiment indexes.[1]

IV. SURVEY REPORT

Table 4.1 Submerges the work accomplished by various researchers in this field-

S NO.	AUTHOR	PAPER TITLE	MODEL	ACCURACY
1.	Rui Ren, Desheng Dash Wu and Tianxiang Liu	Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine**(2018-IEEE)**	1. Support Vector Machine	89.93% with a rise of 18.6%

2.	Bo Zhao, Yongji He, Chunfeng Yuan, and Yihua Huang	Stock Market Prediction Exploiting Microblog Sentiment Analysis**(2016-IEEE)**	1. Support Vector Machine 2. Latent Dirichlet Allocation	performance increases from 64.81% to 69.09%.
3.	Xiaotian Jin, Defeng Guo, Hongjian Liu	Enhanced Stock Prediction using Social Network and Statistical Model**(2014-IEEE)**	1. Support Vector Machine (SVM) regression model 2. Least square regression 3. Ridge regression	82.1%(Bullish) 83.2%(Bearish)
4.	JiangJiao Duan , Jianping Zeng	Mining Opinion and Sentiment for Stock Return Prediction Based on Web-Forum Messages**(2013-IEEE)**	1. Bayesian model	up to between 90.1% and 76.8%
5.	Xi Chen, Zhi-jie He	Prediction of Stock trading signal based on Support Vector Machine**(2015-IEEE)**	1. Support Vector Machine (SVM) 2. Piecewise linear representation(PLR)	72.40%
6.	Mehak Usmani, Syed Hasan Adil, Kamran Raza, Syed SaadAzhar Ali	Stock market prediction Using Machine Learning technique**(2016-IEEE)**	1. Neural Networks 2. Support Vector Machine	80%
7.	Zhihua Jiang, Yanna Chen	BDI Based Stock Prediction**(2016-ICOACS)**	1. Regression	High
8.	Zhihua Jiang, Xixi Pan, Pei Chen	Announcement based stock prediction**(2016-IEEE)**	1. Support Vector Machine (SVM) 2. Naive Bayes	Upto 80%
9.	Vivek Rajput, SarikaBobde	Stock Market Prediction Using Hybrid Approach**(2016-ICCCA)**	1. Sentimental analysis 2. Clustering algorithm(Den clue)	High
10.	Xiaodong Li, Haoran Xie, Yangqiu Song, Shanfeng Zhu, Qing Li, Fu Lee Wang	Does Summarization Help Stock Prediction? News Impact Analysis via Summarization**(2015-IEEE)**	1. Support Vector Machine (SVM) 2. Self-Present Sentence Relevance	54.7%

			(SPSR)	
11.	Ming-Chih Lin, Anthony J. T. Lee, Rung-Tai Kao, Kuo-Tay Chen	Stock Price Movement Prediction Using Representative Prototypes of Financial Reports**(2011-ACM Transactions)**	1. HRK (Hierarchical agglomerative Clusterimg (HAC)and Recursive K-means clustering)	57.7% -76.9% (Depending upon different sectors)
12.	Pei-Chann Chang, Chin-Yuan Fan, and Chen-Hao Liu	Integrating a Piecewise Linear Representation Method and a Neural Network Model for Stock Trading Points Prediction**(2009-IEEE)**	1. piecewise linear representation (PLR) 2. backpropagation neural network (BPN) models	profit improved by 16.5%

IV. CONCLUSION

Stock market research is an area where every one wishes they could gauge the sentiments before investing in it. Though predicting stock movements is not easy. Technologies allowed us predict the movements to a large extent, if not full. The standard classical theory which is based on Efficient Market Hypothesis (EMH) had been in the use for many years. But behavioral finance overtook the classical theory by considering the fact that human sentiment plays an important role in predicting the stock prices. Given that various media on the Internet publish news articles and most of the researchers found that automation of that analysis of huge amount of data could produce informed decisions in a short span of time. Properly analyzing it and employing them with the stock data results a good prediction model. In this research various machine learning algorithm are compared so as to get best of them which can be used for the model.

V. REFERENCES

[1] Rui Ren, Desheng Dash Wu, Tianxiang Liu. "Forecasting Stock Market Movement Direction Using Sentiment Analysis and Support Vector Machine." *IEEE System General*, vol. isuue 99, pp. 1-18, 2018

[2] Bo Zhao, Yongji He, Chunfeng Yuan, and Yihua Huang. "Stock Market Prediction Exploiting Microblog Sentiment Analysis." *International Joint Conference on Neural Networks (IJCNN)* 7 pages(4482-4488), 2016

[3] Xiaotian Jin, Defeng Guo, Hongjian Liu. "Enhanced Stock Prediction using Social Network and Statistical Model." *IEEE Workshop on Advanced Research and Technology in Industry Applications (WARTIA),* pp 1199 – 1203, 2014.

[4] JiangJiao Duan , Jianping Zeng. "Mining Opinion and Sentiment for Stock Return Prediction Based on Web-Forum Messages."*10th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD*) pp. 984-988, 2013.

[5] Xi Chen, Zhi-jie He. "Prediction of Stock trading signal based on Support Vector Machine." *8th International Conference on Intelligent Computation Technology and Automation* pp. 651-654, 2015.

[6] Mehak Usmani, Syed Hasan Adil, Kamran Raza, Syed SaadAzhar Ali. "Stock market prediction Using Machine Learning technique." *3rd International Conference On Computer And Information Sciences (ICCOINS)* pp. 322-327, 2016

[7] Zhihua Jiang, Yanna Chen. "BDI Based Stock Prediction." *Proceedings of ICOACS* pp. 119-122, 2016.

[8] Zhihua Jiang, Xixi Pan, Pei Chen. "Announcement based stock prediction." *International Symposium on Computer, Consumer and Control.* pp. 428-431, 2016.

[9] Vivek Rajput, SarikaBobde. "Stock Market Prediction Using Hybrid Approach." *International Conference on Computing, Communication and Automation (ICCCA)* pages 82-86, 2016.

[10] Xiaodong Li, Haoran Xie, Yangqiu Song, Shanfeng Zhu, Qing Li, Fu Lee Wang. "Does Summarization Help Stock Prediction? News Impact Analysis via Summarization." *IEEE,* 2015.

[11] Ming-Chih Lin, Anthony J. T. Lee, Rung-Tai Kao, Kuo-Tay Chen. "Stock Price Movement Prediction Using Representative Prototypes of Financial Reports." *ACM Trans. Manag. Inform. Syst.* vol. 2, no. 3, Article 19,pp. 1-18,2011.

[12] Pei-Chann Chang, Chin-Yuan Fan, and Chen-Hao Liu. "Integrating a Piecewise Linear Representation Method and a Neural Network Model for Stock Trading Points Prediction." *IEEE Transactions on System, Man and Cybernetics—PART C: Applications and Reviews*, vol. 39, no. 1, pp.80-92, 2009

[13] Burton Gordon Malkiel "A Random Walk Down Wall Street" https://en.wikipedia.org/wiki/A_Random_Walk_Down_Wall_Street