

A Fuzzy based Soft Computing Technique to Predict the Movement of the Price of a Stock

Ashit Kumar Dutta

Department of Computer Science and Information System
Shaqra University
Kingdom of Saudi Arabia

Abstract—Soft computing is a part of an artificial intelligence, and fuzzy logic is the study of fuzziness on data. The combination of these two techniques can provide an intelligent system with more ability and flexibility. The nature of data in the stock/capital market is more complex and challenging to predict the movement of the price of the stock. The study has combined both fuzzy c-means and neural network technique for the prediction of the price of the stock. The research finds an optimum solution to predict the future price of a stock. The comparison of time and space complexity proved that the proposed method is better than the existing methods.

Keywords—Soft computing; fuzzy logic; stock recommendation; fuzzy based soft computing, soft computing systems

I. INTRODUCTION

A wide variety of engineering control processes are using the concept of fuzzy logic (FL). There is a need of expert knowledge to develop an intelligent system. FL concepts were evolved in late 60's from crisp fuzzy sets and reached a high level of perceptions in recent years [6], [7]. A researcher can apply every kind of possibilities ranges from approximate, linguistics to perception, computing with words, natural language processing, computational theory of perceptions, and perception-based probability theory. Logical, relational, set-theoretic and epistemic are the four facets of FL [8], [9]. Sigmoid functions can be used to define fuzzy sets. Professor Zadeh has found many operators in FL and completely replaced the ordinary Boolean logic. Fuzzification and defuzzification are the functions similar to encode and decode functions. Many expert systems were implemented with the concept of FL.

Soft computing (SC) is related to FL used to develop intelligent systems. The purity of thinking, machine intelligence, and fuzziness processing capabilities are the attributes of SC. The SC aims to exploit tolerance for imprecision, uncertainty, and fuzziness to achieve low-cost computation. It can learn from experience, universalize domain, simulation of biological processes, and faster computations [10], [11]. It has the critical role in science and engineering applications [20]-[22]. It is widely used in machine learning, belief networks, chaos theory, wisdom based expert system and probability reasoning. It has feature of extraction of knowledge/information from inaccurate and uncertain data [12], [13] and vast influence of artificial neural network

(ANN), FL and genetic algorithm [14], [15]. The combination of FL and SC is used to develop an accurate expert system [23]-[25].

Stock/capital market is familiar to investors due to its high returns. The non – linear, chaotic, noisy behavior nature of data in the market made difficult to predict the price of a stock [1]. Technical, fundamental, and macroeconomics are the popular methods to analyze the market and predict the movement of a stock. A technical analysis is used to find the price of the stock in a period. Mathematical methods are used in the technical analysis to indicate the buy/sell signals [2], [3]. Fundamental analysis is the technique to analyze the company performance by calculating profit and liabilities. A macroeconomic analysis is a technique to analyze the recent decision taken by the company and its impact on the market [4], [5]. The aim of the research is to build an intelligent system illustrated in Fig. 1 to predict the price of a stock in Indian capital market.

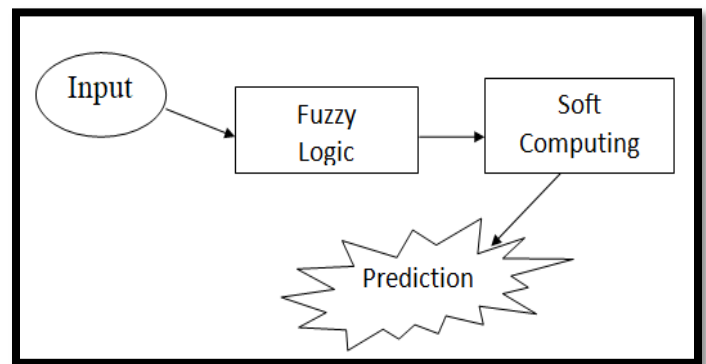


Fig. 1. Stock recommendation systems.

Current and previous day price of the stock, Relative strength index (RSI), the volume of the stock and 30 days average moving price of the stock are the inputs required for the system and FL – SC combined proposed method will predict the future price of the stock [16], [17]. The Nifty – 50 and Pharmacy stocks used in the research. Support vector machine (SVM), Neural networks (NN) are the existing methods used in the research to show the effectiveness of the proposed method. Section 2 elaborates the review of the literature, Section 3 define the methodology of the research and experiment and results provided in Section 4. Finally, Section 5 concludes the research.

II. REVIEW OF LITERATURE

Ibrahim M. Hamed et al. [1], proposed an intelligent system to recommend the price of the stock. They have used multilayer perceptron-based ANN for the research. The model proposed in their research had several stages. Input selection, pre-processing of data and signal detection were some of the critical stages in the model. The algorithm was used in the model update the weights between the neural network neurons to minimise the error of the prediction results. Blind source separation techniques are used in the research for the signal detection and forward to the final stage for the user. ANN is a slow learner, and moreover, the multilayer perceptron leads to taking more time for the prediction of results.

Monica Tirea and Viorel Negra [2] have developed an intelligent stock market analysis system for the prediction of the stock price. The study has taken many attributes and constructed a multi-agent prediction system. Economic, industrial and behavioural are the types of approaches in fundamental analysis. The macro – economical analysis have taken interest rates, earning of the company, profitability and growth rate. The overall structure of the system had three essential parts. They are multi-agent stock trading system, news information system and portfolio optimisation system. The portfolio system has data interpretation level, behavioural level, risk detection level and portfolio optimisation level. The multi-agent stock trading system has used the numerical data with multi-agents. Text mining agent is used to crawl the recent news and forwards to the portfolio optimisation system and modify signals according to the news. The system took more time and space to predict the price and consume more memory comparing to the other new methods.

Sunil Kumar Khatri et al. [3] have proposed a sentiment analysis to predict Bombay stock exchange using ANN. Many forums are there for the investors to participate and share their views related to the market. Data of blogs/forums and E – media were the inputs for the system and pre-processed into four different moods like happy, hope, sad and disappointing. The ANN model was trained with sample data from a detailed set and produces the final prediction from the test data. The output of the ANN model prediction will be positive, negative/neutral. A maximum of 15 hidden layers were used for the prediction of the price of the stock based on the sentiments. The research entirely based on the sentiment of the users in the forum and there was no relation with the real price of the stock. The accuracy is the primary concern of the research as it depends on the sentiment shared in the forum.

Qiaghua wen et al. [4], have proposed an intelligent stock recommendation system using SVM techniques and box theory of stock. The SVM uses statistical learning theory and structural risk minimisation principle. The principle behind the SVM is to map the inputs into a high – dimension space by a non – linear mapping with kernel function to do a linear regression. In this research, the radial basis function is acted as the kernel function of SVM estimators. The selection of the user-prescribed is a vital role in overall performance. The sliding window method is employed in the research to train SVM estimators. The accuracy and computation time of SVM was more compared to other existing methods.

Chin – Cheng Tseng [5] have proposed a portfolio management using hybrid recommendation system. The research has deployed a group of agents with different functionality and groups them under case – based situations. Agents are used to gather, filter, and process various cases in the systems. It does not have any mechanism to deal with uncertainty and complex data. It will help an individual to select a stock based on historical data and not on real-time data. Influence diagram module, decision tree system module, web information agent module, and interface module were the components of the hybrid recommendation architecture. Decision tree is one of the old methods and if the number of data increased then the computation time will be more and create problems in the process of making the decision.

Lei wang et al. [6], have developed a stock recommendation system based on a method called HLP. The method was used to extract data with different frequency and amplitude. The data were extracted from the server will describe the nature of a price of stock and its movement. ANN modules are used to forecast the movement of the stock. The following Table I shows the type of technique and year used in the research.

TABLE. I. TECHNIQUES USED IN THE RESEARCH

S.No.	Technique	Year	Authors
1	Decision Tree	2004	Chiu –Che et.al. ,
2	SVM	2009	QinghuaWen et. al. ,
3	ANN	2011	Ibrahim M. Hamed et.al.,
4	Decision Tree	2013	Zhen Hu et.al.,
5	ANN	2014	Monica Tirea et. al. ,
6	ANN	2014	Sunil Kumar Khatri et. al. ,
7	ANN	2014	Neelima Budhani et.al.,
8	ANN	2016	Amin Hedayati Moghaddam et. al.,

III. PROPOSED METHOD

The proposed method uses fuzzy C-means and ANN model to build an intelligent system to predict the movement of the price of the stock. Fuzzy c-means is similar to K – means algorithm. In this study, Fuzzy C – means and NN are combined into the Fuzzy logic neural network (FL-N) to predict the price of the stock. The fuzzy algorithm used to cluster stock market data, clusters the stock data extracted from National Stock Exchange (NSE) portal. The clusters are given as input to NN. The NN has two input and one output layers. The historical data will be pre-processed for the model for the prediction of its future [18], [19]. Let $\sum D$ be the price data and $f(\sum D)$ is the fuzzification of data. ANN model will take the fuzzified data and compute the future price of the data. Current and previous day price of the stock, Relative strength index (RSI), the volume of the stock and 30 days average moving price of the stock are the inputs. A total of 15 hidden layers were used in NN to produce results. The NN used in the research is the feed forward, back propagation network. It can learn from the failures and fault tolerance. The following algorithm will describe the process of the computation in detail.

Step 1: Start
Step 2: Historic data downloaded from the Nifty 50 and Pharma

Step 3: Initial pre – process of data to normalize it for the fuzzification
 Step 4: Fuzzy C – means clusters the data.
 Step 5: The clusters transformed into NN inputs.
 Step 6: Multi – Layer perceptron NN model receives the transformed input for the training phase.
 Step 7: Learnt NN produces results

Step 8: Goto step 3.
 Step 9: End

Fig. 2 shows the screen of NSEIndia (www.nseindia.com) website having details about NIFTY 50 stocks. Fig. 3 shows the details of NIFTY Pharma stocks. Fig. 4 shows the pre – processing of data extracted from NSEINDIA. The data were normalized by fuzzy plug-ins for Microsoft Excel 2007.

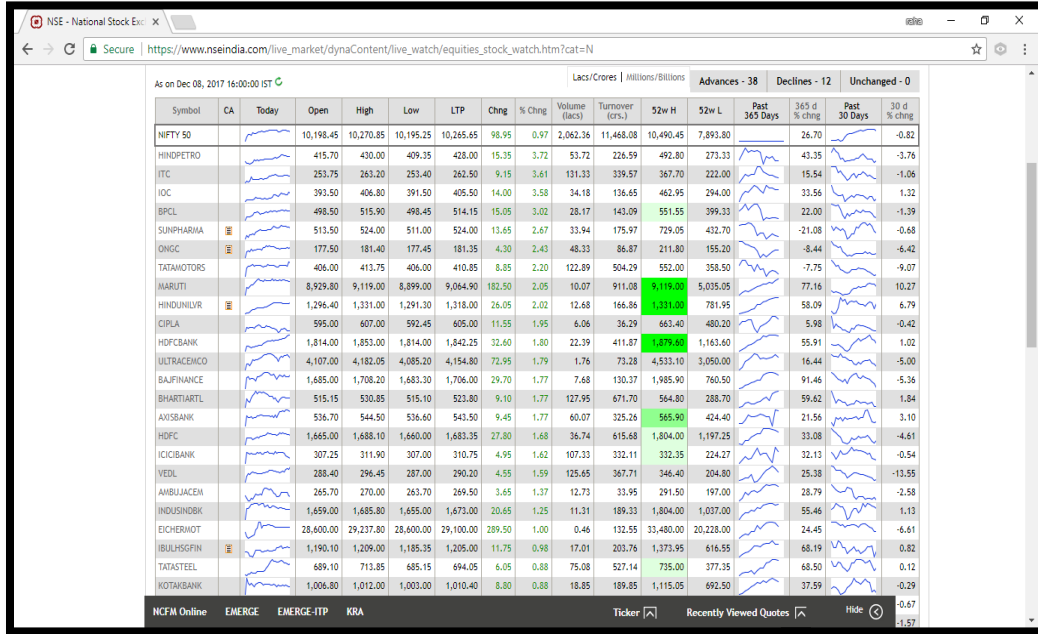


Fig. 2. NSE India – NIFTY 50 Stocks.

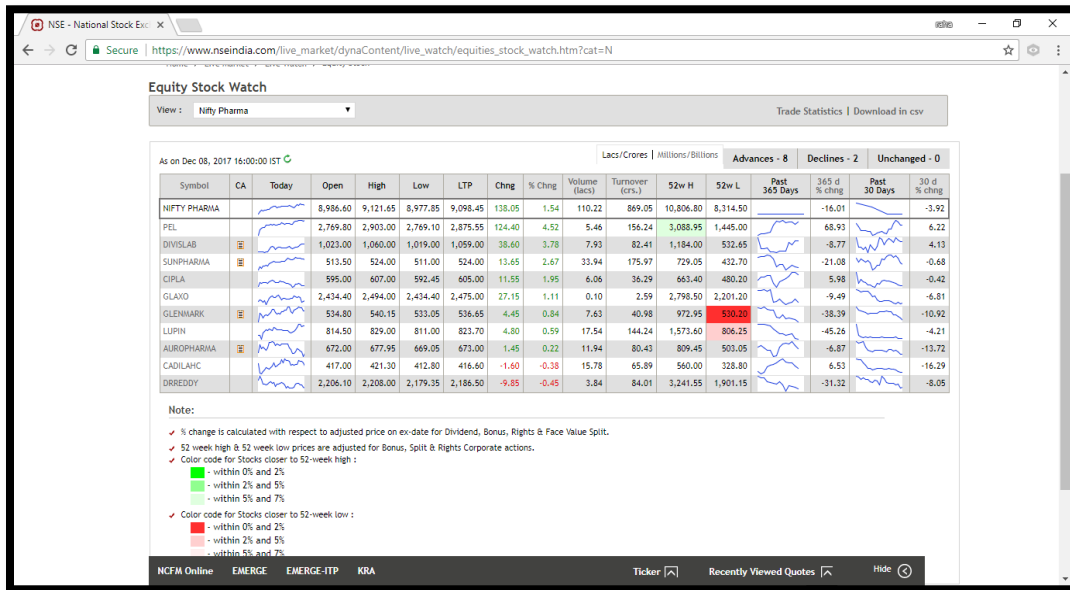


Fig. 3. NSE India – NIFTY PHARMA Stocks.

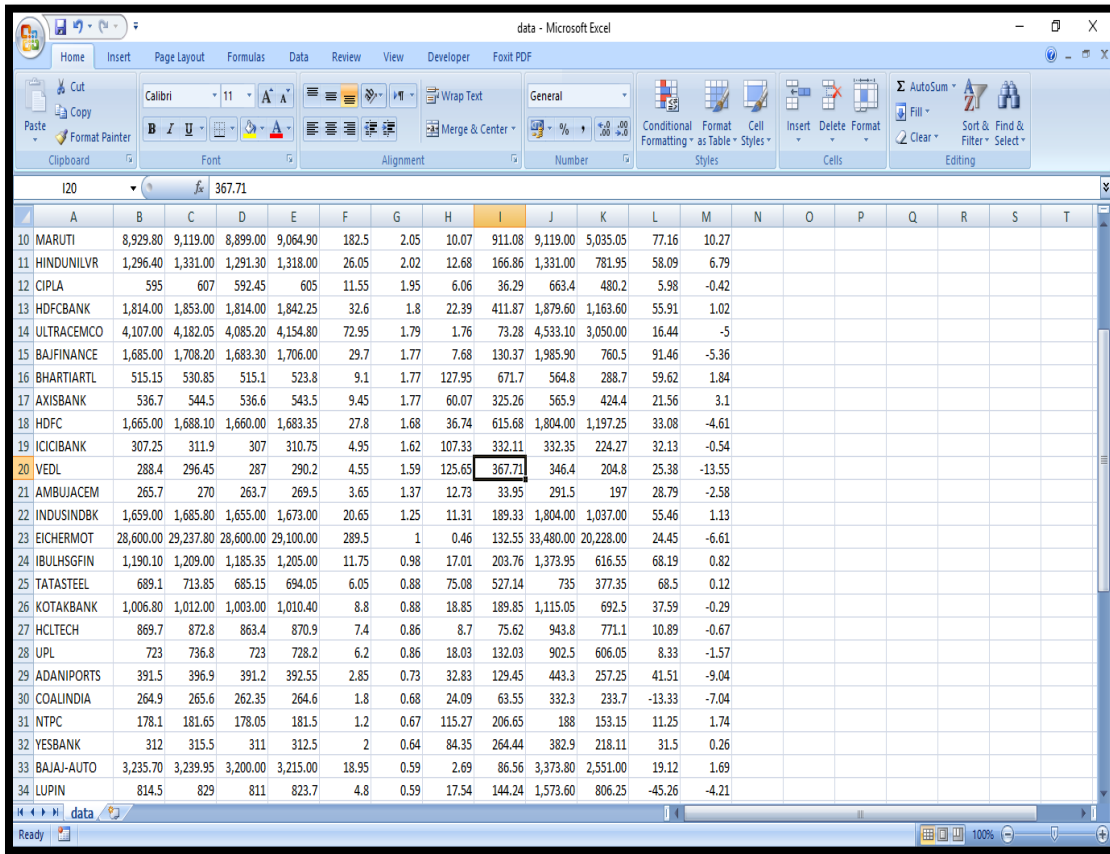


Fig. 4. Pre-processing of data for FL-N.

IV. RESULTS AND DISCUSSION

The research used Windows 10 operating system with I7 processor for the pre – process of data extracted from NSEINDIA. Table II shows the training time of the methods used for 60 days from the data. Usually training time will be more comparing to testing time; every method will take time to learn about the environment. The training data has to be unique and should not be used in testing phase. Comparing to others methods, the proposed method have taken less time in the training phase. Fig. 5 shows the relevant graph of the training phase. Days indicate the data collected from the range of 10 to 60 days. The process of data collection begun from 01-02-2017 to 01-12-2017.

TABLE. II. TRAINING TIME (IN SECONDS) OF METHODS

Days /Methods	Days					
	10	20	30	40	50	60
ANN	0.432	0.326	0.423	0.521	0.563	0.452
SVM	0.621	0.568	0.528	0.612	0.589	0.632
DT	0.745	0.695	0.485	0.569	0.541	0.458
FL - N	0.385	0.215	0.412	0.235	0.346	0.295

Table III shows the testing time taken by the methods and it is evidence that the proposed method have taken less time to produce the results. The FL-N has taken only 0.098 seconds to produce results from 50 days data. Fig. 6 displays the relevant graph of Table III.

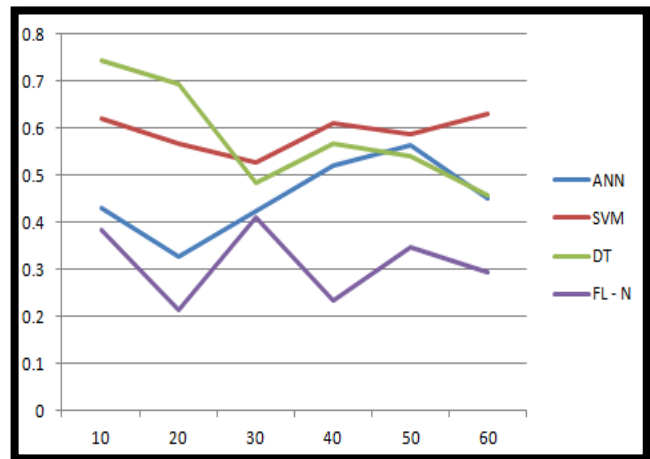


Fig. 5. Training Phase (in seconds).

TABLE. III. TESTING TIME (IN SECONDS)

Days /Methods	Days					
	10	20	30	40	50	60
ANN	0.234	0.256	0.185	0.215	0.265	0.278
SVM	0.285	0.312	0.165	0.178	0.195	0.174
DT	0.195	0.215	0.175	0.185	0.139	0.149
FL - N	0.124	0.138	0.135	0.152	0.098	0.112

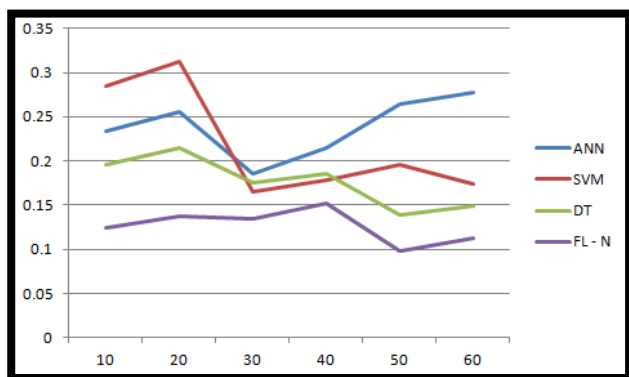


Fig. 6. Testing Phase (in seconds).

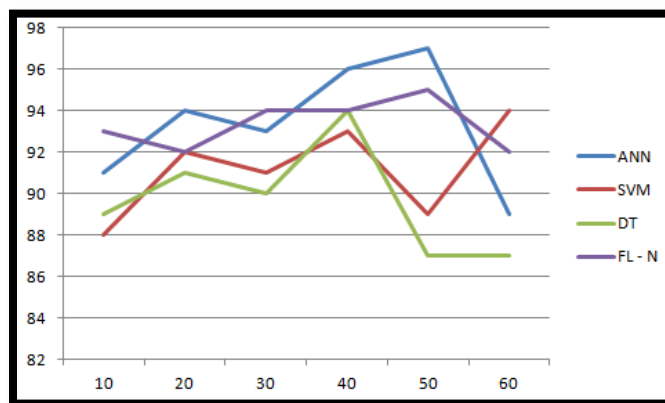


Fig. 8. Accuracy of methods employed in the research (Testing Time).

Tables IV and V show the accuracy for total of 60 days data during training time. It is necessary to measure the accuracy during the training time. It will be useful for the research to find any deviation between training and testing phase. Table IV shows the percentage of accuracy of the methods used in the research. Fig. 7 shows the relevant graph of Table IV. The proposed method has reached an average of 88% for all predictions. DT has also reached the better accuracy comparing to other two methods.

TABLE. IV. ACCURACY (ROUNDED OFF PERCENTAGE) (TRAINING TIME)

Days /Methods	Days					
	10	20	30	40	50	60
ANN	87	86	85	87	85	86
SVM	89	84	88	87	85	86
DT	85	87	86	88	89	87
FL - N	91	89	87	89	90	86

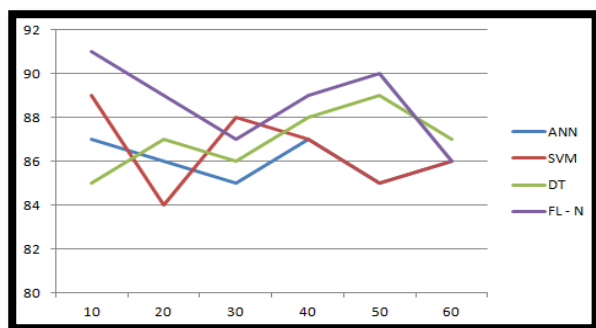


Fig. 7. Accuracy of methods employed in the research (Training time).

Table V shows the percentage of accuracy of the methods used in the research during testing phase. The proposed method has reached an average of 93% for all predictions. DT has also reached an average accuracy of 93% comparing to other two methods but took more time to produce it. Fig. 8 shows the relevant graph of Table V.

TABLE. V. ACCURACY (ROUNDED OFF PERCENTAGE) (TESTING TIME)

Days /Methods	Days					
	10	20	30	40	50	60
ANN	91	94	93	96	97	89
SVM	88	92	91	93	89	94
DT	89	91	90	94	87	87
FL - N	93	92	94	94	95	92

Precision, Recall and F1 measure are the metrics used to indicate the performance of an algorithm and Tables VI, VII and VIII shows the performance of the proposed and other methods. The proposed method has better score than other methods. FL-N has the ability to perform well in complex situation. Even the web data is hard and complex; the FL - N has learnt well about the environment and produced good results. Fig. 9 shows the relevant graph of F1 measure. FL-N and SVM have performed well comparing to DT and ANN.

TABLE. VI. PRECISION

Clusters /Methods	Days						Overall Average
	10	20	30	40	50	60	
ANN	78.5	82.3	84.6	79.8	81.3	83.4	81.65
SVM	82.3	86.4	86.3	83.4	82.3	84.3	84.17
DT	76.5	79.6	79.4	76.2	73.4	79.5	77.43
FL - N	89.3	84.3	87.3	86.3	84.9	85.7	86.3

TABLE. VII. RECALL

Clusters /Methods	Days						Overall Average
	10	20	30	40	50	60	
ANN	79.6	78.6	81.2	75.6	76.4	77.4	78.13
SVM	86.3	85.3	85.3	86.3	85.3	86.3	85.8
DT	79.5	79.4	76.3	78.3	79.6	84.3	79.57
FL - N	85.9	86.3	88.2	87.9	88.6	87.5	87.4

TABLE. VIII. PRECISION, RECALL AND F1 MEASURE

Clusters /Methods	Precision	Recall	F1 Measure
ANN	81.65	78.13	79.85
SVM	84.17	85.8	84.98
DT	77.43	79.57	78.49
FL - N	86.3	87.4	86.85

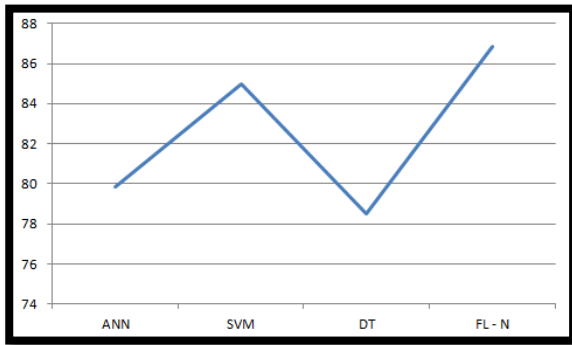


Fig. 9. F1- measure of methods.

The research has measured the overall memory to produce the results during testing time. Training time will take more memory as all methods have to learn the environment. The proposed method has taken 0.3 MB memory to produce the results. DT has taken a maximum of 2 MB to produce the result whereas SVM has occupied 0.7 MB and ANN has taken 0.5 MB. The performance of proposed method is far better than the existing method. ANN and SVM lacks in performance due to the nature of consuming more time and memory. DT has the scalability issue so could not produce better results comparing to the proposed method.

V. CONCLUSION

The objective of the research is to predict the movement of the price of a stock. The research has extracted data from NSEINDIA. NIFTY – 50 and PHARMA data were taken for the purpose of experiment. The experimental results proved that the proposed method has an average accuracy of 88% to predict the stock price. Testing and Training time results shows that the time taken by the proposed method is better than the other existing methods. The overall memory for the prediction is 0.3 MB for the proposed method. The future work of the research is to predict the movement of stocks in Bombay stock exchange.

REFERENCES

- [1] M. Hamed, A. S. Hussein and M. F. Tolba, "An intelligent model for stock market prediction," The 2011 International Conference on Computer Engineering & Systems, Cairo, 2011, pp. 105-110. doi: 10.1109/ICCES.2011.6141021.
- [2] M. Tirea and V. Negru, "Intelligent Stock Market Analysis System - A Fundamental and Macro-economical Analysis Approach," 2014 16th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, 2014, pp. 519-526. doi: 10.1109/SYNASC.2014.75
- [3] S. K. Khatri, H. Singhal and P. Johri, "Sentiment analysis to predict Bombay stock exchange using artificial neural network," Proceedings of 3rd International Conference on Reliability, Infocom Technologies and Optimization, Noida, 2014, pp. 1-5.
- [4] Q. Wen, Z. Yang, Y. Song and Peifa Jia, "Intelligent stock trading system based on SVM algorithm and oscillation box prediction," 2009 International Joint Conference on Neural Networks, Atlanta, GA, 2009, pp. 3341-3347.

- [5] C. C. Tseng, "Portfolio management using hybrid recommendation system," e-Technology, e-Commerce and e-Service, 2004. EEE '04. 2004 IEEE International Conference on, 2004, pp. 202-206.
- [6] L. Wang and Q. Wang, "Stock Market Prediction Using Artificial Neural Networks Based on HLP," 2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics, Zhejiang, 2011, pp. 116-119.
- [7] Amin Hedayati Mogaddam, Moein Hedayati Moghaddam, Morteza Esfandyari, "Stock market index prediction using artificial neural network", Journal of economics, finance and administrative science, vol.21, issue 41, December 2016, pages 89 - 93.
- [8] Zhen hu, Jie zhu, Ken tse, "Stock market prediction using support vector machine", 6th International conference on information managment, innovation management and industrial engineering, China, 23 - 24 November 2013.
- [9] Thira Chavarnakul David Enke "Intelligent technical analysis based equivoolume charting for stock trading using neural networks" Expert Systems with Applicationsvol. 34 pp. 1004-1017 2008.
- [10] David Enke Suraphan Thawornwong "The use of data mining and neural networks for forecasting stock market returns" Expert Systems with Applicationsvol. 29 pp. 927-940 2005.
- [11] Chen An-Sing T. Leungb Mark Daoukc Hazem "Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index" Computers & Operations Researchvol. 30 pp. 901-923 2003.
- [12] Kim Kyoung-jae Boo Lee Won "Stock market prediction using artificial neural networks with optimal feature transformation" Neural Comput & Applicvol. 13 pp. 255-260 2004.
- [13] F. Ham I. Kostanic,"Principles of neuron-computing for science and engineering", New York NY:McGraw-Hill 2001.
- [14] K. Hornik M. Stinchcombe H. Wite "Multilayer Feedforward Networks Are Universal Approximators" Neural Networks vol. 2 pp. 359-366 1989.
- [15] Fredric M. Ham "Ivica Kostanic" in Principles of Neurocomputing for Science and Engineering New York NY:McGraw-Hill Science 2000.
- [16] P. H. Algoet and T. M. Cover. "Asymptotic optimality and asymptotic equipartition properties of log-optimum investment". The Annals of Probability, pages 876-898, 1991.
- [17] T. M. Cover. "Universal portfolios". Mathematical Finance, vol. 1, pages 1-29, 1991.
- [18] T. M. Cover and E. Ordentlich. "Universal portfolio with side information". IEEE Transactions on Information Theory, pages 348-363, March 1996.
- [19] G. Gorry and G. Barnett, "Experience with a model of sequential diagnosis", Computers and Biomedical Research, 1968.
- [20] J. Grass and S. Zilberstein, "Value-Driven Information Gathering", AAAI 97 Building Resource-Bounded Reasoning Systems Workshop, 1997.
- [21] P. J. Gmytrasiewicz and E. H. Durfee, "Elements of a Utilitarian Theory of Knowledge and Action", IJCAI, 1993, pp. 396-402.
- [22] D. Heckerman, E. Horvitz and B. Middleton, "An Approximate Nonmyopic Computation for Value of Information", IEEE Transaction of Pattern Analysis and Machine Intelligence, 1993.
- [23] E. Horvitz and M. Barry, "Display of Information for Time-Critical Decision Making", Proceedings of the Eleventh conference on Uncertainty in Artificial Intelligence, 1995.
- [24] R. A. Howard, "Information value theory", IEEE Transactions on Systems Science and Cybernetics, 1966.
- [25] R. A. Howard, "From Influence to Relevance to Knowledge", Influence Diagrams, Belief Nets and Decision Analysis, 1990, pp. 3-23.