A Survey on Stock market price prediction using data mining techniques

Prateek Purey[1*] , Anil Patidar[2]

[1,2]Computer Science, Acropolis Institute of Research and Technology, RGPV Indore, India

corresponding author; prateekpurey1000@gmail.com tel. +91-9926788853

Abstract— data mining techniques are used in a number of applications such as classification, prediction and others. In this presented work the data mining techniques are investigated for implementing in prediction applications. Therefore this paper provides the study about the stock market price prediction techniques and the recently made contributions in domain of prediction using data mining techniques. The data mining techniques are having the ability to evaluate the historical stock market price trends and can approximate the upcoming market prices. In addition of that a model using available techniques is also presented work.

Keywords— Stock market price, prediction, data mining, survey, improved model.

I. INTRODUCTION

The data mining is an effective tool that can be used for analysis of historical data and recovery of valuable patterns [1]. These recovered patterns can be used for prediction and other intelligent task [2]. The data mining techniques are now in these days used in a number of applications where the large amount of data analysis is required and using the available data decision making, prediction and other kinds of essential analysis is required [3]. The proposed work is intended to investigate about the different data mining techniques which are able to predict precise value according to the trends of the data. Therefore different research efforts and techniques are reported in this paper. In addition of that for improving the technique of prediction a new data model is also reported in this work.

The main aim of the proposed work is to employ the data mining techniques over the stock market historical data analysis. After analysis or learning the model can predict the new prices of the stock market. Therefore the proposed work recommends using the supervised learning techniques. The supervised learning techniques are learn with predefined patterns of data and after training that can predict the close value of stock market. The use of supervised learning is suitable because that are efficient and accurate then the clustering or unsupervised learning techniques [4]. This section provides the overview of the proposed work involved in this paper. In further sections the detailed investigation and design of the proposed stock market price prediction model is provided.

II. LITERATURE SURVEY

This section involves the different research efforts and the techniques that are contributing in stock market price prediction.

Making a keen framework that can precisely anticipate stock cost vigorously has dependably been a subject of extraordinary enthusiasm for some speculators and money related investigators. Anticipating future patterns of money related markets is more noteworthy nowadays particularly after the ongoing worldwide monetary emergency. So brokers who access to a great motor for removing accommodating data toss crude information can meet the achievement. *Reza Hafezi et al [5]* propose another keen model in a multi-operator structure called bat-neural system multi-specialist framework (BNNMAS) to anticipate stock cost. The model performs in a four layer multi-specialist structure to foresee eight long periods of DAX stock cost in quarterly periods. The capacity of BNNMAS is assessed by applying both on essential and specialized DAX stock value information and contrasting the results and the aftereffects of different strategies, for example, hereditary calculation neural system (GANN) and some standard models like summed up relapse neural system (GRNN), and so forth. The model tried for foreseeing DAX stock value a timeframe that worldwide money related emergency was looked to financial matters. The outcomes demonstrate that BNNMAS essentially performs exact and dependable, so it can be considered as an appropriate apparatus for anticipating stock cost extraordinarily in a long haul periods.

Instructions to foresee stock value developments in view of quantitative market information displaying are an alluring subject. Before the market news and stock costs that are generally accepted as two vital market information sources, how to concentrate and adventure the shrouded data inside the crude information and make both precise and quick forecasts at the same time turns into a testing issue. *Xiaodong Li et al [6]* show the plan and design of our exchanging signal mining stage that utilizes extraordinary learning machine (ELM) to make stock value expectation in light of those two information sources simultaneously. Far reaching exploratory examinations amongst ELM and the state-of the-workmanship learning calculations, including bolster vector machine (SVM) and back-spread neural system (BPNN), have been embraced on the intra-day tick-by-tick information of the H-share advertise and contemporaneous news files. The outcomes have demonstrated that (1) both RBF ELM and RBF SVM accomplish higher expectation exactness and speedier forecast speed than BP-NN; (2) the RBF ELM accomplishes comparable precision with the RBF SVM and (3) the RBF ELM has quicker forecast speed than the RBF SVM. Reproductions of a primer exchanging technique with the signs are directed. Results demonstrate that technique with more exact signs will make more benefits with fewer hazards.

The objective of *Thien Hai Nguyen et al [7]* is to construct a model to foresee stock value development utilizing the assessment from online life. Not at all like past methodologies where the general states of mind or estimations are viewed as, are the notions of the particular themes of the organization fused into the stock expectation show. Points and related notions are consequently removed from the writings in a message board by utilizing our proposed technique and also existing theme models. Likewise, this paper demonstrates an assessment of the adequacy of the slant examination in the stock expectation undertaking by means of a huge scale analyze. Contrasting the precision normal more than 18 stocks in a single year exchange, our strategy accomplished 2.07% preferable execution over the model utilizing verifiable costs as it were. Besides, when looking at the strategies just for the stocks that are hard to foresee, our strategy accomplished 9.83% preferred exactness over verifiable value technique, and 3.03% superior to human slant technique.

To contemplate the impact of data on the conduct of securities exchanges, a typical technique in past examinations has been to link the highlights of different data sources into one compound element vector, a strategy that makes it harder to recognize the impacts of various data sources. *QING LI et al [8]* keep up that catching the characteristic relations among numerous data sources is essential for foreseeing stock patterns. The test lies in displaying the unpredictable space of different sources and kinds of data and concentrates the impacts of this data on securities exchange conduct. For this reason, we present a tensor-based data system to foresee stock developments. In particular, our structure models the unpredictable financial specialist data condition with tensors. A worldwide dimensionality-decrease calculation is utilized to catch the connections among different data sources in a tensor, and an arrangement of tensors is utilized to speak to data accumulated after some time. At long last, a tensor-based prescient model to estimate stock developments, which are basically a high-arrange tensor relapse learning issue, is exhibited. Trials performed on a whole year of information for China Securities Index stocks show that an exchanging framework in view of our system beats the exemplary Top-N exchanging methodology and two best in class media-mindful exchanging calculations.

Foreseeing stock trade rates is getting expanding consideration and is essential cost related issue as it adds to the advancement of successful techniques for stock trade exchanges. The determining of stock value development as a rule is thought to be a provocative and fundamental assignment for money related time arrangement's investigation. *Sanjiban Sekhar Roy et al [9]* a Least Absolute Shrinkage and Selection Operator (LASSO) technique in light of a straight relapse display is proposed as a novel strategy to anticipate budgetary market conduct. Tether technique can deliver meager arrangements and performs exceptionally well when the quantities of highlights are less when contrasted with the quantity of perceptions. Examinations were performed with Goldman Sachs Group Inc. stock to decide the effectiveness of the model. The outcomes demonstrate that the proposed display outflanks the edge straight relapse show.

III. PROPOSED WORK

The stock market is a classical research domain of engineering and research. A number of contributions are available in this domain. Most of the data model is used for analysis and prediction of stock market price trends. But in this presented work a new data model with the hybrid technique is proposed for design and implementation. The figure 1 shows the basic overview of the proposed data model which is required to develop for improving the quality of stock market price prediction.
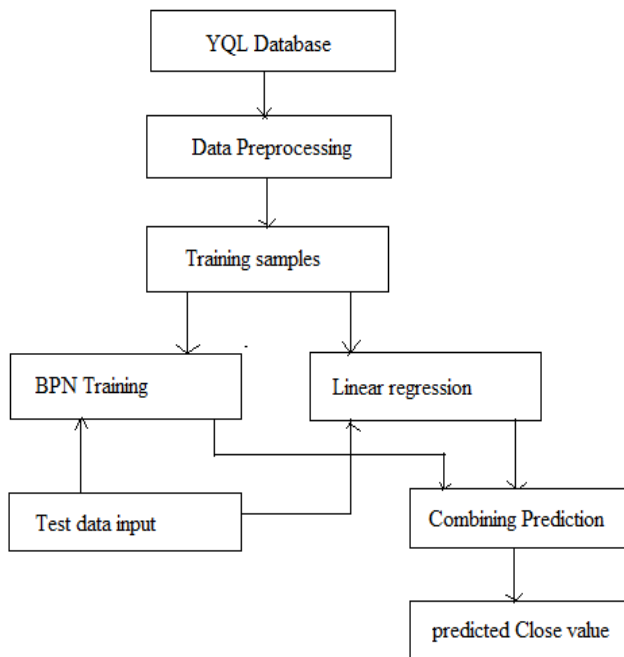
Figure 1 Proposed system architecture

The proposed data model for accurate stock market price prediction is demonstrated in figure 1. The proposed model utilizes the techniques of supervised learning models therefore the entire model works in two major modules in first the model take training from the historical price patterns and in next the model usages the trained data model and current trend samples to predict the future stock market price values. Therefore the YQL (yahoo query language) is used initially for extraction of yahoo stock market historical price database. This data is used in our proposed data model as the initial dataset as training data sample.

In next the data is preprocessed the aim of preprocessing is to improve the quality of data. Using the preprocessing techniques the noise and unwanted data from the initial data samples are removed. In this presented work the data preprocessing technique is employed for removing the missing values from the YQL based extracted data. After improving the data quality the training samples are used with the BPN (back propagation neural network) and linear regression model. Both the models are works in different manner therefore both are trained using the training data samples.

After training of the system the test samples on the trained data models are produced. On the basis of test data samples both the models predict the future values for the input trends. The predicted data is combined in next phase for generating the target prediction values of market close price.

IV. CONCLUSION

The proposed work is a data mining based technique development. That technique can be used for analysis and prediction of stock market price. The paper contains different research efforts and contributions of the researchers in the domain of stock market price prediction. Among most of the researchers are recommending the techniques of data mining for predicting the stock market price. Based on the recommendations and available literature of prediction accuracy a new model for predicting the stock market price is proposed for design and implementation. This technique is implemented in near future and their performance is reported.

REFERENCES

[1] Mondher Bouazizi and Tomoaki Ohtsuki, "A Pattern-Based Approach for Multi-Class Sentiment Analysis in Twitter", 2169-3536, 2017 IEEE, VOLUME 5, 2017

[2] Han, Jiawei, Jian Pei, and Micheline Kamber, "Data mining: concepts and techniques", Elsevier, 2011.

[3] Bharati M. Ramageri, "Data Mining Techniques and Applications", Indian Journal of Computer Science and Engineering, Volume 1 Number 4, pp. 301-305

[4] Dunham, M.H. Data mining introductory and advanced topics. Upper Saddle River, NJ: Pearson Education, New Delhi, 2003. Print. ISBN: 81-7758-785-4, 2006.

[5] Reza Hafezi, Jamal Shahrabi, Esmaeil Hadavandi, "A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price", Applied Soft Computing 29 (2015) 196–210, © 2015 Elsevier B.V.

[6] Xiaodong Li, Haoran Xie, Ran Wang, Yi Cai, Jingjing Cao, Feng Wang, Huaqing Min, Xiaotie Deng, "Empirical analysis: stock market prediction via extreme learning machine", Neural Comput & Applic, DOI 10.1007/s00521-014-1550-z, Springer-Verlag London 2014

[7] Thien Hai Nguyen, Kiyoaki Shirai, Julien Velcin, "Sentiment analysis on social media for stock movement prediction", Expert Systems With Applications 42 (2015) 9603–9611, © 2015 Elsevier Ltd.

[8] QING LI, YUANZHU CHEN, LI LING JIANG, PING LI, HSINCHUN CHEN, "A Tensor-Based Information Framework for Predicting the Stock Market", ACM Transactions on Information Systems, Vol. , No. , Article , Publication date: January 2016.

[9] Sanjiban Sekhar Roy, Dishant Mittal, Avik Basu, and Ajith Abraham, "Stock Market Forecasting Using LASSO Linear Regression Model", Advances in Intelligent Systems and Computing 334, DOI: 10.1007/978-3-319-13572-4_31.

Authors Profile

Mr.Prateek Purey is pursing M.Tech Computer Science and Engineering in the Department of Computer Science and Engineering, Acropolis Institute of Research & Technology, Indore. he received her Bachelor of Engineering Degree from Shree Venkteshwar Institute of Technology, Indore. Her research interests are Data mining, Big data Analytics and Network Security.

Mr. Anil Patidar pursed Bachelor of Engineering and Master of Technology from School of Information Technology, UTD, RGPV, Bhopal, Madhya Pradesh, India. He is currently pursuing Ph.D. from DAVV, Indore and currently working as Assistant Professor in Department of Computer Science & Engineering in Acropolis Institute of Technology & Research, Indore since 2012. His areas of research are Data Analytics, Blockchain, Software Engineering and Agile Development. He has total 9 years of teaching experience and 3 years of research experience.