IJCSE 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[1*], Kaustubh Bhat [2], HarshalKesharwani[3], Pritesh Bhate[4], Khushboo P Khurana [5]

[1*, 2,3,4,5] Department of CSE, Shri Ramdeobaba College of Engineering and Management, Nagpur

[*]*Corresponding Author: piyushj1708@gmail.com, Tel.: +91-9425022291*

Available online at: www.ijcseonline.org

18/May/2018, Published: 31/May/2018

Abstract - The share market is the new business for many individuals and companies.The stock market attracts many as there is a huge profitinvolved. But there can be huge loss as well. The world suffered from many economic crises due to itsdownfall, which may have cost lives and way of living of many. If the share prices couldbe predicted, this can help in avoiding the economic crisis.Many business analysts said that the share market cannot be predicted and it is completelyrandom, but there is no common opinion about that. Many also said that it can bepredicted but with using different measures. In this paper, we have proposed a technique to perform stock market data analysis to find if there exists a relation between stock price change of two companies- TCS and Infosys. This is done using Big data technique by performing sentiment analysis of Tweets from Twitter and finding the correlation. Also, machine learning techniques are applied on the BSE data of the companies to predict the stock prices of next day. The results of allthe factors like sentiments, similar industrial shares, and index shares are combined to get to the conclusion. We found that the share market depends on the numerous factors and considering only certain factorsare insufficient for analysis.

Key Words—Hadoop, Map-reduce, Machine learning, Sentiment analysis, Stock Market

I. INTRODUCTION

Stock market has become a vital part of the global economy. It has been a good investment option for many years due to high returns. Moreover, the development in the technology has made dealing in share market very easy and has attracted more individual inverters.

Shares are the percentage ofinvestment a company gets from individual investors or other companies. People invest in the companies to get thepart of profit a company earns while working with the invested money. The prices of the shares depend on the base price a company declares at release and day-to-dayinvestments.

The prediction of stock prices can assist in maximizing the profits earned and reduce the risk of investment. It is the aspiration of all investors to forecast the market behaviour, so that, it helps them in taking decisions of buying or selling shares.

According to the hypothesis of random walk theory [1], stocks prices are defined randomly and it is impossible to forecast the prices. However, advances in machine learning techniques and proliferation in the amount of data have made forecasting the stock prices possible. A lot of research is being done in this area. Machine learning techniques and data mining tools can be used to perform the task of prediction.

The contributions of this paper are two –fold: (1) to perform stock market analysis and (2) prediction of stock prices using the share market history data.The parameters usedin reference to stock market are Open, Close, High and Low values of the stock. We consider the stocks of TCS and Infosys for analysis. In analysis part, we try to find if there is any correlation between the daily fluctuations of the stock prices of these companies and for prediction of stock value of next day, we apply machine learning techniques.

In this paper we present a review of on-going research in prediction of stock markets and then propose an approach for analysis of stock data and its prediction. In [2], various machinelearning algorithms such as Artificial Neural Networks (ANN), ANN with different feature selection and Support Vector Machines (SVM)are reviewed. The results show thatthe SVM algorithm with Principal ComponentAnalysis (PCA) works best for prediction.

Our technique for prediction of stock markets is based on analysis of tweets from twitter and application of machine learning techniques.The analysis is performed using the

© 2018, IJCSE All Rights Reserved

Hadoop Framework for large, scalable data that might be available for precise and fast calculations.

Rest of the paper is organized as follows, Section I contains the introduction of stock market data analysis and its prediction. Section II describes the scope and enlists various advantages of stock markets prediction. Section IIIcontain the related work in stock market analysis and prediction. Section IV explains the proposed methodology. Section Vdescribes the implementation details and section VI presents the experimental results. Section VIIconcludes research work and providesrecommendations for future work.

II. SCOPE AND ADVANTAGES OF STOCK MARKET PREDICTION

The stock market prediction can help investors in many ways. Valuable suggestions in terms of investment can be provided. Scope of the task of prediction is as follows:

- Predicting open, close, high, low values of stock for future day(s).
- Co-relation between companies: Getting more competitive market where companies might depend on each other (increase or decrease).

Advantages of stock market analysis and prediction are numerous. Some of them are as listed below:

- Short Term Profit Gain: Getting more profit by investing in the company whose stock price is likely to increase the next day or in a short time.
- Long term investment options: Gaining profit in long term.
- Saving the world from the economic crisis in future by knowing the downfall before-hand and taking measures for saving the company assets.
- Saving the company from being bankrupt by knowing the continuous downfall in advance and taking actions in accordance with it.

III. LITERATURE REVIEW

The task of stock markets prediction, in the literature is addressed using techniques of machine learning (ML), natural language processing (NLP) and data mining. These techniques are reviewed in this section. We propose a technique using the combination of NLP and ML to deal with the stock prediction task.

According to Stock market prediction model proposed in [3], a stock can be traded in multiple ways like equity, options, futures, or commodities. There is a need to co-relatethese different types of trade, since a decision made on equity is more stable than a decision made on options. Relating equity with other risk bearing trade types can be of great benefit. Authors tried to find the co-relation between these and found that though Hidden Markov Model (HMM) is the best model for pattern capturing, it does not give optimal results in the case of stock prediction. More observations are needed, such as finding correlation between the currencies of different countries.

In [4], stock market prediction methods using sentiment analysis and data mining are reviewed. Despite a lot research, it was found out that the current stock prediction have many limitations and needs improvement. The final conclusion was that the prediction of stocks is a very complex task and more factors are needed to be considered.

In [5] authors used data mining and artificial intelligence for stock market prediction. They found the need to improve the parameters to achieve accuracy and performance.

A method using HMM is proposed in [6] to predict the close value of a stock for next day. Stocks of four companies were considered for prediction. The system was tested using Mean Absolute Percentage Error (MAPE) performance indicator.

P.Sonami, et.al. have presented a model for stock markets prediction using HMM. Model of HMM is built by training using Baum Welch algorithm. This HMM is used for testing using Maximum a posteriori(MAP) approach. The model selects one best probability value using Viterbi algorithm. Closing value of next day is given predicted by the model. The performance of the model decreases with increase in training data [7].

Elman neural network (ENN) is a local recurrent neural network. It has one context layer that remembers the past states. In [8], ENN along with self-adapting variant PSO algorithm is used for prediction of the opening price of the stock. First the initial weight and threshold value are optimized using PSO algorithm then given as input to ENN for training to build the model for prediction. The results show that the precision and stability of the proposed model are better than the traditional neural network based models.

Multi-layer perceptron(MLP)in feed forward fashion and Elman recurrent network (ERR) are used in [9]. Model based on MLP was developed to predict the stock value of next day based on history data without the knowledge of current market. The Tehran Stock Exchange (TSE) data was used for training and testing. The input parameters considered were low, high and average value in d days. Five performance evaluators (Mean Absolute Deviation, Mean Absolute Percentage Error, Mean Squared Error, and Root

© 2017, IJCSE All Rights Reserved

Mean Squared Error) were used to judge the performance of both the methods. Out of the two methods the authors found that ERR model was better than MLP. But ERR has more errors than MLP.

Stock predictive models using multilayer perceptron are reviewed in [10].

Association rule mining based approach is used in [11] for stock market prediction. Semantic association rule mining. An intelligent data mining technique for rule extraction is presented. An RDF dataset is extracted in triple format and is mined to find the association rules. Ontology knowledge is exploited in the system and MICF measure is used to reduce the search space. Appropriate rules are then generated from thousands of rules. Experiments on stock market data prove the usefulness and efficiency of the approach.

In [12], deep learning methods- Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are used for finding the intraday directional movements of Standard and Poor's 500 index. The titles of financial news and a set of technical indicators are used as input. Complex patterns can be easily detected using deep learning methods. RNN captures the context information from the data and models complex temporal characteristics for prediction of stock market.

Sentiment-Aware Stock Market Prediction using deep learning method is proposed in [13].

In [14], a model based on RNN with Gated Recurrent Units (GRU) is used to predict the instability in the Chinese stock market. Development of model for stock prediction for Infosys is presented in [15].

IV. PROPOSED METHODOLOGY

In this paper, we have done analysis for stocks of two companies- TCS and Infosys. We intend to find whether there is a correlation in the stock prices of these two companies.

The overview of the system is as in figure1. Data is used from two sources- Twitter and BSE share market data. Sentiment analysis is performed on Twitter data to find the view of people regarding the company. Tweets related to TCS are extracted using Twitter API. Then sentiment of each extracted tweet is found (in part-A). These sentiments are then passed as input in Hadoop framework to find the sentiment value for a company for each day. That is, a single value is found by analyzing all sentiments for a day and this is done for all dates available in the data. So for each date we have a polarity about the comment as positive or negative, in the form of values that can be positive or negative.

The data extraction is also performed on stock BSE data to extract data of TCS, Infosys and Sensex overall. Then a difference value is computed for each of the extracted stock values. This difference value D along with sentiment is used to find correlations.

Also, machine learning techniques are used to find the prospect of prediction. The input to machine learning algorithm are the difference data D and raw data of all three- TCS, Infosys and Sensex.

The input stock data contains four values- open, close, high and low. The proposed system is divided into following phases:

Phase1: Data Extraction
Phase 2: Sentiment analysis
Phase 3: Finding Correlation
Phase 4: Prediction using machine learning techniques

These phases are explained next.

A. Data Extraction Engine

For analysis of stock market, we use the tweets from Twitter and the stock history data. The data is extracted and cleaned for use.

Following are the data used for experimentation along with their sources:

- Twitter data for sentiment analysis: The news about the target company i.e. TCS along with user opinions about the company are extracted from Twitter. Twitter data analysis helps to find about the thinking of people about the company, its policies and its products along with any other news that may be appearing in the Tweet. Its sentiment is calculated which serves as the data for further calculations.
  Source: Twitter API
- TCS BSE share market data: TCS is selected as the target company for analysis with the exchange as Bombay Stock Exchange (BSE).
  Source:https://in.finance.yahoo.com/quote/TCS.BO/profile?p=TCS.BO
- Infosys BSE share market data: Infosys is selected as the other company to find the correlation between the stock prices of these two companies. Since, both the companies have similar work profile.
  Source:https://in.finance.yahoo.com/quote/INFY.BO/profile?p=INFY.BO
- Sensex BSE index data: Stock prices of BSE are considered. We also use the index values of BSE overall Sensex.
  Source:https://in.finance.yahoo.com/quote/%5EBSESN/history/

© 2018, IJCSE All Rights Reserved

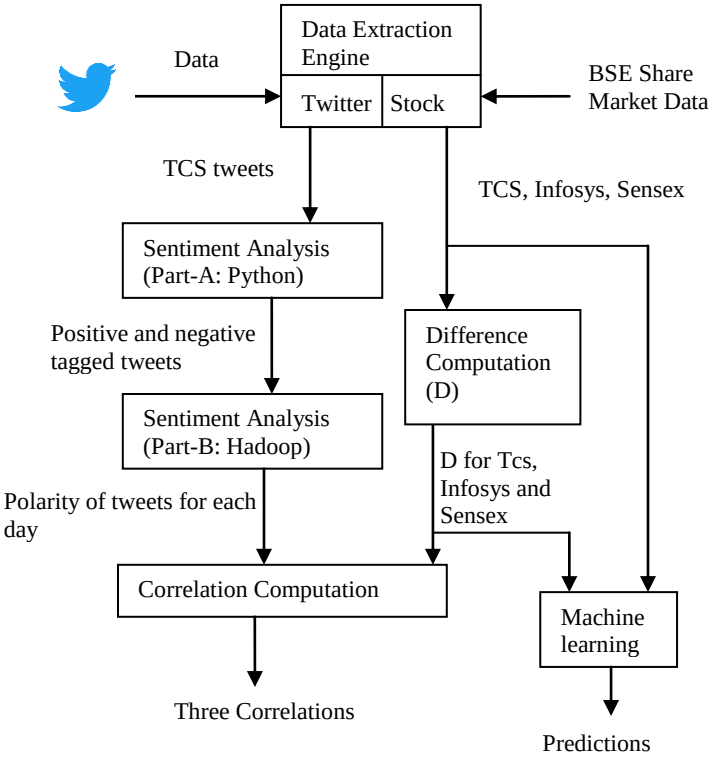

Figure 1. Proposed System Overview

B. Sentiment Analysis

Sentiment analysis is divided into two parts. In Part-A, the sentiment polarity of all tweets extracted are found. Python code is written to perform this task and the IDE used is PyCharm IDE. Then in part-B the output from part-A is considered as input. Map-Reduce technique is used to combine individual sentiment results and find a sentiment for each day. Combining the results using the Hadoop map-reduce framework depends on date as a key for maintaining the resultant data.

The part-B is done in Hadoop framework, since the amount of data and its sentiment may be very large in amount. We have exploited the big data technique to handle the large volume of data.

C. Correlation Computation

Correlation coefficients are used to find how much strong relationship exists between two variables. There are several types of correlation coefficients. We have used Pearson'scorrelation (r). It is a correlation coefficient used to quantitatively measure the extent to which the two variables are related to each other [16]. The formula for Pearson correlation r for two variables x and y is as below,

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{[N\sum x^2 - (\sum X)^2] - [N\sum y^2 - (\sum Y)^2]}}$$

In the formula N is the number of pairs of scores. The value of r lies between -1 and 1. r =1, indicates a positive correlation. For positive increase in onevariable, there is a positive increase of a fixed proportion in the other. r = -1, indicates anegative correlation.For every positive increase in one variable, there is a negative decrease of a fixed proportion in the other. r = 0, indicates no correlation. The two variables are not related to each other [16].

The absolute value of the correlation coefficient defines the strength of the relationship. The larger the number, the stronger the relationship. For example, |-0.95| = 0.95, has astronger relationship than 0.55.

For our method, we compute three correlations as follows,

- TCS D – Infosys D
- TCS D– Sensex D
- TCS D – Sentiment result

Where D means the change in the share price of each day i.e. difference between price at close and at open.

For example, if open for TCS is 1990 and close is 2000 the value of TCS D will be 10.

Similarly, the difference of values of Infosys Open and close is referred as Infosys D. Also, the difference of values of Sensex Open and close is referred as Sensex D.

The complete process is done using the Hadoop map-reduce framework for completing the task in less time with the large data given.

D. Prediction using Machine Learning

Machine learning algorithms have the ability to automatically learn and improve. Programs can be learned from the data [17].In this paper, we try to apply several regression algorithms and comparing their results for prediction of stock prices and select the one with the least error. The error should not be too large for the results and by selecting the algorithm for finding the next day's value for the change ordifference.We have used the Orange tool to apply various machine learning algorithms. It calculates the error values for different algorithms and also the predicted value.

V. IMPLEMENTATION AND RESULTS

The implementation is done using Python for sentiment analysis, Hadoop Map-Reduce framework for computation of sentiment analysis part-B and correlation computation. Orange tool is used to apply machine learning algorithms.

A. Data Extraction and Sentiment analysis

The part-A of sentiment analysis implemented in python and extracts the tweets having tags, TCS, TCS employee, TCS loss, TCS shares. A part of code is shown in figure 2 and figure 3 shows the result of sentiment analysis (part-A) for each tweet and its percentage. Part-B results are shown in figure 4.

© 2017, IJCSE All Rights Reserved

```
from tweepy.streaming import StreamListener
from tweepy import OAuthHandler
from tweepy import Stream
access_token = "888004757331169280-
He7EuBKUg3mvenavU0hS9m2N2kP7IEc"
access_token_secret =
"rXufqRNhBljYXKeCK90Ab30Qyt0IMDUUNxbmtGTwJaP0B"
consumer_key = "fMgbjpoZ1HGcE3o3ucUOAL9rb"
consumer_secret =
"6Yo1I2kDZrxYUfjporMnYnMd8YvMWrdki3RI9LXuxMNYtWQzCb"
class StdOutListener(StreamListener):
    def on_data(self, data):
        print (data)
        filesave = open("F:\\project.json",'a')
        filesave.write(data)
        filesave.close()
        return True
    def on_error(self, status):
        print (status)
if __name__ == '__main__':

    #This handles Twitter authetification and the connection to Twitter
Streaming API
    l = StdOutListener()
    auth = OAuthHandler(consumer_key, consumer_secret)
    auth.set_access_token(access_token, access_token_secret)
    stream = Stream(auth, l)
    stream.filter(track=['arvind', 'kejriwal', 'demonetization'])
```

Figure 2. Extraction of Tweets

```
C:\Python27\python.exeC:/Users/apritesh/PycharmProjects/twitter/sentime
nt_tcs_json.py
                Date                    Sentiment Value
Tue Feb 20 11:05:39 +0000 2018          positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        positive
Sat March 24 11:05:39 +0000 2018        negative
Sat March 24 11:05:39 +0000 2018        negative
Sat March 24 11:05:39 +0000 2018        negative
Sat March 24 11:05:39 +0000 2018        negative
Positive tweets percentage: 77 %
Negative tweets percentage: 22%
```

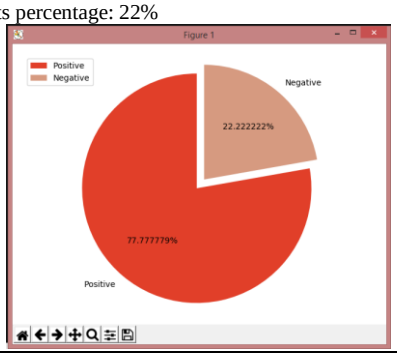

Figure3. Sentiment analysis (part-A)

```
15-03-2018    4.568723270333034    1.2650177473635067 5184.0
   484.0          112.28884078468833
16-03-2018    5.330649401966801    3.067037905544675  7054.5625
   2420.0         86.62839000065513
19-03-2018    5.383495960512723    4.324007874032265
   7194.981516121609    4724.0    70.69534061429832
20-03-2018    5.747520512073462    5.014843666348816
   8199.868409524016    324.0     39.35674975338465
21-03-2018    5.758707088356365    5.537943507137051
   8231.789802133613    7693.0    38.21310531715601
22-03-2018    5.9915789260154915   7.181367842875967
   8910.394855036015    12877.0   45.8997223602654
23-03-2018    6.023481929825224    7.209387623171932
   9005.457355036015    12977.0   44.57383963491303
```

Figure 4. Sentiment analysis (Part-B)

B. Finding the Pearson's coefficient

The Pearson's correlation coefficient requires the mean if the data for its calculation which is calculated using the Map-Reduce framework.After the computation of mean, it is inserted manually in the second reducer and the nextset of mapper and reducer computes the correlation by considering the data input from HDFS. The output is the last entry of the last row and other entries areintermediate temporary results.

We have used the result obtained previously using the mapper and reducer classes and manually entered it in the reducer class. This calculates the Pearson's coefficient for the data given through HDFS input directory.

By computations in Hadoop for Pearson's Correlation coefficient we get results as shown in table 1. The percentage correlation is as follows:

- TCS – Infosys Daily Change - 32.62%
- TCS – Sensex Daily Change - 4.55%
- TCS – Sentiment Daily Change - 44.67%

The values obtained are very less and shows that there is very less correlation between the stock price changes in TCS-Infosys, TCS-sentiment. Moreover, the correlation between TCS and Sensex's negligible.

Table 1.Factors and their Pearson's coefficient

Factors	Pearson's Coefficient
TCS-Infosys daily change	0.3262
TCS-Sensex daily change	0.0455
TCS-Sentiment daily change	0.4456

C. Machine Learning Model results

We have used four machine learning techniques- kNN, Random forest, neural network and linear regression for prediction of stock data. For all these techniques we find the errors- Mean square error (MSE), Root Mean Square Error (RMSE), Mean Absolute error (MAE) and Coefficient of determination (R2). The results are as shown in table 2.

The best result is obtained using Linear Regression with absolute mean error to be nearby25.039 using the random sampling technique with 70% training tuples. In all the models or algorithms, we get the least error (all types) in the

© 2018, IJCSE All Rights Reserved

linear regression and the relationship coefficient as maximum. Thus, linear regression is best to find the predicted values.

We find the best result using the Linear Regression but the error percentage is very large. Hence, the values predicated will not be accurate.

Table 2. Results of different machine learning models

Methods	MSE	RMSE	MAE	R2
kNN	1398.869	37.401	28.295	-0.033
Random Forest	1411.234	37.566	28.095	-0.042
Neural Network	1372.334	37.045	27.784	-0.013
Linear Regression	1161.986	34.088	25.039	-0.142

C. Machine Learning Prediction Results

All the four models kNN, Random Forrest, Neural Network and Linear Regression were used to predict the values of daily change in the stock process of the two companies i.e. TCS and Infosys.The predicted values showed that none of the four models were able to predict the values accurately. However the closest values were predicted by Linear Regression.

D. Visual Results

Fig. 5shows the graphical representation of TCS daily change and corresponding Infosys daily change. Figure 6 shows the graphical representation of TCS daily change and thecorresponding Sensex daily change.

From the graphs, it is clear that TCS and Sensex does not follow anypattern that is there is no correlation between them. The behaviour is completely random with respect to each other. Same is observed from our experimentation results. While the graph ofTCS and Infosys somewhat follow trends but at some of the points they show complete opposite results. The graph in fig. 2 clearly shows that both trend similar except at several points where they go exact opposite to each other and thus, the correlation values become less. But in fig. 3 there is no relation seen as at beginning the graphs are separated and later on they are overlapped with no correspondence and random behaviour.

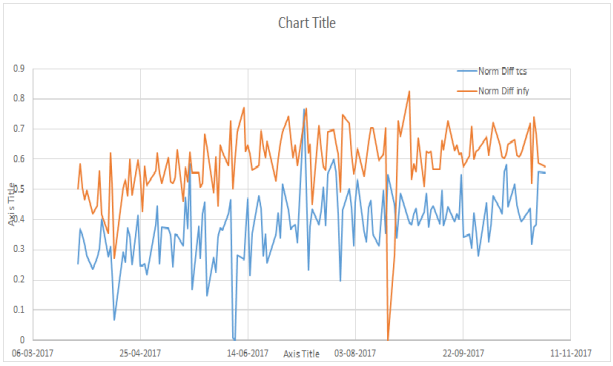

Fig.5 TCS-Infosys difference graph

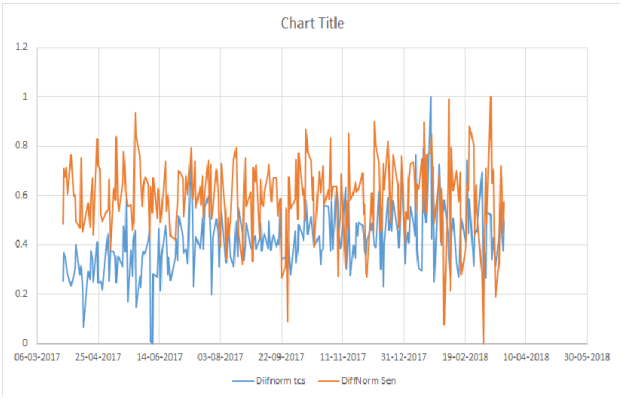

Fig.6 TCS-Sensex difference graph

VI. CONCLUSION

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.

REFERENCES

[1] B. G. Malkiel, “A Random Walk Down Wall Street”, W. W. Norton, New York. 1973.

[2] R. Iacomin, “Stock Market Prediction,”2015 19th International Conference on System Theory, Control and Computing (ICSTCC), October 14-16, CheileGradistei, Romania

[3] S., Kamath, “Stock Market Analysis,” Master's Projects, pp. 326, 2012.

[4] V. Rajput, S.Bobde, “Stock Market Forecasting Techniques: Literature Survey,” International Journal of Computer Science and Mobile Computing, 5(6), June 2016, 500-506.

[5] G. S., Navale, N., Dudhwala, K., Jadhav, P., Gabda, & B. J., Vihangam. , “Prediction of Stock Market using Data Mining and Artificial Intelligence,” International Journal of Computer Science and MobileComputing, 134(12), January 2012, pp 9-11.

[6] A. Gupta and B. Dhingra, "Stock market prediction using Hidden Markov Models," 2012 Students Conference on Engineering and Systems, Allahabad, Uttar Pradesh, 2012, pp. 1-4.

[7] P. Somani, S. Talele and S. Sawant, "Stock market prediction using Hidden Markov Model," 2014 IEEE 7th Joint International Information Technology and Artificial Intelligence Conference, Chongqing, 2014, pp. 89-92.

[8] Z. Zhang, Y. Shen, G. Zhang, Y. Song and Y. Zhu, "Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network," 2017 8th IEEE International

© 2017, IJCSE All Rights Reserved

Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 2017, pp. 225-228.
[9] M. Naeini, H. Taremian and H. Hashemi, "Stock Market Value Prediction Using Neural Networks," International Conference on Computer InformationSystems and Industrial Management Applications (CISIM) 2010.
[10] A. Rao, S. Hule, H. Shaikh, E. Nirwan and P. M. Daflapurkar, "Survey: Stock Predictive Models Using Multilayer Perceptron", International Journal of Computer Sciences and Engineering, Vol.4, Issue.5, pp.111-113, 2016.
[11] S. Asadifar and M. Kahani, "Semantic association rule mining: A new approach for stock market prediction," 2017 2nd Conference on Swarm Intelligence and Evolutionary Computation (CSIEC), Kerman, 2017, pp. 106-111.
[12] M. R. Vargas, B. S. L. P. de Lima and A. G. Evsukoff, "Deep learning for stock market prediction from financial news articles," 2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), Annecy, 2017, pp. 60-65.
[13] Jiahong Li, Hui Bu and Junjie Wu, "Sentiment-aware stock market prediction: A deep learning method," 2017 International Conference on Service Systems and Service Management, Dalian, 2017, pp. 1-6.
[14] W. Chen, Y. Zhang, C. K. Yeo, C. T. Lau and B. S. Lee, "Stock market prediction using neural network through news on online social networks," 2017 International Smart Cities Conference (ISC2), Wuxi, 2017, pp. 1-6.
[15] S. Gour, "Developing Decision Model by Mining Historical Prices Data of Infosys for Stock Market Prediction", International Journal of Computer Sciences and Engineering, Vol.4, Issue.10, pp.92-97, 2016.
[16] J.K.Sharma, "Business Statistics," Pearson Education, India, pp. 454-455, 2007.
[17] P.Domingos, "A Few Useful Things to Know about Machine Learning," Communications of the ACM, Vol. 55 No. 10, October 2012, Pages 78-87.

Authors Profile

Piyush Jain is pursing the final year of Bachelor of Engineering from Shri Ramdeobaba College of Engineering and Management affliated to Rastrasant Tukdoji Maharaj Nagpur University, Nagpur. His research interests include big data analytics and data mining.

Kaustubh Bhat is pursing the final year of Bachelor of Engineering from Shri Ramdeobaba College of Engineering and Management affliated to Rastrasant Tukdoji Maharaj Nagpur University, Nagpur. His research interests include big data analytics and data mining.

Harshal Kesharwani is pursing the final year of Bachelor of Engineering from Shri Ramdeobaba College of Engineering and Management affliated to Rastrasant Tukdoji Maharaj Nagpur University, Nagpur. His research interests include big data analytics and data mining.

Pritesh Bhate is pursing the final year of Bachelor of Engineering from Shri Ramdeobaba College of Engineering and Management affliated to Rastrasant Tukdoji Maharaj Nagpur University, Nagpur.His research interests include big data analytics and data mining.

Khusboo P Khurana is currently working as Assistant Professor in Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management affliated to Rastrasant Tukdoji Maharaj Nagpur University, Nagpur. She has published various research papers in reputed international journals. Her subject of research includes image processing, video processing and Big data analytics.

© 2018, IJCSE All Rights Reserved