International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Predictive Analysis on Stock Market Data

K Prasanna Lakshmi, N V Ganapathi Raju, Anusha Buddaraju, Lahari Devaraju

Abstract: ***In today's digital world data generated is huge and heterogeneous in nature. This data must be analyzed and organized for efficient usage. Predicting unseen future events by analysis is a methodology of progressive analytics. This division uses many modes of operandi staring from statistical modelling, excavating data using data mining, learning and gaining knowledge using machine learning and artificial intelligence. This work presents predictive analytics on stock market data. Predicting stock market performance is the most difficult things to do. Stock prices move up and down every minute due to fluctuations in supply and demand. The altering stock prices makes it arduous for accurate prediction. The predicted values are of immense use to the stock investors. This enables the investors to take right decisions during trading. Investing in right time at an appropriate area is crucial. Stock market prediction accomplishes this task effortlessly. In this paper, we are using data mining techniques. After detailed study on the possible algorithms, k-means was the most suitable clustering algorithm. K-means allows faster computation and produce tighter clusters. To get the finest close results, classification technique is also implemented along with clustering. Decision tree classifier is used to build a model. The upcoming stock values up to five days are predicted. Results are tested on the SBI stock data set which is collected from the national stock exchange limited. A comparative study is conducted using SVM classifier to prove that methodology using decision tree classifier is best suited for predicting the result.***

Index Terms: Prediction, Classification, Clustering, SVM, K-Means

I. INTRODUCTION

The age of economic growth and the endowment of digital technology led to agglomeration of monetary data. The expeditious growth of data foreclosed the capability of human beings to evaluate them manually. In addition to this, financial time series data as in [14][21][20] is further complicated than statistical data due to cyclic variations, seasonal alterations and long-term trends. The endless expansion of such tremendously oscillating and erratic data has put forth the demanding need to encourage automated access to financial data. It enables investors to extract meaningful statistics and facilitates knowledge-driven decisions.Achieving lofty profits is the eventual target of a financier investing in financial market. Stock market, Share market, Exchange-traded funds are few of the investment prospects present in financial market. Investing in stock market is one of the noticeable capitalization way of investors. Venture Capitalist can enhance and gain more profit by investing at proper time. In order to attain optimum stock trading results, the suitable and appropriate time to trade has to be found which mitigates the risk. But the challenge here is that the
stock market is mutable in nature, thereby rendering it most difficult to find the best time to trade. There are some indicators which help in predicting the behavior of stock market prices which is of interest to many market researchers. These indicators are arrived by extrapolation of archived data. Basing the decisions on mere indicators of past data many not always result in profits. Hence additional fine tuning of the basis for predictions is done by various mining techniques [24]. Profitable investments are based on accurate extrapolation of indicators for the future. Successful classification of the abrupt movement in stock values helps to monitor stock data and benefits the investors in efficient trading strategies. The monitoring of the fluctuations of the stock prices of the history will cut down the ambiguity in decision making.

In this study, stock market prediction of upcoming five days is estimated and displayed. The probable trend of the prices of stock in the future, basing on the performance in the past, are estimated by the stock trading experts. The pattern of fluctuation in the past can indicate the likely behavior in the future if correlated accurately by an experienced stock trading expert. This paper aims at precisely predicting the stock values in a conceptual manner. A model is designed for skillful outcomes.

The data set is acquired from the National Stock Exchange of India Limited (NSE). The electronic data provided by NSE is the first of its kind in the country which facilitates the investors for smooth trading.

II. RELATED WORK

Various methodologies related to artificial intelligence and machine learning have been put to use since a decade for analyzing and forecasting of stock market. In table 1 research regarding stock market prediction is shown. KNN, SVM and Naïve Bayes algorithms were used in [1] [22] [23]. Their main aim was to conduct sentimental analytics [25] and construct a model using business news and historic stock values.

Revised Manuscript Received on 22 May 2019.
***** Correspondence Author
K Prasanna Lakshmi*, Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India.
N V Ganapathi Raju, Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India.
Anusha Buddaraju, Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India.
Lahari Devaraju, Information Technology, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

They acquired accuracy ranging from 72.73% to 86.21%. Various numerical attributes present in the day to day news dataset were considered for accurately predicting the futuristic behavior.

Huge data predictive analysis on stock market datasets was conducted and the results of which shown in [2]. Their aim was to prove that stock market prediction can be done in the best way by comparing algorithms used for predictive analysis in order to reduce the error in prediction. The paper mentioned in [3] included social media analysis using stock market prediction. The applications of above findings include add sentiment and discarding of the posts without sentiment. A machine learning model to predict stock market is shown in [4]. The results were tested on various companies including Adobe, Oracle, HP, American Express CO, Bank of New York, Honeywell Company, Hospera, Life technologies and AT & T [4]. It integrates PSO algorithm and LS-SVM for stock price prediction using financial technical indicators [4]. This combination acquires least errors when compared to ANN-BP algorithm. Table 1 depicts the related predictive research work on stock market data.

Table 1: Related research of stock market prediction is shown in the below table

Author	Features	Data set	% of accuracy			Methodology
			KNN	SVM	Naïve Bayes	
[1]Khedr, Ayman E., and Nagwa Yaseen	Sentiment analysis of financial news and historic stock prices.	Yahoo inc	75.86	58.62	86.21	KNN, SVM, Naïve Bayes
		Msft inc	69.80	66.67	72.73	
		Fb inc	72.41	68.97	82.76	
[2] Balaji, S. Naveen, P. Victer Paul, and R. Saravanan	Survey about stock market prediction using big data analytics	Yahoo finance and search twitter API or stream twitter API	5			Sentiment analysis
			86.7			Twitter mood predicts
			Satisfactory results obtained			ANN for future prediction
			Less accurate prediction			Big data approach
			8			Cluster then predict model
[3] Coyne, Scott, Praveen Madiraju, and Joseph Coelho	Forecasting stock prices using social media analysis	Stock Twits data set	65			Multilayer perceptron classifier, sentiment classifier
[4] Hegazy, Osman, Omar S. Soliman, and Mustafa Abdul Salam	Machine learning model for stock market prediction	Yahoo data sets	----			PSO and LS- SVM

The later part of this paper is structured as follows: In Section III, the problem statement is discussed. The proposed methodology is presented in Section IV. It is continued by Section V which is about the experimental results. At the end the conclusions are presented in Section VI.

III. PROBLEM STATEMENT

This paper consists of four modules to detect the accuracy of stock data and predict the upcoming closing stock values. The stock market can help us in making a lot of money, but we might lose all the money if we are driven to invest randomly without knowing the fundamentals of the market. Improper investment could mean great loss to the investor. The lack of guaranteed returns led potential investors to just spectate. It is therefore desirable to have a mechanism that can guide the investors in future ventures. This mechanism assists in predicting upcoming stock values which is favorable for precise decision making. Data mining methods can analyze the stock prices over time and the knowledge gained is used in prediction.

IV. METHODOLOGY

In this paper, stock market prediction is done using clustering along with classification.

A. Clustering

Similar type of data is pooled into a cluster using Clustering technique [18]. This helps in improving the accuracy of prediction as different groups exhibit distinct behavior as in [15]. Hence, a distinct model for individual collection is always advisable. Creation of distinct and dissimilar groups is the goal of clustering concept. The data points of a cluster are tightly placed with a very narrow separating distance. Clustering methodology creates many clusters depending on dissimilarity between data points. Each cluster is distinctly placed for other.

K-Means clustering [21]: K-means is one of the simplest algorithms to solve well known clustering problems. It follows a simple and easy way to classify given data set through certain number of clusters (k) [21]. The 'k' is the number of clusters used in K-Means. It is a threshold given by the user. Cluster center of each cluster will be midpoint (mean) of all points that fall in the same cluster [27]. We calculate Euclidean distance using the formula,

$$dist((x,y),(a,b))=\sqrt{(x-a)^2 + (y-b)^2} \quad (1)$$ [27]

During this process we might come across intra clustering and inter clustering. Intra-cluster distance is the sum of absolute/squared distance between all pairs of points in the cluster or the centroid and all points in the cluster or the "medoid" and all points in the cluster [26]. Inter-cluster distance is the sum of squared distance between all pairs of clusters where distance between two clusters is defined as distance between their centroids/medoids or (spherical clusters) distance between the closest pair of points belonging to the clusters (chain shaped clusters)[26].

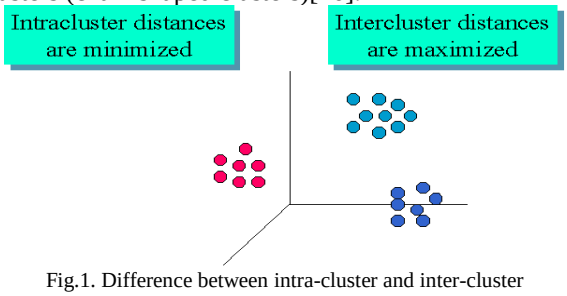


Fig.1. Difference between intra-cluster and inter-cluster

In the direction of finding count of clusters, we are using elbow curve. The optimum number of clusters we get from the elbow curve are given to k-means algorithm. The "elbow" joint of the arm found in the line chart is considered as the value of k. In Fig.1, this value of k is shown as 3 which is considered for further functionality.

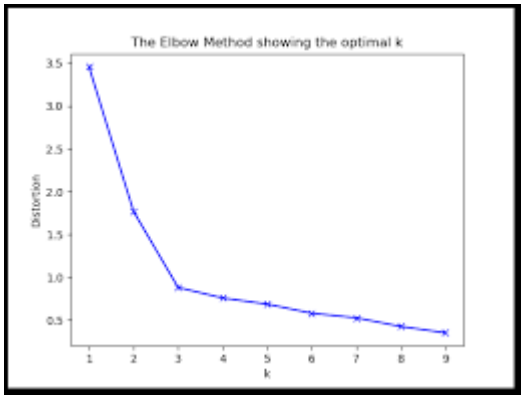


Fig.2. Elbow Curve

In this paper we consider open, close, high, low, volume, turnover, vwap as factors for performing k-means clustering on the dataset. As we have considered 'k' as 3 using the elbow curve, k-means algorithm internally takes 3 random data points on the plane to perform clustering. K- Means clustering algorithm keeps on running till we obtain optimized 3 clusters.

Table 2: Clustering work done by other authors

Author	Methodology	Features
[5] Ferrandez, Sergio Mourelo, et al	K-means and genetic algorithm	Shortest distance prediction
[6] Jain, Vineet, et al	K-means	Crime prediction
[7] Ghorbani, Ali, and Sara Farzai	K-means	Fraud detection
[8] Zulfadhilah, Muhammad, Yudi Prayudi, and Imam Riadi	K-means and log analysis	Cyber profiling

Table 2 shows the previous work done by authors who used clustering algorithms. K-means and genetic algorithm are used in [5] which shows the optimization of truck drone in tandem delivery network. It considers the truck speed, drone speed and the operating space to present the shortest distance for delivery. The maximum used areas on the internet were depicted using clusters. Crime prediction using k-means was shown in [6]. K-means was used in [7]. It included fraud detection in automobile insurance. The fields considered were age of the culprit, time and place of fraud. K-means and log analysis was used in [8] to perform cyber profiling. It forms clusters according to area of theft in the city of New Delhi.

B. Classification

Segregating the given test data and categorizing it based on the knowledge gained by training data is the area of major concern of researchers in the domain of classification. Classifier aids the system in characterizing and evaluating unlabeled data.

Decision Tree: The methodology behind using decision tree is to create a training model as shown in [16][29][18][19]. It can be used to predict test data for class labels by building classifier which contains rules inferred from training data. The algorithm adapted is Classification and Regression Tree (CART) analysis. CART is a non-parametric decision tree learning technique that produces either classification or regression trees, depending on whether the dependent variable is categorical or numerical [28]. In the case of over-fitting, where we get more accuracy in training and less in testing, we use pruning. Pruning is cutting down of irrelevant branches.

Construction of Decision Tree

A tree can be constructed by classifying the data into subsets using an attribute value test. This process is iterated recursively on subsets for further partitioning. This process ends when there is no more splitting possible upon conducting value test. The advantage of this process is that neither a parameter setting nor knowledge in this field is required. Hence provides scope for exploring knowledge.

In this paper we have used DecisionTreeRegressor algorithm to train the model as our target variable is 'close value'. Before we train the model, we must split each of 3 clusters into train and test datasets in 8:2 ratio i.e., 80% of train dataset and 20% of test dataset. Now train the decision tree algorithm on to three train datasets and use this model on test datasets to predict the close value which is our target variable. Then we have plotted the original close values and predicted close values on a line chart just to show the trend followed by the model. Then we have calculated the accuracy of predicted close values against original close values and depicted them on a bar-chart.

Table 3: Classification done by other authors

Author	Methodology	Features
[9] Bhumika, Prof Sukhjit Singh Sehra, and Prof Anand Nayyar	Decision tree, neural network, genetic algorithm	Text classification
[10] Kabakchieva, Dorina	Rule learner, decision tree, neural network, nearest neighbor	Student performance prediction
[11] Saa, Amjad Abu	Decision tree, naïve Bayes	Educational data mining and student performance prediction
[12] Ali, Mohamoud, and Yugyung Lee	Naïve Bayes, logistic regression, random forest	Customer relationship management

Table 3 shows the previous work done by authors who used classification algorithms. Text classification was shown in [9]. The aim of the paper is highlight important algorithms that are employed in text document classification. They considered various algorithms like decision tree, neural network and genetic algorithm in order to measure the performance. Student performance prediction was shown in [10] using data mining classification techniques [20]. The goal of the entire work was to enroll by attracting students with caliber for proficient universities. In [11] performance of students who were considered as customers was predicted using data mining. Various metrics used for this prediction include pre-university performance, personal information and university performance. Emphasis was on creating a model for predicting the grade points of students. CRM suing continuous time-evolving classification was shown in [12]. In this paper they presented sales prediction for automotive CRM based on customer interactions. The predictive capability serves to keep the sales people on the right track.

C. Combination of clustering and classification

Clustering followed by classification is a rare scenario. In this paper, we implemented k-means clustering algorithm followed by decision tree classifier. This methodology was tested, and we were drawn to a conclusion that it is appropriate for stock market prediction.

Table 4: Clustering and classification done by other authors

Author	Methodology	Features
[13] Deepak, Raut Sushrut, Shinde Isha Uday, and D. Malathi	K-means, SVM	Stock market prediction

Table 4 shows the previous work done by authors who implemented clustering and classification. K-means and SVM was used in [13] to predict the stock market data. The methodology was tested on Bombay stock exchange (BSE) index data set. An accuracy of 85% was gained.

D. Flow Chart

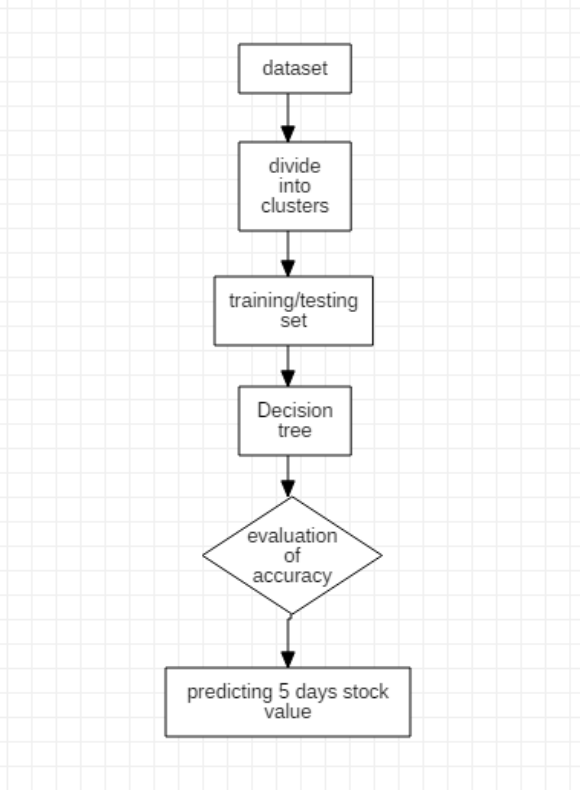


Fig.3. Flowchart which shows the flow of the methodology involved.

The paper includes a data set which contains past stock prices. After importing the data, pre- processing is performed as shown in [17] to omit null or missing values. Fig.3. depicts the flow of methodology. This data is clustered and redirected into training and testing datasets for every cluster formed. Classification is performed on the training data to acquire a model using decision tree classifier. This model is applied on testing data to gain accuracy. In the end, we are predicting the upcoming closing stock values by using forecast functions. For this approach, we are using knowledge gaining algorithms like k-means and decision tree.

V. EXPERIMENTAL RESULTS

Result is calculated on stock data of SBI Company. The data set is acquired from the national stock exchange of India limited [29]. The data resides from January 1, 2000 to February 28, 2019. It includes columns like open, close, high, low, volume and turnover.

A. Clustering

Clustering methodology applied on the dataset yielded three clusters.

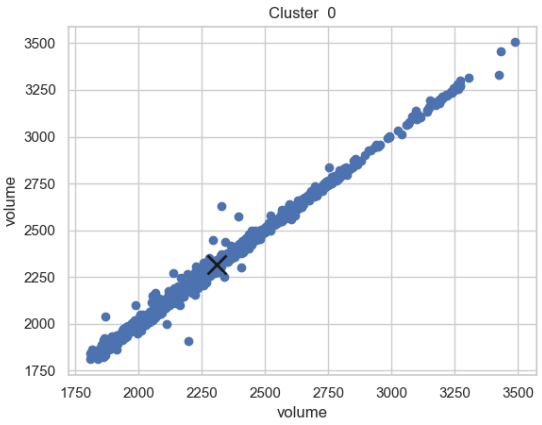


Fig.4. Cluster 0

In Fig.4, the first cluster formed is shown. It shows various number of records that lie between the volumes of 1750 to 3500. Volume is total number of records present. The records are divided into clusters taking into consideration the business, turnover, and number of transactions. There is an inclination in the graph, which denotes that over the years, there is expanded usage of the stock market.

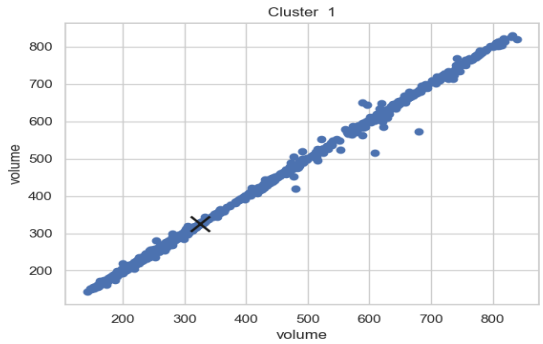


Fig.5. Cluster 1

In Fig.5, the second cluster which is named as cluster 1 is depicted. Here, the number of records is spread across the volume of 200 to 800.

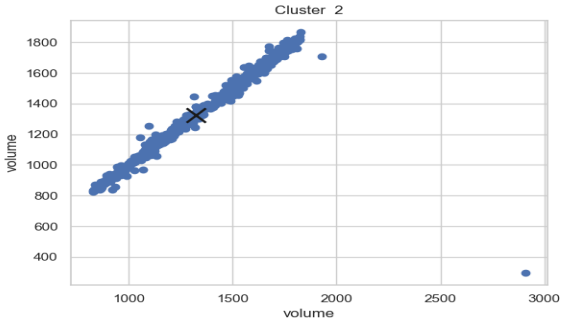


Fig.6. Cluster 2

In Fig.6, the third cluster which is named as cluster 2 is shown. Here, the number of records is spread across the volume of 1000 to 3000.

B. Classification

Using decision tree, closing price values are predicted. The predicted values are compared with original values.

The blue line in the Fig.6, Fig.7 and Fig.8 indicates the original values and the red line indicates predicted values. If the lines are overlapping it implies that the original and predicted values are same. The deviation in the lines present the variation among the values. This graph is plotted for each of the cluster formed. The x-axis and y-axis depict the closing stock value and number of days respectively.

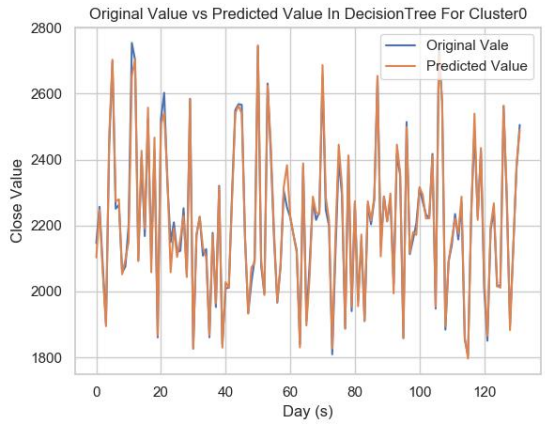


Fig.7. Plotting original and predicted value in decision tree for cluster0

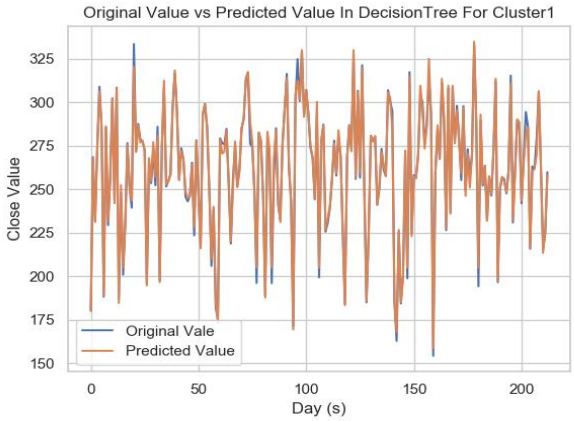


Fig.8. Plotting original and predicted value in decision tree for cluster1

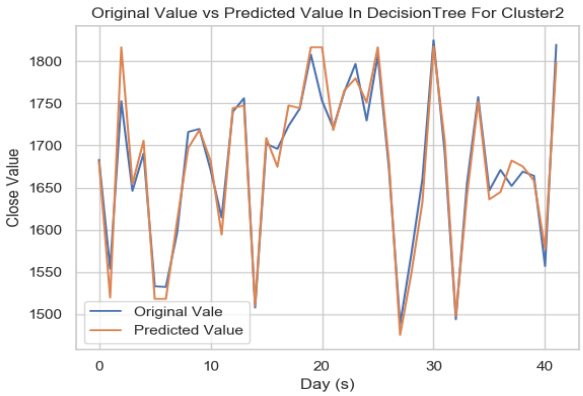


Fig.9. Plotting original and predicted value in decision tree for cluster2

Accuracy: Accuracy is attainted for each cluster based on the performance of the decision tree classifier. The accuracy is predicted by considering the predicted values in place of original values.

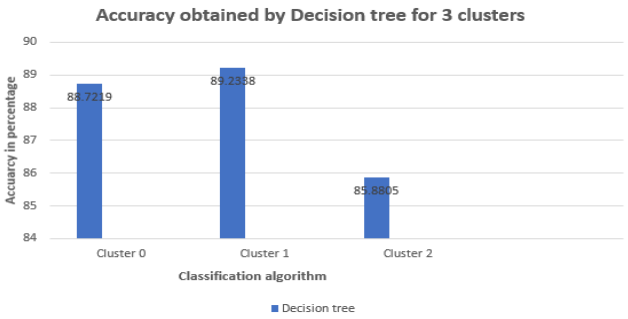


Fig.10. Accuracy predicted on decision tree for each cluster

The accuracy attained by cluster 0 is 88.72%. Cluster 1 gained an accuracy of 89.23% and the accuracy of cluster 2 is 85.88%. The blue bar represents accuracy gained by decision tree.

Comparative study: A comparative analysis is performed using k-means clustering and SVM classifier. In comparison with accuracy obtained by the decision tree classifier, the accuracy gained by using SVM is less. The results generated by using decision tree are more accurate than those achieved by SVM. Hence, the combination of k-means and decision tree classifier is most suited for stock market prediction.

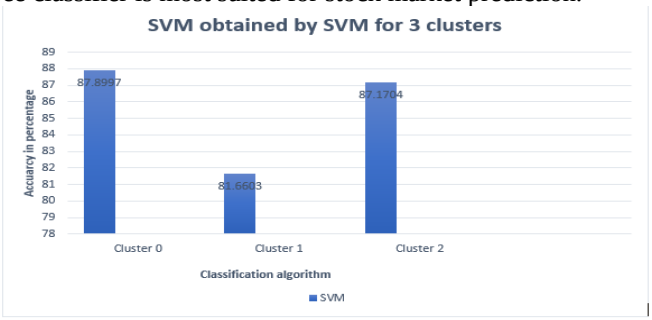


Fig.11. Prediction accuracy for cluster 1

From Fig.11, we can infer that the accuracy attained by cluster 0 is 87.89%. Cluster 1 gained an accuracy of 81.66% and the accuracy of cluster 2 is 82.17%. The blue bar represents accuracy gained by SVM. A clear comparison of Fig.10 and Fig.11 are shown in Fig.12.

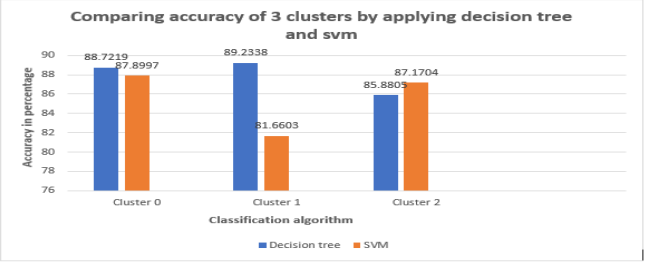


Fig.12. Comparing the accuracy between decision tree and SVM

From Fig.12, we can compare the accuracy of each cluster between decision tree and SVM. The blue bar represents the accuracy of decision tree and the orange bar denotes that of SVM.

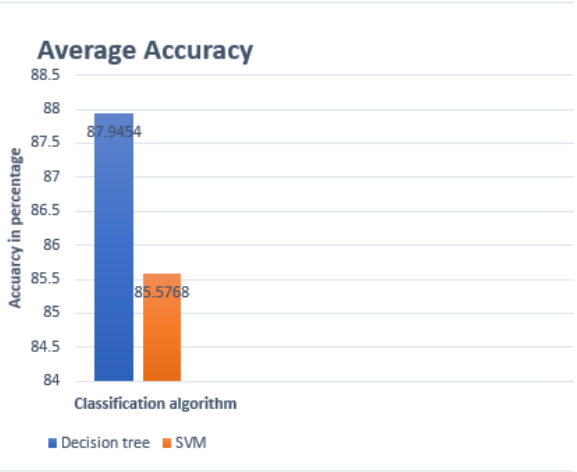


Fig.13. Comparing the average accuracy between decision tree and SVM

From Fig.13, we can see that the average accuracy of decision tree is 87.94% and the average accuracy gained by SVM is 85.57%. The blue bar represents decision tree and the orange bar denotes SVM.

We can clearly state that decision tree has outperformed SVM. Hence, the combination of k- means with decision tree is the best solution for predicting stock values.

Prediction: The upcoming five days closing stock values are predicted. The x-axis and y-axis show the predicted value and the number of days respectively.

Fig.14. Upcoming 5 days closing stock values

User interface: This project includes a UI. The stock name is provided in the name of the stock field. Browse button loads the desired data set. The clustering option forms the clusters from the data set. Classification displays the original and predicted value graph along with accuracy of decision tree. The forecast predicts the upcoming five days closing stock values.

Fig.15. User interface

VI. CONCLUSIONS

This paper presents a predictive analysis model using clustering and classification techniques on stock market data. From the above results, it is noticeable that accuracy up to 87.67% is achieved. The upcoming closing values may be used to track the consistency of the above company. This paper might help stock market investors in later transactions. The implementation of the paper can be expanded further by applying the algorithms on other datasets from different companies. It can also be expanded by using different algorithms.

REFERENCES

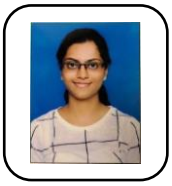
1. Khedr, Ayman E., N.Yaseen. "Predicting stock market behavior using data mining technique and news sentiment analysis." International Journal of Intelligent Systems and Applications 9.7 (2017): 22.
2. Balaji, S. Naveen, P. Victer Paul, and R. Saravanan. "Survey on sentiment analysis based stock prediction using big data analytics." 2017 Innovations in Power and Advanced Computing Technologies (i-PACT). IEEE, 2017.
3. Coyne, Scott, Praveen Madiraju, and Joseph Coelho. "Forecasting Stock Prices Using Social Media Analysis." 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). IEEE, 2017.
4. Hegazy, Osman, Omar S. Soliman, and Mustafa Abdul Salam. "A machine learning model for stock market prediction." arXiv preprint arXiv:1402.7351 (2014).
5. Ferrandez, Sergio Mourelo, et al. "Optimization of a truck-drone in tandem delivery network using k-means and genetic algorithm." Journal of Industrial Engineering and Management (JIEM) 9.2 (2016): 374-388.
6. Jain, Vineet, et al. "Crime Prediction using K-means Algorithm.
7. "Ghorbani, Ali, and Sara Farzai. "Fraud detection in automobile insurance using a data mining based approach." Int. J. Mechatron. Electr. Comput. Technol 8.27 (2018): 3764- 3771.
8. Zulfadhilah, Muhammad, Yudi Prayudi, and Imam Riadi. "Cyber Profiling using Log Analysis and K-Means Clustering." International Journal of Advanced Computer Science and Applications 7.7 (2016).
9. Bhumika, Prof Sukhjit Singh Sehra, and Prof Anand Nayyar. "A review paper on algorithms used for text classification." International Journal of Application or Innovation in Engineering & Management 3.2 (2013): 90-99.
10. Kabakchieva, Dorina. "Student performance prediction by using data mining classification algorithms." International Journal of Computer Science and Management Research 1.4 (2012): 686-690.
11. Saa, Amjad Abu. "Educational data mining & students' performance prediction." International Journal of Advanced Computer Science and Applications 7.5 (2016): 212-220.
12. Ali, Mohamoud, and Yugyung Lee. "CRM Sales Prediction Using Continuous Time- Evolving Classification." Thirty-Second AAAI Conference on Artificial Intelligence. 2018.

13. Deepak, Raut Sushrut, Shinde Isha Uday, and D. Malathi. "MACHINE LEARNING APPROACH IN STOCK MARKET PREDICTION." International Journal of Pure and Applied Mathematics 115.8 (2017): 71-77.
14. https://www.quandl.com
15. https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/
16. https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling- scratch-in-python/
17. https://hackernoon.com/what-steps-should-one-take-while-doing-data-preprocessing- 502c993e1caa
18. K. Prasanna Lakshmi and Dr. C. R. K. Reddy, A survey on different trends in data streams, in Proc. 2010 IEEE International Conference on Networking and Information Technology, 2010, pp. 451–455.
19. K.Prasanna Lakshmi, Dr.C.R.K.Reddy, "Compact Tree for Associative Classification of Data Stream Mining", (IJCSI International Journal of Computer Science Issues, Vol 9, Issue 2, No 2, March 2012), ISSN(online) : 1694-0814
20. Lakshmi, K. Prasanna, and C. R. K. Reddy. "Efficient classifier generation over stream sliding window using associative classification approach." *International Journal of Computer Applications* 115.22 (2015).
21. Kompalli, Prasanna Lakshmi, and Ramesh Kumar Cherku. "Efficient mining of data streams using associative classification approach." *International Journal of Software Engineering and Knowledge Engineering* 25.03 (2015): 605-631.
22. Lakshmi, K. Prasanna, and C. R. K. Reddy. "Fast rule-based heart disease prediction using associative classification mining." *2015 International Conference on Computer, Communication and Control (IC4)*. IEEE, 2015.
23. K. P. Lakshmi and C. R. K. Reddy, "Fast Rule-Based Prediction of Data Streams Using Associative Classification Mining," *2015 5th International Conference on IT Convergence and Security (ICITCS)*, Kuala Lumpur, 2015, pp. 1-5. doi: 10.1109/ICITCS.2015.7292983
24. Lakshmi, K. Prasanna, and C. Ramesh Kumar Reddy. "Efficient Classifier over Stream Sliding Window using Associative Classification." *Proceedings of the International Conference on Data Mining (DMIN)*. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2015.
25. Lakshmi, K. Prasanna et al. "Sentiment Analysis and Prediction using Text Mining." Indian journal of science and technology 10 (2017): 1-11.
26. https://www.inflibnet.ac.in/, https://baadalsg.inflibnet.ac.in
27. V Kishore Ayyadevara. "Pro Machine Learning Algorithms", Springer Nature, 2018
28. Morshed, M.N.. "Computational approach to the identification of novel Aurora-A inhibitors", Bioorganic & Medicinal Chemistry, 20110115
29. https://www.nseindia.com/products/content/equities/indices/historical_index_data.htm

AUTHORS PROFILE

K Prasanna Lakshmi is working as a Professor and Head of Department in Information Technology department, Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad. She received the Bachelor's Degree in Mechanical Engineering (2000), Master's Degree in Computer Science (2002) and Ph.D in Computer Science (2016). He has total 18 years teaching and 8 years research experience. Her research interests include data mining, data stream mining, web mining, Big Data Analysts.

N. V. Ganapathi Raju is working as a Professor in I.T. department, and Associate Dean Alumni Affairs at GRIET, Hyderabad. He received Ph.D from JNTUK, Kakinada. He received M.Tech (C.S.T.) from Andhra University. He has total 18 years teaching and 8 years research experience. His research interests include Information Retrieval and Natural Language Processing, Data Science and Machine Learning. He got UGC minor project grant MRP-4590/14 (SERO/UGC) in March 2014. He published research articles in various International journals. He is Webmaster for GRIET Website. Incharge for TASK, Spoken Tutorials-IITBombay, TEP admin for ISB for GRIET. He started Oracle Academy, CISCO Academy, Red hat Academy, DELL data science academy in GRIET for uplifting skill and knowledge of students.



Anusha Buddaraju is studying Information Technology in Gokaraju Rangaraju Institute of Engineering and Technology. Her research interests include data mining, data stream mining, web mining, Big Data Analysts.

Lahari Devaraju is studying Information Technology in Gokaraju Rangaraju Institute of Engineering and Technology. Her research interests include data mining, data stream mining, web mining, Big Data Analysts.