Gazi University Journal of Science
GU J Sci
28(3):395-403(2015)

Stock Market Prediction Using Clustering with Meta-Heuristic Approaches

S. PRASANNA[1,♠], Ezhil MARAN[2]

[1] *Information Technology and Engineering, Vellore Institute of Technology, Vellore*
[2] *Advanced Sciences, Vellore Institute of Technology, Vellore.*

Received: 11/11/2014 Revised: 15/03/2015 Accepted: 30/05/2015

ABSTRACT

Various examinations are performed to predict the stock values, yet not many points at assessing the predictability of the direction of stock index movement. Stock market prediction with data mining method is a standout amongst the most paramount issues to be researched and it is one of the interesting issues of stock market research over several decades. The approach of advanced data mining tools and refined database innovations has empowered specialists to handle the immense measure of data created by the dynamic stock market. Data mining strategies have been utilized to reveal hidden patterns and predict future patterns and practices in financial markets to help financial investors make qualitative choice. In this paper, the consistency of stock index movement of the well-known Indian Stock Market indices NSE-NIFTY are examined with the assistance of famous data mining strategies known as Clustering. Clustering is the methodology of grouping the alike indices into clusters. It likewise audits three of the meta-heuristics clustering algorithms: PSO-K-Means, Bat Algorithm, and firefly Algorithm. These strategies are implemented and tested against a Stock Market Index Movement Dataset. The performance of the aforementioned procedures is compared based on "integrity of clustering" assessment measures. The investigation is used to the NSE-NIFTY and BSE-NIFTY for the period from January 2011 to April 2014.

Keywords: Data Mining, Clustering, PSO, BAT Algorithm, Firefly Algorithm, Prediction, Stock Index Movement.

1. INTRODUCTION

The rapid development of data accumulations in the science and business applications and in addition they need to examine and concentrate valuable information from this data prompts another era of tools and strategies gathered under the term data mining. Data mining is the important extraction of undeniable, obscure, and possibly significant data from the data and thusly it is climbing as the invaluable information finding procedure is an urgent part in the stock market dissection.

As the dynamic stock market leaves a trail of huge measure of data, securing and examining terabytes of data is constantly challenging for the experts. This could be accomplished by clustering approach in data mining.

Financial markets such as stock market are generating constantly great volume of information needed to analysis and to produce any predicting pattern in any time.Forecasting stocks because of its significance and notoriety among the masses furthermore little and extensive organizations because of monetary profits and its generally safe is a developing theme in exploration. Notwithstanding the danger of falling an excessive amount of quality every offer because of business changes once in a while happens, however once more, the danger is there. These changes which impact on stock value and exchanging volume have a few troubles in foreseeing. The changes impact on the conduct of individuals regarding capital investment funds or venture, the stock cost and the build or reduction of danger for financial specialists. Consequently as a rule, anticipating the stock exchange conduct through systems and different

♠Corresponding author, e-mail: prasannaphd14@gmail.com

strategies is a helpful apparatus to aid financial specialists to act with more prominent sureness and going out on a limb and unpredictability of a venture into thought and know when to purchase the least expensive cost and when to offer to most astounding cost

Clustering is one of the paramount tasks in data mining. Clustering is considered as an intriguing methodology for discovering similarities in data and putting comparable data into clusters [7]. Clustering divides a dataset into several clusters such that the likeness inside a cluster is bigger than that among clusters [4]. Stock market development index clustering is the procedure of collection similar index into the same group and is critical for the accomplishment of Stock Market Prediction. The objective of clustering indexes is to anticipate which months give exceptional yield. On the index clustering, comparative index is clustered focused around returns connected with opening and closing value [8]. Clustering algorithms are utilized widely to compose and arrange data, as well as helpful for data layering and model development. By discovering similarities in data, one can illustrate similar data with fewer images for instance. Additionally, in the event that we can discover clusters of data, we can develop a model of the issue based on those clusters.

Each clustering algorithms are solely capable of focusing on Stock Index Movement data in Indian Stock Market. This focus brings better and more detailed results to the same parts. Meanwhile, in analysing other parts, due to the lack of clustering analyses, it brings challenges to them.

So, each algorithm is capable of doing detailed analyses of some parts of Stock Market Index movement data. To provide comprehensive results and clustering analyses, it must be used several integrated and clustering algorithms. We, in this paper, investigate different types of methods and clustering algorithms. Finally, by using K-means, PSO-K-Means, Firefly K-Means and BAT K-Means for clustering the stock market index movement data, we made clustering vie Matlab software.

The remaining part of this paper is organized as follows. Section 2 reviews relevant literature. Section 3 explains the methodology employed in this study. Section 4 presents the results and discussion. Section 5 concludes this paper.

2. LITERATURE SURVEY

Two measurable operations commonly used in stock market predictions are classification and clustering however the most important part is clustering stock market development indexes. Clustering issues emerge in numerous diverse applications, for example, data compression, data mining and knowledge discovery, pattern recognition and pattern classification keeping in mind the end goal to grouping similar indexes in one cluster so that indexes inside the same clusters are like one another and unique in relation to indexes in different clusters. Clustering is the unsupervised classification of samples (perceptions, data things, or peculiarity vectors) into clusters. The clustering issue has been tended to in numerous settings and via scientists in numerous controls; this reflects its wide request and value as one of the steps in exploratory data analysis [6] . The expression "clustering" is utilized within a few examination groups to portray systems for clustering of unlabeled data.

[1] had exhibited a methodology for identifying clusters of similar indexes and connections among them. The authors apply clustering techniques to discover diverse classifications of stock market indexes, introducing three methods for computing returns inside a graph: convergence, Jaccard and a more unpredictable approach that considers extra distributional measures of indexes in a vector space representation [1]. A customized perspective can overcome vagueness and assignment of characteristic index, providing returns indexes, opening and closing value that relate all the more nearly to their expectation. Particularly, we inspect unsupervised clustering routines focused around meta heuristics for concentrating shared traits between indexes, and utilize the discovered clusters, as intermediate between a returns and values with a specific end goal for opening and closing.

Meta heuristics is an alternate rational system for critical thinking utilizing current methodologies. An advancement in Meta heuristic exploration is the examination of hybridization of Meta heuristic methodologies, for example, PSO, BAT and Firefly and so on. This may additionally come about out in regularly discovering a fine solution with less computational exertion than any viable techniques. Meta heuristic procedures are developed as practical tools and alternatives to more conventional clustering strategies. Among the numerous meta-heuristics procedures, clustering with PSO methods has discovered accomplishment in tackling clustering issues. It is suitable for clustering complex and non-separable datasets. Clustering is one of the broadly utilized data mining systems for stock market data investigation [10]. An extensive number of meta-heuristic algorithms have been developed for clustering. Meta heuristic algorithms have a few deficiencies, for example, the gradualness of their convergence and their sensitivity to instate values. The clustering algorithms arrange stock market index into clusters and the practically related indexes are gathered together in a proficient way [2] .

Ahmed introduced a survey on PSO algorithm and its variants to clustering high-dimensional data [3]. [12] proposed a co-operative clustering algorithm focused around PSO and K-means and he additionally contrasted that algorithm with PSO, PSO with Contraction Factor (CF-PSO) and K-means algorithms[12]. [9] planned to incorporate Particle Swarm Optimization Algorithm (PSOA) with K-means to cluster data and he had demonstrated that PSOA might be utilized to discover the centroids of a client specified number of clusters [9]. [13] proposed an effective hybrid approach based on PSO, ACO and K-means for cluster analysis [13]. Yau-King Lam proposed PSO-based K-Means clustering with upgraded cluster matching of gene interpretation data [15].

3. PHASES OF CLUSTERING STOCK MARKET DATA

Stock Market Data clustering methodology comprises the following steps.

a) Data Extraction
b) Clustering
c) Pattern Analysis

3.1. Data Extraction

This paper examines the daily change of the closing values of NSE-NIFTY and BSE-NIFTY based on the following predictors: Open price, High price, Low price and Close price. NSE-NIFTY and BSE-NIFTY values are obtained from the NSE and BSE websites respectively, for the period from Jan'2011 in April 2014 with a sample of 850 trading days. The data are divided into two sub-samples in the split up of 80:20 where the in-sample or training data spans from Jan' 2011 to Sep' 2013 with 680 trading days and the data for the remaining period from Oct 2013 to April 2014 with 170 trading days are used for out-of sample or test data.

3.2. Clustering

Clustering Stock data is the process of grouping the similar Index values into the same cluster based on return value by applying clustering techniques. Index in the same cluster has been often similar returns. This study compares three clustering techniques: PSO-K-Means, Firefly and BAT algorithm. These techniques are implemented and tested against our Stock Market dataset.

3.2.1. Objective function for clustering

The indeed motto of the clustering algorithm is to hold in the intra-cluster remoteness and also to maximize the inter-cluster distance based on the distance measures, here the aim function is referred as validation measure. To demonstrate it briefly, we have preferred Mean square quantization error (MSQE) and sum of intra cluster distances (SICD) for comparative analysis. MSQE and SICD are briefly explained in Experimental Results

3.2.2. PSO-K-Means clustering

Particle Swarm Optimization (PSO) algorithm is one of the swarm intelligence methods and evolutionary optimization techniques are applied for clustering. PSO is a population-based, globalized search algorithm that utilizes the rules of the social behavior of the swarm and it is an effective, simple, and an efficient global optimization algorithm that can solve discontinuous, multimodal, and non-convex problems . It is computationally efficient and easier to implement when compared with other mathematical algorithms and evolutionary algorithms.

In PSO, N particles are moving around in the D dimensional search spaces. Each particle moves towards the nearest region. Each particle communicates with some other particle and is exaggerated by the best centroid point found by any member of its current centroid value pi. The vector pi for that best neighbour and is denoted by pg. Initialize the particle's location best known position to its initial position: pi ← xi. Then correspondingly update the particles or the position of the centroid value position and their velocity in equation 1 to recognize the best planetary position in equation 2 or of the best centroid value to group the information. These steps are iterated until a termination criterion is satisfied. Ultimately after finding the global best position, best value of cluster centroid is obtained.

$$v_{id} = w * v_{id} + C_1 * rand1 * (P_{id} - x_{id}) + C_2 * rand2 * (P_{gd} - x_{id}) \quad (1)$$

$$x_{id} = x_{id} + v_{id} \quad (2)$$

Where, v_{id} : velocity of particle , x_{id} : current position of particle, W : weighting function,

$w = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} * iter$, w_{min}, w_{max} :

initial and final weight, iter: current iteration , itermax : maximum iteration , c1&c2 : determine the relative influence of the social and cognitive, components p_{id}: pbest of particle i, p_{gd} : gbest of the group. The personal best position of particle is calculated as follows

$$p_{id}(t+1) = \begin{cases} p_{id}(t) & \text{if } f(X_{id}(t+1)) \geq f(p_{id}(t)) \\ X_{id}(t) & \text{if } f(X_{id}(t+1)) < f(p_{id}(t)) \end{cases} \quad (3)$$

The particle to be drawn toward the best particle in the swarm is the global best position of each particle. At the start, an initial position of the particle is considered as the personal best and the global best can be identified with minimum fitness function value.

Algorithm 1: PSO-K-Means Clustering (Proposed Algorithm)

Input: D set of N Indexs, **K** -number of clusters

Output: K overlapping clusters of Indexs from D with associated membership value

Step 1: Assign each vector in the data set to the closest centroid vector

Step 2: Calculate the fitness value for each vector and update the velocity and Particle position, using equations (1) and (2) and generate the next solutions

Step 3: Repeat steps (2) and (3) by one of the following termination conditions is satisfied.

(a) The maximum number of iterations is exceeded
or
(b) The average change in centroid vectors between iterations is less than a predefined value

3.2.3. BAT K-Means

Bat algorithm is swarm intelligence based algorithm which is worked on the echolocation of bats. This algorithm was developed by [14]. It is a new metaheuristic algorithm for solving the many optimization problems. Bats are based on the echolocation behavior of bats. They can find their prey food and also they can know the different type of insects even in a complete darkness. Since we know that micro bats are the insectivore who have the quality of fascinating. These bats use a type of sonar namely as echolocation. They emit a loud sound pulse and detect an echo that is coming back from their surrounding objects. Their pulse variation in properties and will be depend on the species. Their loudness also varies. When they are searching for their prey, their loudness is loudest if they are far away from the prey and they will become slow when they are nearer to the prey. Now for emission and detection of echo which are generated by them, they use time delay. And this time delay is between their two ears and the loudness variation of echoes [11].

The following formulae are used for their position xi and velocities vi when they are updated:

$$f_i = f_{min} + (f_{max} - f_{min})\beta \quad (4)$$
$$vti = vt\text{-}1i + (xti - x^*)fi \quad (5)$$
$$xti = xt\text{-}1i + vti\ , \quad (6)$$

We realize that loudness is diminished when bat thought that it was' prey or food, however the rate of pulse emanation expands. For effortlessness, we utilize loudness A0 = 1 and minimum Amin = 0 that implies a bat discovered their prey and they quit making a sound. Where $\beta \varepsilon [0, 1]$ is an random vector drawn from an uniform distribution. Here x* is the current global best location (solution) which is located after comparing all the solutions among all the n bats. Generally speaking, contingent upon the domain size of the issue of investment interest, the frequency f is assigned to $f_{min} = 0$ and $f_{max} = 100$ in practical implementation. At first, each bat randomly given a frequency which is drawn consistently from [fmin, fmax]. For the local search part, once an answer is chosen among the current best results, another answer for each one bat is produced generally utilizing random walk.

$$x_{new} = x_{old} + \varepsilon At \quad (7)$$

The update of the velocities and positions of bats has some comparability to the methodology in the standard particle swarm optimization as though basically controls the pace and scope of the movement of the swarming particles. To some degree, BA could be considered as an adjusted mix of the standard particle swarm optimization and the intensive local search controlled by the loudness and pulse rate. Besides, the loudness Ai and the rate ri of pulse outflow overhaul in like manner as the iterations proceed

Algorithm 2: BAT-K-Means Clustering (Proposed Algorithm)

Input: D set of N Indexs, K -number of clusters

Output: K overlapping clusters of Indexes from D with associated membership value

Step 1: initialize objective function, population , velocity, maximum iteration , time , pulse rate and loudness
Step 2: Assign each vector in the data set to the closest centroid vector

Step 2: Calculate the fitness value for vector and update the velocity and position, using equations (4) and (5) and generate the next solutions

Step 3: Repeat steps (2) and (3) until one of the following termination conditions is satisfied.

(a) The maximum number of iterations is exceeded
or
(b) The average change in centroid vectors between iterations is less than a predefined value

3.2.4. FIREFLY K-Means

Most of fireflies produced short and rhythmic flashes and have different flashing behavior. Fireflies use these flashes for communication and attracting the potential prey. YANG used this behavior of fireflies and introduced Firefly Algorithm in 2008. In Firefly algorithm, there are three idealized rules: 1) All fireflies are unisex. So, one firefly will be attracted to other fireflies regardless of their sex; 2) Attractiveness is proportional to their brightness. Thus, for any two flashing fireflies, the less brighter one will move towards the brighter one. The attractiveness is proportional to the brightness and they both decrease as their distance increases. If there is no brighter one than a particular firefly, it will move randomly; 3) The brightness of a firefly is determined by the landscape of the objective function. For a maximization problem, the brightness can simply be proportional to the value of the objective function [5].

Algorithm 3: FIREFLY-K-Means Clustering (Proposed Algorithm)

Input: D set of N index, K -number of clusters

Output: K overlapping clusters of tags from D with associated membership value

Step 1: initialize objective function , population , velocity, maximum iteration , time , light indensity and loudness

Step 2: Assign each vector in the data set to the closest centroid vector

Step 2: Calculate the fitness value for vector and update the light intensity and rank the fireflies (4) and (5) and generate the next solutions

Step 3: Repeat steps (2) and (3) until one of the following termination conditions is satisfied.

(a) The maximum number of iterations is exceeded
or
(b) The average change in centroid vectors between iterations is less than a predefined value

The firefly algorithm, there are two important issues, including variation of light intensity and the formulation of the attractiveness. For simplicity, it's assumed that the attractiveness of a firefly is determined by its brightness hitch associated with the objective function of the optimization problem. Since a firefly's attractiveness is proportional to the light intensity seen by adjacent fireflies, we can now formulate the attractiveness of a firefly by:

$$\beta(r) = \beta_0 e^{-\gamma r^2} \quad (8)$$

where, β_0 is the attractiveness at $r = 0$ and γ is the light absorption coefficient at the source. It should be noted that the *ri,j* which is described by equation 2, is the Cartesian distance between any two fireflies *i* and *j* at xi and xj, where, xi and xj are the spatial coordinate of the fireflies *i* and *j*, respectively.

The movement of a Firefly *i*, which is attracted to another more attractive Firefly *j* is determined by:

$$X_i = x_i + \beta_0 e^{-\gamma_{ij}{}^2}\left(x_j - x_i\right) + \alpha\left(rand - \frac{1}{2}\right) \quad (9)$$

where, the second term is the attraction while the third term is randomization, including randomization parameter α and the random number generator *rand* which its numbers are uniformly distributed in interval [0, 1]. For the most cases of implementations, β0 = 1 and α ε [0, 1]. The parameter γ characterizes the variation of the attractiveness and its value is important to determine the speed of the convergence and how the FA behaves. In the most applications, it typically varies from 0.01 to 100.

4. PATTERN ANALYSIS

In this segment, the validity measures such as MSQE and SICD are explained. The Validity Measure is an objective Function of the Clustering algorithms. Clustering algorithms which give the minimum MSQE and SICD value is the algorithm which provides better performance than others.

Mean square quantization error (MSQE)

$$f(X, C) = \sum_{i=1}^{N} Min\left\{\{||X_i - C_l||^2| \quad where \quad l = 1, \dots, K\}\right\} \quad (10)$$

where $||X_i - C_l||^2$ is a Euclidean distance measure between the ith data point of the tag xi and the jth cluster center cl, 1 < i< n, 1 < j< K and is an indicator of the distance of the n tags from their respective cluster centroids [13] and K is the number of clusters.

Sum of intra cluster distance (SICD)

$$J(C_1, C_2, \dots, C_k) = \sum_{i=1}^{k}\left(\sum_{x_j \in C_i}\left\|Z_i - X_j\right\|\right) \quad (11)$$

The Euclidean distance between each data vector in a cluster and the centroid of that cluster is calculated and summed up. Here K is the number of clusters, Z_i represents cluster centroids, X_j is the data vector [12].

5. EXPERIMENTAL ANALYSIS

The experimental data set is collected from NSE-NIFTY, which is a popular stock market index movement data. The information about the data sets contains names of dataset, the number of objects and number of Attributes, which are given in Table 1. In clustering stock market index data, the open, close, high, low, return values were treated as attributes and every month are treated as objects.

Table 1. Dataset Description

S.No	Dataset	Objects	Attributes	Url
1	NSE-NIFTY	850	6	http://www.nseindia.com/products/content/equities/indices/historical_index_data.htm
2	BSE-SENSEX	850	4	http://www.bseindia.com/products/content/equities/indices/historical_index_data.htm

Performance Analysis of clustering algorithms

In this section the performance of Benchmark algorithm K-Means are compared with the meta heuristic algorithm such as PSO-K-Means, BAT-K-Means Algorithm and Firefly-K-Means Algorithm based on MSQE and SICD validity measures.

K-Means: Benchmark algorithm

The performance of K-Means Benchmark algorithm is analyzed based on MSQE and SICD validity measures and the results are shown in Table 2.

Table 2. Performance Analysis of K-Means Algorithm

K-Means Clustering	NSE-NIFTY		BSE-NIFTY	
	MSQE	SICD	MSQE	SICD
K= 2-clusters	18.63	88.96	14.52	74.98
K= 5-clusters	15.95	84.27	11.67	69.27
K= 10-clusters	12.49	83.01	7.98	65.79

Fig. 1 depicts the performance of TRS-K-Means for stock market data sets discussed here based on a MSQE validity measure .

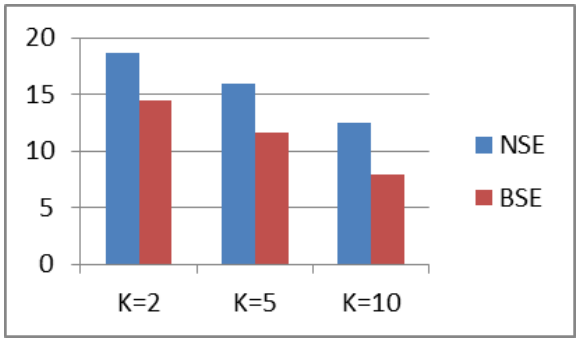


Fig. 1. Performance of K-Means based on MSQE index

Fig. 2 depicts the performance of TRS-K-Means for stock market data sets discussed here based on SICD validity measure.

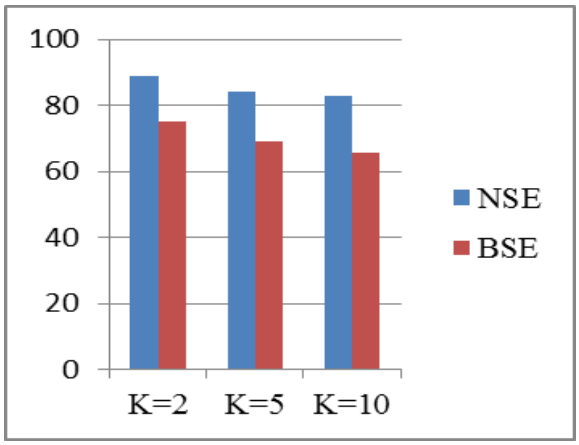


Fig. 2. Performance of K-Means based on SICD index

PSO-K-Means Clustering algorithm

The performance of PSO-K-Means Clustering algorithm is analyzed based on MSQE and SICD validity measures and the results are shown in Table 3.

Table 3. Performance Analysis of K-Means Algorithm

K-Means Clustering	NSE-NIFTY		BSE-NIFTY	
	MSQE	**SICD**	**MSQE**	**SICD**
K= 2-clusters	17.42	84.36	13.57	72.37
K= 5-clusters	15.18	81.57	10.25	68.52
K= 10-clusters	11.98	79.64	7.11	65.42

Fig. 3 depicts the performance of PSO-K-Means for stock market data sets discussed here based on MSQE validity measure.

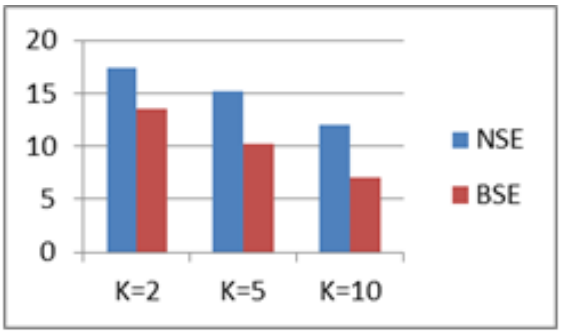


Fig. 3. Performance of PSO-K-Means based on MSQE index

Fig. 4 depicts the performance of PSO-K-Means for stock market data sets discussed here based on SICD validity measure**.**

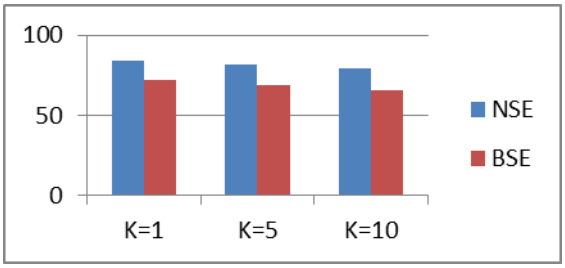


Fig. 4. Performance of PSO-K-Means based on SICD index

BAT-K-Means Clustering algorithm

The performance of BAT-K-Means Clustering algorithm is analyzed based on MSQE and SICD validity measures and the results are shown in Table 4.

Table 4. Performance Analysis of K-Means Algorithm

K-Means Clustering	NSE-NIFTY		BSE-NIFTY	
	MSQE	**SICD**	**MSQE**	**SICD**
K= 2-clusters	16.27	83.57	12.47	70.64
K= 5-clusters	14.52	79.21	9.67	67.41
K= 10-clusters	10.69	77.63	6.58	64.82

Fig. 5 depicts the performance of BAT-K-Means for stock market data sets discussed here based on a MSQE validity measure**.**

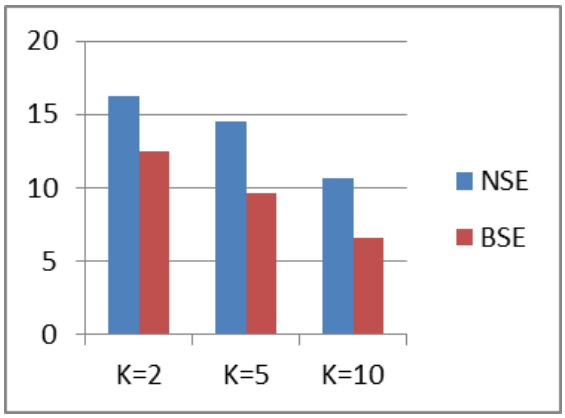
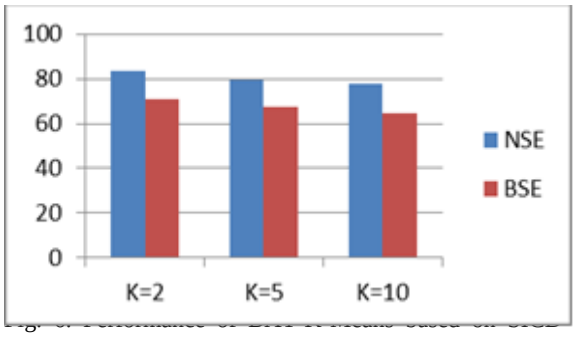


Fig. 5. Performance of BAT-K-Means based on MSQE index

Fig. 6 depicts the performance of BAT-K-Means for stock market data sets discussed here based on SICD validity measure**.**



index

Firefly-K-Means Clustering algorithm

The performance of firefly-K-Means Clustering algorithm is analyzed based on MSQE and SICD validity measures and the results are shown in Table 5.

Table 5. Performance Analysis of K-Means Algorithm

K-Means Clustering	NSE-NIFTY		BSE-NIFTY	
	MSQE	**SICD**	**MSQE**	**SICD**
K= 2-clusters	15.07	81.27	11.12	69.21
K= 5-clusters	13.62	78.29	8.47	66.78
K= 10-clusters	9.48	76.31	6.09	63.92

Fig. 7 depicts the performance of firefly-K-Means for various stock market data sets discussed here based on a MSQE validity measure**.**

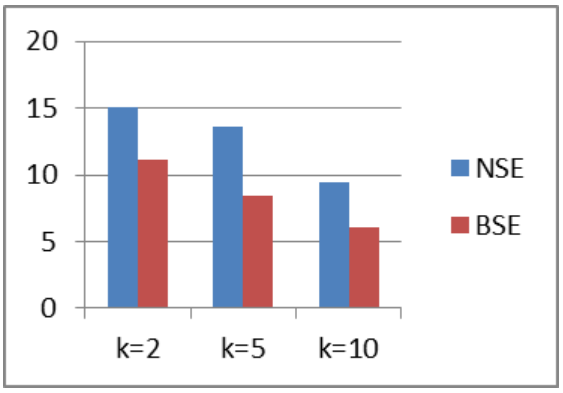


Fig. 7. Performance of BAT-K-Means based on MSQE index

Fig. 8 depicts the performance of firefly-K-Means for various stock market data sets discussed here based on SICD validity measure**.**

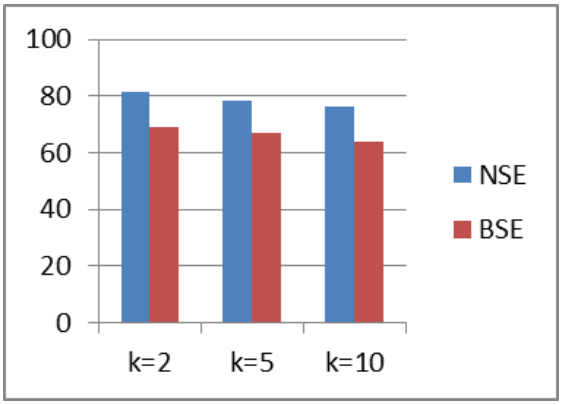


Fig 8. Performance of BAT-K-Means based on SICD index

Comparative Analysis

In this section, the comparative analysis of K-Means Clustering, PSO-K-Means clustering, BAT-K-Means, and Firefly-K-Means clustering for stock market Data using validity measures such as MSQE and SICD measures are provided. This section attempts to carry out comparative analysis of proposed algorithms K-Means clustering, PSO-K-Means, BAT-K-Means and Firefly-K-Means clustering, in terms of the quality of the clusters.

Comparative analysis based on MSQE

The Table 6 shows the comparative analysis of clustering algorithms based on MSQE validity measures for stock market datasets. The experimental results showed that Firefly-K-Means clustering algorithm had shown better results than other clustering algorithms such as K-Means, PSO-K-Means and BAT-K-Means.

Table 6. Comparative analysis based on MSQE

K-Means Clustering	NSE-NIFTY				BSE-NIFTY			
	K-Means	**PSO**	**BAT**	**FIREFLY**	**K-Means**	**PSO**	**BAT**	**FIREFLY**
K= 2-clusters	18.63	17.42	16.27	15.07	14.52	13.57	12.47	11.12
K= 5-clusters	15.95	15.18	14.52	13.62	11.67	10.25	9.67	8.47
K= 10-clusters	12.49	11.98	10.69	9.48	7.98	7.11	6.58	6.09

Comparative analysis based on SICD

The Table 7 shows the comparative analysis of clustering algorithms based on SICD validity measure for stock market datasets. The experimental results showed that Firefly-K-Means clustering algorithm had shown better results than other clustering algorithms such as K-Means, PSO-K-Means and BAT-K-Means.

Table 7. Comparative analysis based on SICD

K-Means Clustering	NSE-NIFTY				BSE-NIFTY			
	K-Means	**PSO**	**BAT**	**FIREFLY**	**K-Means**	**PSO**	**BAT**	**FIREFLY**
K= 2-clusters	88.96	84.36	83.57	81.27	74.98	72.37	70.64	69.21
K= 5-clusters	84.27	81.57	79.21	78.29	69.27	68.52	67.41	66.78
K= 10-clusters	83.01	79.64	77.63	76.31	65.79	65.42	64.82	63.92

Interpretations of Result

Financial markets are highly volatile and generate huge amounts of data on a day to day basis. The present study applied the popular data mining technique for the task of prediction and clustering of the stock index values of NSE-NIFTY and BSE- NIFTY. The goal of stock market data clustering is to predict the period of highly returned investments. In general, we applied the clustering algorithm to stock market data and showed the results based on the stock market index (high return, low return, high risk and low risk). The clustering algorithms were successfully applied to cluster stock market data comprising into two distinct clusters based on the similarity of stock market index profiles and prior share market knowledge. Low level return index are clustered in one group and high level returns are clustered into another one as shown in Fig. 9.

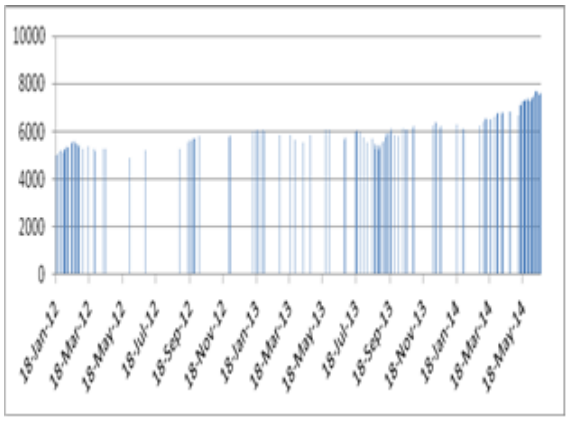


Fig. 9. The higher returns of stock market using clustering the prediction is jan,mar,may, jul,sep , nov of every year gives the high returns and also low risk cluster1

6. CONCLUSION

The clustering problem is a real important problem and has attracted much attention of many researchers. The proposed Clustering algorithm PSO-K-Means, BAT K-Means and FireflyK-Means which are compared with K-Means benchmark algorithm for Stock Market Dataset. The goodness of the clusters is obtained applying the two well-known standards such as MSQE and SICD. The comparative analysis shows that Firefly-K-Means had given the best performance over the other approaches for the Stock Market. This study is a first attempt from the different perspective of personalization and to best of our cognition. This attempt is a new direction in the field of Stock Market Prediction. The results obtained are very encouraging, proving the practical applicability of the stock market. This system adopted with the help of data mining techniques. Data mining techniques are widely applied for knowledge extraction in whole fields. Further studies are recommended in the area of data mining applications in stock markets to gain more useful insights about the predictability of stock markets.

CONFLICT OF INTERESTS

The authors declared that there is no conflict of interests.

REFERENCES

[1] Dattolo A, Eynard D, Mazzola L, “An Integrating Approach To Discover Tag Semantics”, In Proceedings of the 2011 ACM Symposium on Applied Computing, March 21-24, TaiChung, Taiwan (2011).

[2] Dhanalakshmi K, Inbarani HH , "Fuzzy Soft Rough K-Means Clustering Approach For Gene Expression Data", Int. J. of Scientific Engineering and Research 3(10):1-7, (2012).

[3] Esmin A.A, Coelho R.A, Matwin S A, "review on particle swarm optimization algorithm and its variants to clustering high-dimensional data", Artificial Intelligence Review, 1–23, (2013).

[4] Hammouda K A," Comparative Study of Data Clustering Techniques", Technical Report, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada, (2006).

[5] Hassanzadeh, T., Meybodi, M. R. "A New Hybrid Approach for Data Clustering Using Firefly Algorithm and K-means",In: 16thIEEECSI International Symposium on Artificial Intelligence and Signal Processing (AISP), pp. 007 – 011, (2012).

[6] Jain A.K, Murty M.N, and Flyn P.J., "Data Clustering: A Review". ACM Computing Surveys 31(3):264-323, (1999).

[7] Kumar SS, Inbarani HH, "Web 2.0 social bookmark selection for tag clustering", In: Periyar University, (PRIME) Pattern Recognition, Informatics and Medical Engineering (PRIME), Salem, 22-23 Feb 2013, 510- 516, IEEE, (2013a).

[8] Kumar SS, Inbarani HH, "Analysis of mixed C-means clustering approach for brain tumour gene expression data". Int. J. of Data Analysis Techniques and Strategies, 5(2): 214 – 228, (2013b).

[9] Kuo R.J, Wang M.J, Huang T.W, "An application of particle swarm optimization algorithm to clustering analysis". Soft Computing 15(3):533–542, (2011).

[10] Martens D, Baesens B, Fawcett T.. "Editorial survey: swarm intelligence for data mining". Machine Learning 82(1):1–42,(2011).

[11] Monica Sood and Shilpi Bansal, "K-Medoids Clustering Technique using Bat Algorithm", International Journal of Applied Information Systems (IJAIS), 5(8):20-22, (2013).

[12] Neshat M, Yazdi SF, Yazdani D and Sargolzaei M, "A New Cooperative Algorithm Based on PSO and K-Means for Data Clustering". J. of Computer Science 8(2):188-194, (2012).

[13] Taher N, Babak A, "An efficient hybrid approach based on PSO, ACO and k-means for cluster analysis", Appl Soft Comput 10(1):183–197, (2010).

[14] Xin-She Yang, "A new metaheuristic Bat inspired Algorithm". Studies in Computational Intelligence, Springer, (2010) .

[15] Yau K.L, Tsang P.W.M, Leung C.S "PSO-based K-means clustering with enhanced cluster matching for gene expression data", Neural Computing and Application 22(7-8): 1349–1355, (2013).

[16] Lei, Y., He, Z., Zi, Y., "Application of an intelligent classification method to mechanical fault diagnosis", Expert Systems with Applications 36: 9941–9948 (2009).

[17] Toutountzakis, T., Tan, C. K., Mba, D., "Application of acoustic emission to seeded gear fault detection" NDT & E International, 38(1): 27–36 (2005).

[18] Liu, B., Ling, S. F., Gribonval, R., "Bearing failure detection using matching pursuit". NDT&E International, 35: 255–262 (2002).

[19] Yang B. S., Lim D. S., Tan, A. C. C., "VIBEX: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table" Expert Systems with Application, 28(4): 735–742 (2005).

[20] Peng, Z. K., Chu, F. L., "Application of wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography". Mechanical Systems and Signal Processing, 17: 199–221, 2003.

[21] Huang, N. E., Shen, Z., Long, S. R., "The Empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". Proceedings of the Royal Society of London, 454: 903–995. (1998).

[22] Lee, S. K., White, P. R., "Higher-order time-frequency analysis and its application to fault detection in rotating machinery". Mechanical Systems and Signal Processing, 11(4): 637–650 (1997).

[23] Younus, A. MD., Yang, B., "Intelligent fault diagnosis of rotating machinery using infrared thermal image", Expert Systems with Applications 39: 2082–2091 (2012).

[24] Kad, R. S., "IR thermography is a Condition Monitor Technique in industry", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2(3): 988-993 (2013).

[25] Zhang P., Lu B., Habetler T.G., "Active Stator Winding Thermal Protection for Ac Motors", Proceedings of IEEE IAS Pulp and Paper Industry Conference, Alabama, USA, (2009)

[26] Nandi, S., Toliyat, H. A. and Li, X., "Condition monitoring and fault diagnosis of electric motors—a review", IEEE Trans. on energy conversion, 20(4): 719-129, (2005).

[27] Carderock Division Naval Surface Warfare Center, "Handbook of Reliability Prediction Procedures for Mechanical Equipment" (2010).

[28] Barreira, E., de Freitas, V.P., Delgado, J.M.P.Q. and Ramos, N.M.M., "Thermography Applications in the Study of Buildings Hygrothermal Behaviour, Infrared Thermography", Dr. Raghu V Prakash (Ed.), ISBN: 978-953-51-0242-7 (2012).

[29] Stipetic, S., Kovacic, M., Hanic, Z., Vrazic, M., "Measurement of Excitation Winding Temperature on Synchronous Generator in Rotation Using Infrared Thermography", IEEE Transactions on Industrial Electronics, 59 (5): 2288-2298 (2012).

[30] Fantidis J. G., Karakoulidis K., Lazidis G., Potolias C., Bandekas D. V., "The study of the thermal profile of a three-phase motor under different conditions", ARPN Journal of Engineering and Applied Sciences, 8 (11): 892 – 899 (2013).