VOL. 10, NO. 6, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

THREE LAYERED BAR MODEL ARCHITECTURE FOR STOCK MARKET COMPONENT ANALYSIS

S. Sudharsun[1], K. R. Sekar[2], K. S. Ravichandran[2] and J. Sethuraman[2]
[1]Advanced Computing-M.Tech, School of Computing, SASTRA University, India
[2]School of Computing, SASTRA University, India
E-Mail: sekar_kr@cse.sastra.edu

ABSTRACT

Stock market is a place where the companies mobilize money from the people to run their business and in turn benefit people in terms with dividend and profit. Stock market has been an aggregation of both buyers and sellers. As the stock market value increases, the market capital of corresponding firm increases and thus benefiting the investors. Sometimes there may be a chance of downfall in their business which will cause the investors to lose their investment. If the company is not running successfully, the stock price may go down. The reason for investing in the stock market is to earn more profit in a short period of time. Plenty number of stock market shares are available in the existing market. People always find difficulty in choosing a right company shares for their investment. It's a right time for us to make some big analytics, to guide the investors on where to invest their hard earned money. For analysis, umpteen numbers of methodologies are available at our disposal. Two of the methodologies like K-Medoids (Crisp) and Fuzzy K- Means (Soft Computing Techniques) are employed for market analysis. We propose 'BAR Model architecture' for stock market analysis using three layered segments where acronym BAR refers to Budget, Analysis and Result. Budgeting is an entry level to identify the class in the data set. On applying distributed measures on a given data set we get what is called as Budget. After applying the above said methodologies what we get is called Actuals. Both Budget and Actual were compared for variance using Chi-square and ANOVA Test. As the variance we get is very minimal it proves that either methodology is not needed for this kind of application. We come to the conclusion through this paper that the Budget proves to be right. Purity levels of the attributes were measured through Gini Index. This innovative approach will lead us to achieve Predictive Accuracy and Reliability. For the past one decade, this kind of mammoth data collection and analysis have never been reported which has been accomplished in this paper.

Keywords: stock market, K-Medoids (KM), fuzzy K-Means (FK), absolute variance (AV), budget (BG), actuals (AU) and gini index (GI).

INTRODUCTION

Stock market analysis was one of the inevitable conceptual ideologies for investors. For the past twenty years different types of methodologies were employed to know the current and future trends. In the data set plenty number of indisposable attributes like share-holding of promoter group, public share-holding institution, public share-holding non-institution, profit after tax, total assets, pledged shares, dividend yield, dividend, reserves, market capital were considered and companies such as Glenmark Pharma, Asian Paints, ITC, Dabur, Coalgate Palmolive, Hindustan Unilever, Sundaram Fasteners, TCS, Marico, Hindalco, Sesa Sterlite, NMDC, City Union Bank, Jindal Steel and Power Limited, and Sunpharma which come under the face value of 'Rupee One' were taken right from the website www.moneycontrol.com to understand the behaviour of the stock market. The values for these attributes have been taken from the financial month of April 2012 to March 2013. The standalone results of the company for the year ended 2013 was considered in the training set. To find any technical intricacies, K-Medoid methodology was applied to find the average distance between various stocks with the help of R statistical package [18].The crisp and soft computing techniques were used to predict literal niceties and to find the attribute values of the share. Budgeting was calculated by applying distributed measures and normalisation was done in order to find the class attributes like Excellent, Good, Satisfactory, Fair and Poor for supervised learning. Clustering algorithms have gained increasing attention for dissimilarity measurement, dissolution point and isolation robustness [14]. In our referenced paper, evolutionary rough K-Medoid clustering was applied to compare both the results of synthetic as well as real datasets [12]. Clustering was employed for pattern recognition, data mining and machine learning techniques to combine similar objects into different groups [4]. In our dataset the variance was calculated between budgets and actual after implementing methodologies like K-Medoid and K-Fuzzy means. In the same way reference paper proposed K-Medoid and UK-Medoid for resolving accuracy and efficiency factors with less cost [1]. Effective clustering emphasis an initial way to select cluster centres for processing good results [3]. Rat characteristics were identified and differentiated through K-Medoid and Fuzzy K-Means with genetic variants, using sliding window approach these were hypothesized to influence human diseases [2, 36, 11]. Less computation time and complexity were needed for clustering a dataset, ordering needs $0(k\ (n\text{-}k)^2)$ operations and a new bisecting k Medoid algorithm was proposed [10]. Dissimilarity matrix will produce non deterministic result. To overcome the problem K-Medoid clustering has been useful to achieve deterministic result [17]. Genetic algorithm and K-Medoid was employed to find the fitness of the partition of the samples and with less variance [9]. BAT algorithm was used for finding an appropriate centroid and K-Medoid principle has been applied for finding the distance between them [20]. Variance was calculated between budgets and the Actuals produced by the two methodologies like K-Medoid and Fuzzy K-Means to identify the significant difference. Gini index methodology was applied to find the purity of the attributes towards the data set. Performance and accuracy was obtained in biometrics using K-Medoid and Partitioning around Medoids (PAM) with SIFT points [6]. In the coming sections the following are discussed 2) Proposed Methodology 3) Gini-Index 4) Algorithm 5) Distributed Measures 6) Methodology1-(K-Medoids) 7) Methodology2- (Fuzzy K-Means) 8) Budget Variance calculation 9) Chi square distribution 10) ANOVA 11) Accuracy and Reliability Evaluation 12) Results and discussions 13) Research Contributions 14) Conclusions.

Objective

Two different methodologies like K-Means and Fuzzy-K-Means were engaged to find the subtle difference between Budgets versus Actuals through the variance, Chi-square and ANOVA techniques. Level of significance was 5%, which was taken into account for stock market analysis. A new methodology for finding QoS was proposed and we were able to achieve 91.99% and 85.44% as Reliability and Accuracy respectively.

PROPOSED METHODOLOGY

In this paper we propose two methodologies like Crisp Fuzzy K-Medoids and Fuzzy K-Means for analysing the best company to invest our money for a profitable and assured return. In paper[0] they used knowledge discovery in database (KDD) by combining clustering method with soft computing and at last fuzzy clustering algorithm was applied and since the KDD process looked crucial, the K-Medoid based algorithm was used [28, 31]. Original DEA models that deals quantitative data and provides frame work for dealing with qualitative data through fuzzy numbers. Fuzzy extension principle was applied to find α cuts of levelled for fuzzy and factors [35]. Semantics can be represented in high dimensional space where fuzzy membership has been applied to K-Means clustering algorithms, to model the degree of an object belonging to the cluster [32]. K-Medoid algorithm was a cognitive reasoning algorithm for improving and strengthening optimal paradigm and it can also be used to reduce the attenuation [7]. In a uniform distribution data set, data points were randomly distributed and K-Medoid clustering was applied [19]. In the internet technology, security plays a vital role. For that intrusion detection algorithm was used in combination with Fuzzy clustering algorithm and web transactions were handled by using rough K-Means clustering algorithm so that improvement in efficiency can be obtained [37, 38, 40]. Rough Set based attribute clustering for Sample Classification (RSCSC) technique was applied for resolving low and high dissimilarity in a data set to overcome entropy [15]. To provide high performance in clustering, a minimum amount of clusters with high information gives required knowledge. So, applying Fuzzy Gap Statistic in Fuzzy K-Means clustering will resolve the problem [27]. K-Medoid was improved via variance enhanced clustering for pattern recognition and through spatial mining efficiency was improved [13] [24]. Cross sectional characteristics of their component analysis using a triangular Fuzzy time trajectory was done for classification using membership values [39]. Fuzzy K-Means clustering has been a powerful tool for classifying objects into clusters by means of membership degrees and must be equal to one. K-Means clustering model was recommended to relax the constraint [26]. SVD takes large memory time to reduce the size of the lattice and these problems were overcome by using Fuzzy K-Means clustering [25]. Bayesian and Fuzzy K-Means were used to minimize the volume of herbicides in the field of agriculture and for this purpose Hybrid decision making system has been designed which helps in prediction [34]. For classifying a data set SVM will take long time and more iteration, but when SVM was combined with Fuzzy K-Means, helps to achieve high speed and accuracy [30]. The data from Taiwan Telecom was taken as an input for Fuzzy K-Means that avoids illogical answers and saves time, thus improving company’s performance [33]. Based on the Empirical study of the Fuzzy K-Means clustering, Fuzzy partition was proposed through adaptive quadratic distance using interval valued data [22]. In the field of image processing, Mammography CT scan x-ray method has been used for low radiation strength with high resolution to detect tumours in the breast and for this K-Means and Fuzzy C-means Clustering was proposed [16]. A segmentation and stereovision stages were applied with Fuzzy K-Means to match the hemispherical images for forest environments [21]. For thyroid disease, data set provides minimum number of clusters which was obtained using scalar validity measures and several runs were carried out to achieve the optimum results [23].

Table-1. Literature comparison.

Authors	K-Medoids	Fuzzy K-Means	RSCSC	K-Means clustering	Bayesian	SVM	Cluster study	Fuzzy C-Means
Gullo *et al* (2008)	✓	×	×	×	×	×	×	×
Lewis *et al* (2012)	✓	✓	×	×	×	×	×	×
Jinghua *et al* (2009)	×	×	×	✓	×	×	×	×
Madhulatha *et al* (2011)	✓	×	×	✓	×	×	×	×
Fei *et al* (2013)	✓	×	×	×	×	×	×	×
Kisku *et al* (2010)	✓	×	×	×	×	×	×	×
Hao *et al* (2012)	✓	×	×	×	×	×	×	×
Baleon *et al* (2009)	✓	×	×	×	×	×	×	×
Shang *et al* (2006)	✓	×	×	×	×	×	×	×
Kashef *et al* (2008)	✓	×	×	×	×	×	×	×
Sivley *et al* (2013)	✓	×	×	×	×	×	×	×
Peters *et al* (2008)	✓	×	×	×	×	×	×	×
Zhang *et al* (2005)	✓	×	×	×	×	×	×	×
Henning *et al* (2008)	×	×	×	×	×	×	✓	×
Nayak *et al* (2012)	×	×	✓	×	×	×	×	×
Ambarish *et al* (2011)	×	×	×	✓	×	×	×	✓
Zhao *et al* (2013)	✓	×	×	×	×	×	×	×
Alsulaiman *et al* (2013)	✓	×	×	×	×	×	×	×
Velmurugan *et al* (2010)	✓	×	×	✓	×	×	×	×
Sood *et al* (2013)	✓	×	×	×	×	×	×	×
Herrera *et al* (2011)	×	✓	×	×	×	×	×	×
Carvalho *et al* (2010)	×	✓	×	×	×	×	×	×
Azar *et al* (2013)	×	×	×	×	×	×	✓	×
Lai *et al* (2011)	✓	×	×	×	×	×	×	×
Kumar *et al* (2010)	×	✓	×	×	×	×	×	×
Coppi *et al* (2012)	×	×	×	×	×	×	✓	×
Arima *et al* (2008)	×	✓	×	×	×	×	×	×
Barioni *et al* (2008)	✓	×	×	×	×	×	×	×
Junxin *et al* (2013)	×	✓	×	×	×	×	×	×
Ma *et al* (2009)	×	✓	×	×	×	✓	×	×
Li *et al* (2004)	×	✓	×	×	×	×	×	×
Cao *et al* (2004)	×	✓	×	×	×	×	×	×
Hsu *et al* (2011)	×	✓	×	×	×	×	×	×
Tellaeche *et al* (2007)	×	✓	×	×	✓	×	×	×
Lin *et al* (2014)	×	×	×	×	×	×	✓	×
Sivley *et al* (2013)	✓	×	×	×	×	×	×	×
Wu *et al* (2009)	×	✓	×	×	×	×	×	×
Gharehchopogh *et al* (2012)	×	✓	×	✓	×	×	×	×
Coppi *et al* (2002)	×	×	×	✓	×	×	×	×
Bharti *et al* (2010)	×	✓	×	✓	×	×	×	×

Out of the forty reference papers, nineteen papers were dealing with K-Medoids. K-Medoids takes an effort to reduce the distance among the points $$in a cluster and a selected centre point, Lai *et al*. (2011), Sivley *et al*. (2013)] and Velmurugan *et al*. (2010). Fourteen of the papers were related with Fuzzy K-Means which deal the data set by providing membership functions, Li *et al*. (2004), Cao *et al*. (2004) and Ma *et al*. (2009). One paper deals with RSCSC (Rough Set based Attribute Clustering for Sample Classification). In general RSCSC was proved to be capable of finding significant, sufficient and compact patterns by Nayak et al. (2012). Seven papers deal with K-Means Clustering algorithm. K-Means clustering algorithm was the most popular one among the clustering techniques. It tries to divide n number of observations into k clusters and it uses an iterative refinement technique Coppi *et al* (2002), Bharti *et al*. (2010) and Ambarish *et al*. (2011). Bayesian classifier which reduces the probability of wrong classification was applied in one paper and results were obtained with improved accuracy, Tellaeche *et al* (2007). SVM (Support Vector Machine) was applied on one paper to identify pattern and to examine the data, Ma *et al*. (2009). Three papers were dealing with cluster analysis method (cluster study) which combines different clustering algorithms and groups the objects which were of the same kind, Henning *et al* (2008), Azar *et al*. (2013) and Coppi *et al*. (2002). One reference paper deals with Fuzzy C-Means which provides some relaxation by allowing a single piece of data to be available to more than two clusters. Fuzzy C-Means plays a vital role in pattern recognition, Ambarish *et al*. (2011).

Table-2. Technical comparison.

Authors	K-Medoids	Fuzzy K-Means	RSCSC	K-Means	Bayesian	SVM	Cluster study	Fuzzy C-	QoS
				clustering				Means	
Gullo *et al* (2008)	Clusters uncertain	×	×	×	×	×	×	×	Reliability
	data								
Lewis *et al* (2012)	Good tracking	✓	×	×	×	×	×	×	Accuracy
Jinghua *et al*	×	×	×	Metrological	×	×	×	×	Reliability
(2009)				data					
Madhulatha *et al*	Optimal number	×	×	✓	×	×	×	×	Accuracy
(2011)	of clusters								
Fei *et al* (2013)	Clusters patients	×	×	×	×	×	×	×	Robustness
Kisku *et al* (2010)	Fusion of images	×	×	×	×	×	×	×	Accuracy
Hao *et al* (2012)	Text clustering	×	×	×	×	×	×	×	Reliability
Baleon *et al* (2009)	Cryptographic	×	×	×	×	×	×	×	Security
	key algorithm								
Shang *et al* (2006)	Gene analysis	×	×	×	×	×	×	×	Reliability
Kashef *et al* (2008)	Gene analysis	×	×	×	×	×	×	×	Efficiency
Sivley *et al* (2013)	Burden test	×	×	×	×	×	×	×	Accuracy
Peters *et al* (2008)	Less clustering	×	×	×	×	×	×	×	Reliability
	time								
Zhang *et al* (2005)	Spatial clustering	×	×	×	×	×	×	×	Efficiency
Henning *et al*	×	×	×	×	×	×	Handles	×	Robustness
(2008)							ending point		
Nayak *et al* (2012)	×	×	Handles	×	×	×	×	×	Reliability
			gene						
			expression						
			data						
Ambarish *et al*	×	×	×	Breast	×	×	×	Breast cancer detection	Accuracy
(2011)				cancer					
				detection					
Zhao *et al* (2013)	Deterministic	×	×	×	×	×	×	×	Reliability

	solution								
Alsulaiman *et al*	Technical analysis	×	×	×	×	×	×	×	Robustness
(2013)									
Velmurugan *et al*	✓	×	×	Uniform	×	×	×	×	Complexity
(2010)				distribution					
				of points					
Sood *et al* (2013)	Detection of	×	×	×	×	×	×	×	Accuracy
	location of bat								
Herrera *et al*	×	Handles image for	×	×	×	×	×	×	Reliability
(2011)		forest environment							
Carvalho *et al*	×	Quadratic distance	×	×	×	×	×	×	Efficiency
(2010)									
Azar *et al* (2013)	×	×	×	×	×	×	Thyroid	×	Predictive
							disease		Accuracy
Lai *et al* (2011)	Similarity of	×	×	×	×	×	×	×	Efficiency
	data objects								
Kumar *et al* (2010)	×	Lattice reduction	×	×	×	×	×	×	Reliability
Coppi *et al* (2012)	×	×	×	×	×	×	Clusters data	×	Fusion
Arima *et al* (2008)	×	obtaining preferable	×	×	×	×	×	×	Accuracy
		no of clusters							
Barioni *et al* (2008)	Handles the	×	×	×	×	×	×	×	Efficiency
	missing data								
Junxin *et al* (2013)	×	Classification of	×	×	×	×	×	×	Reliability
		sensed images							
Ma *et al* (2009)	×	Exact prediction	×	×	×	High training speed	×	×	Accuracy
Li *et al* (2004)	×	Missing data	×	×	×	×	×	×	Reliability
Cao *et al* (2004)	×	High dimensional	×	×	×	×	×	×	Robustness
		space							
Hsu *et al* (2011)	×	less cost	×	×	×	×	×	×	Low cost
Tellaeche *et al*	×	agriculture	×	×	Exact prediction	×	×	×	Predictive
(2007)									Accuracy
Lin *et al* (2014)	×	×	×	×	×	×	Deals with	×	Reliability
							qualitative		
							data		
Sivley *et al* (2013)	Handles constraint	×	×	×	×	×	×	×	Reliability
Wu *et al* (2009)	×	Clustering web	×	×	×	×	×	×	Accuracy
		transactions							
Gharehchopogh	×	Intrusion Detection	×	✓	×	×	×	×	Reliability
et al (2012)		System							
Coppi *et al* (2002)	×	×	×	Fuzzy time	×	×	×	×	Robustness
				trajectories					
Bharti *et al* (2010)	×	Intrusion Detection	×	✓	×	×	×	×	Efficiency
		System							

In terms with the clustering, K-Medoid algorithm focuses on clustering of the patients, Fei *et al*. (2013). It also helps in finding the optimal number of clusters, Madhulatha *et al*. (2011) and reduces the clustering time, Peters *et al*. (2008). It plays a vital role in text clustering. In terms with data, K-Means clustering algorithm clusters the uncertain data, Gullo *et al* (2008) and handles the missing data, Barioni *et al* (2008). It also provides deterministic solution, Zhao *et al* (2013). Fuzzy K-Means provides a good support to the intrusion detection system, Gharehchopogh *et al*. (2012) and it reduces the cost of the agriculture, Tellaeche *et al*. (2007) when it was applied in this filed. It was much useful in lattice reduction, Kumar *et al* (2010) and the rate of the prediction was high. Rough Set based Attribute Clustering for Sample Classification (RSCSC) helps in handling gene expression data, Nayak *et al*. (2012). K-Means clustering algorithm assures uniform distribution of points and was applied to analyze the metrological data, Jinghua *et al*. (2009). In terms with health related issues it helps in detection of breast cancer. The prediction of the Bayesian classifier was much high. Support Vector Machine combines with the Fuzzy K-Means clustering to improve the training speed and classification accuracy, Ma *et al* (2009). From the cluster study it was found that they deal with qualitative data, Lin *et al* (2014). They also handle endpoints, Henning *et al* (2008). Fuzzy C-Means clustering was useful for detecting the breast cancer, Ambarish *et al*. (2011).

Table-3. Papers published from 2002 to 2014 using different methodologies for various applications.

Year	K-Medoids	Fuzzy K-Means	RSCSC	K-Means clustering	Bayesian	SVM	Cluster study	Fuzzy C-Means
2002	-	-	-	1	-	-	-	-
2004	-	1	-	-	-	-	-	-
2005	1	-	-	-	-	-	-	-
2006	1	-	-	-	-	-	-	-
2007	-	-	-	-	1	-	-	-
2008	3	1	-	-	-	-	1	-
2009	1	2	-	1	-	1	-	-
2010	2	3	-	2	-	-	-	-
2011	2	2	-	2	-	-	-	1
2012	2	1	1	1	-	-	1	-
2013	6	1	-	-	-	-	1	-
2014	-	-	-	-	-	-	1	-

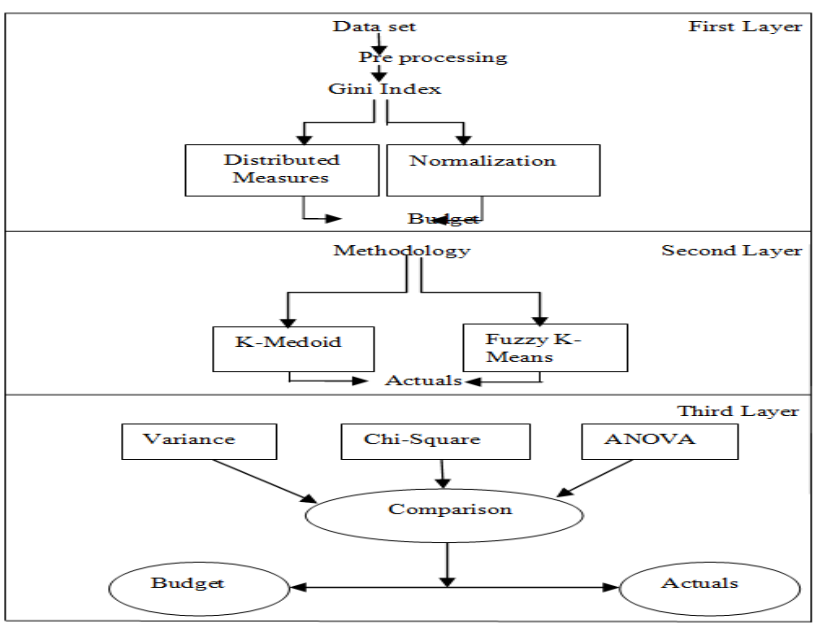

Figure-1. Three Layered BAR model Architecture.

VOL. 10, NO. 6, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-4. Training set.

Companies	SHPPG	PSHI	PSHNI	PAT	TA	PS	DY	DVD	RVS	MC
GLEN	48.31	40.86	10.83	386.11	2832.02	0	0.37	2	2496.09	14460.95
AP	52.79	0	47.21	1050	3782.72	8.8	0.94	4.6	3288.37	48521.02
ITC	0	53.78	46.22	7418.39	22354.25	0	1.63	5.25	21497.67	249245.86
DBR	68.63	24.67	6.7	590.48	1836.36	0	0.88	1.5	1420.49	29679.7
CP	51	26.48	22.52	496.75	489.61	0	2.07	28	475.99	18397.79
HUL	67.25	18.37	14.38	3796.67	2674.02	0	3.24	18.5	2457.77	123475.39
SF	49.53	20.99	29.48	95.06	1403.49	0	2.94	1.4	673.28	999.16
TCS	73.96	21.67	4.37	10413.49	32725.37	1.8	1.01	22	32266.53	425230.05
MRCO	59.69	33.48	6.83	395.9	2647.62	0	0.46	1	1926.95	13984.07
HIND	40.05	42.5	17.45	3397	58117.16	0	1.14	1.4	33239.6	25291.14
SESA	60.65	27.15	12.2	2082.87	17525.33	0	0.05	0.1	12936.88	59841.95
NMDC	80	15.96	4.04	6342	27510.96	0	4.93	7	27114.49	56259.32
CUB	0	29.23	70.77	322	22977.09	0	1.93	1	1593.22	2802.58
JSPL	59.13	27.56	13.31	4002.26	31849.01	0	0.61	1.6	12254.59	24394.8
SUN	63.65	25.95	10.4	516.55	7832.01	0.1	0.44	2.5	7685.32	117590.33

Companies Legend-1 (Column 1): GLEN-Glenmark Pharma, AP- Asian Paints, ITC- Imperial Tobacco Company, DBR- Dabur, CP- Coalgate Palmolive, HUL- Hindustan UniLever, SF- Sundaram Fasteners, TCS- Tata Consultancy Services, MRCO-Marico, HIND-Hindalco Industries, SESA- Sesa Sterlite, NMDC-National Mineral Development Corporation, CUB- City Union Bank, JSPL- Jindal Steel and Power Limited, SUN- SunPharma.

Attribute Legend-2 (Row 1): SHPPG- Share Holding Pattern of Promoter Group, PSHI- Public Share Holding Institution, PSHNI- Public Share Holding Non Institution, PAT- Profit After Tax, TA- Total Assest, PS-Pledged Shares, DY- Dividend Yield, DVD- Dividend, RVS- Reserves, MC- Market Capital. The above training set (Table-4) consisting of 15 companies and 10 attributes with their appropriate instance helps us to make an exact prediction for the future. The attributes such as SHPPG-Share Holding Pattern of Promoter Group, PSHI- Public Share Holding Institution, PSHNI- Public Share Holding Non Institution lies under the term share holding pattern. The addition of these three attributes will result in 100% share holding pattern of particular company. For any prediction the decision variable was needed and the same will be deployed in the class column. So, in this Table the decision variables Excellent, Good, Satisfactory, Fair, and Poor were ascertained through distributed measures.

In the training set, data has been derived by using the following principles. For analysing the Budget and the Actuals we are applying three various methodologies namely Variance, Chi- square Distribution and ANOVA. Using Variance we can gauge in terms with numerical values and the other methods tell us whether the Budget and Actual were equally distributed or not. The tuple (row) for the TCS company has been calculated and explained below: Share Holding: Total number of shares=1957220996, Percentage of shares held by promoter and promoter group= 1447523210/1957220996 = 73.96%, Percentage of shares held by public share holding institution=424137787 /1957220996 = 21.67%, Percentage of shares held by public share holding institution= 85559999/1957220996 = 4.37%. Profit after Tax: Total consolidated profit after year ended March 2013 was 10413.49 Crores. Total Assets: Total assets= tangible assest+ intangible assest =32725.37 Crores. Pledged Shares: Pledged shares= 35233232/1957220996= 1.80%, Dividend: dividend per share was Rs 22. Dividend yield= dividend per share/ price value of one share= 22/2172.62= 1.01%, Reserves: The company reserves stand for rupees 32266.53 Crores, Market Capital: Market capital= price value of one share* total number of shares= 2172.62 *1957220996l= 425230.05 Crores. Similarly all the Tuples were calculated and furnished in the same manner.

Gini-Index

Entropy was calculated in order to find the overall gain of the class and was given by the formulae:

$$1-(C1/\sum \text{class})^2-(C2/\sum \text{class})^2-\ldots\ldots\ldots\ldots (Cn/\sum \text{class})^2 =$$
$$1-(2/15)^2-(2/15)^2-(3/15)^2-(2/15)^2-(6/15) = 0.74 \quad (1)$$

For other attributes, Mean value was taken for each and every column respectively, and the value which was less than or equal to mean value has been taken as c1 and rest of the above values were taken as c2. So the

VOL. 10, NO. 6, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

general formulae to calculate Gini-Index for all attributes were given as:

$$C1/15(1-(\text{No of Excellent}/C1)^2 - (\text{No of Good}/C1)^2-(\text{No of Satisfactory}/C1)^2 - (\text{No of Fair}/C1)^2 - (\text{No of Poor}/C1)^2 + C2/15(1-(\text{No of Excellent}/C2)^2 - (\text{No of Good}/C2)^2-(\text{No of Satisfactory}/C2)^2 - (\text{No of Fair}/C2)^2 - (\text{No of Poor}/C2)^2 \quad (2)$$

Table-5. Gini Index for various attributes.

ENTROPY	SHPPG	PSHI	PSHNI	PAT	TA	PS	DY	DVD	RVS	MC
0.74	0.71	0.71	0.73	0.57	0.54	0.71	0.7	0.6	0.59	0.73

Refer Legend-2 as stated above

Gini Index was applied in order to find the weightage of the attributes. The Entropy value was 0.74. After applying Gini Index to ten attributes, it was found that six of our attributes such as SHPPG, PSHI, PSHNI, PS, DY and MC lie near to our entropy value, while the attributes such as PAT, TA, DVD and RVS were little bit less from the Entropy. Since majority of the attributes lie near to our entropy with elevated purity levels, the predicted answer will have high significance.

Algorithm

a) Training Set was taken from the website www.moneycontrol.com
b) Attributes were measured
c) Gini Index was applied to maintain the weightage of the attribute
d) Budget was obtained by fixing a threshold
e) Based on the threshold, class was fixed

```
Procedure measures (T, R)
T: threshold
R: range from 0-100
        if(T>= 83 && T<= 100)
                write 'Excellent'
                else if(T>=70 && T<=82)
                write 'Good'
                else if(T>=56 && T<=69)
                write 'Satisfactory'
                else if(T>=51 && T<=55)
                write 'Fair'
                else
                write 'Poor'
        end if
end procedure
```

f) The corresponding table was normalised
g) K-Medoid→ Actuals
h) Fuzzy K-Means Actuals obtained by C++ programming

Distance calculation

ARPN Journal of Engineering and Applied Sciences


```
Procedure distance (i, j, max)
    for i= 1 to 10
        if(max<m[i])
            max=m[i];
            j=i;
        end if
            next i
            write "Record r belongs to that particular centroid"
    end for
end procedure
```

Membership calculation

```
if (the value was maximum among 15 Tuples)
    write 'It lies in that particular record'
end if
```

i) variance estimation between budget and K-Medoid Actuals
j) variance estimation between budget and Fuzzy K-Means Actuals
k) distribution tests applied:

a) $$Chi-Square = \frac{\sum(O-E)^2}{E} \quad (3)$$

b) ANOVA

$$\bar{x} = \frac{\bar{X}_1 + \bar{X}_2}{2} \quad (4)$$

$$SSB = \sum_{i=1}^{2} n(\overline{xi} - \bar{x})^2, \quad (5)$$

$$MSB = \frac{SSB}{K-1} \quad (6)$$

$$SSW = \sum_{i=1}^{2} (xi - \overline{xi})^2, \quad (7)$$

$$MSW = \frac{SSW}{N-K} \quad (8)$$

$$F = \frac{MSB}{M} \quad (9)$$

Distributed measures

High values in the column were assigned 100 percentile and other values were measured and distributed based on this. The same principle has been fortified in all the columns respectively. Sum of the ten columns has been summarised and this summarised value has been kept inside Row Total and again the Row Total column was converted to100 Percentage measure and placed in the Total Percentage. In the table PS (Pledged Shares) column appeared twice because highly pledged shares was measured as 'zero' and less Pledged shares by the company will have a reduced percentage accordingly and that scenario was exhibited in two PS Columns. Threshold values fixed for Excellent: 83-100 percentage, Good: 70-82 percentage, Satisfactory: 56-69 percentage, Fair: 50-55 percentage, and Poor: less than 50 percentages. In the above said way the decision variables were declared in the Class Column and it has been taken as the *Budgeted values*. From the table we obtained that Tata Consultancy Services, National Mineral Development Corporation comes under the category of Excellent. ITC, Hindalco Industries grouped into Good. Coalgate Palmolive, Hindustan UniLever, Jindal Steel and Power Limited lies under the shelter of Satisfactory. Sesa Sterlite, City Union Bank lies under the roof of Fair. Glenmark Pharma, Asian Paints, Dabur, Sundaram Fasteners, Marico and Sun Pharma lies under the crowd of Poor.

VOL. 10, NO. 6, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-6. Distributed measure table with the same legend 1 and 2 for rows and columns used.

Companies	SHPPG	PSHI	PSHNI	PAT	TA	PS	PS	DY	DVD	RVS	MC	RT	TP	CLASS
GLEN	60.38	75.97	15.3	3.7	4.87	0	100	7.5	7.14	7.5	3.4	285.76	42.6	POOR
AP	65.98	0	66.7	10.08	6.5	100	0	19.06	16.42	9.89	11.41	206.04	30.71	POOR
ITC	0	100	65.31	71.23	38.46	0	100	33.06	18.75	64.67	58.61	550.09	82	GOOD
DBR	85.78	45.87	9.46	5.67	3.15	0.22	99.78	17.84	5.35	4.27	6.97	284.14	42.35	POOR
CP	63.75	49.23	31.82	4.77	0.84	0	100	41.98	100	1.43	4.32	398.14	59.35	SATS
HUL	84.06	34.15	20.31	36.45	4.6	0	100	65.72	66.07	7.39	29.03	447.78	66.75	SATS
SF	61.91	39.02	41.65	0.91	2.41	0	100	59.63	5	2.02	0.23	312.78	46.62	POOR
TCS	92.45	40.29	6.17	100	56.3	20.54	79.46	20.48	78.57	97.07	100	670.79	100	EX
MRCO	74.61	62.25	9.65	3.8	4.5	0	100	9.33	3.57	5.79	3.28	276.78	41.26	POOR
HIND	50.06	79.02	24.65	32.62	100	0	100	23.12	5	100	5.94	520.41	77.58	GOOD
SESA	75.81	50.48	17.23	20	30.15	0	100	1.01	0.35	38.92	14.07	348.02	51.88	FAIR
NMDC	100	29.67	5.7	60.9	47.33	0	100	100	25	81.57	13.23	563.4	83.99	EX
CUB	0	54.35	100	3.09	39.53	0	100	39.14	3.57	4.79	0.65	345.12	51.44	FAIR
JSPL	73.91	51.24	18.8	38.43	54.8	0.11	99.89	12.37	5.71	36.86	5.73	397.74	59.29	SATS
SUN	79.56	48.25	14.69	4.96	13.47	1.59	98.41	8.92	8.92	23.12	27.65	327.95	48.89	POOR

Refer Legends 1 and 2

Legend-3: EX- Excellent, SATS- Satisfactory, RT- Row Total, TP- Total Percentage.

Normalised table

Table-7. Normalised table for K-Medoids calculation.

Companies	SHPPG	PSHI	PSHNI	PAT	TA	PS	DY	DVD	RVS	MC
GLEN	0.21	0.27	0.05	0.01	0.02	0.35	0.03	0.02	0.03	0.01
AP	0.32	0	0.32	0.05	0.03	0	0.09	0.08	0.05	0.06
ITC	0	0.18	0.12	0.13	0.07	0.18	0.06	0.03	0.12	0.11
DBR	0.3	0.16	0.03	0.02	0.01	0.35	0.06	0.02	0.02	0.02
CP	0.16	0.12	0.08	0.01	0.002	0.25	0.11	0.25	0.004	0.01
HUL	0.19	0.08	0.05	0.08	0.01	0.22	0.15	0.15	0.02	0.06
SF	0.2	0.12	0.13	0.003	0.01	0.32	0.19	0.02	0.01	0.001
TCS	0.14	0.06	0.01	0.15	0.08	0.12	0.03	0.12	0.14	0.15
MRCO	0.27	0.22	0.03	0.01	0.02	0.36	0.03	0.01	0.02	0.01
HIND	0.1	0.15	0.05	0.06	0.19	0.19	0.04	0.01	0.19	0.01
SESA	0.22	0.15	0.05	0.06	0.09	0.29	0.003	0.001	0.11	0.04
NMDC	0.18	0.05	0.01	0.11	0.08	0.18	0.18	0.04	0.14	0.02
CUB	0	0.16	0.29	0.01	0.11	0.29	0.11	0.01	0.01	0.001
JSPL	0.19	0.13	0.05	0.1	0.14	0.25	0.03	0.01	0.09	0.01
SUN	0.24	0.15	0.04	0.02	0.04	0.3	0.03	0.03	0.07	0.08

Refer Legends 1 and 2

For the easy calculation purpose Distributed Measures Table-6 has been normalized between 0 and 1. This was obtained by dividing each Row value by its Row Total. Once the instance values accommodated between 0 and 1, finding the distance measured from the centroid and the respective cost calculation was easy.

Methodology 1- (K-Medoids)

VOL. 10, NO. 6, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Table-8. K-Medoid summarised table.

Companies	E1	G1	S1	F1	P1	E2	G2	S2	F2	P2	CLASS
GLEN	1.1	0.94	0.69	0.83	1.24	0.91	0.79	0.54	0.47	0.36	POOR
AP	1.1	1.2	1.13	1.11	0	1.05	1.31	1.16	1.13	1	POOR
ITC	0.64	0	1.01	0.75	1.2	0.69	0.61	0.66	0.69	0.72	GOOD
DBR	1.09	0.95	0.66	0.82	1.03	0.84	0.82	0.59	0.48	0.31	POOR
CP	0.99	1.01	0	1.9	1.13	0.84	0.92	0.71	0.8	0.69	SATS
HUL	0.71	0.87	0.42	1	0.91	0.52	0.86	0.61	0.68	0.61	SATS
SF	1.2	0.96	0.51	0.63	1.08	0.71	0.93	0.66	0.65	0.54	SATS
TCS	0	0.64	0.99	1.33	1.1	0.51	0.75	0.7	0.77	0.8	EX
MRCO	1.12	1	0.73	0.85	1.18	0.93	0.83	0.58	0.49	0.34	POOR
HIND	0.75	0.61	0.92	0.84	1.31	0.62	0	0.37	0.48	0.67	GOOD
SESA	0.77	0.69	0.799	0.8	1.13	0.62	0.48	0.27	0	0.27	FAIR
NMDC	0.51	0.69	0.84	1.06	1.05	0	0.62	0.51	0.62	0.73	EX
CUB	1.33	0.75	0.82	0	1.11	1.06	0.84	0.79	0.8	0.83	FAIR
JSPL	0.7	0.66	0.71	0.79	1.16	0.51	0.37	0	0.27	0.42	SATS
SUN	0.8	0.72	0.69	0.83	1	0.73	0.67	0.42	0.27	0	POOR

Note: E1, E2, G1, G2, S1, S2, F1, F2, P1 and P2 were explained clearly in 6.1
Legend-4: EX- Excellent, SATS- Satisfactory.

Improvised optimum way of fixing centroids

In general the centroids were fixed in an arbitrary manner. But in our paper the improved optimum way of fixing the centroids were employed. Here companies have been clustered into different categories like Excellent, Good, Satisfactory, Fair and Poor. Each and every category has got many companies rather clustered together. Now we have to decide the centroid for each and every category. For applying K-Medoid, centroids were playing a vital role. Now improvised optimum way of fixing centroids was implemented. We consider E1 and E2 for Excellent, G1 and G2 for Good, S1 and S2 for Satisfactory, F1 and F2 for Fair and P1 and P2 for Poor respectively. In our distributed Table-6, TP (Total Percentile) lies in the range among 83-100 were considered as Excellent. In the Table-8, two companies come under the category of Excellent. Now highest value in this category was called as E1 and the lowest was E2 and these were taken as centroids. Similarly the same process was repeated for G1 and G2, S1 and S2, F1 and F2, P1 and P2. If we follow the above procedures we can able to get ten centroids with fifteen Tuples. In some cases we may have more number of companies appeared in the same categories. At that time the highest total percentile in that category was considered as upper bound centroid and the lowest was lower bound centroid. For example companies like Colgate Palmolive, Hindustan Uni Lever, Sundaram Fasteners and Jindal Steel and Power Limited lies under the category of satisfactory. For the above said companies S1 and S2 were taken as upper bound and lower bound centroids based on total percentile values. The costs were arranged in Column wise and the comparison was made via Row wise. The cost which was red in colour in the above Table-8 shows the least value in the Row and the exact Class group was determined through this least value. Tata Consultancy Services, National Mineral Development Corporation comes under the category of Excellent. Imperial Tobacco Company, Hindalco Industries comes under the roof of Good. Coalgate Palmolive, Hindustan UniLever, Sundaram Fasteners, Jindal Steel and Power Limited lies under the cluster of Satisfactory. Sesa Sterlite, City Union Bank lies under the shelter of Fair. Glenmark Pharma, Asian Paints, Dabur, Marico and Sunpharma lies under the group of Poor.

ARPN Journal of Engineering and Applied Sciences

Table-9(a). Highest cost value in the group Excellent.

CENTROID E1	0.14	0.06	0.01	0.15	0.08	0.12	0.03	0.12	0.14	0.2	ROW COST
GLEN	0.07	0.21	0.04	0.14	0.06	0.23	0	0.1	0.11	0.14	1.1
AP	0.18	0.06	0.31	0.1	0.05	0.12	0.06	0.04	0.09	0.09	1.1
ITC	0.14	0.12	0.11	0.02	0.01	0.06	0.03	0.09	0.02	0.04	0.64
DBR	0.16	0.1	0.02	0.13	0.07	0.23	0.03	0.1	0.12	0.13	1.09
CP	0.02	0.06	0.07	0.14	0.08	0.13	0.08	0.13	0.14	0.14	0.99
HUL	0.05	0.02	0.04	0.07	0.07	0.1	0.12	0.03	0.12	0.09	0.71
SF	0.06	0.06	0.12	0.15	0.07	0.2	0.16	0.1	0.13	0.15	1.2
TCS	0	0	0	0	0	0	0	0	0	0	0
MRCO	0.13	0.16	0.02	0.14	0.06	0.24	0	0.11	0.12	0.14	1.12
HIND	0.04	0.09	0.04	0.09	0.11	0.07	0.01	0.11	0.05	0.14	0.75
SESA	0.08	0.09	0.04	0.09	0.01	0.17	0.03	0.12	0.03	0.11	0.77
NMDC	0.04	0.01	0	0.04	0	0.06	0.15	0.08	0	0.13	0.51
CUB	0.14	0.1	0.28	0.14	0.03	0.17	0.08	0.11	0.13	0.15	1.33
JSPL	0.05	0.07	0.04	0.05	0.06	0.13	0	0.11	0.05	0.14	0.7
SUN	0.1	0.09	0.03	0.13	0.04	0.18	0	0.09	0.07	0.07	0.8

Refer Legend 1.

From the Table-9(a) E1 represents the Highest Cost in the group Excellent. Since Excellent group is having fortunately two Tuples with highest and lowest cost and were taken as E1 and E2 respectively. The green colour instance cost of TCS taken from the normalised Table-7 was considered highest cost value in the segment of excellent group. The remaining 14 Tuples in the training set cost were subtracted from the centroid E1 and the same will be furnished above. If there were three values in a particular cluster, then two values were selected and among those two values, one will act as upper bound and the other will be a lower bound value. Row Cost was the summation of ten rows. In the same way Row Cost was calculated for each upper bound and lower bound values in Excellent, Good, Satisfactory, Fair and Poor. In the above Table-9(a) since the values of TCS was zero it was made red in colour and the reason for this was that the TCS value has been taken as the centroid.

Table-9(b). Lowest cost value in the group Excellent.

CENTROID E2	0.18	0.05	0.01	0.11	0.08	0.18	0.18	0.04	0.14	0	ROW COST
GLEN	0.03	0.22	0.04	0.1	0.06	0.17	0.15	0.02	0.11	0.01	0.91
AP	0.14	0.05	0.31	0.06	0.05	0.18	0.09	0.04	0.11	0.04	1.05
ITC	0.18	0.13	0.11	0.02	0.01	0	0.12	0.01	0.02	0.09	0.69
DBR	0.12	0.11	0.02	0.09	0.07	0.17	0.12	0.02	0.12	0	0.84
CP	0.02	0.07	0.07	0.1	0.08	0.07	0.07	0.21	0.14	0.01	0.84
HUL	0.01	0.03	0.04	0.03	0.07	0.04	0.03	0.11	0.12	0.04	0.52
SF	0.02	0.07	0.12	0.11	0.07	0.14	0.01	0.02	0.13	0.02	0.71
TCS	0.04	0.01	0	0.04	0	0.06	0.15	0.08	0	0.13	0.51
MRCO	0.09	0.17	0.02	0.1	0.06	0.18	0.15	0.03	0.12	0.01	0.93
HIND	0.08	0.1	0.04	0.05	0.11	0.01	0.14	0.06	0.05	0.01	0.62
SESA	0.04	0.1	0.04	0.05	0.01	0.11	0.18	0.04	0.03	0.02	0.62
NMDC	0	0	0	0	0	0	0	0	0	0	0
CUB	0.18	0.11	0.28	0.1	0.03	0.11	0.07	0.03	0.13	0.02	1.06
JSPL	0.01	0.08	0.04	0.01	0.06	0.07	0.15	0.03	0.05	0.01	0.51
SUN	0.06	0.1	0.03	0.09	0.04	0.12	0.15	0.01	0.07	0.06	0.73

Refer Legend 1.

ARPN Journal of Engineering and Applied Sciences

From the Table-9(b), E2 represents the lowest Cost in the group Excellent. The green colour instance cost of NMDC was taken from the normalised Table-7 and was considered as lowest cost value in the segment of excellent group. The remaining 14 Tuples in the training set were subtracted from the centroid E2 and the results were displayed in the above table. Similarly this process was repeated for all the decision variables in the class.

Methodology 2- (Fuzzy K-Means)

Fuzzy K-Means clustering algorithm was applied in the field of digital image segmentation to accelerate the convergence of the outcome [29]. Fuzzy K-Means methodology has been useful in the normalised table, (Table-no) for finding the clusters like Excellent, Good, Satisfactory, Fair and Poor. Here ten centroids were considered right from the fifteen Tuples of the training set like E1 and E2, G1 and G2, S1 and S2, F1 and F2, P1 and P2. This represents highest and lowest values in their respective segments as explained above in section 7.1. Fuzzy K-Means algorithm as follows:

a) To find the distance between fifteen Tuples versus ten centroids.
b) Membership function was taken for each and every fifteen Tuples.
c) The highest membership value has been considered and the tuple lie in the cluster.
d) In the membership formulae bias value was used which was normally taken from 1 to 9. But, in our membership function, it was taken as bias b=2 for the ease of operations.

Distance Formulae for Fuzzy K-Means

$$d\,(x,c_1)= \text{sqrt}((x_1-x_2)^2+(y_1-y_2)^2) \tag{10}$$

Membership function formulae for Fuzzy K-Means

$$\mu c1(x) = \frac{(1/\text{d}(\text{x, c1}))^{\text{b-1}}}{1/\ \text{d}(\text{x, c1}) + \text{d}(\text{x, c2}) + \text{d}(\text{x, c3})} \tag{11}$$

Individual membership measures were calculated and adhered

$$\mu c_2(x) = \frac{(1/d(x,c_2))}{\sum_{i=1}^{n} 1/d(x,c_i)}$$ similarly n^{th} membership function was denoted as

$$\mu c_n(x) = \frac{(1/d(n,c_2))}{\sum_{i=1}^{n} 1/d(n,c_i)} \tag{12}$$

Calculation for Fuzzy K-Means

Table-10. Tuple 1 which acts on ten centroids was shown below:

TUPLE E1	**0.21**	**0.27**	**0.05**	**0.01**	**0.02**	**0.35**	**0.03**	**0.02**	**0.03**	**0.01**

Fifteen Tuples were taken from the Table-7, and the ten centroids as E1and E2, G1and G2, S1and S2, F1and F2, P1 andP2 respectively.

Table-11. Ten centroids for fuzzy calculations.

E1	**0.14**	**0.06**	**0.01**	**0.15**	**0.08**	**0.12**	**0.03**	**0.12**	**0.14**	**0.15**
E2	0.18	0.05	0.01	0.11	0.08	0.18	0.18	0.04	0.14	0.02
G1	0	0.18	0.12	0.13	0.07	0.18	0.06	0.03	0.12	0.11
G2	0.1	0.15	0.05	0.06	0.19	0.19	0.04	0.01	0.19	0.01
S1	0.16	0.12	0.08	0.01	0.002	0.25	0.11	0.25	0.004	0.01
S2	0.19	0.13	0.05	0.1	0.14	0.25	0.03	0.01	0.09	0.01
F1	0	0.16	0.29	0.01	0.11	0.29	0.11	0.01	0.01	0.001
F2	0.22	0.15	0.05	0.06	0.09	0.29	0.003	0.001	0.11	0.04
P1	0.32	0	0.32	0.05	0.03	0	0.09	0.08	0.05	0.06
P2	0.24	0.15	0.04	0.02	0.04	0.3	0.03	0.03	0.07	0.08

Note: E1, E2, G1, G2, S1, S2, F1, F2, P1 and P2 were explained clearly in 6.1

Table-12. The corresponding distances between tuple1 and ten centroids was shown.

d(T1, E1)	**0.0049**	**0.0441**	**0.0016**	**0.0196**	**0.0036**	**0.0529**	**0**	**0.01**	**0.0121**	**0.0196**
d(T1, E2)	0.0009	0.0484	0.0016	0.01	0.0036	0.0289	0.0225	0.0004	0.0121	0.0001
d(T1, G1)	0.0441	0.0081	0.0049	0.0144	0.0025	0.0289	0.0009	0.0001	0.0081	0.01
d(T1, G2)	0.0121	0.0144	0	0.0025	0.0289	0.0256	0.0001	0.0001	0.0256	0
d(T1, S1)	0.0025	0.0225	0.0009	0	0.000324	0.01	0.0064	0.0529	0.000676	0
d(T1, S2)	0.0004	0.0196	0	0.0081	0.0144	0.01	0	0.0001	0.0036	0
d(T1, F1)	0.0441	0.0121	0.0576	0	0.0081	0.0036	0.0064	0.0001	0.0004	0.000081
d(T1, F2)	0.0001	0.0144	0	0.0025	0.0049	0.0036	0.000729	0.000361	0.0064	0.0009
d(T1, P1)	0.0121	0.0729	0.0729	0.0016	0.0001	0.1225	0.0036	0.0036	0.0004	0.0025
d(T1, P2)	0.0009	0.0144	0.0001	0.0001	0.0004	0.0025	0	0.0001	0.0016	0.0049

Legend -5: T1 refers to tuple 1, while E1and E2, G1andG2, S1andS2, F1andF2, and P1andP2 correspond to Excellent, Good, Satisfactory, Fair and poor respectively.

Table-13. Table to find the highest membership value.

Row Sum (A)	**Sqrt** (Row Sum) (B)	**1/Sqrt** (Row Sum) (C)	**1/Sqrt** (Row Sum)/ ∑(C)
0.1684	0.410365691	2.436850892	0.069802645
0.1285	0.358468967	2.789641763	0.079908202
0.122	0.349284984	2.862991672	0.082009282
0.1093	0.330605505	3.02475302	0.086642872
0.0962	0.310161248	3.224129401	0.092353931
0.0562	0.237065392	4.218245406	0.120829997
0.132481	0.363979395	2.74740827	0.07869844
0.03389	0.184092368	5.432055713	0.155599121
0.2922	0.54055527	1.84994959	0.052991086
0.025	0.158113883	6.32455532	0.181164424

Fifteen Tuples have been taken into account and each tuple was represented as r1 to r15. These tuples were made to act on ten centroids C1 to C10. A tuple say r1 was taken and applied on ten centroids so that we get the respective distances which was denoted as d(x,c1) and this was shown in tabulation as Row Sum(A). Square root was applied to this result which was shown in tabulation as Sqrt(Row Sum)(B) and then the reciprocal has been taken for the obtained result and was denoted as 1/d(x, c1) which was shown in tabulation as 1/Sqrt (Row Sum) (C). Summation was done from 1/d(x, c1) to 1/d(x, c15) and this corresponding result was divided with 1/d(x, c1) for the first iteration. In the same manner all tuples were made to act on ten centroids and the maximum value which was obtained in that particular tuple was taken as the highest membership function and that maximum value denotes, in which centroid it lies. The value which was red in colour in the above tabulation shows that it was the value with highest membership and therefore r1 lies in centroid C10. C10 denotes that r1 lies under the cluster of Poor from Table-13. This process was repeated by making remaining fourteen tuples to act on ten centroids and the highest membership value was calculated and with this result we can identify where the centroid lies. This calculation of Fuzzy K-Means was implemented using C++ program and the result has been obtained and the corresponding tabulation has depicted below:

Table-14. Tuples clustered in the respective centroids.

Tuples	**r1**	**r2**	**r3**	**r4**	**r5**	**r6**	**r7**	**r8**	**r9**	**r10**	**r11**	**r12**	**r13**	**r14**	**r15**
Centroids	C10	C2	C4	C10	C10	C5	C10	C2	C10	C6	C6	C6	C3	C8	C8
Class	P	E	G	P	P	S	P	E	P	S	S	S	G	F	F

Legend-6: E- Excellent, G- Good, S- Satisfactory, F- Fair and P-Poor.

VOL. 10, NO. 6, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

Tuples like r1, r4, r5, r7 and r9 lie in centroid C10 while tuples such as r10, r11 and r12 lie in centroid C6, and tuples r2 and r8 lie in centroid C2, and tuples r6 and r3 lie in centroid C5 and C4, r13, r14 and r15 lie in centroid C3, C8 and C8, respectively.

Budget variance calculation

Here 83-100 (excellent), 70-82 (good) and 51-55 (fair) have two frequencies each. 56-69 (satisfactory) and 0-50 (poor) have got frequencies three and six respectively referred from the Table-6. Mean=183+152+187.5+106+150/15.

Table-15. Threshold vs. frequency in budget.

Threshold	83-100	70-82	56-69	51-55	0-50
Frequency	2	2	3	2	6

Mean=51.9.The budgeted variance comes around 597.23.

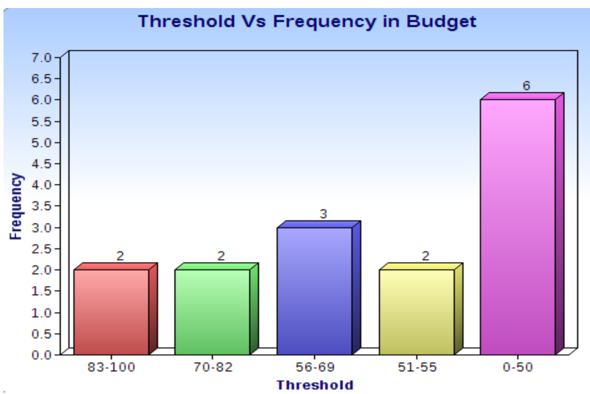

Figure-2. Threshold vs. Frequency in Budget.

K-Medoid for Actuals

Here 83-100 (excellent), 70-82 (good) and 51-55 (fair) have two frequencies each. 56-69 (satisfactory) and 0-50 (poor) have got frequencies four and five respectively referred from Table-8. Mean=183+152+187.5+106+150/15.

Table-16. Threshold vs. Frequency in K-Medoid is shown.

Threshold	83-100	70-82	56-69	51-55	0-50
Frequency	2	2	4	2	5

Mean=51.9.The budgeted variance comes around 554.98.

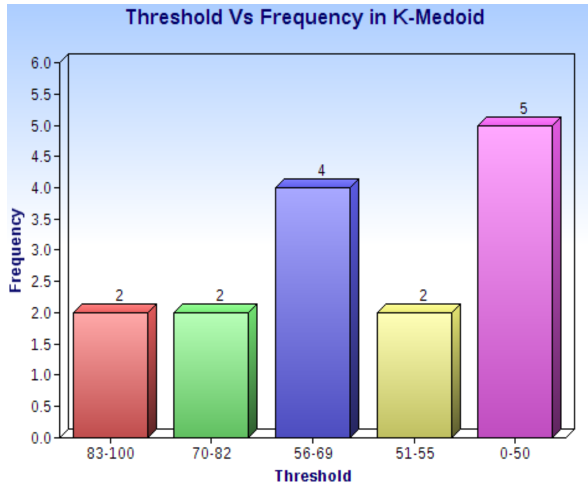

Figure-3. Threshold vs. Frequency in K-Medoid.

Variance between Budget and K-Medoid was 7.07% and the same as calculated as 554.98/597.23*100 = 7.07%

Fuzzy k-means for Actuals

Here 83-100 (excellent), 70-82 (good) and 51-55 (fair) have got two frequencies each. 56-69 (satisfactory) and 0-50 (poor) have got four and five frequencies respectively referred from Table-14. Mean=183+152+187.5+106+150/15.

Table-17. Threshold vs. Frequency in Fuzzy K-Means was shown.

Threshold	83-100	70-82	56-69	51-55	0-50
Frequency	2	2	4	2	5

Mean=51.9.The budgeted variance comes around 554.98.

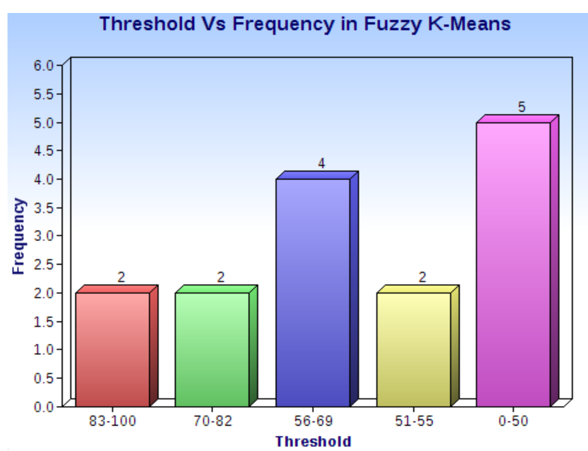

Figure-4. Threshold vs. Frequency in Fuzzy K-Means.

Variance between Budget and Fuzzy K-Means was 7.07% and the same is calculated as 554.98/597.23*100 = 7.07%.

VOL. 10, NO. 6, APRIL 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

Chi- square distribution

Both one and two way of Chi-Square distributions are as follows:

One way chi-square distribution

Chi- square distribution was applied for budget vs. K-Medoid and the results were obtained

Table-18. Budget vs. K-Medoid Actuals.

	Ex	Good	Sats	Fair	Poor
Budget	2	2	3	2	6
Actuals	2	2	4	2	5

Legend-7: Ex-Excellent and Sats-Satisfactory

In Table-14 Budget and Actuals refers to Expectation (E) and Observed (O)

$\sum$ (O-E) $^{2/}$E= 0.50, Significance level = 5% = 0.05, Degree of freedom = n-1 = 5-1 = 4

H_o was the Hypothesis which was equally distributed. The Tabulated value was found to be 9.49, from the chi- square table. Chi square value was found to be 0.50, which was less than the tabulated value. So Hypothesis H_o was accepted. Chi- square distribution was applied for Budget vs. Fuzzy K-Means Actuals and the same results were achieved.

Two way Chi-square distribution

Table-19. K-Medoid vs. Fuzzy K-Means.

	Ex	Good	Sats	Fair	Poor
K-Medoid	2	2	4	2	5
Fuzzy K-Means	2	2	4	2	5

Refer Legend no-7.

Chi square value was applied for K-Medoid and Fuzzy K-Means Actuals were found to be **0**(Zero). The tabulated value was 3.84. Chi square value was less than tabulated value. So Hypothesis H_o was accepted.

ANOVA

Table-20. One way ANOVA on Budget vs. K-Medoid Actuals.

	Ex	Good	Sats	Fair	Poor
Budget	2	2	3	2	6
Actuals	2	2	4	2	5

Refer Legend no-7.

Hypothesis H_0 was equally distributed. The formulae as shown below

$$\bar{x} = \frac{\bar{X}_1 + \bar{X}_2}{2},\ SSB = \sum_{i=1}^{2} n(\overline{xi} - \bar{x})^2,\ MSB = \frac{SSB}{K-1},$$

$$SSW = \sum_{i=1}^{2} (xi - \overline{xi})^2,\ MSW = \frac{SSW}{N-K} \tag{13}$$

F= MSB/MSW, Degree of freedom= N-1 and Significance level = 5%

Table-21. Table to find the Fishers test.

SSB 0	**K-1** 1	**MSB** 0	**F 0** (Zero)
SSW 20	**N-K** 8	**MSW** 2.5	

Fisher's test was 0(zero). But the tabulated value was 5.32. Since the fisher's value was less than the tabulated value our hypothesis H_0 was accepted.

Legend-8: SSB- Sum of Squares Between, SSW- Sum of Squares Within , MSB- Mean Square Between, MSW- Mean Square Within and F- Fishers ratio.

The above process was also repeated for Budget vs. Fuzzy K-Means Actuals and similar results were obtained and the frequencies of the tuples distributed equally.

Table-22. Two ways ANOVA on K-Medoid and Fuzzy K-Means Actuals.

Source	SS	Df	MS	F-Test
Columns	SSC 16	C-1 4	MSC 4	Infinity
Rows	SSR 0	R-1 1	MSR 0	Infinity
Residuals	SSE 0	C-1*R-1 4	MSE 0	Infinity
Total	SST 16			

Hypothesis H_0 was equally distributed or not found due to the infinity.

ARPN Journal of Engineering and Applied Sciences

Legend-9: SSC- Sum of Square of Columns, SSR- Sum of Square of Rows, SSE- Sum of Square of Residuals, SST- Sum of Squares of Total, C- Column, R- Row, MSC- Mean Square Column, MSR- Mean Square Row, MSE- Mean Square Residuals and F- Fishers Test.

MSE was 0(zero) since there was no variation between the Actuals. Hypothesis H_0 was not able to be ascertained because of the value infinity.

Accuracy and reliability evaluation

Table-23. Accuracy and reliability.

Decision variables	Accuracy	Reliability
Excellent	91.99	85.44
Good	79.79	80.78
Satisfactory	61.79	51.8
Fair	51.66	51.32
Poor	42.07	39.08

Table-24. Membership ranking for the attributes.

Attributes	SHPPG	PSHI	PSHNI	PS	PAT	RVS	MC	TA	DY	DVD
Membership Ranking	100	90	80	70	60	50	40	30	20	10

Refer Legend 2 for attribute expansion.

Accuracy and reliability were calculated for different decision variables like excellent, good, satisfactory, fair and poor. And the same was deployed in the above Table-23. Excellent Accuracy was calculated by taking the mean value of the Total percentile in that segment. Similarly the process was repeated for other decision variables. Reliability was calculated according to the basis of attributes importance. Vital attributes were given ranking from 1 to n. First rank attribute was given value 100 as membership and the following attributes were given 90, 80, and 70 and so on.

$$Accuracy = \frac{\sum DV}{TNDV} \quad (14)$$

Where DV refers to the Decision Variables and $TNDV$ denotes Total Number of Decision Variables

$$\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} C_{ij}}{N} = \text{Mean (Column)} \quad (15)$$

Where C_{ij} was the column instance, N=10, n and m = 10.

$$\text{Reliability} = \frac{MM_S V}{Mean(Column)} * MDV \quad (16)$$

Where $MM_S V$ stands for Mean membership value, MDV Denotes Mean Decision Variables.

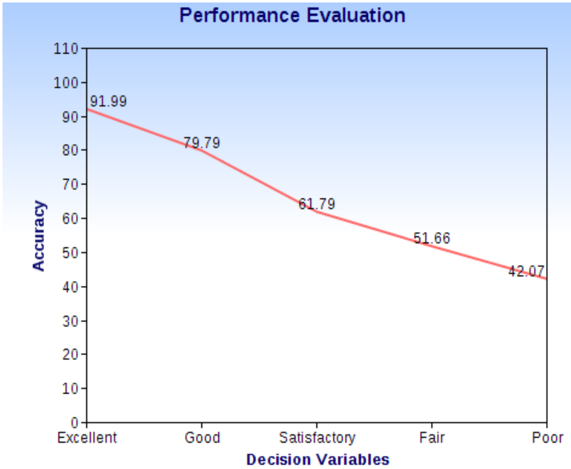

Figure-5. Performance evaluation for accuracy.

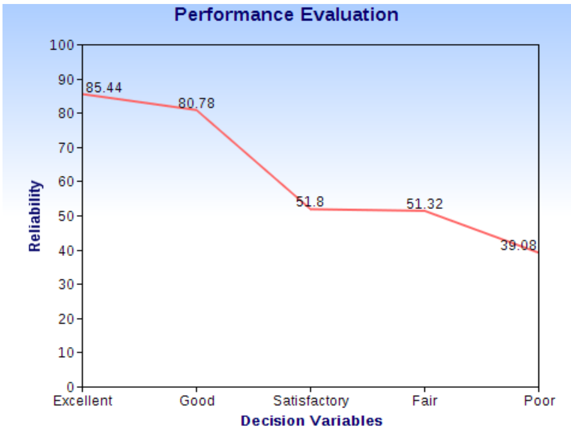

Figure-6. Performance evaluation for reliability.

RESULTS AND DISCUSSIONS

In Table-1 literature comparison for the forty reference papers were done while in Table-2 technical comparison was done for the same forty papers. In Table-4 training set was considered for analysis with each firm having various attributes. In Table-5 Gini index was applied in order to check the purity level of the attributes. Distributed measures were applied so as to find the Budget and thus a class was fixed in Table-6 based on the total percentile value. For the purpose of easier calculation, normalization was done and was shown in Table-7. K-Medoid summarisation was done in Table-8, through which we can find the clusters. The calculation of upper bound and lower bound of the excellent was shown in the tabulation 9a and 9b and explanation has been given in the previous paragraphs. For Fuzzy methodology a tuple was made to act on ten centroids and the corresponding distances between tuple1 and ten centroids was revealed in Table-12. Fuzzy membership value was calculated through the formulae and the highest membership value taken there for ascertaining in which class the firm will get clustered. In the same manner, each tuple was made to act on ten centroids and thus Table-14 was obtained which explains the clustering of the Tuples in respective centroids. Variation in budget was calculated and budgeted variance came around 597.23. In similar manner the variance of the Actuals of both the methodologies like K-Medoid and Fuzzy K-Means came around 554.98. Variance between Budget and K-Medoid and Variance between Budget and Fuzzy K-Means was found to be a minimum of 7.07%. So, the Budget that itself predicts the right clustering. Through one-way and two-way of Chi-Square and ANOVA distributions, we found that our hypothesis H_0 was accepted to the extent. Accuracy and Reliability for the decision variables were calculated and shown in Table-23.

Research contribution

This paper brings up the Accuracy and Reliability of Ad-hoc component commodities in stock market and put a great effort in substantiating it through different techniques. The message for the investors has been narrated that investments of our hard earned money made in excellent and good segments were very much sure of a safe return and at the same time reasonable benefits were obtained through them. Here the values 91.99 and 85.44 were noted as the Accuracy and Reliability in terms with Excellent. For decision variable like Good, 79.79 and 80.78 were the values denoting Accuracy and Reliability respectively. Excellent companies were Tata Consultancy Services (TCS) and National Mineral Development Corporation (NMDC) while the Good companies were Imperial Tobacco Company (ITC) and HINDUSTAN UNILEVER (HUL).

CONCLUSIONS

Stock market analysis has been a real need for the people of our society and research papers bring up the hidden ideologies of the intricate behind the stock market product. According to the Quality of Service (QoS) of the product, investment should be made for an assured return. The above findings in this research article were the eye opener to the investors in this segment which will improve the economics of our country in the long run. Other techniques are also available and possible to impart and incorporate for the same dataset domain.

REFERENCES

[1] Francesco Gullo, Giovanni Ponti, Andrea Tagarelli. 2008. Clustering Uncertain Data via K-Medoids. Lecture Notes in Computer Science. 5291: 229-242.

[2] Rory Lewis, Chad A.Mello, Andrew M. White. 2012. Tracking Epileptogenesis Progressions with Layer Fuzzy K-Means and K-Medoid clustering. International Conference on Computational Science. 9: 432-438.

[3] Huang Jinghua, Wang Zhenchong, Yuan Mei, Bao Youwen. 2009. Meteorological Data Analyze Base on K-Means algorithm. Computational Intelligence and Design. 2: 60-63.

[4] Tagaram Soni Madhulatha. 2011. Comparison between k -Means and k Medoids Clustering Algorithms. Lecture Notes in Computer Science. 198: 472-481.

[5] HangyingFei, Nadine Meskens. 2013. Clustering of Patients Trajectories with an Auto-Stopped Bisecting K-Medoids Algorithm. Journal of Mathematical Modelling and Algorithms in Operations Research. 12(2): 135-154.

[6] Dakshina Ranjan Kisku, Phalguni Gupta, Jamuna Kanta Sing. 2010. Feature Level Fusion of Face and Palmprint Biometrics by Isomorphic Graph-Based Improved K-Medoids Partitioning. Lecture Notes in Computer Science. 6059: 70-81.

[7] Zhan Gang Hao. 2012. Text Clustering Method Based on K-Medoids Social Evolutionary Programming. Lecture Notes in Computer Science. 148: 473-477.

[8] H.A. Garcia-Baleon, V. Alarcon-Aquino, O. Starostenko. 2009. K-Medoids-Based Random Biometric Pattern for Cryptographic Key Generation. Lecture Notes in Computer Science. 5856: 85-94.

[9] Weiguo Shang, Xiaohui Liu. 2006. A Genetic K-Medoids Clustering Algorithm. Journal of Heuristics. 12(6): 447-466.

[10] Rasha Kashef, Mohamed S. Kamel. 2008. Efficient Bisecting K-Medoids and Its Application in Gene Expression Analysis. Lecture Notes in Computer Science. 5112: 423-434.

[11] R. MichaelSivley, Alexandra E. Fish, William S. Bush. 2013. Knowledge-Constrained K-Medoids Clustering of Regulatory Rare Alleles for Burden Tests. Lecture Notes in Computer Science. 7833: 35-42.

[12] Georg Peters, Martin Lampant, Richard Weber. 2008. Evolutionary Rough K-Medoid Clustering. Lecture Notes in Computer Science. 5084: 289-306.

[13] Qiaoping Zhang, Isabelle Cauloigner. 2005. A New and Efficient K-Medoid Algorithm for Spatial Clustering. Lecture Notes in Computer Science. 3482: 181-189.

[14] Christian Hennig. 2008. Dissolution Point and Isolation robustness: Robustness Criteria for general Cluster Analysis Methods. Journal of Multivariate Analysis. 99(6): 1154-1176.

[15] Rudra Kalyan Nayak, Debahuti Mishra, Kailash Shaw, Sashikala Mishra. 2012. Rough Set based Attribute Clustering for Sample Classification of Gene Expression Data. International Conference on Modelling Optimization and Computing. 38: 1788-1792.

[16] NaliniSingh Ambarish G Mohapatra, Gurakalyan Kanungo. 2011. Breast Cancer Mass Detection in Mamograms using K-Means and Fuzzy C-means Clustering. International Journal of Computer Applications. 22(2): 15-21.

[17] Yuzhen Zhao, Xiyu Liu, Hua Zhang. 2013. The K-Medoids Clustering Algorithm with Membrane Computing. Telekominka: Indonesian Journal of Electrical Engineering. 11(4): 2050-2057.

[18] Eng Talal Alsulaiman. 2013. Classifying Technical Indicators using K-Medoid Clustering. The Journal of Trading. 8(2): 29-39.

[19] T. Velmurugan and T. Santhanam. 2010. Computational Complexity between K-Means and K-Medoids Clustering Algorithms for Normal and Uniform Distributions of Data Points. Journal of Computer Science. 693): 363-368.

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

[21] P. Javier Herrera, Gonzalo Pajares, Maria Guijarro. 2011. A segmentation method using Otsu and Fuzzy K-Means for stereovision matching in hemispherical images for forest environments. Applied Soft Computing. 11(8): 4738-4747.

[22] Francisco de A.T. de Carvalho, CamiloP. Tenorio. 2010. Fuzzy K-Means clustering algorithms for interval-valued data based on adaptive quadratic distances. Fuzzy Sets and Systems. 161(23): 2978-2999.

[23] Ahmad TaherAzar, Shaimaa Ahmed El- Said, Aboul Ella Hassanien. 2013. Fuzzy and hard clustering analysis for thyroid disease. Computer methods and programs in Biomedicine. 111(1): 1-16.

[24] Por-Shen Lai, Hsin-Chia Fu. 2011. Variance enhanced K-Medoid clustering. Expert system with applications. 38(1): 764-775.

[25] Ch. Aswani Kumar, S. srinivas. 2010. Concept lattice reduction using Fuzzy K-Means Clustering. Expert systems with applications. 37(3): 2696-2704.

[26] Renato Coppi, PierpaoloD'Urso, Paolo Giordani. 2012. Fuzzy and Possibilistic Clustering for Fuzzy data. Computational Statistics and Data Analysis. 56(4): 915-927.

[27] Chinatsu Arima, Kazumi Hakamada, Masahiro Okamoto, TaizoHanaj. 2008. Modified Fuzzy Gap statistic for estimating preferable number of clusters in Fuzzy K-Means clustering. Journal of Bioscience and Bioengineering. 105(3): 273-281.

[28] Maria Camila N. Barioni, HumbertoL. Razente, Agma J.M. Traina, Caetano Traina Jr. 2008. Accelerating K-Medoid based algorithms through metric access methods. Journal of System and Software. 81(3): 343-355.

[29] Dong-junxin, Yen-Wei Chen. 2013. SOR Based Fuzzy K-Means Clustering Algorithm for Classification of Remotely Sensed Image. Lecture Notes in Computer Science. 7951: 375-382.

[30] Li-juan Ma, Ming-hu Ha. 2009. Support Vector Machines Based on Sectional Set Fuzzy K-Means Clustering. Advances in Soft Computing. 54: 420-425.

[31] Dan Li, JitenderDeogun, Wiliam Spaulding, Bill Shuart. 2004. Towards Missing Data Imputation: A Study of Fuzzy K-Means clustering method. Lecture Notes in Computer Science. 3066: 573-579.

[32] Guihong Cao, Dawei Song, Peter Bruza. 2004. Fuzzy K-Means clustering on a High Dimensional Semantic Space. Lecture Notes in Computer Science. 3007: 907-911.

[33] Ming-Jia Hsu, Ping-Yu Hsu, BayarmaaDashnyam. 2011. Applying Agglomerative Fuzzy K-Means to reduce the cost of Telephone Marketing. Lecture Notes in Computer Science. 7027: 197-208.

ARPN Journal of Engineering and Applied Sciences

[34] Alberto Tellaeche, Xavier-P BurgosArtizzu, Gonzalo Pajares, Angela Ribeiro. 2007. A Vision Based Hybrid Classifier for Weeds Detection in Precision Agriculture through the Bayesian and fuzzy k-means paradigms. Advances in Soft Computing. 44: 72-79.

[35] Pei HuangLin. 2014. A General framework of dealing with qualitative data in DEA: A Fuzzy number approach studies in Fuzziness and soft computing. Studies in Fuzziness and Soft Computing. 309: 61-87.

[36] R. Michael Sivley, Alexandra E. FISH, William S. Bush. 2013. Knowledge-Constrained K-Medoids Clustering of Regulatory Rare Alleles for Burden Tests. Lecture notes in computer science. 7833: 35-42.

[37] Rui Wu, Peilin Shi. 2009. Clustering Web Transactions Using Fuzzy Rough K-Means. Communications in Computer and Information Science. 30: 231-240.

[38] Farhad Soleimanian Gharehchopogh, Neda Jabbari, ZeinabGhaffariAzar. 2012. Evluation of Fuzzy K-Means and K-Means clustering Algorithms in Intrusion Detection Systems. International Journal of Science and Technology Research. 1(11).

[39] Renato Coppi, PierpaoloD'Urso. 2002. Fuzzy K-Means Clustering models for triangular Fuzzy time trajectories. Statistical Methods and Applications. 11(1): 21-40.

[40] Kusumbharti, Shweta Jain, SanyamShukla. 2010. Fuzzy K-Mean Clustering via Random Forest for Intrusion Detection System. International Journal on Computer Science and Engineering. 2(6): 2, 97-2200.