Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 5): pp. 5597-5603
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.05.629

Agent-MB-DivClues: Multi Agent Mean based Divisive Clustering

Mohammed Ali Shaik, Research Scholar, Department of Computer Science & Engineering, Dr. APJ Abdul Kalam University, Indore, India, niharali@gmail.com
Dhanraj Verma, Professor, Department of Computer Science & Engineering, Dr. APJ Abdul Kalam University, Indore, India, dhanrajmtech@gmail.com

Abstract- The conventional way of performing data search comprises of disadvantages specifically in environment induced and non-stationarity while performing advanced prediction. In this context we propose an agent based MB-Divclues hierarchicalclustering method which is based on Multi Agent System (MAS) approach which enables us with decentralized control over the framework to attain the similarity among objects to structure various categorical attributes and measuring of relative distance among various numeric values by calculating the arithmetic mean which is assigned to the rootnode. In this paper we propose a novel divisive hierarchy that is based on the construction and implementation of a non-binary tree using multi agents. And to predict the forecasting error and be a step ahead In this approach we use a system of several adaptive forecasting agents, where the novel multi-agent forecasting system allows an agent to perform minimal training sequences and to attain accurate results to forecast by employing single prediction algorithm.

Keywords: MAS, Classification, Clustering, MB-Divclues, forecasting, prediction.

I. INTRODUCTION

The process of performing classification and clustering are the vital techniques for partitioning the data objects that comprises of maximum attributes intomeaningful disjoint subgroups [5] as the objects in each group are almost related to each other with respect to the attribute values pertaining to a objectsover an object group [6].

When a system requires essential requirements to utilize the initial needs using a dynamic approach that performs the problem tackling in a parallel process or approach, we need to use the Multi Agent Systems (MAS) [1]. To handle the multiple agents which are the software entities that comprises of heterogeneous capabilities that are implementedas services for performing the decision making process in an independent from over ever probable agent which still exists in the framework [2].

There exists a major supervised categorization between classification and clustering by analyzing the aspects as a single unit by tagging the training data for categorizing the membership for performing training or constructing the basic model [9]. The major issue in analyzing the clusters by analyzing the existence of clusters over a specific region and resolving the outliers over a specific data set [10]. This process may be further used to merging the clusters into a single tagged cluster [14]. For implementing the supervised classification model by analyzing the clusters that may be related to any specific domain [7]. The primary necessity of performing the collection of various objects to form homogenized teams as the preferable set to partition the grouping of data objects with aspects such as similarity or distinct structure that partitions in sorting the natural teams [8].

The process of clustering or to perform unsupervised classification which is performing splitting of data objects into data sets called as clusters of related objects [13]. In any attained cluster all the objects must be related to one another and must be distinct from the rest of the objects in rest of the clusters [12]. The superior the resemblance within a group and the different in between other groups leads us to the better clustering process [11].

The implementation of clustering is being performed over most of the fields as most of the algorithms classify the clustering types over the planned numerical data as it comprises of appropriate mixed knowledge for collecting the data [9,22]. The major aspect of getting with this paper is to unify the distance measure to illustrate the numeric knowledge which adopts the generated distance metricsfor evaluate and estimates the conditional probability while defining the relationshipbetween various groups [10,23].

In this paper we will provide an acceptable solution to the predominant question of the process of obtaining the minimal computational complexity at agent level when implemented over the synthetic dataset using the process of natural groups, for obtaining the probable efficiently and effectively we propose the “optimal”classification scheme over any agent to identify the objects using the arithmetic mean value [12,24]. The enhancing process of clustering requires most of the substantial modifications in the implementation of proposed algorithm in almost all aspects. For calculating the distance between two objects or pointed by two distinct agents using hierarchical clustering technique the data set that is considered is a synthetic numerical data set. Based on this process the mean value is obtained over every iteration with the time complexity of $\frac{O(n)}{N}$ which yields to 1 because the mean value obtained is n times by consuming $O(n\log_n)$ of time, which is lesser than the existing agglomerative clustering algorithms that we have [13].

I. Clustering Algorithms

The process of cluster analysis was initially proposed in numeric domains by clearly specifying the distance aspect which is further extended and categorized based on the properties of data [3,29]. Though most of the real time data in present world comprises of combination of categorical and continuous facts as aoutcome based on these aspects the demand factor of performing clustering and analyzing s data over diverse areas is growing [4,30].

The clustering process and its analysis pare is treated to be the considerable area of research over many decades and there exists many methodologies used to be covered thoroughly [16,27]. Most of the novel methods are still in the inception phase and some other are to be completed. And some of the most popular clustering algorithms that are available in the present eras with the similar complexity [14,28].

K-means[1]: K-means is the primary methodology that is well established in preforming the agglomeration strategies to the best fit in the process [15,26]. Moreover conventionally most of the process exclusively being utilize the basic data with regardto most of the existing items that tend to identify the basic memory utilization functions accurately [17,25]. The major strategy adopted and is identified is the K-which implies to be a subsequent process that evolves in K groups as it is further depicted by the attained mean value of the objects being pointed by the agents [10].

Nearest Neighbor Algorithm[2]: this algorithm is associated with the process of formulating distinct aspects that relay over the linking technique which is determined to be the closer neighboring formula [2]. Whereas the basic implementation is imparting the formula with a threshold value obtained in an iterative fashion to identify the closest cluster as the closet value used to determine and confirm the existence of clusters that are to be replaced with the clusters that merely exist [18].

Divisive ClusteringAlgorithm[3]: by implementing the dissentious agglomeration [4] that is based on the distinct things that are initiated at the initial place over the cluster where the initial cluster splits various objects in a iterative fashion till they attain their own clusters. The splitting process is bused to disseminate various clusters that are considered in distinct parts which are merely nearer as an alternative process for attaining the data objects [19].

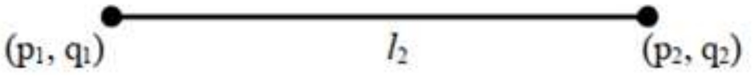
BIRCH Algorithm[4]: this algorithm is intended to be an agglomeration because the size aspect is oversized with respect to its quantity that comprises of numerical information by performing the integration part of stratified agglomeration process which is the basically tiny in an distinctive agglomeration substation which is equivalent for reiterating the process of partitioning for overcoming two distinct aspects [20].The process of overcoming the basic dual issues for clustering the agglomerative methods for performing a) measurability b) nonexistence to undo the process to wipe with the prior step [21].

ROCK (Robust agglomeration mistreatment links) [5]: this algorithm is considered to be the stratified the agglomerative for formulating the process of exploring throughout the data links for obtaining the information that comprises of definitiveattributes [22].

CURE formula[6]: The major objective of the CURE agglomeration algorithm which is used for handling various outliers that comprises of laminated elementsthat partitions various data element [23].

Chameleon [7]: this is the stratified agglomerative formulation which uses the dynamic modeling process the identifies the similarity that exist among various pairs of clustered data [24]. The absolute process that derives the support that identifies the weakness and determines various agglomeration algorithms in a straightforward manner are considerably: ROCK [27] and CURE [26]. The distance the is attained with the HC that is widely implemented using unsupervised data that analyze distinct authors that comprises of fewer and are basically uncertain in a distinct dataset [27].

Distance attained between two distinct points say point P_1, P_2 is 5, that provides us with: D(P_1, P_2) =5,



$L_2=((P_2-P_1)^2 +(Q_2-Q_1)^2)^{0.5}$ (1)

Based on equation let use implement and example for attaining the result: Let us say O =(1 0, 1 3), D=(2 0, 1 5)

$$\sqrt{(20-10)^2+(15-13)^2}$$

$=\sqrt{(100+4)}$

=10.19

$D(A, B) = \min \sum_{i=1}^{n} ||X_i - X_j||^q$ (2)

Where A and B are considered to be the tow pair of elements that comprises of a cluster denoted by D(a,b) where the distance attained between both the elements and the distance function nature is clearly defined by an integer Q (Q=2), for anyspecific data set that comprises of numeric values [30].

II. PROPOSED SYSTEM

Agent-MB-DivClues

Agent-MB-DivCluesis divisive hierarchical agglomeration algorithms which is supported and implemented by MAS in a monothetic bi-partitioned approach that allows an agent to implement the dendogramin a hierarchy which is in a tree traversal format. This algorithm performs clustering based on categories over numerical data where every cluster is handled by an individual agent. The MB-DivClues is hierarchical clustering that inverse the hierarchical agglomeration [7] and the method where the strategy initiates the total information for performing the scaling of a cluster into two or more sub clusters, this scaling process is performed by an agent in a continuous process till a single object is left over with the agent in a cluster, further the square measure is implemented which comprises of:

- Mono_Thetic: An agentpartitions the process of cluster victimization that comprises of an individual element handled by an agent which is associated by the degree level attribute for attaining the dissimilarity while the data is selected for performing the partitioning process [6].
- Poly_Thetic: An agent scales the partitions for performing the cluster victimization over all the attributes over the dual clusters for isolating a well designed to sustain the distance measure among various items [8].

A typical polythetic divisive methodology works in the following process [2]:

1. Attain the distance measure to verify the distance attained by two or more agents that holds unique data objects along with the predefined threshold value representing the probable distance measure attained.
2. Construct a distance vector quantization matrix for processing the agentsamong variouscomparison pairs of agents processing objects that denotes distinct objects I a specific data cluster.
3. Classifying the maximum distance measure between pair of agents over object-based agents over an agent group that is clustered.
4. When the distance measured between two distinct objects in a cluster is attained with the smallest value that that of the pre-defined threshold value over the distinct clusters or a specific cluster which comprises of distinct implementation process that is based on divided and stop approach, or else the rest of the agents can continue the processing task.
5. Use the agent process pairing over distinct objects for performing the seeding of seeds with K-means methodology to classify into novel clusters.
6. Halt the agent process when there exists one single agent in a cluster or else continue with above steps.

In this process the maximum attained procedure will tend to provide solution to the following dual concerns:

- Which cluster has to be divided by an agent in the next process?
- What is the process adopted to split the cluster?

Agent-MB-DivClues Algorithm:

Step 1:Start
Step 2: Initialize and formulate all objects in a single cluster
Step 3:An Agent calculates the arithmetic mean by implementing the distance matrix
Step 4: The attained mean value is initialized to the root node
Step 5: If the attained object distance is lesser than that of the attained mean value thenagent has tocreate a new cluster by assigning an agent who has to place the data objects into the newly created or assigned cluster and allocate an agent to the left of parent node.
Step 6: If the Object distance attained is greater than the attained mean value thencreate a new cluster and place the objects in new cluster and allocate an agent to the right of parent node.
Step 7: If the Object distance attained is equal to the attained mean value thenincrement the count of the node being pointed by the agent.
Step 8: Goto step 5 by calculating the mean value till single element cluster is attained.
Step 9: Stop

Computing the search complexity:

In order to initiate the implementation of pessimistic scenario that evaluates the basic searched value from the initial or the base node of the tree to its very own leaf node will provide us with the maximum complexity of O(nlog$_n$) for an agent. The search operation performed by an agent over a specific time period will be almost equivalent with the height of the tree being generated over a basic paired to perform the search operation over trees. This generates us with n probable number of hubs that require almost O(log n) computing cost. And the probable mean value required for performing each iteration is considered to be the probable time complexity that is $\frac{O(n)}{N}$which yields to 1because the mean value obtained is n times that of the complexity attained is in a iterative of O (n) possibility. And similaryour proposed algorithm requiresO(nlog$_n$) of the time complexity which is the smallest value attained than that of the present day “Agglomerative clustering algorithms” which merely depends on the binary search tree implementation lacks load balancing. The methodology provides us with the authentication that the objects or elements can be attained or consumes almost lesser or likeO (log n).

Experimental Valuation:

Example 1: The implementation of “Agglomerative algorithm”by purely based on the Euclidean distance measure to formulate a cluster and the Agent MB-Divclues algorithm by utilizing the attained mean value for the randomly take 6 agents which denotes the data objects.

Table 1: Distance matrix between the data set

Distance	**Obj$_1$**	**Ob$_2$**	**Obj$_3$**	**Obj$_4$**	**Obj$_5$**	**Obj$_6$**
Obj$_1$	0.0	0.62	5.32	3.21	4.05	3.04
Obj$_2$		0.0	4.65	2.35	3.12	2.15
Obj$_3$			0.0	2.02	1.32	2.25
Obj$_4$				0.0	0.98	0.43
Obj$_5$					0.0	1.02
Obj$_6$						0.0

We have 6 objects that have been pointed by six distinct objects i.e. [Obj$_1$, Obj$_2$, Obj$_3$, Obj$_4$, Obj$_5$, Obj$_6$] which are kept in a single groupby a merging agent. Initially we had six groupings where the primary goal is to combine them with the maximum utilization of cycles to attain a single group with all the six agent’s data. In each of the attained progression for the cycle an object identifies the frequent uprises the join groups in this process Obj$_4$ and obj$_6$ are grouped as the minimal separation is of 0.5. and this process of attaining the initial separation of lattice by assembling agents to perform grouping of Obj$_4$ and Obj$_6$ as illustrated in Table2.

Distance	**Obj$_1$**	**Ob$_2$**	**Obj$_3$**	**Obj$_4$,Obj$_6$**	**Obj$_5$**
Obj$_1$	0.0	0.62	5.32	?	4.05
Obj$_2$		0.0	4.65	?	3.12

Obj$_3$			0.0	?	1.32
Obj$_4$, Obj$_6$				0.0	?
Obj$_5$					0.0

In this way the process of precisive linkage principle is being implemented for utilizing the single linkage where an agent has todetermine the least distance attained between exceptional data objects in between two or more agents or data groups that requires to categorize the aspects such as: lattice [12], distance among agent groups (Obj$_4$, Obj$_1$) and bunch [13].

d(Obj$_4$, Obj$_6$) →Obj$_1$ min (d$_{Obj4,Obj1}$, d$_{Obj6,Obj1}$) = min (3.61, 3.14) = 3.14 (1)

The distance attained between these two clusters (Obj$_4$, Obj$_6$) and cluster Obj$_2$is:

d(Obj$_4$,Obj$_1$) →Obj$_2$ min (d $_{Obj4,Obj1}$, d $_{Obj6,Obj2}$) =min (2.92, 2.13) = 2.13 (2)

The distance attained between these two clusters (Obj$_4$, Obj$_6$) and cluster Obj$_3$is: 2.12

The distance between cluster (Obj$_4$, Obj$_6$) and cluster ∈-0.95

The simplified distance matrix is denoted in the Table 3 that is attained form equation 1.

Table 3: From equation (1)

Distance	**Obj$_1$**	**Ob$_2$**	**Obj$_3$**	**Obj$_4$, Obj$_6$**	**Obj$_5$**
Obj$_1$	0.0	0.62	5.32	3.14	4.05
Obj$_2$		0.0	4.65	2.13	3.12
Obj$_3$			0.0	2.12	1.32
Obj$_4$, Obj$_6$				0.0	0.95
Obj$_5$					0.0

Table 3 denotes the attained distance matrix that is denoted by e,and is further discovered to avoid the separation process in between group Obj$_1$ and Obj$_2$which is evaluated presently to 0.71. hencewe consider the object group denoted by Obj$_2$towards a specific batch name denoted as (Obj$_1$, Obj$_2$) constructed based on the probable input distance matrix in 6x6 formatto attain the Obj$_3$ and then further cluster (Obj$_1$, Obj$_2$). The resultant value of the minimum possible distance is illustrated to be 4.65which is almost same as equation 1.

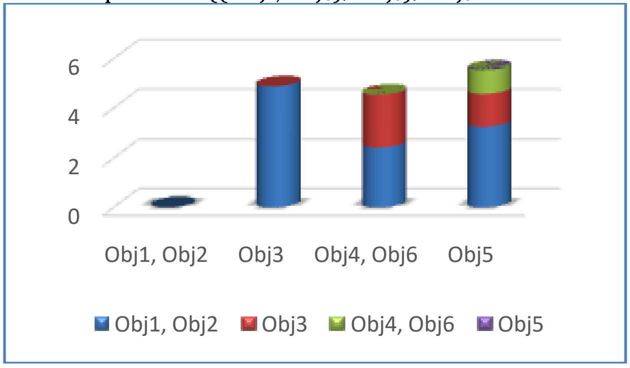
Cluster (Obj$_1$, Obj$_2$) and the cluster (Obj$_4$, Obj$_6$) minimum distance is 2.42 cluster

Cluster (Obj$_1$, Obj$_2$) and the cluster ∈is related to the minimum distance which is resultant to be 2.12 for further performing the distance upgradation by an agent over the matrix.

Table 4: cluster implementation to attain the distance among various objects.

Distance	**Obj$_1$, Obj$_2$**	**Obj$_3$**	**Obj$_4$, Obj$_6$**	**Obj$_5$**
Obj$_1$, Obj$_2$	0	4.88	2.42	3.25
Obj$_3$		0.0	2.12	1.32
Obj$_4$, Obj$_6$			0.0	0.95
Obj$_5$				0.0

In above table 4 the distance matrix generated over the cluster in the range of (Obj$_4$, Obj$_6$) we have merged the data agents such as ((Obj$_4$, Obj$_6$), Obj$_5$) that tends to endure this process over the cluster that denotes the Obj$_3$, so the attained cluster denoting by the agent related to Obj$_3$is almost a closer cluster that comprises of ((Obj$_4$, Obj$_6$), Obj$_5$), Obj$_3$where the clusters are combined at the end of this phase.



Agent-MB-DivClues Algorithm using mean

Distance	**Obj$_1$**	**Ob$_2$**	**Obj$_3$**	**Obj$_4$**	**Obj$_5$**	**Obj$_6$**
Obj$_1$	0.0	0.62	5.32	3.21	4.05	3.04
Obj$_2$		0.0	4.65	2.35	3.12	2.15
Obj$_3$			0.0	2.02	1.32	2.25
Obj$_4$				0.0	0.98	0.43

Obj$_5$					0	1.02
Obj$_6$						0

Mean value of $\bar{X}$=$\sum_{i=1}^{N} \frac{d(x_y, y_y)}{N}$, where X is root -2.54

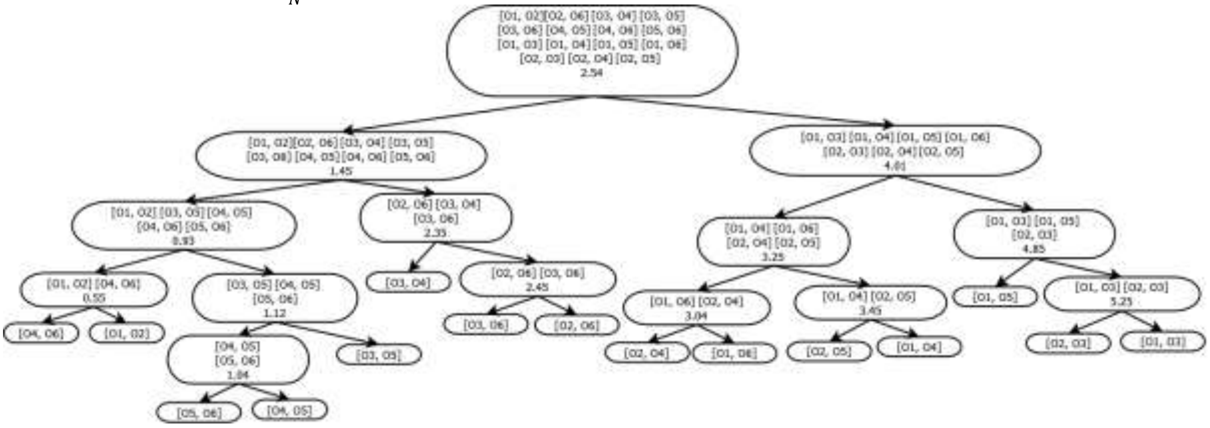


Fig. Agent-MB-DivClues Algorithm implementation using proposed algorithm

In order to perform the search operation over distinct objects in a cluster by means of attaining the mean valuewhich is considered to be the key element. Which is being verified with the root node where the key node agent value is equivalent when comparison is performed with the root node agent then that location is called as the hub node. Whereas on the contrary if the key element does not match with the root element then traversal is performed to its left sub tree if the value is lesser or to the right sub tree if the value is greater than the element, that is only two possibilities as the value does not match with the node. This process continues till either the value is matched or we reach to the leaf node and further there exists no subtree.

III. CONCLUSION

In this paper we have examined the single link hierarchical clustering by creating and implementing agent-MB-DivClues clustering algorithm over pure numeric synthetic data set using six agent objects in this paper. We have examined the distance measure acquired by each agent using single linkage over numeric data sets for generating the agent based mean value of hierarchical clustering methods. The results that are attained boosts us as it yields the optimal performance. And further we compared the classical method like K-means over hierarchical clustering and we have identified the running time of our proposed algorithm is faster than agglomerative (SLHC) algorithm due to implementation of multi agent system. The future scope of this paper is to store clusters in a n-dimensional arrays for reducing further time complexity.

REFERENCES

[1] C. C. Aggarwal and C. Zhai, 2012Mining Text Data

[2] M. Kepa, J. Szymanski 2015 Two stage SVM and kNN text documents classifier In: Pattern Recognition and Machine Intelligence, Kryszkiewicz M. (Ed.), Lecture Notes in Computer Science Vol. 9124 pp. 279 -289

[3] Mohammed Ali Shaik and Dhanraj Verma 2020 Enhanced ANN training model to smooth and time series forecast IOP Conf Ser: Mater Sci Eng Vol (981) 022038 https://doiorg/101088/1757-899X/981/2/022038

[4] R. C. Barik and B. Naik 2015 A Novel Extraction and Classification Technique for Machine Learning using Time Series and Statistical Approach Computational Intelligence in Data Mining vol. 3 pp. 217-228

[5] Mohammed Ali Shaik Dhanraj Verma P Praveen K Ranganath and Bonthala Prabhanjan Yadav 2020 RNN based prediction of spatiotemporal data mining IOP Conf Ser: Mater Sci Eng Vol (981) 022027 https://doiorg/101088/1757-899X/981/2/022027

[6] Babu, C.N., Reddy, B.E., 2014. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. Appl. Soft Comput. 23, 27–38.

[7] Shaik MA 2019 A survey on text classification methods through machine learning methods Int J Control Autom vol 12(6) pp 390-396

[8] Sampath Kumar T, Manjula B, Srinivas D. 2017 A new technique to secure data over cloud. J Adv Res Dyn Control Syst vol 2017(Special Issue 11) pp 391-396

[9] Mohammed Ali Shaik and Dhanraj Verma 2020 Deep learning time series to forecast COVID-19 active cases in INDIA: a comparative study IOP Conf Ser: Mater Sci Eng Vol (981) 022041 https://doiorg/101088/1757-899X/981/2/022041

[10] P. Praveen, C. J. Babu and B. Rama, Big data environment for geospatial data analysis, 2016 International Conference on Communication and Electronics Systems (ICCES), Coimbatore, pp 1-6

[11] Mohammed Ali Shaik, 2019 A Survey of Multi-Agent Management Systems for Time Series Data Prediction, International Journal of Grid and Distributed Computing (IJGDC), Vol 12(3) pp 166-171

[12] Praveen P, Jayanth Babu C. 2019 Big Data Clustering: Applying Conventional Data Mining Techniques in Big Data Environment. Lect Notes Networks Syst vol 74 pp 509-516

[13] Mohammed Ali Shaik, P.Praveen, Dr.R.Vijaya Prakash 2019 Novel Classification Scheme for Multi Agents Asian Journal of Computer Science and Technology Vol.8 (S3) pp. 54-58

[14] Shaik, M.A 2018 Protecting Agents from Malicious Hosts using Trusted Platform Modules (TPM) Proceedings of the International Conference on Inventive Communication and Computational Technologies, (ICICCT 2018) pp. 559-564

[15] C. H. Wan, L. H. Lee, R. Rajkumar, and D. Isa 2012 A hybrid text classification approach with low dependency on parameter by integrating K-nearest neighbor and support vector machine Expert Syst. Appl vol. 39(15) pp 11880 -11888

[16] Shaik, M.A 2019 A survey on text classification methods through machine learning methods International Journal of Control and Automation vol 12(6) pp 390-396

[17] Hamilton, J.D., 1994. Time Series Analysis, first ed Princeton University Press, Princeton.

[18] Hollander, M., Wolfe, D.A., Chicken, E., 1999. Nonparametric Statistical Methods. John Wiley& Sons, Hoboken, NJ.

[19] Ali Shaik, M 2020 Time series forecasting using vector quantization International Journal of Advanced Science and Technology vol 29(4) pp. 169-175

[20] Shaik, M.A., Sampath Kumar, T., Praveen, P., Vijayaprakash, R 2019 Research on multi-agent experiment in clustering International Journal of Recent Technology and Engineering vol 8(1 Special Issue 4) pp. 1126-1129

[21] Behera, H.S., Dash, P.K., Biswal, B., 2010. Power quality time series data mining using S-transform and fuzzy expert system. Appl. Soft Comput. 10 (3), 945–955.

[22] A. Chaudhuri 2014 Modified fuzzy support vector machine for credit approval classification IOS Press and Authors vol. 27(2) pp 189-211

[23] E. Baralis, L. Cagliero, and P. Garza 2013 EnBay: A novel pattern-based Bayesian classifier Tkde vol. 25(12) pp 2780-2795

[24] X. Fang 2013 Inference-Based Naive Bayes: Turning Naive Bayes Cost-Sensitive vol. 25(10) pp 2302-2314

[25] L. H. Lee, R. Rajkumar, and D. Isa, "Automatic folder allocation system using Bayesian-support vector machines hybrid classification approach," Appl. Intell., vol. 36, no. 2, pp. 295-307, 2012.

[26] Ravi Kumar R, Babu Reddy M, Praveen P 2019 An evaluation of feature selection algorithms in machine learning Int J Sci Technol Res vol 8(12) pp 2071-2074

[27] De Livera, A.M. Hyndman, R.J. Snyder, R.D. 2011 Forecasting time series with complex seasonal patterns using exponential smoothing J. Am. Stat. Assoc vol 106 pp 1513–1527

[28] Box, G.E.P., Jenkins, G., 1990. Time Series Analysis, Forecasting and Control. *Holden-Day Incorporated*.

[29] Chandra, D.R., Kumari, M.S., Sydulu, M., 2013. A detailed literature review on wind forecasting *International Conference on Power, Energy and Control, ICPEC* pp. 630-634.

[30] Babu, C.N., Reddy, B.E. 2014 A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data. *Appl. Soft Comput* vol 23 pp 27–38