

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/326239105>

Analysis of student academic performance using clustering techniques

Article in International Journal of Pure and Applied Mathematics · January 2018

CITATIONS

22

READS

1,975

2 authors:



[Karthikeyan Govindasamy](#)

Amrita Vishwa Vidyapeetham

5 PUBLICATIONS 42 CITATIONS

[SEE PROFILE](#)



[Velmurugan Thambusamy](#)

Dwaraka Doss Goverdhan Doss Vaishnav College

126 PUBLICATIONS 1,365 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Ph.D project [View project](#)



An Implementation of Substitution Techniques in Cloud Security using Playfair and Caesar algorithms [View project](#)

ANALYSIS OF STUDENT ACADEMIC PERFORMANCE USING CLUSTERING TECHNIQUES

K. Govindasamy¹, T. Velmurugan²

¹Research Scholar, VELS University, Chennai, India.

²Associate Professor, PG and Research Department of Computer Science,
D. G. Vaishnav College, Chennai, India.

E-Mail: ¹mphilgovind@gmail.com, ²velmurugan_dgvc@yahoo.co.in

Abstract: Student's performance is an essential part in higher learning institutions. Predicting student's performance becomes more challenging due to the large volume of data in educational databases. Clustering is one of the method in data mining to analyze the massive volume of data. It categorizes data into clusters such that objects are grouped in the same cluster when they are similar according to specific metrics. This paper is designed to study and compare four clustering algorithms. The algorithms used for the research is k-Means, k-Medoids, Fuzzy C Means (FCM) and Expectation Maximization (EM). The main advantage of clustering is that interesting patterns and structures can be found directly from very large data sets with little or none of the background knowledge. The performance of the clustering algorithms is compared based on the factors: Purity, Normalized mutual information(NMI) and time taken to form cluster.

Keywords: Educational Data Mining, k-Means Algorithm, k-Medoids Algorithm, Fuzzy C Means Algorithm, Expectation Maximization Algorithm.

1. Introduction

Data mining is a process of extracting previously unknown, valid, potential useful and hidden patterns from large data sets. As the amount of data stored in educational data bases is in increasing rapidly. In order to get required benefits from such large data and to find hidden relationships between variables using different data mining techniques developed and used. Clustering is most widely used techniques in data mining. The aim of clustering is to partition students in to homogeneous groups according to their characteristics and abilities [1].

Usually educational organizations used to collect huge amount of data which would be relevant to faculty members, students, etc. But the importance of data that is collected is unknown. The data that are used in generating simple queries or traditional reports may be in significant, which will not contribute to the process of inference/decision making in the educational organizations. The collected data may also contain such insignificant data. Also the volume and complexity of the collected data may be very high such that it is not easy to handle. If that is the case then the collected data may not be used and memory is occupied unnecessarily. The available data can be made usable if and only if it is converted into useful information by exploiting potentiality of the collected data. A wide range of data mining algorithms is used to extract useful information from potential data gathered in various educational organizations.

There are increasing research interests in education field using data mining. Application of Data mining techniques concerns to develop the methods that discover knowledge from data and used to uncover hidden information. The discovered knowledge can be used to better understand students' behavior, to assist instructors, to improve teaching, to evaluate and improve e-learning system, to improve student academic performance; to improve curriculums and many others benefits [2].

This study investigates the educational domain of data mining. This paper performs a comparative analysis of four clustering algorithms namely k-means algorithm, k-Medoids algorithm, Fuzzy C Means algorithm and Expectation Maximization algorithm. The performance of these clustering algorithms is compared in terms of purity, normalized mutual information and time taken to form a cluster. The student data was collected from different private Arts and Science colleges. The collected academic data was grouped according to their similar characteristics, forming clusters.

The rest of the article is organized as follows. Section 2 discusses about various research articles related to data mining techniques for predict clustering students' performance. Section 3 explores the basic concepts of k-Means algorithm, k-Medoids algorithm, FCM algorithm and EM algorithm in detail. The clustering results of each algorithm were examined in detail and compared with each other to evaluate the performance of the algorithms in section 4. Finally, concludes the research work.

2. Related Work

In educational data mining various research have been done in predicting students' performance using different data mining techniques such as clustering, classification, neural networks, etc. Some of the methodologies from different research articles were discussed in this section. Educational Data Mining (EDM) is the field of study concerned with mining educational data to find out interesting patterns and knowledge in educational organizations. In [3], the study explores multiple factors theoretically assumed to affect students' performance in higher education, and finds a qualitative model which best classifies and predicts the students' performance based on related personal and social factors. In [3] four decision tree algorithms was used on the collected student's data, namely, C4.5 decision tree, ID3 decision tree, CART decision Tree, and CHAID.

Durairaj et al., [4] propose Educational Data mining for Prediction of Student Performance Using Clustering Algorithms. They predicting the students' performance, used weka data mining through clustering, which paved way to strategic management tool. In [5] Prashant et al., examined the clustering analysis in data mining that analyzes the use of k-means algorithm in improving students academic performance in higher education and presents k-means clustering algorithm as a simple and efficient tool to monitor the progression of students performance.

Shiwani and Roopali [6], had proposed a work to evaluate the performance of students of Digital Electronics of university institute of engineering and technology. The researcher had applied unsupervised learning algorithms such as K-means and Hierarchical clustering using WEKA tool as an open source tool. The paper [7] focuses on the study of data mining techniques applied to small data sets concerning higher education institutions, concludes that the use of these techniques in real-

life situations is useful and promising, and can provide administrators with precious tools for decision. Clustering is used in [8] for analyzing data concerning the evaluation of courses taken by students, linked to their results in the corresponding exams. The work presented in [9] reviews different clustering algorithms applied to educational data mining context while [10] is an interesting review of recent educational data mining development whose contents are in turn analyzed by a data mining approach.

Sarala et al., discussed [11] the applications of data mining in educational institution to extract useful information from the huge data sets and providing analytical tool to view and use this information for decision making processes by taking real life examples. The paper in [12] focuses on set up a clustering algorithm which is most suitable for predicting students performance in educational data mining. The objective of this research work is to gain an insight into how clustering analysis can be done in educational domain and to highlight the potential characteristics of the clustering algorithms within the educational data set. In [13] a new model was used to predict the student performance using a neural network. The model helps to accurately predict students at risk of dropping and reduce dropout rates. This comparison between planned and actual performance indicates that the model works in the estimation of student performance. A Research work done by Veeramuthu et al. [14], had designed a model to present as a guideline for higher educational system to improve their decision making processes. The authors aim to analyze how different factor affect a student learning behavior and performance using K-means clustering algorithm. A work done by Sivaram et al., [15] had surveyed the applicability of clustering and classification algorithms for recruitment data mining techniques that fit the problems which are determined. A study has been made by applying K-means, fuzzy C-means clustering and decision tree classification algorithms to the recruitment data of an industry.

3. Clustering Algorithms

Clustering is process, grouping a set of physical or abstract objects into classes of similar objects. A cluster is a collection of data objects that are similar to one another with in the same cluster and are dissimilar to the objects in other clusters. Data clustering is alternatively referred to an unsupervised learning and statistical data analysis. Cluster analysis is an important human activity. Cluster analysis has been widely used in numerous applications including pattern recognition, data analysis, image processing and market research. Clustering is a descriptive task that seeks to identify homogenous group objects based on the values of their attributes. Clustering has many requirements like scalability, dealing with different types of attributes, discovery of clusters with arbitrary shape, minimal requirements for domain knowledge to determine input parameters, ability to deal with noisy data, high dimensionality, interpretability and usability. Clustering techniques can be broadly classified into many categories; partitioning, hierarchical, density-based, grid-based, model-based algorithms.

3.1 The k-Means Algorithm

k-Means is one of the simplest unsupervised learning algorithms used for clustering. Given D , a data set of n objects, and k , the number of clusters to form, a partitioning algorithm organizes

the objects into k partitions ($k \leq n$), where each partition represents a cluster. The clusters are formed to optimize an objective partitioning criterion, such as a dissimilarity function based on distance. The algorithm is composed of the following steps:

Step 1: Place k points into the space represented by the objects that are being clustered. These points are present initial group centroids.

Step 2: Assign each object to the group that has the closest centroid.

Step 3: When all objects have been assigned, recalculate the positions of the k centroids.

Step 4: Repeat steps 2 and 3 until the centroids no longer move.

This produces a separation of the objects into groups from which the metric to be minimized can be calculated. The k -means simple clustering algorithm that has been improved to several problem domains.

3.2 The k -Medoids Algorithm

The k -Medoids algorithm is related to the k -Means algorithm and the medoid shift algorithm. Both the k -Means and k -Medoids algorithms are partition (breaking the dataset up into groups). k -Means attempts to minimize the total squared error, while k -medoids minimizes the sum of dissimilarities between points labeled to be in a cluster and a point designated as the center of that cluster. In contrast to the k -Means algorithm, k -Medoids chooses data points as centers (medoids or exemplars). k -Medoids is also a partitioning technique of clustering that clusters the data set of n objects into k clusters with k known *a priori* [16]. The algorithm is composed of the following steps:

Step 1: Using Euclidean distance as a dissimilarity measure, compute the distance between every pair of all objects as follows

$$d_{ij} = \sqrt{\sum_{a=1}^p \|x_{ia} - x_{ja}\|^2} \quad (1)$$

$i=1, \dots, n; j=1, \dots, n$

Step 2: Calculate P_{ij} to make an initial guess at the centers of the clusters.

$$P_{ij} = \frac{d_{ij}}{\sum_{i=1}^n d_{ij}} \quad (2)$$

$i=1, \dots, n; j=1, \dots, n$

Step 3: Calculate $\sum_{i=1}^n P_{ij} (j = 1 \dots n)$ at each objects and sort them in ascending order. Select k objects having the minimum value as initial group medoids.

Step 4: Assign each object to the nearest medoid.

Step 5: Calculate the current optimal value, the sum of distance from all objects to their medoids.

Step 6: Replace the current medoid in each cluster by the object which minimizes the total distance to other objects in its cluster.

Step 7: Assign each object to the nearest new medoid.

Step 8: Calculate new optimal value, the sum of distance from all objects to their new medoids. If the optimal value is equal to the previous one, then stop the algorithm. Otherwise, go back to the Step 6.

3.3 The FCM Algorithm

Fuzzy C-Means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method is frequently used in pattern recognition. FCM algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More data is near to the cluster center and its membership towards the particular cluster center. Clearly, summation of membership of each data point should be equal to one [16]. After each iteration membership and cluster centers are updated according to the formula:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{ij}}{d_{ik}} \right)^{\frac{2}{m-1}}} \quad (3)$$

$$v_j = \frac{\left(\sum_{i=1}^n \mu_{ij}^m x_i \right)}{\left(\sum_{i=1}^n \mu_{ij}^m \right)}, \forall j = 1, 2, \dots, c \quad (4)$$

where,

'n' is the number of data points.

'v_j' represents the j^{th} cluster center.

'm' is the fuzziness index $m \in [1, \infty]$.

'c' represents the number of cluster center.

' μ_{ij} ' represents the membership of i^{th} data to j^{th} cluster center.

' d_{ij} ' represents the Euclidean distance between i^{th} data and j^{th} cluster center.

Main objective of fuzzy c-means algorithm is to minimize:

$$J(U, V) = \sum_{i=1}^n \sum_{j=1}^c \mu_{ij}^m \|x_i - v_j\|^2 \quad (5)$$

where,

' $\|x_i - v_j\|$ ' is the Euclidean distance between i^{th} data and j^{th} cluster center.

Steps for Fuzzy c-means clustering

Let $X = \{x_1, x_2, x_3, \dots, x_n\}$ be the set of data points and $V = \{v_1, v_2, v_3, \dots, v_c\}$ be the set of centers.

Step 1: Randomly select 'c' cluster centers.

Step 2: Calculate the fuzzy membership ' μ_{ij} ' using:

$$\mu_{ij} = \frac{1}{\sum_{k=1}^c \frac{\mu_{ik}^{m-1}}{d_{ik}^{m-1}}} \quad (6)$$

Step 3: Compute the fuzzy centers ' v_j ' using:

$$v_j = \frac{\left(\sum_{i=1}^n \mu_{ij}^m x_i \right)}{\left(\sum_{i=1}^n \mu_{ij}^m \right)}, \forall j = 1, 2, \dots, c \quad (7)$$

Step 4: Repeat step 2 and 3 until the minimum ' J ' value is achieved or $\|U^{(k+1)} - U^{(k)}\| < \beta$.
Where,

' k ' is the iteration step.

' β ' is the termination criterion between $[0,1]$

' $U = (\mu_{ij})_{n \times c}$ ' is the fuzzy membership matrix.

' J ' is the objective function.

3.4 The EM Algorithm

The EM algorithm is an efficient iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. In ML estimation, wish to estimate the model parameter(s) for which the observed data are the most likely. The iteration of the EM algorithm consists of two processes. They are E-step and M-step. In the expectation, or E-step, the missing data are estimated given the observed data and current estimate of the model parameters. This is achieved using the conditional expectation, explaining the choice of terminology. In the M-step, the likelihood function is maximized under the assumption that the missing data are known. The estimate of the missing data from the E-step is used in lieu of the actual missing data. Convergence is assured since the algorithm is guaranteed to increase the likelihood at each iteration. The algorithm is composed of following steps.

Step 1: Initialization

Step 2: E-Step: This step is responsible to estimate the probability of each element belong to each cluster.

Step 3: M-Step: This step is responsible to estimate the parameters of the probability distribution of each class for the next step.

Step 4: Convergence Test: After each iteration is performed a convergence test which verifies if the difference of the attributes vector of iteration to the previous iteration is smaller than an acceptable, given by parameter.

4. Experimental Results

This section explains the performance evaluation of proposed approach. The soil nutrients is implemented using Java (version 1.7), and the experiments are performed on a Intel(R) Pentium machine with a speed 2.13 GHz and 2.0 GB RAM using Windows 7 32-bit Operating System.

4.1 Data Set Description

Various departments of student data is collected from private Arts and Science Colleges. More than 1531 student's details are collected with their performance in Seminar and Assignments. The data is mainly used for evaluating the performance of various clustering algorithms to predict the academic performance of the students in their end of the semester examinations. Table 1 shows the data set attribute description.

Table 1: Description of Student Data set

Attribute	Description
`S.N	Student Serial Number
Name	Name of the Student
Sex	Gender Male/Female
Branch	B.A., (Eng,& Tam), BBA, BCA, B.Com, B.Sc., (CS & Maths)
SSLC Mark	10 th total marks
HSS Mark	12 th total marks
Medium	Studies in school- English/Tamil
Location	Student Native City
FS	Family Size
FT	Family Type
FAI	Family Annual Income
FQ	Father Qualification
MQ	Mother Qualification
LT	Location Type (Village, Town)
PSG	Previous Semester Grade (Average, Good, Excellent)
SemP	Seminar Performance (for 3 years)
Att	Student Attendance
ESG	End Semester (Average, Good, Excellent)

Figure 1: Sample Data Set

S.N	Name	Sex	DOB	Age	Branch	10TH	12TH	Sem	Medium
1	Ajith. A	Male	10.05.1998	19	B.Sc CS - II	340			
2	Ajitha. E	Female	02.06.1998	18	B.Sc CS - II	328			
3	Arul Jothi. A	Female	14.03.1997	19	B.Sc CS - II	339			
4	Arunkumar . V	Male	15.04.1998	18	B.Sc CS - II	387			
5	Balaji. J	Male	05.05.1998	18	B.Sc CS - II	260			
6	Bharathi. P	Female	15.03.1998	19	B.Sc CS - II	323			
7	Ezhilmathi. U	Female	12.06.1996	21	B.Sc CS - II	299			
8	Girija. K	Female	13.04.1998	18	B.Sc CS - II	240			
9	Indhumathi. V	Female	09.03.1998	18	B.Sc CS - II	314			
10	Indhumathi. V	Female	15.06.1998	18	B.Sc CS - II	407			
11	Karunakaran. P	Male	15.03.1997	19	B.Sc CS - II	323			
12	Sasikala. S	Female	15.06.1998	19	B.Sc CS - II	402			
13	Sathosh Kumar. S	Male	29.03.1998	18	B.Sc CS - II	350			

4.2 Metrics of Cluster Evaluation

A clustering algorithm is evaluated using (i) some internal evaluation measure like cohesion, separation, or the silhouette-coefficient (addressing both, cohesion and separation), (ii) some external evaluation measure like accuracy, precision, or recall with respect to some given class-structure of the data. In some cases, where evaluation based on class labels does not seem viable, (iii) careful (manual) inspection of clusters shows them to be a somehow meaningful collection of apparently somehow related objects.

The proposed clustering algorithms are evaluated using Purity, Normalized mutual information (NMI) and time taken to form cluster. *Purity* is a simple and transparent evaluation measure. *Normalized mutual information* can be information-theoretically interpreted. To compute purity, each cluster is assigned to the class which most frequent in the cluster and then accuracy of this assignment is measured by counting the number of correctly assigned documents and dividing by N.

$$purity(\Omega, C) = \frac{1}{N} \sum_k \max_j |w_k \cap c_j| \quad (8)$$

Where $\Omega = \{w_1, w_2, \dots, w_k\}$ is the set of clusters and $C = \{c_1, c_2, \dots, c_j\}$ is the set of classes.

Bad clustering have purity values close to 0, a perfect clustering has a purity of 1.

Normalized Mutual Information or *NMI* is computed as follows:

$$NMI(\Omega, C) = \frac{I(\Omega, C)}{\sqrt{H(\Omega) + H(C)}} \quad (9)$$

Where I is the mutual information

$$I(\Omega, C) = \sum_K \sum_J P(w_k \cap c_j) \log \frac{P(w_k \cap c_j)}{p(w_k)p(c_j)} \quad (10)$$

$$= \sum_K \sum_J \frac{|w_k \cap c_j|}{N} \log \frac{N |w_k \cap c_j|}{|w_k| |c_j|} \quad (11)$$

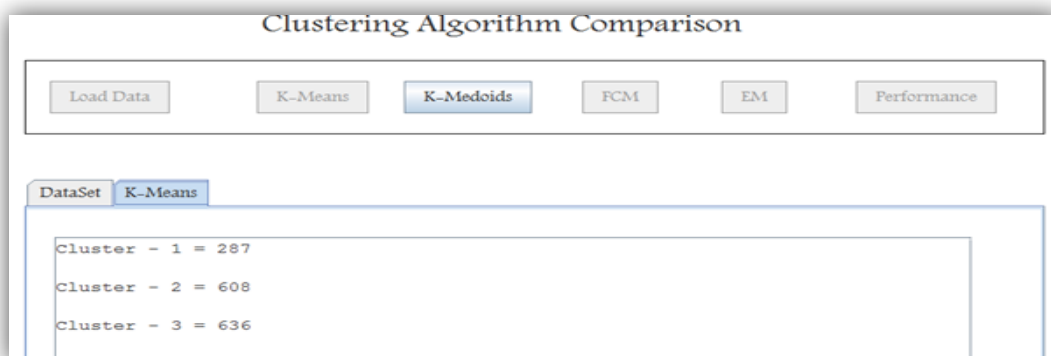
Where $P(w_k)$, $P(c_j)$ and $P(w_k \cap c_j)$ are the probabilities of a document being in cluster w_k class c_j and in the intersection of w_k and c_j .

H is entropy,

$$\begin{aligned} H(\Omega) &= -\sum_k P(w_k) \log P(w_k) \\ &= -\sum_k \frac{|w_k|}{N} \log \frac{|w_k|}{N} \end{aligned} \quad (10)$$

NMI is always a number between 0 and 1.

Figure 2 The Results of K-Means Algorithm



4.3 Performance of Clustering Algorithm

The cluster quality is evaluated using the number of clusters, execution time, purity, and NMI. Distribution of requirements data set among the clusters: The total number of clusters is three (Average, Good, Excellent). Table 2 shows the total number of requirements that are distributed when k Means, k-Medoids, FCM and EM algorithm are applied.

Figure 3 Clustering Algorithm Comparison

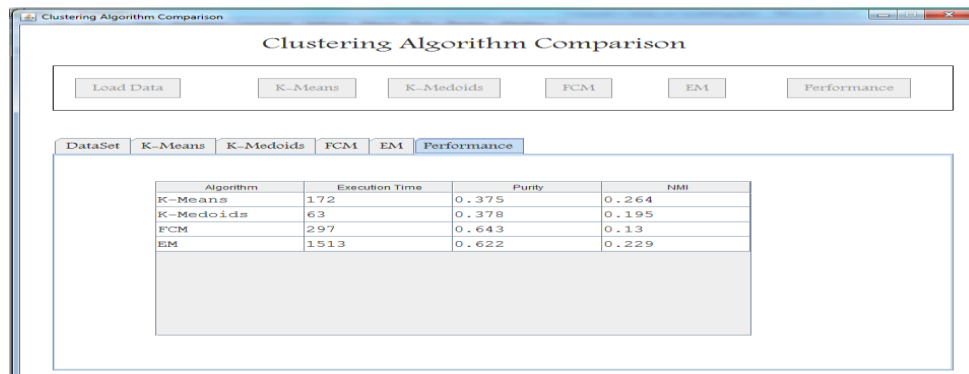


Table 2 Distribution of requirements in clusters

Clustering Algorithm	Average	Good	Excellent	Total
k-Means	287	608	636	1531
k-Medoids	613	615	303	1531
FCM	709	126	696	1531
EM	231	684	616	1531

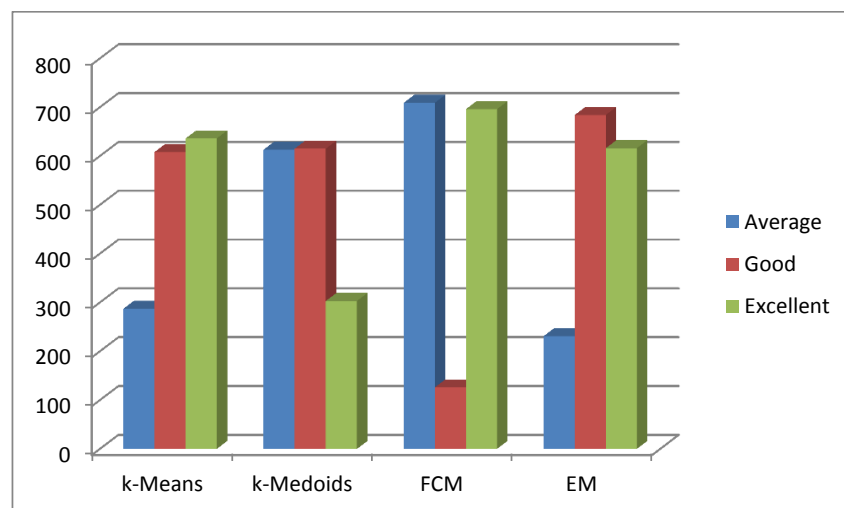


Figure 4 Cluster Distribution Comparisons

Figure 1 shows the distribution of cluster comparison. The distribution shows that the data points in cluster-1 uniformly distributed except k-Means algorithm.

5. Results and Discussion

Table 3 Execution Time comparison

Clustering Algorithm	Execution Time in ms
k-Means	128
k-Medoids	110
FCM	250
EM	560

Table 3 shows the execution time of clustering algorithms. The time consumption of FCM is less compared to the EM. The lowest execution time is in K-Medoids. In figure 2, the x axis represents the clustering algorithm and y-axis represent the time in milliseconds.

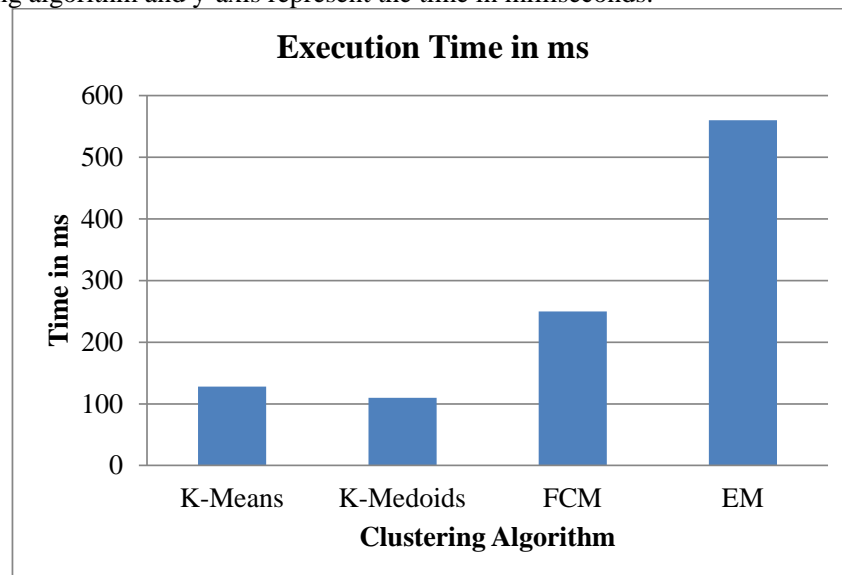
**Figure 5 Comparison of Execution time**

Table 4 and Figure 3 show the comparison of purity and NMI values.

Table 4 Comparison of Purity and NMI Values

Algorithm	Purity	NMI
k-Means	0.375	0.264
k-Medoids	0.374	0.199
FCM	0.624	0.071
EM	0.664	0.032

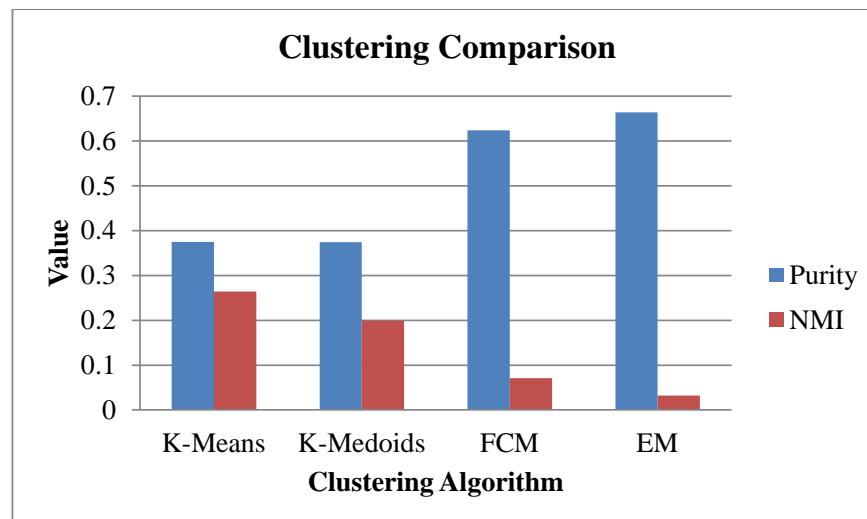


Figure 6 Purity and NMI Comparison for Clustering Algorithm

From the comparison the purity value of EM and FCM is more compared to the k-Means and k-Medoids algorithms. The NMI value of EM and FCM is less compared to the k-Means and k-Medoids algorithms. From the comparison the clustering algorithm FCM and EM is better compared to k-Means and k-Medoids in terms of distribution purity, and NMI but these algorithms take more execution time.

6. Conclusion

The research work has put an effort to reveal that the clustering techniques serve as a powerful tool in educational data mining. Here various clustering algorithms are discussed and by using these algorithms, student's performance is evaluated. In this research work, clustering algorithms k-Means, k-Medoids, FCM and EM were examined and compared based on the performance of the algorithms using student data set. The taken parameters of student data set are evaluated and the results are analysed. The parameters purity, NMI and etc are analysed in this work. The clustering algorithms are evaluated using execution time, purity and NMI. The result shows that FCM and EM algorithm performs well compared with other two clustering algorithms.

Reference

- [1] Sreenivasarao, Vuda, and Capt Genetu Yohannes. "Improving academic performance of students of defence university based on data warehousing and data mining" *Global Journal of Computer Science and Technology*, 2012, Vol. 12(2), pp 29-36.

- [2] Romero, Cristobal, and Sebastian Ventura. "Educational data mining: A survey from 1995 to 2005.", *Expert systems with applications*, 2007, Vol. 33, pp. 135-146.
- [3] Saa, Amjad Abu. "Educational Data Mining & Students' Performance Prediction.", *International Journal of Advanced Computer Science and Applications*, 2016, Vol. 7(5), pp. 212-220.
- [4] Durairaj, M., and C. Vijitha., "Educational Data mining for Prediction of Student Performance Using Clustering Algorithms." , *International Journal of Computer Science and Information Technologies* , 2014, Vol. 5(4), pp. 5987-5991.
- [5] Saxena, Prashant Sahai, and M. C. Govil., "Prediction of Student's Academic Performance using Clustering.", Special Conference Issue: National Conference on Cloud Computing & Big Data., 2014,
- [6] Rana, Shiwani, and Roopali Garg., "Evaluation of student's performance of an institute using clustering algorithms.", *International Journal of Applied Engineering Research*, 2016, Vol.11(5), pp. 3605-3609.
- [7] Natek, Srečko, and Moti Zwilling., "Student data mining solution–knowledge management system related to higher education institutions.", *Expert systems with applications*, 2014, Vol. 41(14) pp. 6400-6407.
- [8] Campagni, Renza, Donatella Merlini, and M. Cecilia Verri. "Finding Regularities in Courses Evaluation with K-means Clustering.", CSEDU - 6th International Conference on Computer Supported Education, 2014, Vol. 2, pp. 26-33.
- [9] A. Dutt, S. Aghabozrgi, M.A.B. Ismail, H. Mahroeian, "Clustering algorithms applied in educational data mining", *International Journal of Information and Electronics Engineering*, 2015, Vol. 5, pp. 280-291.
- [10] Peña-Ayala, Alejandro. "Educational data mining: A survey and a data mining-based analysis of recent works." *Expert systems with applications*, 2014, Vol.41 (4), pp.1432-1462.
- [11] Sarala, V., and J. Krishnaiah. "Empirical Study Of Data Mining Techniques In Education System.", *International Journal of Advances in Computer Science and Technology (IJACST)*, 2015, pp. 15-21.
- [12] C.Anuradha, T.Velmurugan, R. Anandavally, "Clustering algorithms in educational data mining: a review ", *International Journal of Power Control and Computation(IJPCSC)* Vol 7. No.1 – 2015 pp.47-52
- [13] Shirodkar, Jateen Shet, and Viren Pereira., "Determining Students Performance Using the Tool of Artificial Neural Network.", *International Journal of Innovative Research and Development*, 2016, Vol. 5 No. 2, pp. 314-318.
- [14] Veeramuthu, P., Dr R. Periyasamy, and V. Sugasini., "Analysis of Student Result Using Clustering Techniques." *IJCSIT*, *International Journal of Computer Science and Information Technologies*, 2014, Vol. 5, No. 4, pp. 5092-5094.
- [15] Sivaram, N., and K. Ramar., "Applicability of clustering and classification algorithms for recruitment data mining." , *International Journal of Computer Applications*, 2010, Vol. 4, No. 5, pp. 23-28.

- [16] Velmurugan. T and T. Santhanam, “Computational Complexity between K-means and K-medoids clustering algorithms for normal and uniform distributions of data points”, Journal of Computer Science, Vol. 6, Issue 3, 2010, pp.363-368.
www.thescipub.org/fulltext/jcs/jcs63363-368.pdf

