K-mean and EM Clustering algorithm using attendance performance improvement Primary school Student

Manmohan Singh, Harish Nagar, Anjali Sant

Department of Computer Science and Engineering,. Mewar University, Rajasthan, India

ABSTRACT

Based on the clustering methods such as centroid based, distribution based and density based clustering. Cluster includes groups in with small distance among the cluster members. The performance of student's multilevel of optimization formulated by using clustering. In centroid based clustering, clusters are represented by a central vector. The number of clusters is fixed to k, k-means clustering gives a formal definition as an optimization problem. The clustering model most closely related to statistics is based on distribution model. Experiments attempts to improve the accuracy by using the method of data mining using weak Tool.

Keywords: Centroid based, Density based Cluster, K-means algorithm, EM Clustring Algorithem

I. INTRODUCTION

The main objective of primary education institutes is to provide quality education to its students and to improve the quality of managerial decisions. One way to achieve highest level of quality in primary education system is by discovering knowledge from educational data to study the main attributes that may affect the students' performance [5]. The discovered knowledge can be used to offer a helpful and constructive recommendations to the academic planners in primary education institutes or School to enhance their decision making process, to improve students' academic performance and trim down failure rate, what is the basic Reason, and how to better understand students' behavior, to assist instructors, to improve teaching and many other benefits Improved educational data mining uses many techniques such as centroid based, and density based clustering. Cluster includes group with small distance among the cluster of Group members.

II. METHODS AND MATERIAL

Clustering Algorithm:

Data Clustering is unsupervised and statistical data analysis technique. It is used to classify the same data into a homogeneous group of primary school students it is used to operate on a large data-set to discover hidden pattern and relationship helps to make decision quickly and efficiently. Cluster analysis is used to break down a large set of data into subsets called clusters. Each cluster is a collection of data objects that are similar to one another. They are placed within the same cluster but are dissimilar to objects in other clusters. Following algorithms are used in education mining in Clustering.

1. K-Mean Clustering Algorithm:

K-Means is a non-hierarchical clustering method that seeks to partition the data into the form of one or more clusters [3, 8]. This method partitions the data into clusters so that the data having the same characteristics are grouped into one cluster and the data that have different characteristics grouped into another cluster. In general, K-Means algorithm is as follows:

1. If prev-grade=C, school type = private
School- area= urban and attendance=irregular then
final-grade=low.

2. If prev-grade=B, school type = private
School- area = rural and attendance=regular then final-grade= B

3. If prev-grade=A, , school type = private,
School- area= urban and attendance=regular , then final-grade=A

4. If prev-grade=B, , school type = government ,
School- area= urban and attendance=regular, then final-grade=A

5. If prev-grade=B, school type = government
School- area = rural and attendance=regular then final-grade= B

Primary school academic performance is measured by internal and external assessment, attendance, school type and area of school. External assessments are previous grade and final current grade. So, by taking the assessment and previous exam grade and by using data clustering technique, we can predict what will be the final grade of a student the proposed models try to identify/find out the weak students before final school exam in order to save them from serious harm. Teachers can take appropriate steps at the right time to improve the performance of primary students in final exam or next exam.

2. Expectation Maximization (EM) Algorithm:

Expectation Maximization Algorithm (EM-clustering) is used to cluster the given data. An EM algorithm is a mixture based algorithm that finds maximum likelihood estimates of parameters in probabilistic models. In our student primary education case, we used EM-clustering to group students according to their performance [7, 8]. It gives Mean of each cluster for each primary attribute. Using these results we can divide students into groups and guide them according to their previous Result, School Type and school attendance. Expectation Maximization Algorithm in educational data mining. It showed how useful data mining can be in primary education particularly to improve student performance. We used students' data from database result improvement. We collected all available data including their usage of our proposed area study of Betul District. We applied data mining techniques. Also we clustered the student into group using EM-clustering. Finally, using outlier analysis, we detected all outliers in the data.

Each one of this knowledge can be used to improve the performance of primary students.

III. RESULTS AND DISCUSSION

1. S **K-means clustering Algorithm Based on Student Attendance:**

Here, we have clustered the data based on the attendance in School It means that the students in the school are regular or irregular.
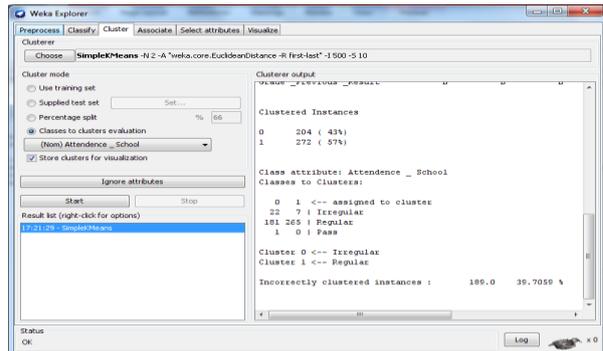
Figure 1.1 : Applying K-means clustering Algorithm Based on Student Attendance

2. Expectation Maximization Algorithm Based on Student Attendance
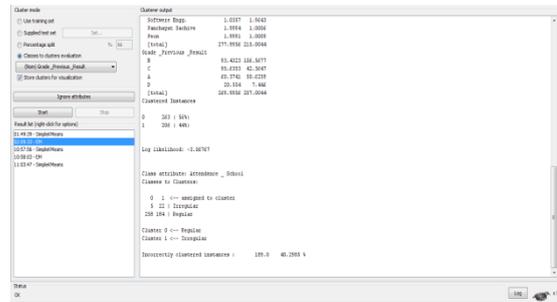
Figure 1.2 : Applying EM clustering Algorithm Based on Student Attendance

Table 1.1 Performance Compression Approach Based on Student Attendance- K-Mean and EM Algorithm (Accuracy)

	Cluster 1	Cluster 2	incorrect cluster
K-means	61%	39%	35%
EM	56%	44%	40%
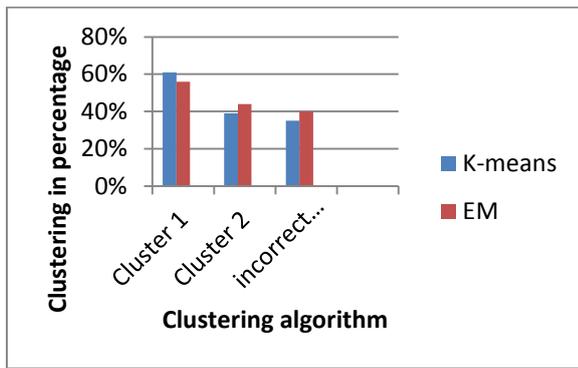
Figure 1.3 : Performance Compression Approach Based on Student Attendance- K-Mean and EM Algorithm (Accuracy)

data objects as initial centroids and applying the k-means algorithm and EM Algorithms we obtain the following results from the student data set consisting of 442 students.35 percentages of the and 40 percentages students incorrect cluster category , cluster 2 ,39 percentages of the students and 44 percentages cluster 2 category of (Highest percentage of Grade, Medium attendance in percentages), 22in percentages of the students fall in the category of (Medium percentage of Grade, Highest attendance in percentages) 61 percentages of the students attendance and 56 percentages attendance of student cluster 1 the category of (highest percentage of grade , highest attendance in percentages). Analyze the problem of students and overcome the lacking points so that the students can perform well than what they are performing after attending most of the usual classes' attendance. Now, the second group of students requires proper attention of the faculty which motivates them to attend the classes and time to time mentoring.

IV. REFERENCES

[1] Anand Kumar N V, Uma G V 2009, Improving Academic Performance of Students by Applying Data Mining Technique, European Journal of Scientific Research ISSN 1450-216X Vol.34 No.4 , pp.526-534.

[2] Ayesha S, Mustafa T, Sattar A.R., and.Khan, M.I 2010 Data Mining Model for Higher Education System, European Journal of ScientificResearch, Vol.43, No.1, pp.24-29

[3] Brijesh Kumar Baradwaj & SaurabhPal, 2011"Mining Educational Data to analyses Student's performance", (IJACSA) International Journalof Advanced Computer Science and Applications, Vol. 2, No. 6, 2011, p 63-69.

[4] Bray M 2007, the Shadow Education System: Private Tutoring And Its Implications For Planners, (2nd ed.), UNESCO, Paris, France.

[5] Banerjee, Abhijit, Rukmini Banerji, Esther Duflo, and MichaelWalton. 2012. Effective Pedagogies and a Resistant Education System:Experimental Evidence on Interventions to Improve Basic Skills in Rural India. MIT

[6] Chandra, E. and Nandhini, K. 2010 .Knowledge Mining from Student Data', European Journal of Scientific Research, vol. 47, no. 1, pp. 156-163.

[7] Carlos Márquez-Vera, Cristóbal Romero Morales, and Sebastián Ventura Soto. 2013, 'Predicting School Failure and Dropout by Using Data Mining Techniques' , IEEE Journal Of Latin-American Learning Technologies, Vol. 8, No. 1, Feb 2013 .

[8] Dréze, Jean and Kingdon, Geeta, 2000. "School Participation in Rural India," Review of Development Economics.

[9] Delavari, N., Beikzadeh, M.R. (2004). A new model for using data mining in higher education system, 5th international Conference on Information Technology based Higher education and training: ITEHT (04), Istanbul, Turkey, 31st m\May-2nd June 2004.

[10] Das J,S, dercon , J Habyarimana ,P Krishna ,K Muralidharana and V.sundararaman 2013 " School Input ,hose hold substitution and test scores" American Economic Journal Applied Economics 5(2), p-29-29.

[11] Ghaida Abu, D, and Stephan Klasen (2004), The Costs of Missing the Millennium Development Goal on Gender Equality." World Development32 (7): 1075–107

[12] Han, J., & Kamber, M. (2006). Data mining: Concepts and techniques (2nd ed.). Boston, MA: Elsevier.

[13] Johnstone, James and Jiyono. 1983. "Out-of-school Factors and Educational Achievement in Indonesia." Comparative Education Review, 27(2), pp. 278-295.

[14] Jaiwei Han and Micheline Kamber, 2008 "Data Mining Concepts and Techniques", Second Edition Morgan Kaufmann Publishers.

[15] James, Estelle, Elizabeth M King, and Ace Suryadi. 1996. "Finance, Management, and Costs of Public and Private Schools in Indonesia." Economics of Education Review, 15(4), pp. 387-98

[16] Jayaraman, Rajshri, Dora Simroth, and Francis De Vericourt. 2010. The Impact of School Lunches on Primary School Enrollment: Evidence from India's Mid-Day Meal Scheme. Indian Statistical Institute.

[17] Kotsiantis, S., Pierrakeas, C., Pintelas, P. 2004. Preventing student dropout in distance learning systems using machine learning techniques, In International Conference on Knowledge-Based Intelligent Information & Engineering Systems, Oxford, 3-5.

[18] Khan Z, Khan N 2005. "Scholastic Achievement of Higher Secondary Students in Science Stream", Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87.