International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

Prediction of Students Vulnerabilities using Clustering Technique

Madhvi A. Bera[1], Jinesh H. Jagani[2], Darshil H. Patel[3], Isha M Hirpara[4], Prarthi K Thakkar[5]

[1]*Assistant Professor, Department of Computer Engineering, Indus University Ahmedabad, India*
[2]*Student, Department of Computer Engineering, Indus University Ahmedabad, India,*
[3]*Student, Department of Computer Engineering, Indus University Ahmedabad, India,*
[4]*Student, Department of Computer Engineering, Indus University Ahmedabad, India,*
[5]*Student, Department of Computer Engineering, Indus University Ahmedabad, India,*

Abstract - The academic advancement of students' performance is an issue of utmost importance. In recent times, the biggest adversity that educational institutions are facing is the proliferated growth of data, and the usage of this data to improve the quality, is continuously getting complex. A systematic step-by-step algorithm for analyzing their performance, based on cluster analysis, needs to be applied on the data, in order to get meaningful information. In the following paper, algorithms of data clustering like K-means, K-medoids, Fuzzy C-means and Expectation Minimization have been discussed. A study on the comparison of K-means and K-medoids clustering algorithm has also been done, in order to examine their respective efficiency. In this manner, an institution can have a better understanding of students' vulnerabilities in order to achieve better management.

Key Words: mining, K-medoids, k-means,Fuzzy C-Means

1. INTRODUCTION

Data mining is a procedure which extracts formerly hidden, plausible, greatly useful and unknown patterns from large amount of data. The amount of data in the field of education is increasing day-by-day. Thus, in order to get desired advantages from such a large amount of data, and to find concealed relations between various entities, numerous data mining techniques have been developed and are being used. There are various techniques for clustering of data, which include techniques like k-means clustering, fuzzy c-means clustering, hierarchical clustering, etc. In this paper, we have discussed upon k-mean clustering algorithm for scrutinizing students' academic progress' data. It remains the most efficient, unsupervised data mining algorithm, for the clustering, or segregation of students on the basis of their academic performance. It operates on the notion of Euclidean distance and eventually calculates the cluster centroids. The clustering will further help both educators, as well as students, in improving their performance, respectively.

The student's performance plays a significant role in the progress of any successful institution. As new institutes are emerging day by day and also the number of students is increasing, institutes are becoming performance oriented and correspondingly their goals and strategies for achievements are getting firm. The performance of the student during a given semester is evaluated by the final result of analysis, and steps are taken accordingly to improve students' performance. In this paper, various data mining methods such as K-means, K-medoids, Fuzzy C-means (FCM), Expectation minimization, etc. are presented, that are used in evolution and progress of student's performance. K-means is iterative, non-deterministic, unsupervised method. The method has been proven efficient in producing good clustering results, due to its simplicity and ability to solve faster. K-medoids method is one of the partitioning methods. It uses the medoids i.e. the most centrally located object in the cluster, and is less sensitive to the out liners as well as it is more time consuming compared to the K-means approach. Fuzzy C-means is a clustering method in which each data point can belong to more than one cluster.FCM is mostly used in pattern recognition. In this algorithm, according to the distance between the cluster center and data point, membership is assigned to each data point, corresponding to each cluster center. Summation of membership of every data point should be a single entity. Expectation minimization is another iterative process. In the presence of missing and hidden data, expectation minimization (EM) algorithm is used to calculate the Maximum Likelihood (ML). ML is then used to estimate the missing or hidden parameters of the cluster from all of the observed parameters.

2. LITERATURE REVIEW

Govindaswamy and T.Velumurugan carried out a study on various clustering algorithms. The main objective of this paper was to present a comparison between various algorithms used for student evaluation system. The aim is to create groups of students according to their characteristics and performance. There often exists some hidden relationships between various variables which usually cannot be simply pointed out by us which is served by these clustering algorithms. So, further in this method we take three clusters (Good, Better, Excellent) and then we evaluate these clusters on the basis of no. of clusters generated, execution time, purity and NMI. The methods taken into consideration were k-means, k-Medoids, Fuzzy C Means(FCM) and Execution Minimization (EM). The results generated through this research were that FCM and EM performed well in all the factors taken into consideration rather than other two algorithms.

Oyelade O.J, Oladipupa O.O and Obagbuwa I.C, suggested a way of combining the k-means clustering method with a deterministic model. The main objective of this paper is to plan a good academic progression for students and to ensure a proper evaluation for the students to enter a certain institution and in turn to help the academic planner for the selection of the former. The aim of this paper primarily is to present k-means clustering algorithm as a very proficient tool to evaluate a student's academics in a higher institution. So, as mentioned the combination of k-means clustering algorithm and the deterministic model was done and applied to a data set of a private school with nine courses offered and the number of students were 79. And it was further observed that this combination of two models was successful in overcoming some of the limitations from the existing model and further it served as a good benchmark to evaluate a student's performance in higher institution.

Prerna joshi and Pritesh jain suggested an approach in which k-means and k-medoids are combined for the proper examination of the student's academic performance in a particular institution. The main objective of the paper is to make the screening of the students for the academic area very efficient and easy. The aim is to present the two algorithms and display a proficient method to evaluate a student's performance in a specific institution. Further, k-means and k-medoids algorithms are applied and the comparison is presented with the decision tree algorithm which is being currently used and the number of clusters taken for the same is six.

Table- 1:

	Current Method	Proposed Method	
Parameter	Decision tree	K-means	K-medoids
All subject percentage	Above 68%	Above 71%	Above 73%

In sum, the grouping calculation serves a great k-medoid in comparison to k-means for the purpose of on screen for the evaluation of the progression of the students in the higher institution. It even likewise upgrades the decision making towards the faculty members should screen the candidate's execution semester wise towards upgrading on the future scholastic achievements.

Zhongxiang Fan, Yan Sun and Hong Luo have studied k means clustering for improving the knowledge of college about students' progression in order to maintain proper management. The problem of outliers influencing the result is solved using K-means algorithm, which is based on grid density. The method proposed for removing the outliers is as follows:

i. Sort the points in the data set according to the values.
ii. Now traverse all the points (P).
iii. For the current point traverse the other points front (P+1) and back (P-1) in range, until all dimensions are judged.
iv. The point is considered as an outlier if the grid density of the point is less than the threshold assumed.

This method reduced the time taken by traditional k-means algorithm and provided faster results. The next step after removing the outliers is selecting the initial cluster centers. The traditional way selected the cluster center randomly. The new algorithm instead divides the data into k segments; the average value of each segment will be the coordinate of the cluster center of that segment. This will increase the distance between the cluster centers as well as between the clusters. The large difference will lead to better results in less time than the traditional way.

Md. Hetayul Islam shovon and Mahfuza Haque have studied and discussed the use of decision tree and data clustering for the improvement of the students' GPA in engineering and Science University. When clustering the data they divided the clusters into 3 main categories: good, medium and low standard student. The decision tree is a tree where the root and internal node have a question and according to that the arcs are emanated from each node. Each arc represents all the possibilities of answers to all the questions in the node. The techniques discussed in the paper are very helpful in tracking a student's progress. This proves helpful in reducing the failure ratio by implementing appropriate measures to improve the education quality. These techniques do not provide the best results in terms of accuracy. In the future the techniques can be improved by finely tuning the parameters and changing the algorithm for the better use of it in this aspect.

Ramdayal Tanwar and Dr. Rajeev Kumar Gupta have analysed and compared the use of k means clustering and decision tree in order to bring some hidden information from student's database, which can be helpful in improving the student's performance. They took the same sample database and applied both k means clustering and decision tree algorithm in order to determine which one is better suited for predicting the student's performance more accurately. It was found that the single decision tree was relatively inexact than random forest of decision trees, but random forest of decision trees are harder to implement than single decision trees. Finally it was concluded that decision trees were rather inefficient and less stable -than k-means. Hence using k-means and constantly improving the algorithm will be very useful for the teaching institutes to develop student's based on their performance.

3. ANALYSIS

From 50 data tuples, we have shown 10 here.

Table - 2: students result

Roll No.	Name	Average	Grade
1	Sharveel	48	A
2	Savan	41.2	B
3	Deepa	44.5	B

4	Shubham	45	A
5	Dhruvin	45.6	A
6	Binny	45.6	A
7	Jenil	44.9	B
8	Rajat	40.2	B
9	Sanjana	48.8	A
10	Raviraj	39.3	C

Here we apply K-means clustering algorithm on the given data, eventually classifying them into 4 classes or grades- 'A', 'B', 'C' and 'D'. The above table 2 shows the relation of average marks and grade.

Table - 3: No. of students based on grade and percent

Grade	Number	Percent
A	26	52
B	16	32
C	4	8
D	4	8

In Table 3, we can see that grade A contains 26 students, which constitute 52% of the total students. Grade B contains 16 students which constitute 32%, and so on.

4. RESULTS

The graphical representation of number of students belonging to a particular grade, and their percentage, is given as Chart- 1

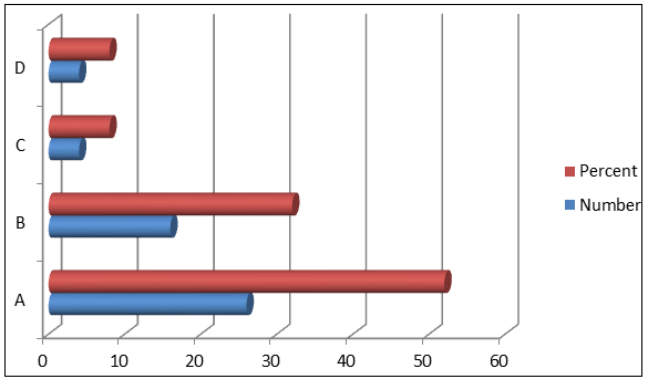


Chart -1: Relation between grade and percentage

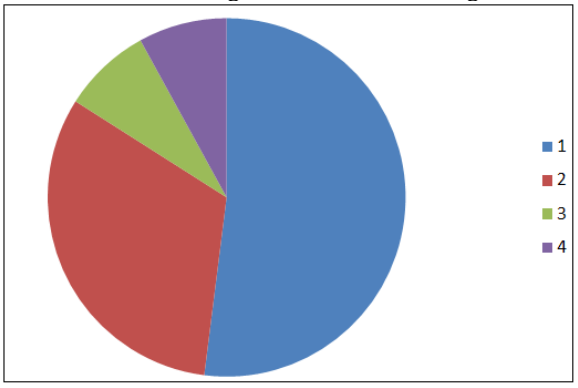
After applying the K-means clustering algorithm, we group the students into four grades- A, B, C and D see in table - 4.

Table - 4: students group based on grade

Class	Average	Number	Percent
A	45-60	26	52
B	40-44.9	16	32
C	35-39.9	4	8
D	0-34.9	4	8

Graphical representation of the above mentioned table, can be viewed in a pie chart as:

Chart - 2: Percentage of Students' Average marks



5. CONCLUSION

K-means remains one of the most applicable algorithms in case of data clustering. In this study, we have briefly defined the K-means algorithm and examined the result of students in order to enhance the future teaching system on the basis of students' academic performance. In this paper, we have compared the predictive capability of two different clustering algorithms – K-means and K-medoids. To conduct this analysis, we have used the K – means clustering algorithm, using Weka as the statistical tool. Thus, the K-means algorithm serves well for the purpose of clustering of students on the basis of academic aspects.

REFERENCES

[1] Ramdayal Tanwar1, Dr. Rajeev Kumar Gupta2 "Analysis of Students Performance using Modified K-Means Algorithm (Machine Learning Techniques) International Journal for Research in Applied Science & Engineering Technology (IJRASET) ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177 Volume 7 Issue X, Oct 2019

[2] Deepshikha Aggarwal, Deepti Sharma "Application of Clustering for Student Result Analysis" International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-6C, April 2019

[3] Anil Kumar Pandey, Sachin Saxena "Implementation clustering approach for prediction of Academic Performance" International Journal of Innovations in Engineering and Technology (IJIET) ISSN: 2319-1058,Volume 12 Issue 2 January 2019

[4] Ei Ei Phyo, Ei Ei Myat "Efficient K-Means Clustering Algorithm for Predicting of Students' Academic Performance" International Journal of Engineering Trends and Applications (IJETA) – Volume 5 Issue 6, Nov-Dec 2018

[5] Prerna Joshi, Pritesh Jain "Prediction of Students Academic Performance Using K-Means and K-Medoids Unsupervised Machine Learning Clustering Technique" © June 2018 IJSDR | Volume 3, Issue 6 ISSN: 2455-2631

[6] K. Govindasamy, T.Velmurugan "ANALYSIS OF STUDENT ACADEMIC PERFORMANCE USING CLUSTERING TECHNIQUES" International Journal of

Pure and Applied Mathematics Volume 119 No. 15 2018, 309-323 ISSN: 1314-3395

[7] Zhongxiang Fan, Yan Sun and Hong Luo "Clustering of College Students Based on Improved K-means Algorithm" Journal of Computers Vol. 28, No. 6, 2017, pp. 195-203 doi:10.3966/199115992017122806017

[8] Snehal Bhogan, Kedar Sawant, Purva Naik, Rubana Shaikh,Odelia Diukar, Saylee Dessai " PREDICTING STUDENT PERFORMANCE BASED ON CLUSTERING AND CLASSIFICATION" IOSR Journal of Computer Engineering (IOSR-JCE)e-ISSN: 2278-0661,p-ISSN:2278-8727, Volume 19, Issue 3, Ver. V (May-June 2017), PP 49-52

[9] Mr. Shashikant Pradip Borgavakar, Mr. Amit Shrivastava " Evaluating Student's Performance using Means Clustering" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Vol. 6 Issue 05, May - 2017

[10] Mrs. Biradar Usha "Knowledge Discovery to Analyze Student Performance using k-mean Clustering depend upon various mean values input methods: A Case Study" International Journal of Advanced Research in Computer Science Volume 6, No. 2, March-April 2015

[11] J. James Manoharan, Dr. S. Hari Ganesh, M. Lovelin Ponn Felciah "Discovering Students' Academic Performance Based on GPA using Means Clustering Algorithm" 2014 World Congress on Computing and Communication Technologies

[12] Rakesh Kumar Arora, Dr. Dharmendra Badal "Evaluating Student's Performance Using k-Means Clustering" IJCST Vol. 4, Issue 2, April - June 2013 ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

[13] Md. Hedayetul Islam Shovon, Mahfuza Haque "An Approach of Improving Student's Academic Performance by using K-means clustering algorithm and Decision tree" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol.3, No. 8, 2012

[14] Oyelade, O. J, Oladipupo, O. O, Obagbuwa, I. C "Application of k-Means Clustering algorithm for prediction of Students' Academic Performance" (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, o. 1, 2010