



ORIGINAL RESEARCH PAPER

Medicine

ENHANCED STUDENT PERFORMANCE ANALYSIS IN GRADUATION USING K MEAN CLUSTER

KEY WORDS: clustering; K-means algorithm; data mining; overall performance; patterns;

P Indira Menon

Vedavyasa Institute of Technology, Malappuram, Kerala, India

S.Kavitha

Vedavyasa Institute of Technology, Malappuram, Kerala, India

ABSTRACT

Education is not the filling of a pail , but the lighting of a fire –William Butler Yeats . Concerning with the importance of the education, the key aspect is needed to measure the achievement levels in higher education. Challenges are faced by universities and technical organizations in the analysis of students performance. The major challenges are in admission of freshers and outgoing students placement. The ability to monitor the progress of students academic performance is a critical issue to the academic community The ranking of the university depends on academic performance and placement of the students. Apart from academic performance there are various other factors which help in understanding the overall performance of the students. Data Mining is an analytic process designed to explore data in search for consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data. Hence analysis is focused on the implementation of a clustering data mining technique and a method for acquiring student overall performance during their entire term. K-means is one of the simplest unsupervised learning algorithms used for clustering. So system is analyzed the performance of students using k-means clustering algorithm . Instructor plays a vital role in the performance of the student .Instructor less attention towards students affects their performance. The data mining of the technique is used here is to understand the performance of student and group the students under various categories as a student need to consistently improve and instructor to motivate them to compete vie in today's world.

I. INTRODUCTION

The student's performance plays vital role in success of any institution. With the significant raise in number of students and institutions, institutions are becoming increasingly performance oriented and are accordingly setting goals and developing strategies for their achievements. K-means clustering technique is used here to analyse the performance of students.This results will help the academic planners in evaluating the performance of students during specific semester and steps that need to be taken to improve students' performance from next batch onwards. These collected information has to be converted to useful form for better decision making in any university , there are volumes of student data like attendance, marks etc With traditional approach of grouping students based on their average scores, it is difficult to obtain a comprehensive view of the state of the students' performance and simultaneously discover important details from their time to time performance [1] Oyelade. Performance analysis plays a very important role in higher institutions and student mark is an important factor. The most important parameter that is used to judge a student performance in the college are marks .Marks of each student is analysed in order to find the faculty performance .The other parameters like projects completed , internships and skill set also play a vital role in employment opportunity for a student Dr.K.Kala [5] . Therefore grouping of students based on these parameters is required to obtain a comprehensive view of the performance of the student and simultaneously ascertain details from their time to time performance. This paper is focussed on the implementation of a data mining technique and a method for acquiring student overall performance during their entire term . In the analysis, uses K-means clustering algorithm for categorizing students in different clusters. It will help the student teachers to focus on improvement strategies by way of monitoring the performance of the student.

LITERATURE REVIEW

A. Data Mining Techniques

Data mining is the process of finding information from the data set. Data mining obtains knowledge from the given data . It deals with various kind of patterns. Data mining technique are broadly classified as: Supervised and Unsupervised learning techniques. K.Shanmuga Priya [3] The Learning techniques are broadly divided as classification, prediction , association , clustering . The classification technique are Decision tree, Nadve Bayes etc. The prediction techniques are Linear Regression , Multi linear Regression etc. The association technique are Apriori , Frequent Patter Tree and K Mean, DBSCAN , hierarchical are some of the clustering techniques. Some data mining research work has been

conducted for the performance of students. In the analysis presents k-means clustering algorithm as a simple and efficient tool to monitor the progression of students' performance in higher institution Oyelade [1]. Another work is Ahmed [2] the attribute is Overall semester marks , practical lab, attendance , paper presentation , end semester marks Their usage is student performance improvement.

B. Clustering Technique (K-Mean)

Clustering is concerned with grouping the similar objects together in a cluster and grouping the dissimilar objects in a different cluster. The center of a cluster is known as centroid may be part of this cluster or an imaginary point . Clustering has widespread applications such as in data mining , pattern recognition , compression , compression and machine learning . The clustering algorithm used in this analysis in K-means algorithm. K-Means algorithm splits the number of observation into 'K' clusters where every observation belongs to the cluster with the nearest mean. This algorithm is NP-Hard to optimize , but efficient heuristic algorithm can be applied to converge quickly to local optimum. The number of clusters should be pre specified to apply this algorithm , ie state the value of k. The performance of this algorithm largely depends on the value of 'K',

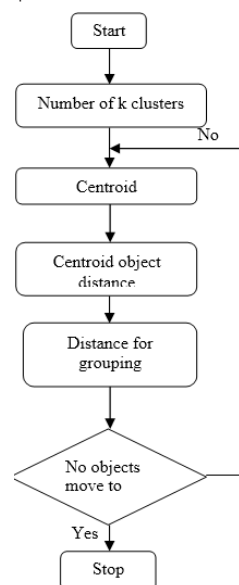


Figure 1: Flowchart of K Means Clustering

After K value is selected , the algorithm is applies as follows:

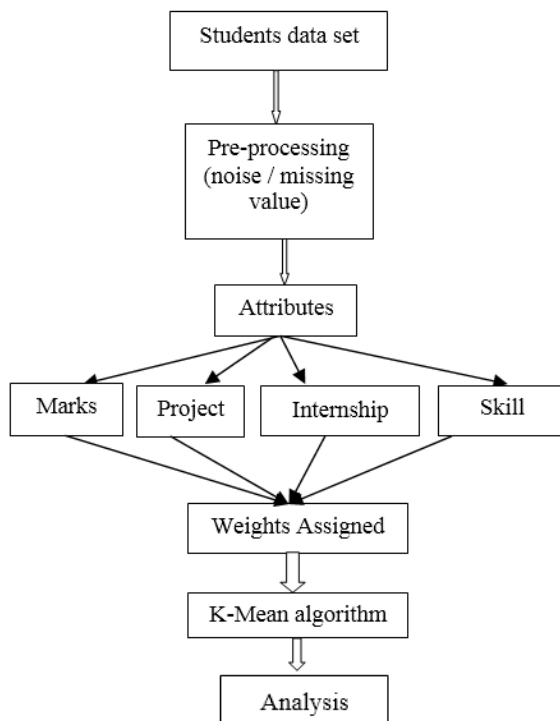
1. Select the initial cluster centers from the given instances randomly equal to the value of 'K'.
2. Now assign all the instances to the closest cluster center.
3. Now every cluster center is updated by taking mean of constituent instances.
4. After assigning all objects , the position of K centroid is recalculated .

Repeat 2nd and 3rd steps till there is no further changes in assignment of instances to cluster.

The value of 'K' with highest average silhouette width is chosen for the data set.It is calculated by using the following procedures: "For each data i , let $a(i)$ is average dissimilarity of ' i ' with all other data within the same cluster and $b(i)$ be the lowest average dissimilarity of ' i ' is not a member .It is formulated as: $S(i) = b(i) - a(i) / \max\{a(i), b(i)\}$

II. METHODOLOGY

Firstly , data mining project is collection of data .The data for section of Computer science student was considered.Collected data was pre-processed and the missing values are removed .Data were normalised and suitable weights were assigned to relevant attributes of the data .The number of clusters that are needed in k-means clustering algorithm are calculated by taking the maximum of silhouette measure among all the values of 'k'.K-means algorithm is applied on the pre-processed data set for obtaining the cluster.


Figure 2: Architecture of Student Performance Analysis

IV. EXPERIMENTAL SETUP

A. Data-set: Collection of data was done through students performa . The data set was preprocessed for missing and noise value . A sample of 59 data was obtained for processing . To evaluate the performance of students the various parameters considered are graduation marks , skill sets, projects , internships backlog and guide . The analysis was further extended by considering Xth and XIIth marks of the student. The data set was preprocessed and made error free.

B. Attribute Selection: The attributes like Xth , XIIth and graduation marks along with projects,internships and skill were considered as relevant and Date of Birth , Gender , 10th passing year , 12th passing were ignored(irrelevant) .

C. Weights for the attributes were assigned as follows

1. Marks weight : Marks play a vita, role in the performance of the student . The aggregated marks of Xth , XIIth , B.tech was combined for the mark score. Marks were moderated to 1-10 and further a overall scale of 1-4 was considered .

2. Skills weight : Knowledge of computer language is an important requirement to determine suitable candidate during placement . Therefore suitable weights were added and a cumulative weight score for the skill set was calculated and was normalised in the scale [0-2].Skills like C-1, C++ - 2, Java-3,Python - 3, HTML-2, CSS-1, SQL-3.shown in table 1.

Table 1: Skill Set

C	C++	JAVA	PYTHO	SQL	CSS	HTML
1	2	3	3	3	1	2

3. Back Paper / Backlog Weight : Weights were assigned to students having backlog during their course work .These weights were further normalized in the scale of [0-1] for processing . Number of backs are 0 , 1, 2or more and weights assigned are 10,5,0. Shown in table 2.

Table 2:Back Paper

Number of backs	0	1	2 or more
Weights assigned	10	5	0

4. Projects weight : This attribute tell us about the number of projects that are completed by a student . The scale for this attribute is[0-2].

5. Internships weight : This attribute explains the real time industry experience of a student and a score of [0-1] was given.

E. Identification of 'K' : This value should be chosen such that it should reflect some characteristics of the data set . The acceptable value of 'K' was calculated using silhouette measure. The value of K was taken as 3,4,5,6,7,8. The index of k=5 was found to have high silhouette measure.

V. RESULT & ANALYSIS

A. Analysis of selected attributed

The attributes selected were marks, projects, internships, skills set and the overall performance of students based on the index given in table 3.

Table 3: Student Performance Index

7.1 and above	Excellent
7.0-6.51	Good
6.50-5.51	Average
5.50-4.51	Below Average
4.5 and below	Need Improvement

VI. CASE STUDIES

A. ACADEMIC PERFORMANCE (MARKS)

Tenth Overall Performance

The performance of 10th is shown in table 3

Table 3: Tenth Cluster Measure

Cluster	Cluster size	Overall Performance
0	27	9.0
1	14	6.51
2	5	5.51
3	10	4.51
4	3	1.0

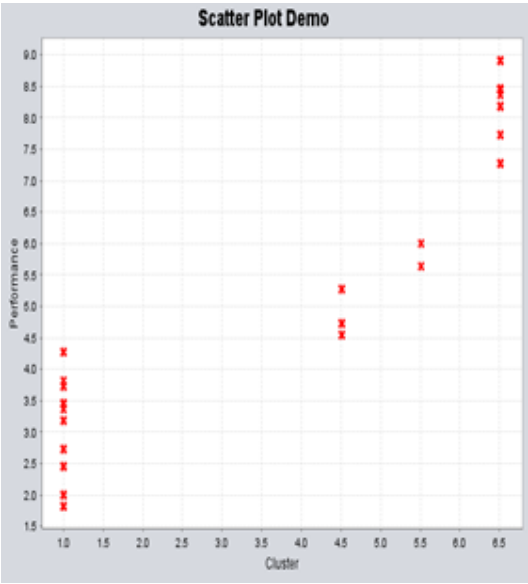


Figure 4: 10th performance graph

Tenth performance is shown in fig 3 ,In cluster 0,cluster size is 27 and their overall performance figured is 9.0 that is “excellent”. In cluster 1,cluster size is 14 and their overall performance figured is 6.51 that is “good”. In cluster 2,cluster size is 5 and their overall performance figured is 5.51 that is “average” .cluster size 3 have 10 below average students which is 4.51 that is below average cluster 4 have 3 students which is 1.0 means need improvement.

Twelfth Overall Performance

The performance of 12this shown in table 4

Table 4: Twelfth Cluster Measure

Cluster	Cluster size	Overall Performance
0	20	10
1	18	6.51
2	8	5.51
3	10	4.51
4	3	1.0

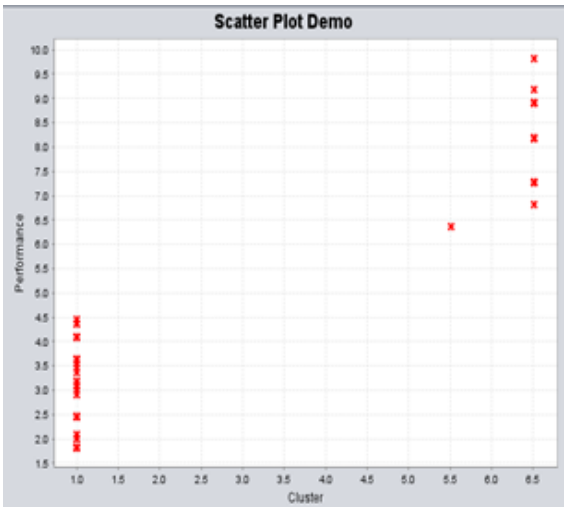


Figure 5: 12th performance graph

Twelfth performance is shown in fig 4 In cluster 0,cluster size is 20 and their overall performance figured is 10 that is “excellent”. In cluster1,cluster size is 18 and their overall performance figured is 6.51 that is “good”. In cluster 2,cluster size is 8 and their overall

performance figured is 5.51 that is “average” .cluster size 3 have 10 below average students which is 4.51 cluster 4 have 3 students which is 1.0 means need improvement

Graduation Overall Performance

The performance of graduation is shown in table 5

Table 5: Graduation Cluster Measure

Cluster	Cluster size	Overall Performance
0	16	8.45
1	8	6.53
2	12	5.51
3	20	4.51
4	3	1.0

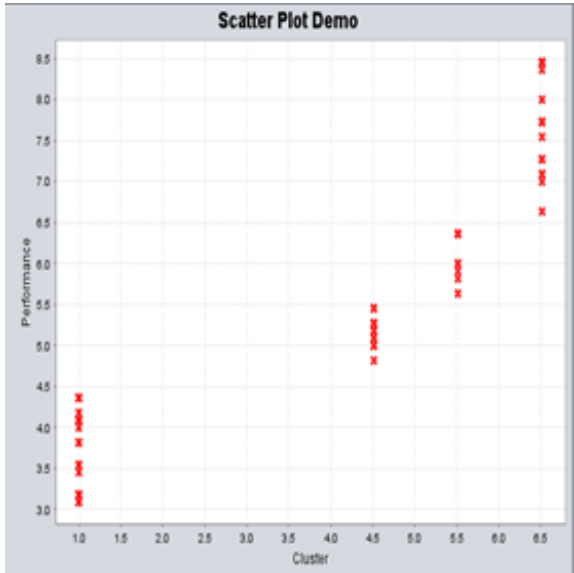


Figure 6: Btech performance graph

Graduation performance is shown in fig 5 In cluster 0,cluster size is 20 and their overall performance figured is 10 that is “excellent” . In cluster1,cluster size is 18 and their overall performance figured is 6.51 that is “good” . In cluster 2,cluster size is 8 and their overall performance figured is 5.51 that is “average” .cluster size 3 have 10 below average students which is 4.51 cluster 4 have 3 students which is 1.0 means need improvement

Overall Performance Score

Based on the above index measures the student performance score was calculated . This include cluster, cluster size, and overall performance The results are shown in table 6

Table 6: Overall Cluster Measure

Cluster	Cluster size	Overall performance
0	26	8.45
1	15	6.51
2	5	5.51
3	11	4.51
4	2	1.0

In cluster 0,cluster size is 26 and their overall performance figured is 8.45 that is “excellent”. In cluster 1,cluster size is 15 and their overall performance figured is 6.51 that is “good”. In cluster 2,cluster size is 5 and their overall performance figured is 5.51 that is “average” .cluster size 3 have 10 below average students which is 4.51 that is below average cluster 4 have 3 students which is 1.0 means need improvement.

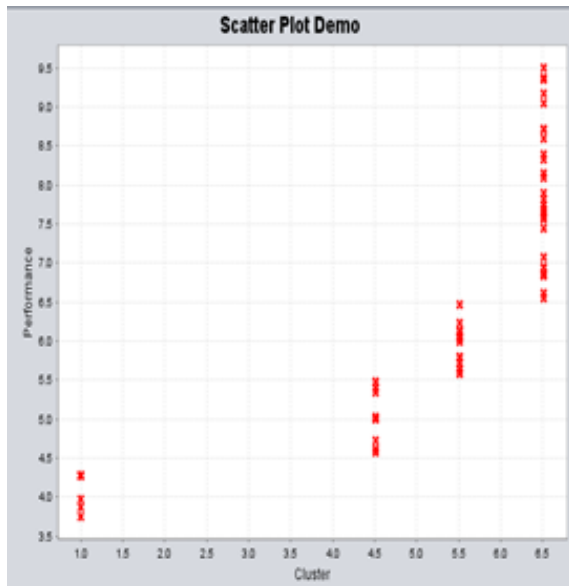


Figure 3: Overall performance plot

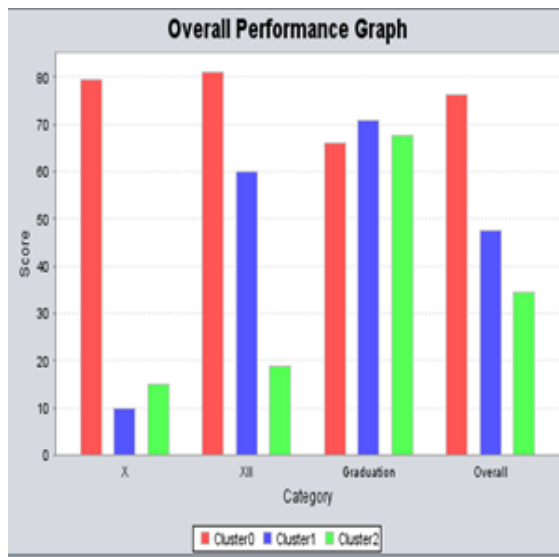


Figure 7: Academic Performance of students

Academic performance of students can be analysed using overall performance. Shown in fig 6 In cluster 0, cluster size is 20 and their overall performance figured is 10 that is "excellent". In cluster1, cluster size is 18 and their overall performance figured is 6.51 that is "good". In cluster 2, cluster size is 8 and their overall performance figured is 5.51 that is "average". cluster size 3 have 10 below average students which is 4.51 cluster 4 have 3 students which is 1.0 means need improvement The faculty performance graph shown in fig 7, here faculty percentage is according to students scores of their subjects.

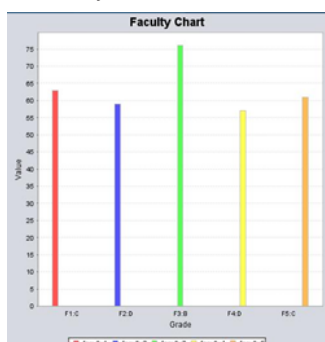


Figure 8: Faculty performance graph

B. COMPARATIVE ANALYSIS

The performance of students based on graduation without motivation and graduation with motivation. It was found that the number of students with motivation shown greater performance than students without motivation in fig 8. Therefore, projects, internships, and skills set should also be considered in addition to graduation marks for grouping students in different categories which in turn helps in enhancing the student performance.

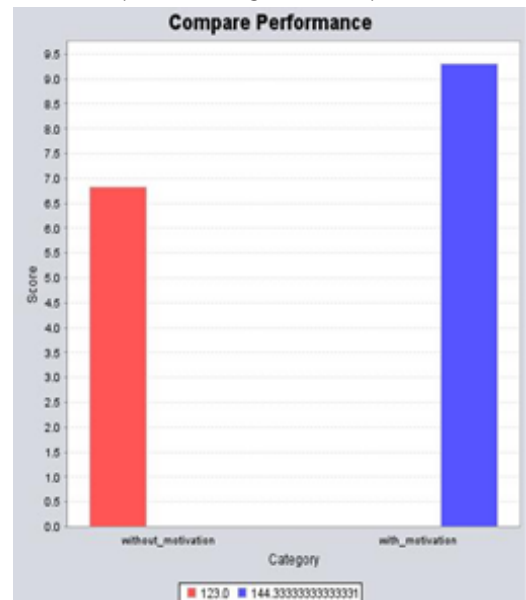


Figure 9: Compare performance graph

VII. CONCLUSION

Clustering analysis is used to understand the student's behaviour. Data mining algorithm served as a good standard to understand whether there is a consistent improvement in students performance. Showed how useful data mining can be used in higher education particularly to improve graduate students' performance. The analysis is helpful during the admission and placement process. The parameters taken for this analysis are: Xth, XIIth, Graduation marks, projects, internships, skill-set and guide. Based on the clustering methods such as centroid based, cluster includes groups with small distance among the cluster members. Also clustered the students into groups using K-Means clustering algorithm. The clustering algorithm implemented is K-Means as it is simple to implement and due to its computational efficiency. It was found that the number of students with motivation shown greater performance than students without motivation. Students are motivated by instructor giving more attention on improving the skill sets. Therefore, this clustering algorithm serves as a good benchmark to monitor the progression of students' performance in higher institution. It also enhances the decision making by academic planners to monitor the candidates' performance semester by semester by improving on the future academic results in the subsequent academic session. The work can be further extended using other clustering technique like fuzzy. Furthermore, the objects within the clusters can be ranked to get an enhanced students performance analysis.

REFERENCES

- [1] Oyelade, O. J., Oladipupo, O. O., Obagbuwa, I. C., (2010) Application of k-Means Clustering Algorithm for prediction of Students' Academic Performance. (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 1.
- [2] Abeer Badr El Din Ahmed , Ibrahim Sayed Elaraby (2014) Data Mining: A Prediction For Student's Performance Using Classification Method, World Journal of Computer Application and Technology 2(2):43-47.
- [3] K. Shanmuga Priya , A.V. Senthil Kumar, (2013) Improving the Student's Performance Using Educational Data Mining, Int. J. Advanced Networking and Applications Volume: 04 Issue: 04 Pages: 1680-1685
- [4] Shiwani Rana, Roopali Garg (2016) Evaluation Of Student's Performance Of An Institute Using Clustering Algorithms, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 5 (2016) pp 3605-3609.
- [5] Dr. K. Kala, (2017) Student's Skills Evaluation Techniques Using Data Mining, IJSR

Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 19, Issue 1, Ver. IV.

- [6] Saurabh Pal, (2012) Mining educational data to reduce dropout rates of engineering students, International Journal of Information Engineering and Electronic Business (IJIEEB), vol. 4, no. 2, pp. 1.
- [7] Chady El Moucary, Marie Khair, Walid Zakhem, (2011) Improving Student's Performance Using Data Clustering And Neural Networks In Foreign-Language Based Higher Education, The Research Bulletin Of Jordan ACM, Vol. 11(111).
- [8] Bindiya M Varghese, Jose Tomy J, Unnikrishnan A, Poulse Jacob K,(2013) Clustering Student Data To Characterize Performance Patterns, International Journal of Advanced Computer Science and Applications, Special Issue on Artificial Intelligence (IJACSA).
- [9] J.K. Jothi Kalpana, K. Venkatalakshmi,(2014) Intellectual Performance Analysis Of Students By Using Data Mining Techniques, International Conference on Innovations in Engineering and Technology (ICIET) Vol 3.
- [10] V Sreenivasarao, (2012) Improving Academic Performance Of Students Of Defence University Based On Data Warehousing And Data Mining, Global Journal Of Computer Science And Technology Volume 12 Issue 2 Version 1.0.