Student Educational Performance Analysis Using Clustering Algorithm

ISSN (e) 2520-7393
ISSN (p) 2521-5027
Received on 3rd Dec, 2018
Revised on 25th Dec, 2018
www.estirj.com

Anam Farooque[1], Dr.Tariq J. S Khanzada[2], Arbab Ali Samejo[3]

[1]*Student at IICT Department, Mehran University of Engineering and Technology, Jamshoro, Sindh Pakistan*
[2]*Professor at Computer System Department, Mehran University of Engineering and Technology, Jamshoro*
[3]*Assistant Professor, Computer System Department, Mehran University of Engineering and Technology, Jamshoro*

Abstract: Analysis is most typical and challenging tasks in any industry specially education and learning field. All educational institutions and organizations have taken deep research to modernized student performance evaluation method and procedure to get better university ranking and student's placements. The major mediums of collecting data are admission and placements process. Currently, they are based on conventional model where analysis depends upon statistical approach to segregate student's marks. It basically not full reflection of students performance, so we proposed a model which uses data mining unsupervised techniques by implementing K-mean algorithm to get better reflection of their performance with future support of automation [1]. We applied methodology on different dataset to help in knowledge discovery.

Keywords: *Data mining, Supervised Learning approach, Unsupervised Learning approach, K-means, Lloyds Algorithm*

1. Introduction

Student's KPI is indicated to student's achievement, performance and level of interest and ability in academic period and commonly it is named as Grading. It is most common and standardized techniques to identify student's capabilities and understanding related to that course or academic period and used in various purposes like admissions and interviews. But unfortunately, this system is not truly represented student's performance in overall respects [2][5]. This system is just evaluation of marks which is obtained during examination hall. It means that it is not reflect the true picture of student's performance, skills and capabilities [2]. Data mining is the process of finding unhidden patterns and their relations from large data and transform into knowledge-based format and able to show in various dimensions and different perspectives using artificial intelligence and statistical or mathematical methods or algorithms. Data mining helps in finding organizational behavior, future layouts, and next generation trends by using predictive models which is more useful than retrospective models. The main purpose of predictive models is to find the needs of decisions, what is the best decision should be taken, what are the policies should be taken and what are the rules and regulations should be enforced overall organization [3][4]. Although each and every thing in life is depends upon data and information. After modeling or reorganizing it and extract knowledge from it, it is become a sample standard in that specific domain. The most common techniques in data mining are classification of data and information gathered from different resources. It is the process in which data is map with several predefined classes and patterns to known its nature and origin. To analyzing the basis of data, data mining functions is divided into two algorithms. One is unsupervised and second one is supervised. By using these methods, also extract the relationship between classified and non-classified data models. Clustering is commonly used method in unsupervised data mining approach. It is helps to extract objects and its behavior from finite set of data and information [6]. It is easy to categorizing and grouping the data. It identifies similar attributes, processed characteristics, natural behavior and relationships with its other objects like what characteristics plays major role in student academic performance? How to evaluate those characteristics to achieve student weak points? The methods to clustering on raw data are K-means, principal component analysis and polar ordination.

2. Related Work

Data mining in education institutes is a recent analysis field and this area of analysis is gaining popularity due to its potentials to academic institutes. Data Mining is utilized in academic field to reinforce our understanding of learning method to specialize in distinctive, extracting and evaluating variables associated with the educational process of students. A researcher [1] applied the K-means approach on student's semester results and it was based on student analysis using semester result as an attribute and attained 60% of accuracy results. The most common techniques in data mining are classification of data and information gathered from different resources. It is the process in which data is map with several predefined classes and patterns to known its nature and origin. Another author [2] has used Classification and Clustering approach on student data to study the understudy's overall performance via taking result as a parameter. Researcher [3] applied Hierarchical approach on student's marks to Evaluate between various DM approaches and Implementation of hierarchical clustering approach over understudy's data. An author [4] described the evaluation of Predictive clustering approach and Euclidean distance and Implementation of K-means approach with different

Corresponding author Email address: *anamarain07@gmail.com*

courses results. Clustering Approach helps to extract objects and its behavior from finite set of data and information. It is easy to categorizing and grouping the data and identifying similar attributes, characteristics and natural behavior and relationships with its other objects like in retail business, what is the best product which is feasible or suitable to sales offers / discounted offers and promotions [6].

3. Methodology

The goal of this research is to create a model to analyze and predict students' academic performance using factors of students based on their previous results (e.g. CGPA, activities and skills). The approach will study and identify the attributes used in analyzing students' performance and apply K-means clustering Algorithm to group student data based on their attributes. To determine the factors that have impact on Student's Performance.

3.1 System model

Figure 1.1 represent proposed model flow diagram.

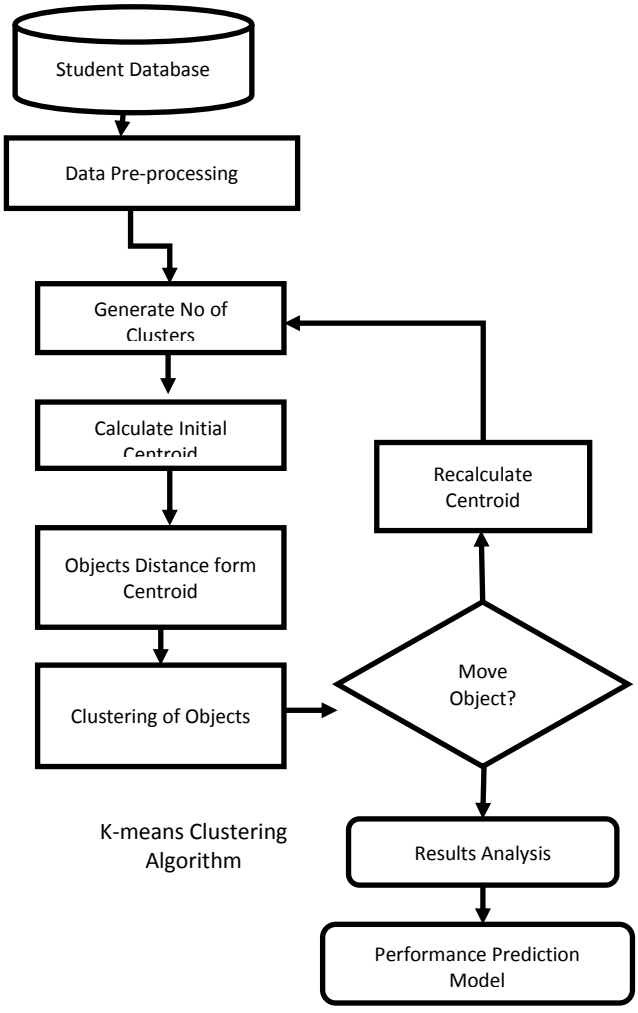

Figure.1. Student Performance Prediction Model

3.2 K-Means Clustering Algorithm

K-means clustering algorithm works on N number of inputs and produces K number of results in form of clusters [6].This algorithm works on inputs which includes X dataset of N points with K parameter, represent the number of clusters needs to be create. The results come with K number of cluster centroids and labeled the X number of points which are unique cluster. All points in cluster are closer according to distance to their centroids against any other centroids.

3.2 Dataset

UCI Machine Learning Repository has collections of data related to students, approximately sample of 395 dataset have been collected for this purpose [16]. This dataset contains student's previous semester details. Following are the main attributes of each course selected by attendees or students.

No	Variables	Description	Values
1	CGPA	Past results	numeric: 1 to 4
2	Failures	Total number of past class failures	numeric
3	Attendance	Total number of absents	numeric
4	Age	Student's age	numeric: from 15 to 22

Table 1: Main attributes of each course

4. Results and Discussion

In this paper, student academic performance dataset is used to apply K-means algorithm techniques. We applied the model on the data set and three different Test Performances were taken to group the data into k no of clusters and find best partition of Cluster. We further categorized the data with one of two class labels "Success" and "Failure" based on Performance Table.

Class Label	CGPA	Performance
Success= 0	CGPA < 2	Failure
Success=1	CGPA >= 2	Success

Table 2: Performance Table

4.1 Test # 01 Performance Results

Results in Table 3 for cluster number (k)=5, in cluster 0, the cluster size is 121 and the performance is 22.03%. In cluster 1, the cluster size is 63 and the performance is 9.62%. In cluster 2, the cluster size is 3 and its output is 0.76%. In cluster 3 the cluster size is 17% and the performance is 2.53%.The size of the cluster in cluster No 4 is 191 and the performance is 33.15%.

Analysis of selected attributes:

The attributes selected were ma set and the overall performance based on the index given below

Cluster No	Cluster Size	Success	Failure
Cluster 0	121	22.03%	8.86%
Cluster 1	63	9.62%	6.33%

Cluster 2	3	0.76%	0.00%
Cluster 3	17	2.53%	1.77%
Cluster 4	191	33.16%	14.94%

Table 3: Results for Cluster (k) =5

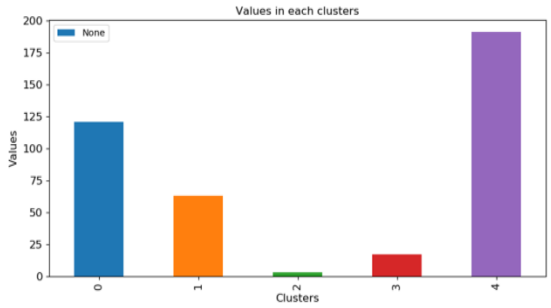

Figure 2: Test Performance 01 Bar chart

4.2 Test # 02 Performance Results

In table 4 for cluster number (k)=4, cluster size in cluster 0 is 100 and performance is 16.96%. The cluster size in cluster 1 is 5 and the output is 7.01%. In cluster 2, the size of the cluster is 244 and the performance is 41.27%. The size of the cluster in cluster 3 is 46 and the performance is 8.35%.

Cluster No	Cluster Size	Success	Failure
Cluster 0	100	16.96%	8.35%
Cluster 1	5	7.01%	0.25%
Cluster 2	244	41.27%	20.51%
Cluster 3	46	8.35%	3.29%

Table 4: Results for Cluster (k) =4

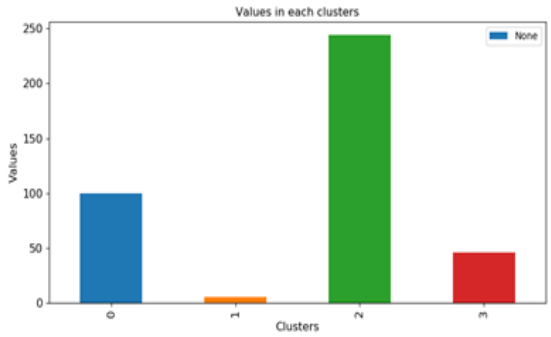

Figure 3: Test Performance 02 Bar chart

4.3 Test # 03 Performance Results

For table 5 for cluster number (k)= 3. The cluster size in cluster 0 is 309 and the performance is 51.6%. The cluster size in cluster 1 is 81 and the output is 13.9 percent. The cluster size in cluster 2 is 5 and the output is 1.3 percent.

Cluster No	Cluster Size	Success	Failure
Cluster 0	309	51.6%	26.5%
Cluster 1	81	13.9%	6.6%
Cluster 2	5	1.3%	0.0%

Table 5: Results for Cluster (k) =3

Figure 4: Test Performance 03 Bar chart

4.4 Results Accuracy:

For all test performances the final accuracy of k-means clustering algorithm is either the fraction (default) or the count (normalize=False) of correct predictions. For Test Performance 01 we have got accuracy 42%, similarly for Test Performance 02 and 03 we have got accuracy 48% and 35%.

Test Performance	Accuracy
Test 01	41.01%
Test 02	47.59%
Test 03	21.01%

Table 6: Results Accuracy

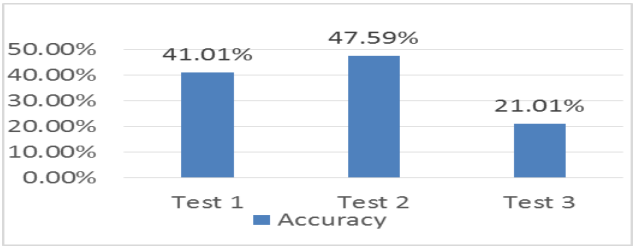

Figure 5: Results Accuracy Bar chart

5. Conclusion

This research paper provides us unsupervised analysis of student educational performance using K-mean algorithm rather than conventional modal so that better evaluation and comparison upon student educational data reflect actual figure to help in decision process. To implement K-mean algorithm technique, we use dataset of student academic results. We have Performed 3 different Test Performance and produces the numerical interpretation of results for performance evaluation. In 3 Test Cases, for k=3, k =4 and k = 5 we have classified students' data into 3,4 and 5 No of Clusters and predicted 67% overall success and 33% failure. Numerical equation is used to reduce the complexity and repetition of same procedures. The model has capacity to do extendibility in sense of enhancement and support of large and complex data. The model has successfully reduced the issues of conventional model. The model gives support to make academic planner and monitoring policy and can automate the overall performance evaluation procedure. That's it has flexibility to integrate with major educational systems. With the help of this, it's made easy to do future analysis and prediction student academic performance and make appropriate steps in this regard.

6. Future Work

This model may be extending to integrate with big data and its applications and support its architecture as well as there is also need of such these work in future.

Acknowledgement

This research work was conducted and supported by the Department of Computer System, Mehran University of Engineering and Technology, Jamshoro Pakistan. Authors greatly acknowledge and appreciate their excellent unconditional support for using laboratories and other facilities.

References

[1] J. Jamesmanoharan, S. H. Ganesh, M. L. P. Felciah and A. K. Shafreenbanu, "Discovering Students' Academic Performance Based on GPA Using K-Means Clustering Algorithm," 2014 World Congress on Computing and Communication Technologies, Trichirappalli, 2014, pp. 200-202.

[2] S. Rana and R. Garg, "Application of Hierarchical Clustering Algorithm to Evaluate Students Performance of an Institute," 2016 Second International Conference on Computational Intelligence & Communication Technology (CICT), Ghaziabad, 2016, pp. 692-697.

[3] I. Singh, A. S. Sabitha and A. Bansal, "Student performance analysis using clustering algorithm," 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence), Noida, 2016, pp. 294-299.

[4] K. Shaukat, I. Nawaz, S. Aslam, S. Zaheer and U. Shaukat, "Student's perfor-mance in the context of data mining," 2016 19th International Multi-Topic Conference (INMIC), Islamabad, 2016, pp. 1-8.

[5] A. M. de Morais, J. M. F. R. Araújo and E. B. Costa, "Monitoring student per-formance using data clustering and predictive modelling," 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, Madrid, 2014, pp. 1-8.

[6] D. V. Paul, C. Nayagam and J. D. Pawar, "Modeling Academic Performance using Subspace Clustering Algorithm," 2016 IEEE Eighth International Conference on Technology for Education (T4E), Mumbai, 2016, pp. 254-255.

[7] Kuzilek, Jakub, et al. "OU Analyse: analysing at-risk students at The Open University." Learning Analytics Review (2015): 1-16.

[8] Siemens, George, et al. Open Learning Analytics: an integrated & modularized platform. Diss. Open University Press, 2011.

[9] Amirah Mohamed Shahiri, Wahidah Husain, Nur'aini Abdul Rashid, A Review on Predicting Student's Performance Using Data Mining Techniques, Procedia Computer Science, Volume 72, 2015.

[10] A. M. de Morais, J. M. F. R. Araújo and E. B. Costa, "Monitoring student per-formance using data clustering and predictive modelling," 2014 IEEE Frontiers in Education Conference (FIE) Proceedings, Madrid, 2014, pp. 1-8.

[11] O.J. Oyelade, O.O. Oladipupo, I.C. Obagbuwa, "Application of k means clustering algorithm for prediction of students academic performance" Inter-national Journal of Computer Science and Information Security (IJCSIS), 7 (1) (2010), pp. 292-295

[12] Kumar, V. 2011. An Empirical Study of the Applications of Data Mining Techniques in Higher Education. International Journal of Advanced Computer Science and Applications, 2(3): 80–84.

[13] Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012. International Journal of Data Mining & Knowledge Management Process, 2(5), 13-24.

[14] Chen, S. Y., & Liu, X. (2004). The contribution of data mining to information science. Journal of Information Science, 30(6), 550-558. doi:10.1177/0165551504047928

[15] Siguenza-Guzman, L., Saquicela, V., Avila-Ordóñez, E., Vandewalle, J., & Cattrysse, D. (2015). Literature Review of Data Mining Applications in Academic Libraries. The Journal of Academic Librarianship, 41(4), 499-510. doi:10.1016/j.acalib.2015.06.007

[16] P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7

About Authors

Engr. Anam Farooque got B.E (computer Systems) from MUET, Jamshoro in 2012 now he near to complete M.E in (Computer and Information Engineering) from MUET,Jamshoro. She is working as a junior software engineer at Basecamp Data Solution.
Email: anamarain07@gmail.com

Dr. Tariq J.S Khanzada received his Ph.D (Wirless Communication), University of Magdeburg, Germany at 2010. He got his M.E in (Communications Systems & Networking) and B.E in (Computer Systems) from MUET, Jamshoro. He is working as a Professor at the Department of Computer Systems and Software Engineering at MUET, since 2004 and as a lecturer since 2001. He is also a coauthor of a practical book on computer workshop published in 2004. He is also an author of many international research publications.
Email: tariq.khanzada@faculty.muet.edu.pk

Engr. Arbab Ali Samejo working as an Assistant professor at the Department of Computer Systems at MUET for many years. He got his B.E in (Computer Systems) at MUET Jamshoro. He is also an author of some research publications.
Email: arbab.samejo@faculty.muet.edu.pk