IJASRM
International Journal of Advanced Scientific Research and Management, Special Issue 5, April 2019
www.ijasrm.com
ISSN 2455-6378

# Analysis of the Performance of Libraries and Laboratories using K-means Clustering

**Dr. Gauri Shanker Kushwaha**

Faculty of Science & Environment,Mahatma Gandhi Chitrakoot Gramodaya Vishwavidhyalaya Chitrakoot, Satna (M.P.)

## Abstract

In modern-day scenario, educational organization collects large quantity of scholar mastering behavior facts. This look at provides a proof-of-idea analytics device that may discover at-threat college students alongside their studying journey. Educationalist can promote from the early detection of at-risk college students by way of information elements which may additionally cause now not a success of student. K-means clustering approach is utilized by WEKA 3.8.1. This paper presents an analysis of the performance of institutions that are providing higher education on the basis of student responses of required facilities of libraries and laboratories.

***Keywords:*** *K-means clustering, WEKA 3.8.1, Data Generation Methods, Instances.*

## 1. Introduction

Educational institution is sort of way of delivery which requires pleasant environment to run efficiently. Thus, the fine of group relies upon on its faculty, pedagogy, curriculum layout, strategic planning, exam pattern, resources and guidelines. The evolution of technology and the extensive unfold of globalization for the use of rapid and ubiquitous assets of statement and data (facts) together with Internet has modified the current role of schooling. Higher education institutions are nucleus of research and future improvement performing in a competitive environment with the prerequisite undertaking to generate, collect and share information and knowledge. The chain of producing information inside and among external groups (which includes groups, different universities, partners, groups) is taken into consideration crucial to reduce the constraints of inner assets and can be appears that evidently advanced with the use of data mining technologies.

Educational data mining indicates the influential factors of higher educational institutions, including identifying student characteristics and the dimensions of getting to know reviews that honestly have an effect on data units. Research uses academic data mining to address academic records, translates the mining outcomes from the angle of getting to know analytics, and explores the influential factors imposed by students studying, organizational characteristics, student behaviors, coaching nice, duty of institutional leaders and excellent teaching gadgets.

The algorithm is powerful in generating clusters for plenty convenient packages. But the computational complexity of the unique k-means set of rules is very excessive, mainly for huge facts sets. Moreover, this set of rules outcomes in distinctive varieties of clusters depending at the random size of preliminary centroids. Several tries have been made by researchers for improving the performance of the k-means clustering on the set of generated rules.

The WEKA workbench includes a set of visualization technique and algorithms for records analysis and predictive modeling, collectively with Graphical User Interfaces (GUI) for this functionality. WEKA encompasses filters, classifiers, clusters, associations, and attribute selection. The visualization tool in WEKA allows datasets and the predictions of classifiers in a pictorial form.

These consequences additionally facilitate in decision making and answering certain questions like whether the college v/s student's ratio is giving satisfactory outcomes or there is an exchange needed in the teaching technique. There are numerous strategies of facts mining like class, clustering, association rule mining and many others. Every method has its personal importance in keeping with his function and in this paper clustering approach has been used for further observe. [1] This paper is aim to apply data mining algorithms on institutional datasets (collected in the form of responses from students) and analyzed in the field of Educational Data Mining.

## 2. K – Means Clustering and Analysis of Generated Data

K – Means clustering is a set of rules to categorize the objects based on attributes/functions into K number of instances. K is effective integer quantity. By minimizing sum of squares of distances

between statistics and the corresponding cluster centroid grouping is achieved and the purpose is to classify the information. The fundamental step of k-manner clustering is easy and less complicated to use. In the beginning a number of K cluster determined and assumed that the centroid or center of these clusters. Any random objects can be chosen as the preliminary centroid or the primary K objects in sequence that can be initial centroid and algorithm will do the under given steps until convergence iterate till regular:

1. Determine the centroid coordinate
2. Determine the distance of each object to the centroid
3. Group the object based on minimum distance

With the help of k-means algorithm data will be analyzed.

The data set used for this study is based on the "higher educational quality" available in comma-separated format student_feedback.csv. This paper assumes that appropriate data preprocessing has been performed. As an illustration of performing clustering in WEKA, its implementation of the K-means algorithm to cluster the respondents in this educational data set, and to characterize the resulting respondents segments. Below given figure shows the main WEKA Explorer interface with the data file loaded.

**Figure 1: Preprocessing of Higher Educational Dataset**

K-means simply allow numerical values for attributes, if so, it may be vital to transform the statistics set into the same old spreadsheet layout and convert precise attributes to binary. It may also be imperative to normalize the values of attributes which are measured on extensively distinctive scales. While WEKA provides filters to perform all of those preprocessing responsibilities, they're now not essential for clustering in WEKA. This is due to the fact WEKA.

Simple K-Means algorithm robotically handles a combination of express and numerical attributes. Furthermore, the set of rules automatically normalizes numerical attributes by computing distance. The WEKA Simple K-Means algorithm makes use of Euclidean distance to compute distances between instances and clusters.

To carry out clustering, we've selected the “Cluster” tab in the Explorer by click on the "Choose" button. After clicking on the “Choose” button a dropdown list will displayed that displays clustering algorithms, where we choose "Simple K-Means". Next, click on the textual content container to the proper of the "Choose" button for editing the clustering parameter, shown in Figure.

**Figure 2: Popup window for editing the clustering parameters**

In the window we enter 5 instead of the default values of 2 as the number of clusters. The seed value is used in generating a random number which is, in turn, used for making the initial assignment of instances to clusters. Note that, in general, K-means is quite sensitive to how clusters are initially assigned. Thus, it is often necessary to try different values and evaluate the results.

Once the options have been specified, run the clustering algorithm. Make sure that in the "Cluster Mode" panel, the "Use training set" option is selected, and click "Start". Then by right click the result set in the "Result list" panel and view the results of clustering in a separate window. This process and the resulting window is shown in Figures 3.

IJASRM
International Journal of Advanced Scientific Research and Management, Special Issue 5, April 2019
www.ijasrm.com
ISSN 2455-6378

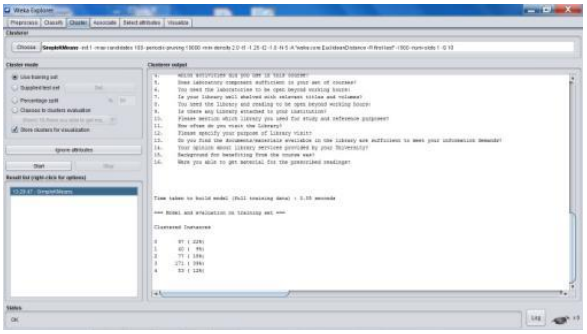

**Figure 3: Representing cluster instances from result buffer**

The result window indicates the centroid of every cluster in addition to data on the number and percent of times assigned to distinct clusters. Cluster centroids are the imply vectors for every cluster, each dimension in the centroid represents the imply value for that measurement in the cluster.

## 3. Results & Discussion

The three different dimensions to be had x-axis, y-axis, and coloration are decided on for acquiring the cluster wide variety and any of the alternative attributes. Different combos of clusters will bring about a visual rendering of different relationships inside each cluster. In the below given result, we have chosen number of instances recorded as the x-axis, Name of the institute as the y-axis, and the shade (cluster) characteristic as the color measurement. This will result in a visualization of the distribution of each example in each cluster. As an example of appearing clustering in WEKA, K-means algorithm used to create the cluster of the respondents on this set inside dataset, and to characterize the resulting respondents segments. Below given figure 4 shows the every cluster through visualization interface of performance of the libraries and laboratories in higher educational institutions.

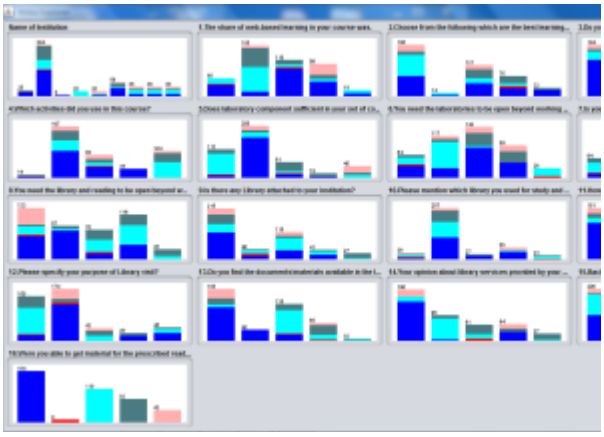

**Figure 5: Visualization of all attributes of student feedback**

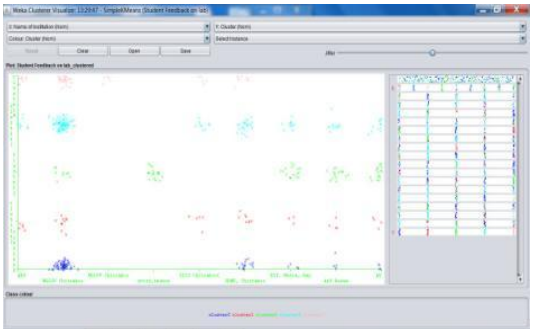

**Figure 6: Each cluster through Visualization**

From the result it is obtained that in the analysis of educational dataset cluster 3 represent maximum number of instances that represents the 'Bad' performance of the laboratory usage and available/required facilities.

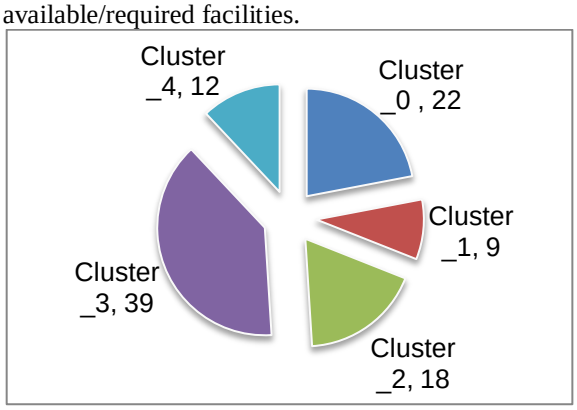


**Figure 7: Percentage of Instances from analyzed datasets**

The records collected from eight different institutions that are providing higher education in India. The above given graph also represents the recorded responses (maximum recorded responses in each cluster) in each cluster, here cluster were categorized in five different categories as Cluster 0 (represents the maximum response in the favor of 'very good' performance), Cluster 1 (represents the maximum response in the favor of 'good' performance, Cluster 2 (consist the response in terms of 'neutral'), Cluster 3 (contains the responses in the favor of 'bad' performance) and Cluster 4 (characterized as the responses into 'very bad' performance. The analyzed dataset of the student's responses represents the maximum instances in Cluster 3 that replicates the performance of higher educational institutions into 'bad' situation. The facts includes library and laboratory facility supplied through organization, that is again bone of scholar educational service, information of 438 students with 16 attributes.

## 4. Conclusion

Educational Data Mining used in this method to get better effects and effortlessly recognize and expect the effects from the information. Educational data mining reproduces/produces the platform for computerized analysis of the recorded datasets from primary sources like students/public. The algorithm categorizes five clusters including different instances. Each includes a number of respondent that subject to an instance. On the basis of above educational dataset analysis in the perspective of student feedback it is concluded that quality of libraries and laboratories in higher educational institutions of the selected institutions was bad.

**References**

[1] Gauri Shanker Kushwaha and Bharat Mishra (2017), Data Mining Algorithm used in Educational datasets, International Journal of Scientific Research, Volume 6, Issue 2, 676-677.

[2] P. Gill, et al., (2008), Methods of data collection in qualitative research: interviews and focus groups , British Dental Journal, Volume 204, No. 6, 291-295.

[3] May K M., (1991), Interview techniques in qualitative research: concerns and challenges. In Morse J M (ed) Qualitative nursing research, Newbury Park: Sage Publication, 187-201.

[4] Britten N., (1999), Qualitative interviews in healthcare in Pope C, Mays N (eds) Qualitative research in health care. 2nd ed. pp 11-19. London: BMJ Books.

[5] Morgan D L., (1998), The focus group guide book. London: Sage Publications.

[6] Bloor M, et al., (2001), Focus groups in social research. London: Sage Publications.

[7] Stewart D W and Shamdasani P M, (1990), Focus groups. Theory and practice. London: Sage Publications.

[8] Dr. Bharat Mishra and Gauri Shanker Kushwaha, (2016), A Review of Quality Factors of Higher Education, IOSR Journal of Research & Method in Education, Volume 6, Issue 4, 62-68.