

Data Mining Techniques Applying on Educational Dataset to Evaluate Learner Performance Using Cluster Analysis

Minimol Anil Job

Abstract—Due to the advancement of technology in this digital era, academic institutions are bringing out graduates as well as generating enormous amounts of data from their systems. Hidden information and hidden patterns in large datasets can be efficiently analyzed with data mining techniques. Application of data mining techniques improves the performance of many organizational domains and the concept can be applied in the education sectors for their performance evaluation and improvement. Understanding the business value of the collected data it can be used for classifying and predicting the students' behavior, academic performance, dropout rates, and monitoring progression and retention. This paper discusses how application of data mining can help the higher education institutions by enabling better understanding of the student data and focuses to consolidate clustering algorithms as applied in the context of educational data mining.

Index Terms—Data Mining; Educational Data Mining; Clustering, Cluster Analysis; Performance; Evaluation, Prediction.

I. INTRODUCTION

Data mining is the knowledge discovery process, which involves discovering hidden patterns, messages and knowledge within large datasets and the process of analyzing outcomes or behaviors in various business domains. Knowledge discovery and data mining can be considered as tools for decision-making as well as organizational effectiveness. Data mining is a systematic process of extracting relevant knowledge using various techniques from large structured and unstructured data. There are a variety of different data mining techniques and approaches available such as clustering, classification, and association rule mining etc. One of the most challenging tasks of the higher educational institutions is the provision and utilization of up to date information for their sustainability, to monitor student performance and to measure institutional effectiveness. In this paper, the researcher analyses and presents the use of data mining techniques in higher education sectors to evaluate learner performance. The available data in higher education institutions can be evaluated using data mining techniques and it will lead into discovering hidden patterns, messages and knowledge. Based on the results from the applied techniques the institution management can take measures to allocate relevant resources more effectively, make effective decisions on educational academic activities to improve students' performance, increase students' learning behavior,

increase student's retention rate, and increase students' consistency in various academic activities.

II. LITERATURE REVIEW

Data Mining (DM) is described as a process of discover or extracting interesting knowledge from large amounts of data stored in multiple data sources such as file systems, databases, data warehouses etc. [3],[18]. Defining characteristic of data mining is 'Big data' [50]. Data mining is defined as "the process of discovering "hidden message," patterns and knowledge within large amounts of data and of making predictions for outcomes or behaviour" [30]. 'Pattern' is a single record that consists of input and output. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining functionalities are classified into two broad categories as descriptive and predictive ones [2] [21]. The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data [48]. The implementation of data mining methods and tools for analyzing data available at educational institutions, defined as Educational Data Mining (EDM) [22] is a relatively new stream in the data mining research. Educational data mining is a research area falls under data mining. A survey of the application of data mining techniques to various educational systems is given in Romero and Ventura [41]. In another work of Romero et al. [42],[44] on educational data mining, the application of various data mining techniques on data collected from the activities of students who use Moodle e-learning course management system is discussed.

Brijesh Kumar Baradwaj and Saurabh Pal [9] describes the main objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on. Romero and Ventura, in 2010 [43] published a paper in IEEE, which listed most common tasks in the educational environment resolved through data mining and some of the most promising future lines. Educational Data Mining community remained focused in North America, Western Europe, and Australia/New Zealand. They mentioned that there is a considerable scope for an increase in educational data mining's scientific Influence. They also suggested developing more unified and collaborative studies [43]. There are increasing research interests in using data mining in education. This new emerging field, called Educational

Published on November 21, 2018.

M. A. Job is Assistant Professor, Arab Open University, Kingdom of Bahrain. (e-mail: m.aniljob@aou.org.bh)

Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments [4],[24].

Educational Data Mining uses many techniques such as Decision Trees, Neural Networks, Naïve Bayes, K- Nearest neighbor, and many others. Han and Kamber [21] describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process [34]. Data mining tools predict future trends and behaviors, allowing institution to make proactive, knowledge-driven decisions [11]. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer institution questions that traditionally were too time consuming to resolve [10],[40].

III. APPLYING DATA MINING IN EDUCATION

A. Knowledge Discovery Process

Following are the steps in Data mining process [25],[26]:

- Understand application domain
- Create target dataset
- Data cleaning and transformation
- Apply data mining algorithm
- Interpret, evaluate and visualize patterns
- Manage discovered knowledge

Fig. 1 shows data mining as a step in an iterative knowledge discovery process.

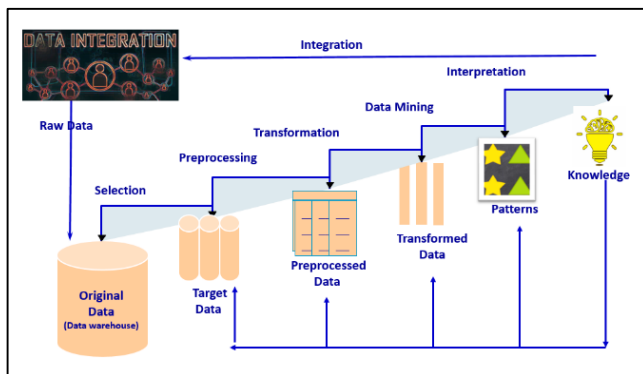


Fig. 1. Step in an iterative knowledge discovery process

B. Educational Data Mining

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [28],[35]. The emergent field of EDM examines the unique ways of applying data mining techniques education institutions academic problems. Fig. 1 shows the concept of EDM.

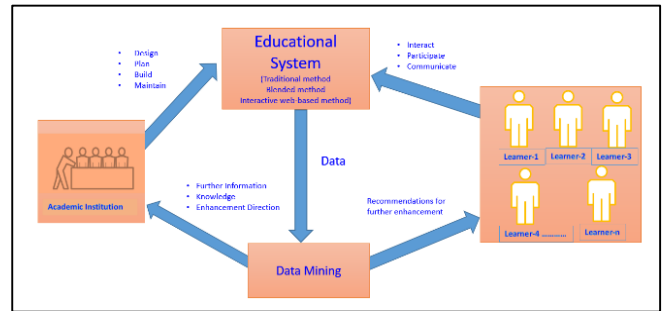


Fig. 2. Concept of Educational Data Mining

The major components of a data mining system are data source, data warehouse, data mining engine, knowledge base and pattern evaluation module [30]. It shows that our system is using the historical data from the data warehouse server and then training the data after applying various pre-processing techniques [23]. Data mining tasks can be done in two different ways; predictive or descriptive. Data processing undergoes two types of functions namely clustering and classification. We can apply clustering in predictive model or classification in descriptive model to the dataset.

C. Classification

Classifying data into a fixed number of groups and using it for categorical variables is known as classification [33]. Classification can be classified into two types: Supervised and Unsupervised. When the objects or cases are known in advance is called supervised classification whereas unsupervised classification means the objects or cases are not known in advance. The following algorithm can be used for classification model [19],[1].

- Decision/Classification tree
- K-nearest neighbour classifier
- Rule-based methods,
- Statistical analysis, genetic algorithms
- Bayesian classification
- Neural Networks
- Memory-based reasoning
- Support vector machines

D. Clustering

Clustering is grouping similar objects. Clustering is defined as a process of grouping a set of physical or abstract object into a class of similar objects [38]. According to Larose [29] cluster does not classify, estimate or predict the value of target variables but segment the entire data into homogeneous subgroups. Heterogeneous population is classified into number of homogenous subgroups or clusters are referred as clustering [6]. Furthermore, clustering task is an unsupervised classification. Clustering is a process where the data divides into groups called as clusters such that objects in one cluster are very much similar to each other and objects in different clusters are very much dissimilar to each other. [14]. Finding groups of objects such that the objects in a group will be similar or related to one another and different from or unrelated to the objects in other groups. Clustering is the technique of arrangement of similar objects into a class. By this similar objects are placed in a class and many classes with different object sets can be constructed. Various clustering methods like Partitioning

method, Hierarchical method, Model-based method, Grid-based method, Density-based method, Constraint based method can also be used and the clustering can be done to the dataset [49][5],[7]. Some of the fields has been undergoing clustering techniques in the fields such as data mining, image processing, text mining, machine learning and pattern recognition [25]-[27]. Educational data mining is concerned with developing new methods to discover knowledge from educational database in order to analyse student trends and behaviours towards education. In the educational sector, for example, it can be helpful for course administrators and educators for analyzing the usage information and students' activities during a course to get a brief idea of their learning [4]. Visualization of information and statics are the two main methods that have been used for this task. Studies show that data mining was first implemented for marketing outside higher education but it has parallel implications and value in higher education [37],[45].

In every higher education institution, marketing is part of student relationship management. Within educational institutions, marketing concerns the service area, enrollment, annual campaign, alumni, and college image. Combined with institutional research, it expands into student feedback and satisfaction, course availability, and faculty and staff hiring. A university service area now includes on-line course offerings, thus bringing the concept of mining course data to a new dimension. Data mining is quickly becoming a mission critical component for the decision making and knowledge management processes [21],[35].

In summary, data mining can be applied in classifying students based on academic achievements, knowledge, gender, age, semester-wise grades and monitoring of progression and regression.

IV. METHODOLOGY

The approach begins from the collection of data of enrolled students, and later the pre-processing procedures are tested to the dataset. The data pre-processing approach is utilized to make the data more deserved for data mining. Then the dataset is classified into a training data set and a testing data set. The dataset which is trained is utilized to compose the classification model. The dataset which is treated as testing dataset is either utilized to examine the assessment of the developed order demonstrate or to contrast the forecasts with the known target values.

To evaluate undergraduate student academic performance; the following data mining algorithms can be used to identify the significant variables that affects and influences the performance of students depending on the significant variables. Since the study is focused on undergraduate students' performance evaluation the learner in the given below table describes an undergraduate student.

TABLE I: SUGGESTED CLUSTERING ALGORITHMS IN EDUCATIONAL DATA MINING

Significant variable	Algorithm	Data Source
Evaluate learner's academic performance	A combination of Data Mining methods like ANN (Artificial Neural Network), Farthest First method based on k-means clustering	Student data of the programs offered by the institution
To identify the significant variables that affects and influences the performance of a learner	Fuzzy C-Means clustering method	Academic dataset from the institution which can contain data like faculty, academic program, demographic data, method of study (part time or full time), academic achievements data etc.
How to teach a basic writing skills and mathematical skills courses to learner's who fail in the placement tests	K-means Clustering method	Registration data of newly admitted students with placement tests results and allocated courses
To predict the potentiality of learner performance who can fail during an online assignments in a Learning Management System (LMS)	Expectation Maximization, Hierarchical Clustering, Simple k-Means and X-Means Clustering	Online assessment test results data from tutors
To develop profiles of learner behavior from their activities in an online learning environment and also to create click-stream server data from LMS	Two clustering methods, Hierarchical clustering and Non-Hierarchical Clustering method (k-means clustering)	LMS activities data receive from log entries and results of forum activities
To analyze the web log data files of (LMS)	Markov Clustering (MCL) algorithm for clustering the learner' activity and a Simple K-Means algorithm for clustering the registered courses	The data set from the registration department for the registered students and the LMS log files from the server
To map out the approaches to teaching profiles of teachers in higher education	A hierarchical cluster analysis was performed on the questionnaire data	Questionnaires and Interviews conducted in the institution among academicians
To compare the emotional intelligence of students.	K-means clustering is applied on questionnaire data	Questionnaires and Interviews conducted in among learners.

ANN (Artificial Neural Network)- Neural networks are useful for data mining and decision-support applications due to the fact that people are good at generalizing from experience [45],[8]. Neural networks bridge this gap by modeling, on a computer, the neural behavior of human brains. Neural networks are useful for pattern recognition or data classification, through a learning process [17],[32]. Neurons work by processing information. They receive and provide information in form of spikes. An artificial neural network is composed of many artificial neurons, which are linked together according to a specific network architecture [47],[36]. The objective of the neural network is to transform the inputs into meaningful outputs.

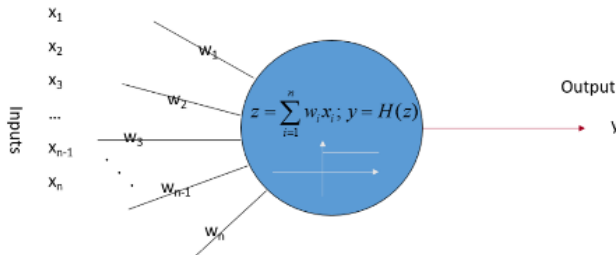


Fig. 3. The McCulloch-Pitts model

In 1943, McCulloch and Pitts first presented a mathematical model (M-P model) of a neuron. Since then many artificial neural networks have developed from the well-known M-P model [36],[32].

Farthest first algorithm proposed by Hochbaum and Shmoys 1985 has same procedure as k-means, this also chooses centroids and assign the objects in cluster but with max distance and initial seeds are value which is at largest distance to the mean of values. Here cluster assignment is different, at initial cluster we get link with high Session Count, like at cluster-0 more than in cluster-1, and so on. Farthest first actually solves problem of k-centre and it is very efficient for large set of data. In farthest first algorithm we are not finding mean for calculating centroid, it takes centroid arbitrary and distance of one centroid from other is maximum [12].

Fuzzy c-Means Clustering performs clustering by iteratively searching for a set of fuzzy clusters and the associated cluster centres that represent the structure of the data as best as possible [15]. The algorithm relies on the user to specify the number of clusters present in the set of data to be clustered.

K-means algorithm is a partitioned clustering approach, which is used when you have unlabeled data (i.e., data without defined categories or groups) [20]. The goal of this algorithm is to find groups in the data, with the number of groups represented by the variable K. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided [24]. Data points are clustered based on feature similarity. The results of the K-means clustering algorithm are: the centroids of the K clusters, which can be used to label new data and the labels for the training data (each data point is assigned to a single cluster)

The name comes from representing each of k clusters C_j by the mean (or weighted average) c_j of its points, the so-called centroid. While this obviously does not work well with categorical attributes, it has a good geometric and statistical sense for numerical attributes [45]. The sum of discrepancies between a point and its centroid expressed through appropriate distance is used as the objective function. For example, the L2-norm based objective function, the sum of the squares of errors between the points and the corresponding centroids, is equal to the total intra-cluster variance

$$E(C) = \sum_{j=1:k} \sum_{x_i \in C_j} \|x_i - c_j\|^2$$

The sum of the squares of errors can be rationalized as a (negative of) log likelihood for a normally distributed

mixture model and k-means algorithm can be derived from a general probabilistic framework [35],[16].

Hierarchical clustering is a method of cluster analysis which seeks to build a hierarchy of clusters. In data mining hierarchical clustering works by grouping data objects into a tree of cluster [31],[16]. Hierarchical techniques produce a nested sequence of partitions, with a single, all-inclusive cluster at the top and singleton clusters of individual objects at the bottom.

X-Means clustering algorithm, an extended K-Means which tries to automatically determine the number of clusters based on BIC scores. Starting with only one cluster, the X-Means algorithm goes into action after each run of K-Means, making local decisions about which subset of the current centroids should split themselves in order to better fit the data. The splitting decision is done by computing the Bayesian Information Criterion (BIC) [38]-[40].

The Markov Cluster Algorithm (MCL) is well-recognized as an effective method of graph clustering [46]. It involves changing the values of a transition matrix toward either 0 or 1 at each step in a random walk until the stochastic condition is satisfied.

In data mining, expectation-maximization (EM) is generally used as a clustering algorithm (like k-means) for knowledge discovery [48]. EM Algorithm on Clustering Features takes each clustering feature as a data object. It describes a sub cluster of data items more accurate, and so less sensitive to the data summarization procedure.

V. PROPOSED APPROACH

A. Application of k-Means Clustering algorithm in educational data set

1) K-means:

- Partitional clustering approach
- Each cluster is associated with a centroid (center point)
- Each point is assigned to the cluster with the closest centroid
- Number of clusters, K, must be specified [13]

The K-means algorithm accepts the number of clusters to group data into, and the dataset to cluster as input values. It then creates the first K initial clusters (K= number of clusters needed) from the dataset by choosing K rows of data randomly from the dataset. The K-Means algorithm calculates the Arithmetic Mean of each cluster formed in the dataset. The Arithmetic Mean of a cluster is the mean of all the individual records in the cluster. In each of the first K initial clusters, there is only one record. The Arithmetic Mean of a cluster with one record is the set of values that make up that record [20],[4].

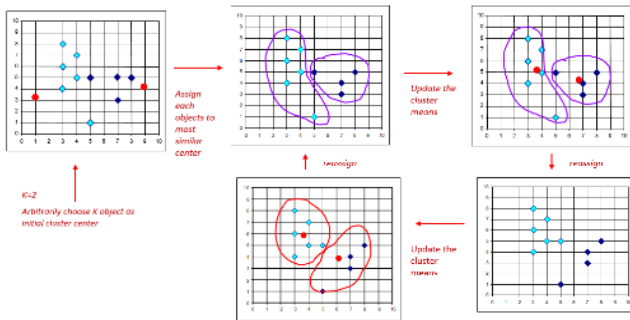


Fig 4. Example K-means clustering method

- 1: Select K points as the initial centroids
- 2: repeat
- 3: From K clusters by assigning all points to the closest centroid
- 4: Recompute the centroid of each cluster
- 5: until The centroids don't change

Fig. 5. Algorithm: Basic K-means Algorithm

- Step 1: Accept the number of clusters to group data into and the dataset to cluster as input values
- Step 2: Initialize the first K clusters
 - Take first k instances or
 - Take Random sampling of k elements
- Step 3: Calculate the arithmetic means of each cluster formed in the dataset.
- Step 4: K-means assigns each record in the dataset to only one of the initial clusters
 - Each record is assigned to the nearest cluster using a measure of distance (example: Euclidean distance).
- Step 5: K-means re-assigns each record in the dataset to the most similar cluster and re-calculates the arithmetic mean of all the clusters in the dataset.

Fig. 6. Generalized Pseudocode of Traditional k-means

K-means clustering algorithm used in this research as a simple and efficient tool to monitor the progression of students 'performance. A model has been applied here on a data set of academic result of one semester. The result generated from this data set in various clusters are shown in below given tables. Table-II shows the performance criteria used in evaluation of the results. Table II shows the dimension of the data set (Student's mark) in the form N by M matrix, where N is the rows (number of students) and M is the column (number of courses) taken by each student. The overall performance is evaluated by applying deterministic model in Equation.

$$\frac{1}{N} \left(\sum_{j=1}^N \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \right)$$

where N = the total number of students in a cluster and n = the dimension of the data

The group assessment in each of the cluster size is evaluated by summing the average of the individual scores in each cluster. The results generated are shown in the following tables.

TABLE II: PERFORMANCE CRITERIA

Grade	Criteria
>90	Excellent
80-89	Very Good
70-79	Good
60-69	Average
50-59	Below Average
< 50	Poor

TABLE III: STATISTICS OF THE DATA

Student's Mark	Number of Students	Total number of courses registered
Data	72	4

2) Using cluster number, K=3

In Table IV, for k = 3; in cluster 1, for a selected cluster size 24 the overall performance is 67, in cluster 2, for a selected cluster size 18 the overall performance is 48.5 and in cluster 3, for a selected cluster size 30 the overall performance is 58.6.

TABLE IV: DATA FOR K = 3

Cluster No.	Cluster Size	Overall Performance
1	24	67
2	18	48.5
3	30	58.6

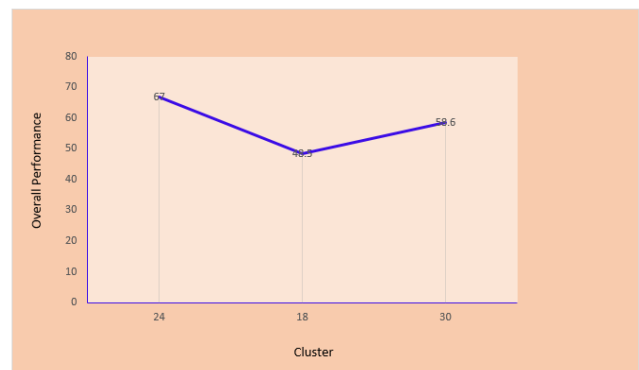


Fig. 6. Overall Performance versus cluster size for k = 3

In Fig. 6, the overall performance for a selected cluster size 24 is 67%, in cluster 2, for a selected cluster size 18 the overall performance is 48.5% and in cluster 3, for a selected cluster size 30 the overall performance is 58.6%. We can analyze and conclude that, 24 out of 72 total selected students shows an "Average" performance (67%), 18 out of 72 total selected students shows performance in the criteria of "Poor" performance (48.5%) and while the remaining 30 out of 72 total selected students shows an "Below Average" performance (58.6%).

3) Using cluster number, K=4

Table V shows the sample data, in table 1, for k = 3; in cluster 1, for a selected cluster size 24 the overall performance is 67, in cluster 2, for a selected cluster size 18 the overall performance is 48.5 and in cluster 3, for a selected cluster size 30 the overall performance is 58.6.

TABLE V: DATA FOR K = 4

Cluster No.	Cluster Size	Overall Performance
1	8	52.5
2	14	70.5
3	22	61.4
4	28	45.5

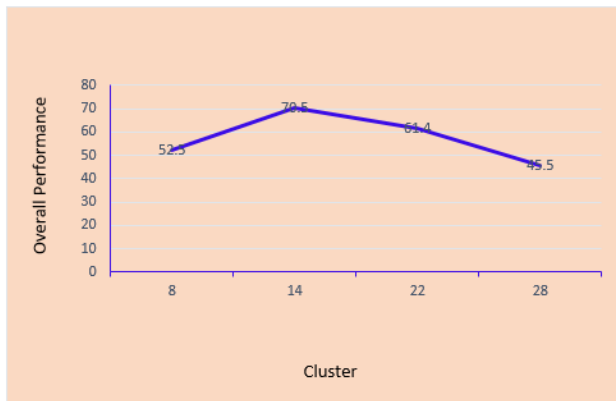


Fig. 7. Overall Performance versus cluster size for $k = 4$

In Fig. 7, the overall performance for a selected cluster size 8 is 52.5%, in cluster 2, for a selected cluster size 14 the overall performance is 70.5% and in cluster 3, for a selected cluster size 22 the overall performance is 61.4% and for a selected cluster size 28 the overall performance is 45.5%. We can analyze and conclude that, 8 out of 72 total selected students shows a “Below Average” performance (52.5), 14 out of 72 total selected students’ shows performance in the criteria of “Good” performance (70.5). 22 out of 72 total selected students shows performance in the criteria of “Average” performance (61.4) and while the remaining 28 out of 72 total selected students shows an “Poor” performance (45.5).

4) Using cluster number, $K=5$

Table VI shows the sample data, in table 1, for $k = 5$; in cluster 1, for a selected cluster size 7 the overall performance is 51.66, in cluster 2, for a selected cluster size 11 the overall performance is 80.3. While in cluster 3, for a selected cluster size 17 the overall performance is 61.25, in cluster 4, for a selected cluster size 15 the overall performance is 42.9 and in cluster 5, for a selected cluster size 22 the overall performance is 66.5.

TABLE VI: DATA FOR $K=5$

Cluster No.	Cluster Size	Overall Performance
1	7	51.66
2	11	80.3
3	17	61.25
4	15	42.9
5	22	66.5

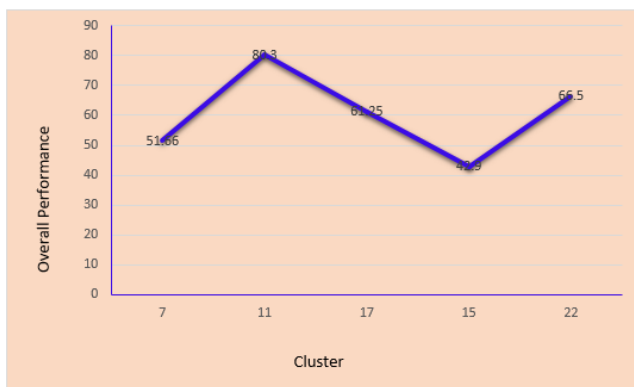


Fig. 8. Overall Performance versus cluster size for $k = 5$

In Fig. 8, the overall performance for a selected cluster size 7 is 51.66%, in cluster 2 for a selected cluster size 11 the overall performance is 80.3%. In cluster 3, for a selected cluster size 17 the overall performance is 61.25%, for a selected cluster size 15 the overall performance is 42.9% and for a selected cluster size 22 the overall performance is 66.5%. We can analyze and conclude that, 7 out of 72 total selected students shows a “Below Average” performance (51.66), 11 out of 72 total selected students’ shows performance in the criteria of “Very Good” performance (80.3). 17 out of 72 total selected students shows performance in the criteria of “Average” performance (61.25), 15 out of 72 total selected students shows performance in the criteria of “Poor” performance (42.9) and while the remaining 22 out of 72 total selected students shows an “Average” performance (66.5).

VI. CONCLUSION

Data mining is a powerful analytical tool that enables educational institutions to predict and analyze the performance of the students. With the ability to reveal hidden patterns in large databases, higher education institutions can build models that predict with a high degree of accuracy the behavior of population by dividing the available up to date data samples into specific clusters. By acting on these predictive data mining models, educational institutions can effectively address issues such as students’ progression, transfers and retention. The newly identified data patterns can also be utilized to monitor the graduates’ aspects as well as the alumni facts. In this paper the researcher identified various data mining algorithms that can be used to identify the significant variables that affects and influences the performance of students in higher education institutions. The study was basically focused on the performance evaluation of undergraduate students. At the end, the researcher demonstrated a technique using k-means clustering algorithm and combined with the deterministic model on a data set of undergraduate students’ examination marks in four registered courses in one semester. This is done for each student for total number of 72 students. A numerical interpretation of the results for the performance evaluation has been produced as a result. As the result of the analysis, it has been concluded that clustering algorithm serves as a good benchmark to monitor the students’ performance in higher institution. The clustering algorithm used in this paper is k-means algorithm. The use of clustering data mining algorithm also helps the institution management to monitor the students’ semester by semester performance and the results can be used in identifying measures to improve future academic semester performances.

REFERENCES

- [1] Aggarwal, C. Charu and Yu, S. Philip. “Data Mining Techniques for Associations, Clustering and Classification.” in Zhong, Ning and Zhou, Lizhu (Eds.) methodologies for knowledge discovery and data mining, third pacific Asia Conference, PAKDD, Beijing, China, April 26-28, 1999 proceedings, Springer, New York

- [2] Alex Berson, Stephen Smith, and Kurt Thearling, An Overview of Data Mining Techniques Excerpted from the book Building Data Mining Applications for CRM
- [3] Anand, S., and Buchner, A. 1998. *Decision Support Using Data Mining*. Financial Times Pitman
- [4] Baker, R.S.J.D. and Yacef, K. (2009) The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1, 3-16.
- [5] Berkhin, Pavel. "A survey of clustering data mining techniques." *Grouping multidimensional data*. Springer Berlin Heidelberg, 2006. 25-71.
- [6] Berry, J.A. Michael and Linoff S. Gordon., "Data Mining Techniques for Marketing, Sales, and Customer Relationship Management." Second Edition, Wiley Publishing, 2004
- [7] Berthold, M. "Fuzzy Logic." In M. Berthold and D. Hand (eds.), *Intelligent Data Analysis*. Milan: Springer, 1999
- [8] Bradshaw, J.A., Carden, K.J., Riordan, D., 1991. Ecological —Applications Using a Novel Expert System Shell. *Comp. Appl. Biosci.* 7, 79–83.
- [9] Brijesh Kumar Baradwaj, Saurabh Pal, Data mining: machine learning, statistics, and databases, 1996.
- [10] Cabena, P., Hadjini, P., Stadler, R., Verhees, J., and Zanasi, A. 1998. *Discovering Data Mining: From Concepts to Implementation*, Prentice Hall Saddle River, New Jersey
- [11] D. E. Rumelhart and J. L. McClelland, *Parallel Distributed Processing*, vols. 1, 2. Cambridge, MA: MIT Press, 1986.
- [12] D. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
- [13] Dan Pelleg and Andrew Moore. X-means: Extending K-means with Efficient Estimation of the Number of Clusters. *ICML*, 2000.
- [14] Daniel B.A. , Ping Chen Using Self-Similarity to Cluster Large Data Sets[J]. *Data Mining and Knowledge Discovery*, 2003, 7(2):123-152.
- [15] Danon, L., Diaz-Guilera, A., and Arenas, A. Effect of Size Heterogeneity on Community Identification in Complex Networks, *J. Stat. Mech.* P11010 (2006)
- [16] Douglas H. Fisher. Iterative optimization and simplification of hierarchical clustering. *Journal of Artificial Intelligence Research*, 4:147–179, 1996.
- [17] Esehrieh S. Jingwei Ke, Hall L. O. etc. Fast accurate fuzzy clustering through data reduction *IEEE Transactions on Fuzzy systems*, 2003, 11(2):262-270.
- [18] Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (Eds.), 1996. *Advances in Knowledge Discovery and Data Mining*, AAAI Press, Cambridge
- [19] Gorunescu, Florin. *Data Mining: Concepts, Models and Techniques*, Springer-Verlag Berlin Heidelberg, 2011
- [20] Guo, W. William, "Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction." *Expert System with Applications*, 37 (2010) 3358-3365
- [21] Han, J., Kamber, M. & Pei, J. (2011). 'Data Mining: Concepts and Techniques,' *Morgan Kaufmann*, San Francisco, CA.
- [22] Han, J., Kamber, M. 2012. *Data Mining: Concepts and Techniques*, 3rd ed, 443-491
- [23] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," *Morgan Kaufmann*, 2000.
- [24] J. O. Omolehin, J. O. Oyelade, O. O. Ojeniyi and K. Rauf, "Application of Fuzzy logic in decision making on students' academic performance," *Bulletin of Pure and Applied Sciences*, vol. 24E(2), pp. 281-187, 2005
- [25] Jiawei Han and Micheline Kamber (2006), *Data Mining: Concepts and Techniques*, The MorganKaufmann/Elsevier India
- [26] Kantardzic, Mehmed. *Data Mining: Concepts, Models, Methods and Algorithm*, Second Edition, John Wiley and Sons, New Jersey, 2011
- [27] Kriegel, Hans-Peter; Kröger, Peer; Sander, Jörg; Zimek, Arthur (2011). "Density-based Clustering". *WIREs Data Mining and Knowledge Discovery* 1 (3): 231–240. doi:10.1002/widm.30.
- [28] Kumar, Varun, and Anupama Chadha. "An empirical study of the applications of data mining techniques in higher education." *International Journal of Advanced Computer Science and Applications* 2.3 (2011).
- [29] Larose, T. Daniel. *Discovering knowledge in data: An Introduction to Data Mining Techniques*, John Wiley and Sons, New Jersey, 2005
- [30] Luan, Jing. *Data mining and its applications in higher education, new directions for institutional research*, No. 113, 2002, Springer
- [31] M. Srinivas and C. Krishna Mohan, "Efficient Clustering Approach using Incremental and Hierarchical Clustering Methods", 2010 IEEE
- [32] Mano, M.Morris and Charles R.Kine. *Logic and Computer Design Fundamentals*, Third edition. Prentice Hall, and 2004.p.73
- [33] Nisbet, Robert and Elder, John and Miner, Gary., *Handbooks of Statistical analysis and Data Mining Applications*, Academic Press Publications, 2009
- [34] Nkitaben Shelke, Shriniwas Gadage, "A Survey of Data Mining Approaches in Performance Analysis and Evaluation", (2015), *International Journal of Advanced Research in Computer Science and Software Engineering*
- [35] Peña-Ayala, Alejandro. "Educational data mining: A survey and a data mining-based analysis of recent works." *Expert systems with applications* 41.4 (2014): 1432-1462. Publishers, London
- [36] R. J. Mammone and Y. Y. Zeevi, *Neural Networks: Theory and Applications*. San Diego, CA: Academic, 1991
- [37] Rajni Jindal and Malaya Dutta Borah, "A SURVEY ON EDUCATIONAL DATA MINING AND RESEARCH TRENDS, *International Journal of Database Management Systems (IJDBMS)* Vol.5, No.3, June 2013.
- [38] Raymond T. Ng and Jiawei Han. Efficient and effective clustering methods for spatial data mining. In *Proc. of the VLDB Conference*, Santiago, Chile, September 1994
- [39] Renza Campagni, Donatella Merlini, Renzo Sprugnoli, Maria Cecilia Verri "Data mining models for student careers", at *Science Direct Expert Systems with Applications*, pp55085521, 2015, www.elsevier.com.
- [40] Rokach, Lior. "A survey of clustering algorithm," in O.Maimon, L. Rokach (Eds.), *Data Mining and Knowledge Discovery Handbook*, 2nd Edition, Springer, 2010
- [41] Romero, C. & Ventura, S. (2007). "Educational Data Mining: A Survey from 1995 to 2005," *Expert Systems with Applications*, 33(1), 135-146.
- [42] Romero, C., Ventura, S. & Garcia, E. (2008). "Data Mining in Course Management Systems: Moodle Case Study and Tutorial," *Computers & Education*, 51(1), 368-384.
- [43] Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, IEEE Transactions on 40, no. 6 (2010): 601-618
- [44] Romero, Cristóbal, Sebastián Ventura, Pedro G. Espejo, and César Hervás. "Data Mining Algorithms to Classify Students." In *EDM*, pp. 8-17. 2008.
- [45] S. Archana, Dr. K. Elangovan —Survey of Classification Techniques in Data Mining, *International Journal of Computer Science and Mobile Applications* vol 2, Issue 2, February 2014, p.g. 65-71
- [46] Soni, Neha, and Amit Ganatra. "Comparative study of several Clustering Algorithms." *International Journal of Advanced Computer Research (IJACR)*(2012): 37-42. *Systems with Applications*, Vol. 33, 2007, 135-146.
- [47] The McCULLOCH- Pitts Model by Samantha Hayman
- [48] U. Fayadd, Piatetsky, G. Shapiro, and P. Smyth, *From data mining to knowledge discovery in databases*, AAAI Press / The MIT Press, Massachusetts Institute of Technology. ISBN 0–262 56097–6, 1996.
- [49] Veenman C.J., Reinders M.J.T., Backer E. A maximum variance cluster algorithm, [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(9):1273-1280.
- [50] Weiss, M. Sholom and Indurkha, Nitin. *Predictive Data Mining: A practical guide*, Morgan Kauffmann Publishers, USA, 1998