Clustering Using Performance Analysis In Educational Data Mining

M.Sivamani[1], S.M jagatheesan[2]
[2] Associate professor
[1, 2] Gobi Arts And Science College, Gobichettipalayam.

Abstract- *Data mining techniques are used to extract useful knowledge from raw data. The extracted knowledge is valuable and significantly affects the decision maker. Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique and increasingly large-scale data that come from educational settings and using those methods to better understand students, and the settings which they learn in. Whether educational data is taken from students' use of interactive learning environments, computer-supported collaborative learning, or administrative data from schools and universities, it often has multiple levels of meaningful hierarchy, which often need to be determined by properties of the data itself, rather than in advance. Issues of time, sequence, and context also play important roles in the study of educational data. The increase of technology use in educational systems has led to the storage of large amounts of student data, which makes it important to use Educational Data mining to improve teaching and learning processes. Educational Data Mining is useful in many different areas including identifying at-risk students, identifying priority learning needs for different groups of students, increasing graduation rates, effectively assessing institutional performance, maximizing campus resources, and optimizing subject curriculum renewal. Educational Data Mining (EDM) is an emerging trend which is concerned about the effective use of huge data coming from educational systems to improve and optimize teaching-learning process, and also there are various techniques and methods used to predict and analyze the student's behavior and their performance. we have developed a model to Analyze the student's performance by using clustering model with the help of k-means algorithm. We have focused on social media perspective data, which is strongest source of Educational Data Mining and nowadays, it is more widely used to be as an important factor in the field of Educational Data Mining.*

Keywords- Classification, Clustering, Educational Data Mining (EDM), Machine Learning.

I. INTRODUCTION

The data mining algorithms have proved their usefulness in various application domains such as Credit Card frauds, Sports, Health Care, Banking, Education and Insurance. The data mining techniques are used by researchers in the Educational domain [1] and it is known as Educational data mining. The Data mining techniques are applied to extract the hidden patterns from the educational data. The hidden patterns are helpful to the educational institutions to make the decisions. The decisions taken by academia are useful to improve the performance of the weaker students. The classification or clustering techniques are used by most of the researchers to know the implicit patterns from the educational data. The data mining techniques are broadly classified as supervisory and un supervisory techniques. The supervisory techniques are helpful for the data already contain known class labels. On the other hand, un supervisory techniques are also called as clustering techniques. The clustering techniques gained interest by researchers. Clustering is broadly used in many applications such as market research, pattern recognition, data analysis, image processing, academic performance and intrusion detection. Educational Data mining focuses on analyzing data generated in an educational setup by the various intra-connected or disparate systems to develop model for improving learning experience and institutional effectiveness. Data mining also sometimes referred to as knowledge discovery in databases (KDD) is a known field of study in life sciences and commerce but the application of data mining to educational context is limited [2]. Just like commercial e-commerce based websites are using recommender systems that collect user browsing data and recommend similar products there have been efforts to apply the same in the educational context but they have not been successful as they are highly domain dependent Various methods have been proposed, applied and tested in data mining field and its argued by some researchers that these generic methods or algorithms are not suitable to be applied to this emerging field of study. It's proposed that educational data mining methods must be different from the standard data mining methods because of multi-level hierarchy and non-independence in educational data [1]. Institutions are increasing being held accountable for student success [3]

Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. The training data are analyzed by classification algorithm. The classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier. 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:

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

Clustering

Clustering can be said as identification of similar classes of objects. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes. Classification approach can also be used for effective means of distinguishing groups or classes of object but it becomes costly so clustering can be used as preprocessing approach for attribute subset selection and classification. 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. 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, and Constraint based method can also be used and the clustering can be done to the dataset [7]. Some of the fields have been undergoing clustering techniques in the fields such as data mining, image processing, text mining, machine learning and pattern recognition [27]. Educational data mining is concerned with developing new methods to discover knowledge from educational database in order to analyze student trends and behaviors 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] 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].Data mining can be applied in classifying students based on academic achievements, knowledge, gender, age, semester-wise grades and monitoring of progression and regression.

Predication

Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. In data mining independent variables are attributes already known and response variables are what we want to predict. Unfortunately, many real-world problems are not

simply prediction. Therefore, more complex techniques (E.g., logistic regression, decision trees, or neural nets) may be necessary to forecast future values. The same model types can often be used for both regression and classification. For Example, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.

Association rule

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.

Neural networks

Neural network is a set of connected input/output units and each connection has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the remarkable ability to derive meaning from complicated or imprecise data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are well suited for continuous valued inputs and outputs. Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs.

Decision Trees

Decision tree is tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

Nearest Neighbor Method

A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset (where k is greater than or equal to 1). Sometimes called the k-nearest neighbor technique.

II. APPLYING DATA MINING IN EDUCATION

Following are the steps in Data mining process :

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

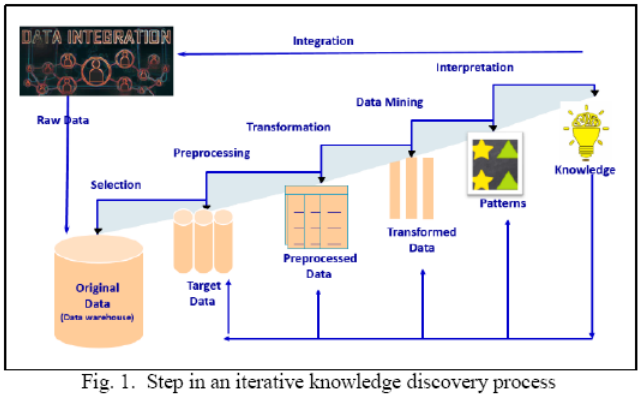

Fig. 1. Step in an iterative knowledge discovery process

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

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]. The emergent field of EDM examines the unique ways of applying data mining techniques education institutions academic problems.

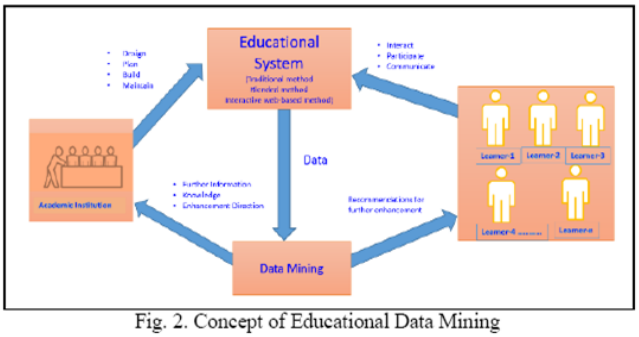

Fig. 2. Concept of Educational Data Mining

Fig. 2 shows the 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.

III. MACHINE LEARNING

Machine learning aims at building an intelligent system which will be intelligent enough to determine a decision or calculate output based on new inputs after passing the learning phase and being fed with a set of training data. According to the definition of Tom Mitchell: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E".

Learning can be a supervised learning where the correct output in the training set is made available. Supervised learning is used to solve regression or classification problems. Supervised as the learning Application of machine learning includes classification and regression. Examples of classification problems include identifying an email as a spam, face recognition and hand writing recognition while regression problems include building a model for a system that can be used to predict the output value of the system for a given input. The other type of learning is unsupervised learning where the exact output is unknown. This type of learning is used typically to solve clustering problems. The machine leaning techniques is a supervised learning algorithm for producing decision tree , it was proposed by Ross Quinlan as an extension of the basic ID3 algorithm. It is considered a statistical classifier as it can deal with both continuous and discrete attributed and data set with missing attributes values. The standard algorithm is as follows:

1) Read set S of examples described by continuous or Discrete attributes.
2) Identify base case.
3) Find the attribute which has the highest informational Gain (Abest).
4) Divide S into S1, S2, S3... according to the values of Abest.
5) Repeat the steps for S1, S2, S3 etc ..

K-Means Clustering

K-means is an unsupervised learning algorithm. It is one of the partitioning clustering procedures. It is dependent on distance-based that split "n" dataset into the specific predetermined number of clusters in which each cluster is associated with the centroid and then each point in the dataset follows the cluster in the nearest centroid. The basic k-means algorithm which is standard and simple as given below:

1) Select K points as the initial Centroids repeat
2) Form K clusters by assigning data points to nearest Centroid
3) Recalculate the Centroid of each cluster
4) Until the Centroids do not change.

Implementation of K-Means Clustering Algorithm

The K-Means is the unsupervised algorithm used for clustering the data objects. The K-Means clustering algorithm partitions the „n" data objects into „k" clusters (groups), in which each data object belongs to the cluster with the nearest mean. The data objects in each group are highly cohesive and the object in other group is disjoint. The K-means algorithm creates „k" distinct group of elements using the sum of squares. The input parameter for the algorithm is number of centroids. Then the distance between each element with each centroid is calculated. The distances calculated for a data element with each centroid is compared and assigned the data element to nearest centroid. In this way, all the data elements are assigned to one of the centroid. Initially, K clusters are formed by assigning the data elements to respective centroid. Then recalculate the centroid of assigned data elements in each cluster. Once again with the new centroids calculate the distance between each data element with the new centroids and reassign the data element to centroid which is near. This process continues until the no data element assigns to any new centroid, that means the centroids of „n-1" iteration is equal to centroid of „n" iteration.

The distance measure in K-means clustering is Euclidean distance. Suppose the elements are X=(x1,x2,x3…) and Y=(y1, y2,y3,…).

D (X, Y) = (x1-y1)2 + (x2-y2)2 +….
∑√∑_(i-1)^n▒〖xi-yi2〗

The distance between each data element and the centroid is calculated. And the data element is assigned to the centroid with the minimum distance. The centroid is the mean of all data points in that group. Each centroid with set of the data elements is called as cluster.

IV. RELATED WORK

Many researchers have contributed to the field of data mining in education. The researchers will give an overview on a few representative works. Abu Tair and El-Halees [3], gave a case study from the higher education stage. The main purpose of their study is to show how useful data mining can be in the educational domain in order to discover many kinds of knowledge by applying the graduate student data set on the different educational data mining techniques by using Rapid Miner software to discovered classification, clusters, association rule, and outlier detection then gave description to their importance in the education domain. M. Sukanya, S. Biruntha, S. Karthik and T. Kalaikumaran [4] applied the Bayesian classification technique on the existing higher education student. The main goal of their study is to predict the number of upcoming students in the next year based on the valid number of enrolled students in the previous years. This study helps decision makers to manage the number of resources and staffs they need to administer the outcomes of a student. This study helps also the teachers to know at early stage the students that need more attention to facilitate taking the correct action at the suitable time to reduce the failure in the academic approach and improve the student's academic performance. Md. Hedayetul Islam Shovon and Mahfuza Haque [5] implemented a k-means cluster algorithm. The main goal of their study is to help both the instructors and the students to improve the quality of the education by dividing the students into groups according to their characteristics using the application which have been implemented. Er. Rimmy Chuchra, M. tech [6] gave a case study from the higher education university. Their study was based on applying the educational data mining techniques on the existing student data set from the database university to discovered cluster, decision tree and neural network to show how to evaluate the performance of students with the usage of these techniques. Brijesh Kumar Bhardwaj and Saurabh Pal [7] applied Baysian classification on the student database from the higher education stage. This study aimed at identifying those students which needed more attention to reduce the drop out ratio and take action at a right time which helped to improve the performance of the students and the instructors. Md. Hedayetul Islam Shovon and Mahfuza Hague [8] applied a hybrid procedure that was based on Decision Tree and Data clustering from Data Mining methods. The main goal of their study is to predict the Grade Point Average which helped the teachers to reduce the drop out ratio to improve the performance of the students and the academics.

Framework Description

The proposed framework uses both classification and clustering techniques to suggest recommendations for a certain department for a student or an educational dataset that is required. This framework shall include attributes representing:-

- Student academic level before entering college
- Department chosen by student
- Student grade in the first year

Classification Phase

In this phase, a classification algorithm is applied on the educational dataset to find an efficient classifier. The role of the classifier is to output the department recommended for the student. The steps in this phase are as follows:

1) Remove all the records for the student who failed in his/her first year
2) Use this training dataset, and apply different classification algorithm with the Department attribute as the class.
3) Record the set of rules for the classification algorithm with highest F-Measure

Clustering Phase

In this phase, a clustering algorithm is applied on the educational dataset to divide student records into a number of clusters based on marks' similarity. The steps in this phase are as follows:-

1) Remove all attributes regarding the Department chosen by student.
2) Remove all attributes regarding first year grade.
3) Choose the number of clusters.
4) Use K-means algorithm to identify the clusters.
5) Identify the distribution percentage of each department along all clusters
6) Record the set of rules.

Request an Output from the System

A user can ask the system to acquire a recommendation for a certain educational department. The steps of this phase can be summarized as follows:

1) The new student will enter his/her data
2) The purposed system will read the data and validate its soundness
3) Predict the cluster (Xcluster) according to rules declared by clustering phase

4) Output the department with the highest percentage rate in Xcluster
5) Predict the department according to rules declared by classification phase. If both predict the same department, the output will be one choice, otherwise the output will be two choices where the first choice will be the one with the highest accuracy and the other will be the second choice.

V. CASE STUDY

The Student Data used in this case study is obtained from "Cairo Higher Institute for Engineering, Computer Science, and Management" (CHI which is located in Cairo, Egypt. The institute has four departments:

1) Management Information System (MIS)
2) Computer Science (CS)
3) Architecture Engineering (AE)
4) Computer Engineering (CE)

VI. RESULT AND DISCUSSION

Clustering is defined as the grouping of unsupervised set of objects based on their similar characteristics [10]. The term cluster is defined as the group of objects with similar characteristics contained within clustering. Clustering is similar to classification, the only difference in it is that clusters are not predefined where in classification groups are predefined. Set of objects within a one cluster is similar to each other but different than set of objects resides within another clusters. Results obtained by clustering are dynamic due to Cluster model is broadly classified into following categories depicted by Figure 2.

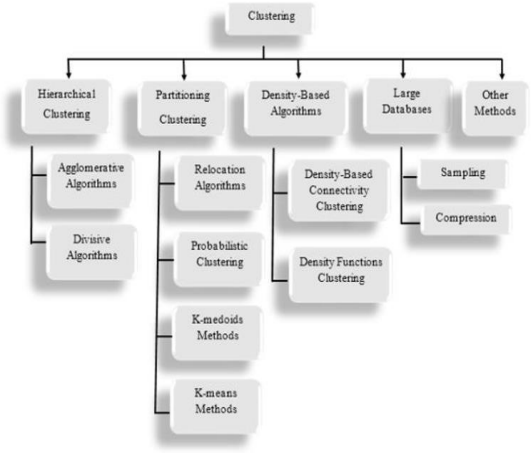

Figure 2. *Classification of Clustering in Data mining*

we are using k-means clustering technique to identify student's performance. K- means methods are most frequently used clustering technique, which can partition the unlabeled [12] data sets into 'k' number of clusters automatically [9]. Here the term unlabeled data means that the available data set is currently does not belongs to any group i.e. it does not have preexisting groups available because as we are familiar that k-means is the type of clustering which contained unsupervised type of learning. K-means algorithm are simple to implement [13] due the reason that k-means provides the methods which provides approximate, scalable solutions. This algorithm is mainly used to discover the groups based on the number assigned to the k variable in order to represent the groups [12].

We are using Euclidean distance calculation technique of k-means clustering to discover the number of clusters in the data set. In this algorithm,

- Initial clusters are randomly chosen.
- • Iteratively, items are moved among sets of clusters until the desired set is reached.
- • High degree of similarity among elements in a cluster is obtained.

Analysis of data is most important part of decision making. In Educational Data Mining, various data mining techniques and methods are used to predict and uncover hidden trends and patterns and making accuracy based predictions through a higher level of analytical sophistication in the process of counseling students.

To keep track of student's details, various types of educational software's are available which may varies in subjects [14] and also now for the ease of use and availability, IT companies associated to educational domain are building the educational applications that are beneficial for both students and teachers used as teaching-learning tools. These teaching-learning tools provide functionalities to both teacher and students. Due to these tools teachers are able to posts assignments, share lecture notes and presentation slides, tutorials, videos, and so on. The educational companies integrate the education applications with the games. These games are the instructional games. Most popular example of Educational Data Mining is Module which is open source online learning management system and anything produced by them is freely available all over the globe. Module faculties are able to give the assignments to the students, they can chat, organize aptitude tests, and much features are available that improves the learning-teaching procedure. Educational Data Mining is most widely used to evaluate and interpret teaching-learning process, student's performance and so on. Some of the activities are mentioned below to better understand about Educational Data Mining. Predicting and analyzing useful data consists of performance, learning, interest, needs of students

from raw data is very important in educational community. This extracted data involves various techniques available in data mining. We took two clusters, k=2 , When the iteration process is completed, there is no object move between groups. Clustering is done with the iteration process and complexity depends on the number of objects. Finally we obtain the group result, which is shown in the following Table 2.

Result of Clustering as shown below

Subject	No of students(Attribute1)	Placements (Attribute2)	Group(Result)
Tamil	50	22	1
English	30	18	1
mathematics	25	10	2
Science	60	35	2

Every object must be belonged to any of the group. This grouping result can be used in many ways. The result can be used to find the trends in the university about subjects, and courses which are preferred by the universities and companies. The whole process can be automated with the data mining tools and software. The aim of clustering is to create n-groups of students of homogeneous levels with respect to learning.

VI. CONCLUSION

Recent advances in the use of technology in educational field create huge data. Data mining can be applied on this data, for understanding and Analyze student's performance, and is known as educational data mining. This prediction can be taken into account for improving the effectiveness of teaching-learning process. we have developed a model to predict the student's performance by using clustering model with the help of k-means algorithm. It is observed that, the model built using Educational Data Mining techniques incorporate students overall behavior and pedagogy to analyze the knowledge and teaching and learning outcomes. This research is a starting attempt to use data mining functions to analyze and evaluate student academic data and to enhance the quality of the higher educational system. Data Mining is directly associated with use of technology for accessing data and to give result as required in a desired way. Cluster analysis is a popularly used data analysis method in number of areas. We make use of data mining process in a student's database using k-means clustering algorithm to predict student's placement behavior. K-means is a well known partitioning based clustering technique that attempts to find a user specified number of clusters represented by their centroids. Cluster Analysis was performed to group the data into clusters based on its similarities. With the help of data mining methods, such as clustering algorithm, it is possible to discover the key characteristics from the students' performance and possibly use those characteristics for future prediction. We hope that the information generated after the implementation of data mining technique may be helpful for educational institutions.

REFERENCES

[1] Shyamala K., Rajagopalan S. P.,"Data Mining Model for a better Higher Educational System", Information Technology Journal, Vol. 5, No. 3, pp. 560-564, 2006.

[2] Heikki, Mannila,"Data mining: machine learning, statistics, and databases", IEEE, 1996.

[3] U. Fayadd, Piatesky, G. Shapiro, P. Smyth,"From data mining to knowledge discovery in databases", AAAI Press / The MIT Press, Massachusetts Institute of Technology, 1996.

[4] Sun Jigui, Liu Jie, Zhao Lianyu,"Clustering Algorithms Research", Jounal of Software. 19.

[5] Suman, Pooja Mittal P., and M. Pooja. "A Comparative Study on Role of Data Mining Techniques in Education: A Review." *International Journal of Emerging Trends & Technology in Computer Science* 3.3 (2014): 65-9.

[6] Tair, Mohammed M. Abu, and Alaa M. El-Halees. "Mining educational data to improve students' performance: a case study." *International Journal of Information* 2.2 (2012): 140-146.

[7] Bhise, R. B., S. S. Thorat, and A. K. Supekar. "Importance of data mining in higher education system." *IOSR Journal Of Humanities And Social Science (IOSR-JHSS) ISSN* (2013): 2279-0837.

[8] Romero, Cristóbal, and Sebastián Ventura. "Educational data mining: a review of the state of the art." *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 40.6 (2010): 601-618.

[9] Priya C.,Sana S., "Educational Data Mining: A Critical Study ",BVIMSR'S Journal of Management Research ISSN: 0976-4739.

[10] http://bigdata-madesimple.com/what-is-clustering-in-data-mining/

[11] T. M. Mitchell. Machine Learning. McGraw-Hill, New York, 1997.

[12] Shi Na, Liu Xumin, and Guan Yong. Research on k-means clustering algorithm: An improved k-means clustering algorithm. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on, pages 63–67, 2010.

[13] A. M. El-Halees M. M. Abu Tair. Mining educational data to improve students' performance: A case study. International Journal of Information and Communication Technology Research, 2(2): 140–146, April 2011.

[14] S. KarthikM. Sukanya, S. Biruntha and T. Kalaikumaran. Data mining:Performance improvement in education sector using classification and clustering algorithm. In

Proceedings of the International Conference on Computing and Control Engineering, ICCCE 2012, 2012.

[15] Mahfuza Haque Md. Hedayetul Islam Shovon. Prediction of student academic performance by an application of k-means clustering algorithm. International Journal of Advanced Research in Computer Science and Software Engineering, 2(7): 353–355, July 2012.

[16] M. tech Er. Rimmy Chuchra. Use of data mining techniques for the evaluation of student performance: a case study. International Journal of Computer Science and Management Research, 1(3): 425–433, October 2012.

[17] Nisanci, M., et al., The prediction of the electric field level in the reverberation chamber depending on position of stirrer. Expert Systems with Applications, 2011. 38(3): p. 1689-1696.

[18] Zhao, Y. and Y. Zhang, Comparison of decision tree methods for finding active objects. Advances in Space Research, 2008. 41(12): p. 1955-1959.

[19] Quinlan, J.R., C4. 5: Programming for machine learning. Morgan Kauffmann San Mateo, CA, USA 1993.

[20] Kou, G., Y. Peng, and G. Wang, Evaluation of clustering algorithms for financial risk analysis using MCDM methods. Information Sciences, 2014. 275: p. 1-12.

[21] Banitaan, S., A.B. Nassif, and M. Azzeh. Class Decomposition Using K-Means and Hierarchical Clustering. in 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA). 2015. IEEE.

[22] Japkowicz, N. Supervised learning with unsupervised output separation. In International Conference on Artificial Intelligence and Soft Computing. 2002.

[23] Ashfaq, R.A.R., et al., Fuzziness based semi-supervised learning approach for intrusion detection system. Information Sciences, 2016. 378: p. 484–497.

[24] Wang, G., et al., A Bayesian network model for prediction of weather-related failures in railway turnout systems. Expert Systems with Applications, 2017. 69: p. 247-256.

[25] Lichman,M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml].2013.

[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] Pena-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] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," User modeling and user-adapted interaction, vol. 4, no. 4, pp. 253–278, 1994.

[37] R. Baker et al., "Data mining for education," International encyclopedia of education, vol. 7, pp. 112–118, 2010.