Educational Data Mining for Student Learning Pattern Analysis using Clustering Algorithms

Kamal Bunkar, Sanjay Tanwani

Abstract: The exponential increase in universities' electronic data creates the need to derive some useful information from these massive amounts of data. The progression in the data mining field causes it conceivable to educational data to improve the nature of educational processes. This study, thus, uses data mining methods to study the learning behavior and performance of university students. It focused on two aspects of the performance of the students. First, predicting students' learning behavior at the end of a complete year of the study program. Second, predict student performance with the help of the data model proposed by this study. Finally, provide course material recommendations using the data mining algorithm. Three data mining algorithms were considered which are K-Means, FCM, and KFCM., and maximum accuracy of 90.22% was achieved by KFCM. The study indicates that in terms of time and memory usages K-means algorithm give better results. This creates an opportunity for identifying students that may graduate with poor results or may not graduate at all, so early intercession might be possible.

Keywords: Data Mining, Educational Data mining, Clustering Algorithm, learning behavior, student learning pattern.

I. INTRODUCTION

Data mining techniques are there to help with exploring the data and obtaining useful patterns. They are also used for various data-related applications and their improvements. Data mining techniques offer users various kinds of algorithms for analyzing the data [1]. Based on data formats available (i.e. structured or unstructured) the algorithms are selected to employ. There are two main kinds of algorithms available supervised and unsupervised learnings [2]. These algorithms are supporting various kinds of data analysis strategies such as prediction, classification, clustering, association mining, and many more. According to the different application areas of these algorithms and methodologies identified, such as web mining, educational data mining, and others [3]. In this work the data mining techniques are used with the educational sector data therefore, that is termed as Educational Data Mining (EDM) [4].

The proposed work mainly focuses on exploring the educational data for empowering students learning, enhancing the teacher's productivity, obtaining the student learning and study patterns, and recommending the compatible study or course material is the primary aim of the proposed work. In this context, the work is further subdivided into three main modules. In the first module, the clustering algorithm is employed on the student's historical performance records so that it can obtain a group of weak, average, and efficient students. Through analyzing these student groups we can identify the learning behavior of students and then can offer them the most compatible course or study material. Additionally, that also works as feedback to educators and institutions for optimizing their educational strategy or methodology and offering more suitable resources to their students. In the second module, we offer a predictive technique with the students' performance data for understanding the future improvement and growth of the students. Finally, a recommendation model is implemented that combines the clustered data and the predictive student performance data for understanding the student's learning behavior and then offers the most compatible learning material. Using these three modules the proposed data model is developed. The EDM systems are not only offering just employment of data mining algorithms it also offers to explore the patterns and hidden information in raw academic data. In this context the four aim of the proposed work is established:

1. To understand the learning methodology of the students.
2. To improve the fecundity of teachers and educators.
3. To predict the performance of students and to make compatible resources available and to predict the future growth of students.
4. To recommend and offer the best suitable course materials for the appropriate learning to the students.

Among the first two goals are accomplished using the clustering algorithms where the three groups of students are prepared for demonstrating the student learning behavior and providing feedback to the educators or teachers on the basis of the teachers are preparing the new strategy to deal with below-average low performing students. Secondly, for the accomplishment of objective three, the predictive methodology is used where the decision tree algorithms are used to classify the student's performance and predicting their future growth. Further based on the learning patterns a recommendation model is developed which are used to offer relevant course materials to the students.

II. EDUCATIONAL DATA MINING (EDM)

EDM is the application of data mining techniques to educational data, and so, its objective is to analyze educational and academic data. In order to resolve educational research issues. DM can be defined as the process involved in extracting interesting, interpretable, useful, and novel information.

Kamal Bunkar*, Ph.D Scholar, School of Computer Science and IT DAVV, Indore, India.
Prof. Sanjay Tanwani, Professor and Head of Department, Department of Computer Science and IT DAVV, Indore, India.

Educational Data Mining for Student Learning Pattern Analysis using Clustering Algorithms

EDM is concerned about creating strategies to investigate one of a kind sorts of information in instructive settings. What's more, utilizing these strategies to all the more likely get understudies and the settings in which they learn.On one hand, the increase in e-learning as well as student databases has created large repositories reflecting how students learn.

On the other hand, the use of the Internet in education has created a new context known as e-learning in which large amounts of information about teaching-learning are generated and available. EDM seeks to use these data repositories to better understand learners and learning, to develop computational approaches that combine data and theory to benefit learners [5]. EDM has emerged as a research area for researchers from different and related areas:

- Offline education tries to transmit knowledge and skills face-to-face contact and also study psychologically on how humans learn. Psychometrics and statistical techniques have been applied to data like student behavior, performance, and curriculum.
- E-learning and Learning Management System (LMS). E-learning gives online guidance and LMS likewise gives correspondence, joint effort, organization, and announcing instruments.Web Mining techniques have been applied to student data stored by these systems in log files and databases [6].
- Intelligent Tutoring (ITS) and Adaptive Educational Hypermedia System (AEHS) are an alternative to the just-put-it-on-the-web approach by trying to adapt teaching to the needs of each particular student. Data Mining has been applied to data picked up by these systems, such as log files, user models, etc.

The EDM converts raw data coming from educational systems into useful information that could potentially have a great impact on educational research and practice. This process does not differ much from other application areas of data mining like business, medicine, etc. because it follows the same steps as the general process: pre-processing, mining and post-processing [7].

A. Challenges of EDM

EDM allows, to discover new knowledge based on students' usage data to help validate/evaluate education systems, to improve quality of education and to lay the groundwork for an effective learning process. Some similar ideas were already successfully in e-commerce, in order to determine clients' interests to increase online sales. However, there has been less progress in this direction in Education, this situation is changing and there is currently an increasing interest in applying data mining to the educational environment. There are some important issues that differentiate the application of DM to education from how it is applied in other domains [8]:

- **Objective.** The objective of data mining in each application area is different. For example, in business the objective is to increase profit, which can be measured in term of money, a number of customers, and customer loyalty. But EDM has both objectives, such as improving the learning process and guiding students; as well as pure research objectives, such as achieving a deeper understanding of educational phenomena [9].
- **Data.** In education there are many types of data available. These data are specific to the educational area and so have semantic information, relationships with other data, and multiple levels of hierarchy. For example

model used in ITS and AEHS, represents the relationships among the concepts of a subject in the graph; and the q-matrix shows relationships between questions of a quiz system and the concepts of the test [10].

- **Techniques.** Educational data and problems have some special characteristics that require to be treated in a different way. Although most of the traditional techniques can be applied directly, others cannot and have to be adapted to the specific educational problem. Specific mining techniques can be used for specific educational problems [11].

B. EDM Users and Objectives

EDM involves different groups of users or participants. Different groups look at educational information from different angles according to their own mission, vision and objectives. For example, knowledge discovered by EDM can be used not only to help teachers to manage their classes, understand students' learning and teaching methods, but also to support a learner's reflections on the situation and provide feedback to learners. There are actually more groups involved with many more objectives [12]:

- **Learners:**
  - To personalize e-learning;
  - To recommend activities to learners and resources, and learning tasks that could improve their learning;
  - To suggest interesting learning experiences to the students;
  - To suggest path pruning and shortening or simply links to follow, to generate adaptive hints, to recommend courses, relevant discussions, books, etc [13].
- **Teachers/Instructors:**
  - To get feedback about instruction,
  - To analyze learning and behavior to detect which students require support,
  - To predict student performance, to classify learners into groups,
  - To find a learner's regular and irregular patterns, frequently made mistakes, effective activities,
  - To improve the adaptation and customization of courses etc [13].
- **Educational Researchers:**
  - To evaluate and maintain courseware,
  - To improve student learning,
  - To evaluate the structure of course content and effectiveness in the learning process,
  - To automatically construct student models and tutor models,
  - To compare data mining techniques to be able to recommend the most useful one for each task,
  - To develop specific data mining tools for educational purposes etc [14].
- **Organizations/Universities:**
  - To enhance the decision processes in higher learning institutions,
  - To streamline efficiency in the decision making process,
  - To achieve specific objectives,

o To suggest certain courses that might be valuable for each class of learners,
o To find the most cost-effective way of improving retention and grades,
o To select the most qualified applicants for graduation,
o To help to admit students who will do well in university etc [14].

- **Administrators:**
o To develop the best way to organize institutional resources (human and material) and their educational offer,
o To utilize available resources effectively
o Toenhance educational program and determine the effectiveness of the distance learning approach
o Toevaluate teacher and curricula,
o To set parameters for improving efficiency and adapting it to use etc. [15].

C. Educational Task

There are many issues in educational environments that have been resolved through DM For example improving student, improving domain, support, research into learning and learners and methods: prediction, clustering, relationship mining, distillation of data and discovery. Applications dealing with the assessment of the student are learning performance, applications that provide course adaptation and learning recommendations based on learning behavior, approaches dealing with the evaluation of learning material and educational courses, applications that involve feedback to both teacher and students, and developments for detection of typical students' learning behaviors. There some categories come from different research communities and they also use different techniques [16].

A. Analysis and visualization of data: The goal of the investigation and representation of information is to feature helpful data and bolster dynamic.For example, it can help educators to analyze the course activities and usage information to get a view of learning. Statistics and visualization are the main techniques that have been used for this task [17].

B. Providing feedback for supporting instructors: The objective is to provide feedback, to support course, in decision making about student's learning, organize resources, and enable them to take appropriate action. It is important to point out this task is different than analyzing and visualizing tasks. Moreover, providing feedback, hidden and interesting information extraction in data. Several techniques have been used in this task, and association rule mining has been the most common. It reveals relationships among variables and presents in the form of rules [18].

C. Recommendations for students: The objective is to make recommendations with respect to their activities, like URLs to visits, the next task or problem to be done etc. also able to adapt learning contents, interfaces and sequences. Several techniques have been used for this task commonly are association rule mining, clustering and sequential pattern mining [19].

D. Predicting student performance: The objective is to estimate the value of a variable that describes the student. In education the values normally predicted based on performance, knowledge, score or mark. This value can be a numerical or categorical value. Relapse examination and Classification is a method where singular things are set into bunches dependent on quantitative data with respect to at least one attribute. The expectation of an understudy's presentation is one of the most seasoned and well-known applications, and various procedures have been likewise applied (neural systems, Bayesian systems, rule-based frameworks, relapse, and relationship investigation) [20].

E. Student Modeling: The aim is to develop cognitive models of students, including modeling of their skills and knowledge. DM has been applied to consider user characteristics (motivation, satisfaction, learning styles, affective status, etc.) and behavior to automate the construction of student models. Different DM algorithms have been used for this task (mainly, Bayesian networks) [21].

F. Detecting undesirable student behaviors: The objective of detecting undesirable student behavior or detect students who have some problems such as: erroneous actions, low motivation, misuse, cheating, dropping out, failure etc. mainly classification and clustering have been used [21].

G. Grouping students: To create groups of students according to their features. The clusters of students can be used by the instructor to build a personalized learning system, to promote group learning, to provide adaptive content etc. The techniques used are classification and clustering [22].

H. Social network analysis: Social Networks Analysis aims at studying relationships between individuals, instead of individual attributes. A social network is considered to be a group of people, an organization that are connected by social relationships like friendship, cooperative relations, for information exchange. Different techniques have been used to mine social networks, but collaborative filtering is the most common. Collaborative filtering is a method of making predictions about the interests of a user by preferences. It can produce recommendations by computing the similarity between students' preferences, so this task is directly related to the previous task of recommendations [23].

I. Developing concept maps: To construct concept maps is to help instructors in the process of developing it. A concept map is a graph that shows relationships between concepts in a hierarchal structure. The association rules and text mining have been used [24].

J. Constructing courseware: To construct courseware to help instructors to carry out the development process of courseware and learning contents. It tries to promote the reuse and exchange of existing learning resources among different users [24].

K. Planning and scheduling: The aim of planning and scheduling is to enhance the educational process by planning future courses, helping with student course scheduling, planning resource allocation, helping in the admission and counseling processes, developing curriculum, etc. Different techniques have been used for this task mainly association rules [25].

III. CLUSTERING ALGORITHMS

The EDM process is used the clustering algorithm for identifying the student groups. That may help to understand the student's weakness, required improvements, and promotion of group study. Therefore the clustering algorithms are playing an essential role in EDM.

This section includes the algorithm used in preparing the proposed EDM model for student's learning behavior identification. In the literature, there are various clustering algorithms exits among some frequently used algorithms namely k-means, FCM, and a recently available KFCM is employed in this work. The k-means algorithm is implemented for cluster the dataset available for higher education student's performance. In this context the predefined centroids (k=3) are produced for preparing the group of students. The classical k-means algorithm is given in table 1. [26]

Table 1 k-means clustering

Input: dataset objects to be cluster $(x_1, x_2 \dots x_n)$, Number of clusters $k = 3 \ (i.e. \ low, mid, and \ high)$;
Output: $k = 3$ clusters and dissimilarity between each object and its centroid;
Process: 1. Arbitrarily select $k = 3$ objects as initial centroid $(m_1, m_2, m_{k=3})$;   2. Calculate the distance between each object x_i and three centroids,   3. Assign instances to the nearest centroid label, distance calculated as: $$d(x_i, m_i) = \sqrt{\sum_{j=1}^{d} (x_i - m_{j1})^2}, i = 1 \dots N, j = 1 \dots k$$ $d(x_i, m_i)$ is the distance between data i and cluster j.   4. Calculate the mean of instances in each cluster to update centroid, $$m_i = \frac{1}{N} \sum_{j-1}^{n_i} x_{ij}, i = 1,2,\dots,K$$ N_i is the number of instance in considered cluster i;   **5.** Repeat till the function E converged,   **6.** Return $(m_1, m_2, m_{k=3})$

Similarly the FCM (Fuzzy C-Means) clustering is split the data into the required number of groups. It is also an unsupervised learning technique. Let there is a vector which contains data instances $V = \{x_1, x_2, x_3, \dots, x_n\}$. Where x_i is the individual instance. Now we need to create k number of data segments then we need a distance or similarity function to make decisions of cluster assignments. In the FCM algorithm membership function is used. The categories are $C = c_1, c_2, \dots, c_k$. Here the membership function is defined as:

$$M_{i,j} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\|x_i - c_i\|}{\|x_i - c_k\|}\right)^{\frac{2}{n-1}}}$$

The membership function describes the relationship among data instances and the centroid. Using this function the decision is made to assign a label to the data. But for deciding the stopping criteria an objective function is required. When the objective function report values tend to zero then the algorithm stops. The following equation is used as an objective function.

$$O_n = \sum_{i=1}^{N} \sum_{j=1}^{k} M_{i,j}^{j} \|x_i - c_j\|^2$$

Where, O_n is the objective function, $M_{i,j}^{j}$ is the membership function used to establish the relationship, finally $\|x_i - c_j\|^2$ is the Euclidean distance function.

However, FCM is suitable for clustering. But to enhance the performance and dealing with nonlinear data we considered the kernel function. It is an optimization process. The fuzzy membership function used for clustering and as an objective function is used as a classical algorithm. But we found that the RBF function can be used for the unknown patterns of data and nonlinear structures. Therefore by using the RBF kernel function is used to modify the membership function such that.

$$M_{i,j} = \frac{1}{\sum_{k=1}^{c} \left(\frac{rbf(x_i, c_i)}{rbf(x_i, c_k)}\right)^{\frac{2}{n-1}}}$$

And objective function becomes

$$O_n = \sum_{i=1}^{N} \sum_{j=1}^{k} M_{i,j}^{j} rbf(x_i, c_j)$$

The Euclidean distance measuring function is replaced with the RBF kernel function. It is expressed as

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - y_j\|^2\right)$$

Which can be rewrite kernel as:

$$rbf(x, y) = \exp\left(-\gamma \left(\sqrt{\sum_{i=1}^{N} (x_i - y_i)^2}\right)\right)$$

IV. PROPOSED WORK

Student behavior analysis is mean of student learning pattern analysis. That helps to understand the learning ability of the students. Therefore using such kind of pattern analysis we can estimate how the different group of students is learning effectively and which group of students is weak. Therefore we need to group the students according to their current performance. In ML and DM for grouping of data according to the similarity of their features the clustering algorithms were used. Using the clustering algorithm a model for student learning pattern analysis is prepared. Figure 1 shows the data model for student learning pattern analysis. The different functional components are discussed in this section.

Student dataset: The work is intended to discover the student learning behavior according to their existing performance in different subject performance.

Therefore a university student's dataset is used here as an experimental dataset.

The cluster analysis producing the groups of students, who have a similar performance therefore according to their learning behavior (i.e. low, mid and high) the efforts for improvements can be applied. Here we have used Genius University students data obtained from kaggle repository.
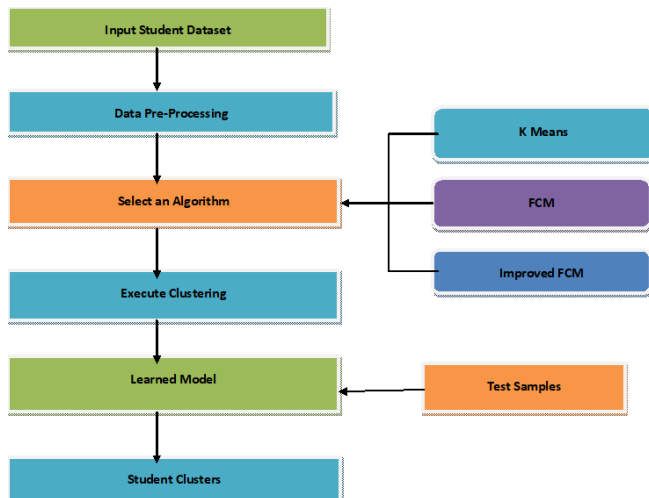

Figure 1 Student Behavior Analysis

Data preprocessing: the data preprocessing is an essential step of data mining where the data is optimized for improving the performance of learning algorithms. That is involved in cleaning those data instances which may produce conflicts during the decision making or can influence the actual target values. To preprocess the 2D vector data the following function can be used:

Table 2 Preprocessing Algorithm

Input: Dataset D

Output: Preprocessed data P

Process:

1. $[row, col] = ReadDataset(D)$
2. $for(i = 1; i \le row; i + +)$
   a. $for(j = 1; j \le col; j + +)$
      i. $if(D_{i,j} == null)$
         1. $D. RemoveRow(i)$
      ii. Else
         1. $P. AddRow(i)$
      iii. End if
   b. End for
3. End for
4. Return P

Select an algorithm: For computing the students learning behavior significant efforts are made. Among most of the authors is favoring the use of clustering algorithms or unsupervised algorithms. Thus here we implemented three clustering algorithms for performing clustering. These algorithms are namely k-means clustering, fuzzy c-means clustering and a modified fuzzy c mean clustering algorithm. The modified algorithm usages the Gaussian kernel function

for measuring the distances in place of a normal distance measuring function

Execute clustering: There are three clustering algorithms were implemented. The system user selects one of them for performing the experiments. The preprocessed data is used with the clustering algorithm and the groups of students are created according to their learning pattern similarity.

Learned model: The unsupervised learning algorithms after training are returning the selected centroids. Additionally, the relevant data instances which belong to the selected centroids. These centroids are further used for categorizing data according to the distance or their membership values.

Test samples: the test samples are prepared using the random selection of data instances from the initially available dataset. That test sample contains 30% of instances that are used for training.

Student clusters: the test data clustered in this phase according to their selected centroids. According to the correctly categorized data the performance of the implemented algorithms were calculated.

The above discussed process can be summarized using the algorithm steps the table 3 shows the key steps of the proposed algorithm.

Table 3 Proposed Algorithm

Input: student dataset D, Number of Clusters k = 3 i.e. {low, mid, high}, Selected algorithm A, Test Samples T

Output: Student clusters C

Process:

1. $D_n = ReadDataset(D)$
2. $P_m = preProcessData(D_n)$
3. $[centroid, index] = A. createCluster(P_m, k)$
4. $T_o = readTestData(T)$
5. $for(i = 1; i < o; i + +)$
   a. $C = A. Predict(T_i)$
6. $end\ for$
7. Return C

V. SIMULATION AND RESULTS

The understudy's presentation dataset is assessed utilizing three distinctive grouping calculations in particular k-means, FCM (fluffy c means), and an improved FCM calculation. These bunching calculations are utilized for making understudies bunches as per their presentation in three principle classes for example low, medium and high. These gatherings of understudies are useful for setting up the showing procedure for various entertainer understudies. That upgrades the educator's profitability just as understudy execution. In light of the completed grouping, the presentation of these three calculations is estimated and revealed in this segment.

Educational Data Mining for Student Learning Pattern Analysis using Clustering Algorithms

A. Accuracy

Accuracy is the sign of the decency of any data mining calculation. This is the proportion of the rightness of the calculation. The following is composed of the recipe through which we can ascertain exactness.

$$Accuracy(\%) = \frac{total\ correctly\ Classified\ Samples}{total\ samples\ to\ classify} X100$$

Table 4 Comparative Accuracy (%)

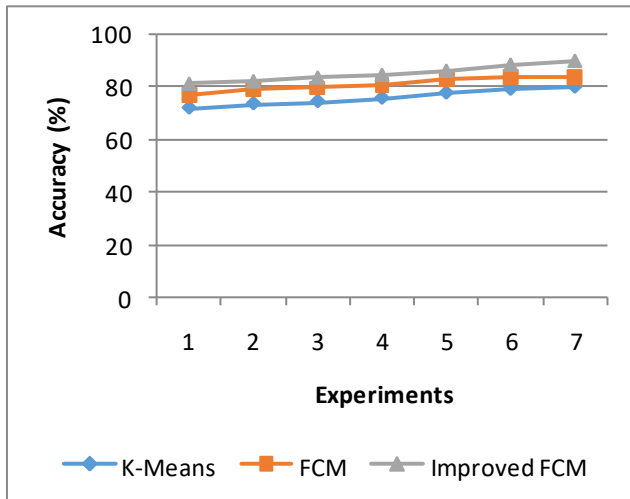| Exp. No	K-Means	FCM	Improved FCM
1.	72.02	76.76	81.25
2	73.69	78.82	82.21
3	74.55	79.52	83.66
4	75.87	80.41	84.74
5	77.88	82.64	86.36
6	79.36	83.54	88.74
7	80.2	83.63	90.22

Figure 2 Comparative Accuracy (%)

The accuracy of clustering algorithms in a comparative manner is reported in figure 2 and table 4. The different experiments are carried out with a diverse set of testing datasets. Additionally, the captured results are reported in both tables and figures. According to the obtained results the k-means clustering is efficient in terms of clustering creation but produces less accurate results as compared to FCM and improved FCM. Both algorithms are providing accurate results as compared to k-means clustering. Therefore in further experiments, the improved FCM algorithm is used.

B. Memory usage

Memory is a basic presentation boundary traditionally. That indicates the computational requirements of the algorithms. In JAVA technology it is computed on the basis of the following formula:

$$memory\ used = total\ memory - free$$

The memory usages of the implemented clustering algorithms for finding the student's learning behavior is reported in figure 3 and table 5. The X-axis of this diagram includes the different experiments conducted with the system

and the Y axis shows the required amount of memory in terms of KB (kilobytes).
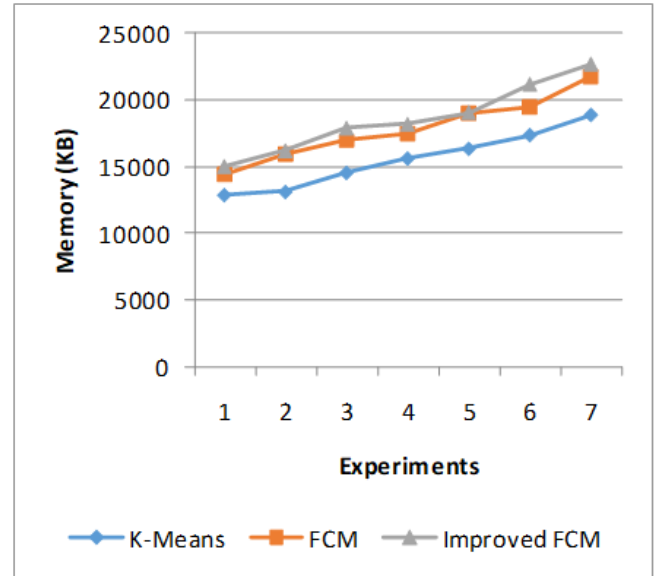

Figure3 Comparative memory consumption (KB)

According to the measured results, the improved FCM algorithm is consuming the higher memory as compared to the other two implemented algorithms. In recent literature, we observe that memory is not much significant due to the low cost of memory resources.

Table 5 Comparative memory consumption (KB)

Experiment No	K-Means	FCM	Improved FCM
1	12872	14388	15031
2	13106	15922	16194
3	14530	16928	17882
4	15605	17432	18174
5	16331	18994	19048
6	17359	19472	21135
7	18864	21743	22638

C. Time usages

To process a measure of information calculation requires a measure of time, this time necessity is named here as time uses. That is determined utilizing the accompanying recipe:

$$time\ required = end\ time - start\ time$$

Table 6 Time consumption (MS)

Experiment No	K-Means	FCM	Improved FCM
1	26	56	75
2	48	105	150
3	70	220	260

486

4	150	300	300
5	200	350	450
6	300	450	550
7	350	500	650

The time consumption of all three algorithms for students learning pattern analysis is reported in table 6 and their line graph is provided in figure 4. The X axis shows the experiments and the Y axis shows the time consumed in terms of milliseconds. The observation said the large number of instances the large consumption. Therefore according to the performance of k-means is efficient as compared to FCM and improved FCM algorithm. but in terms of accuracy the improved FCM is the winner.
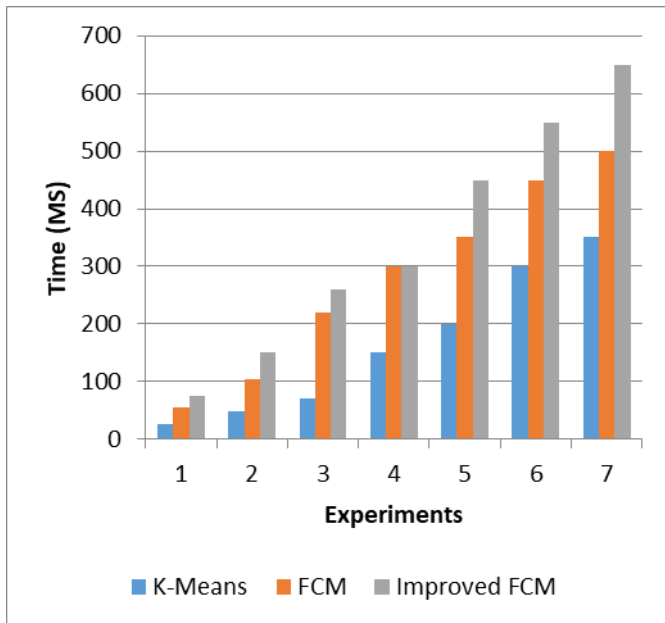

Figure 6.3 Time consumption (MS).

VI. CONCLUSION & FUTURE WORK

The aim of the proposed work is to investigate the EDM processes, existing challenges, ongoing techniques used identification. Therefore we proposed three modules of the work to be implemented and experiment. Initially, a model for student learning patterns or behavior analysis is proposed. That technique usage the clustering algorithm to the learning behavior identified. This obtained outcome is compared with the help of k-means, FCM, and a kernel-based FCM algorithm. The input student performance data is produced to these algorithms and the model's performance is evaluated. According to found results, the k-means clustering algorithm is efficient in terms of time and memory requirements. But the kernel-based FCM is accurate enough as compared to k-means. In this presented work the accuracy is the key parameter for data mining employment. Thus in further experiments, the KFCM (Kernel-based FCM) is being used. Along these lines, the main goal of the work for finding the model is to acquire understudy learning designs that are practiced effectively.

Sooner rather than later, we give two other information models. The first is utilized for experimentation of managed learning approach in this way CART (grouping and relapse tree) calculation and C4.5 is planned to be utilized. Moreover utilizing the productive classifier the understudy execution expectation framework is proposed. Further to support the client (understudy/instructors) the model is reached out to give a valuable understudy execution expectation and course material suggestion device that help the teachers to offer reasonable assets to their understudies.

REFERENCES

1. Mrs. Bharati M. Ramageri, "DATA MINING TECHNIQUES AND APPLICATIONS", Indian Journal of Computer Science and EngineeringVol. 1 No. 4 301-305.
2. Lidong Wang, Chery Ann Alexander, "Machine Learning in Big Data", International Journal of Mathematical, Engineering and Management SciencesVol. 1, No. 2, 52–61, 2016
3. Cristóbal Romero, Member, IEEE, Sebastián Ventura, Senior Member, IEEE, "Educational Data Mining: A Review of theState-of-the-Art", Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews
4. P. MeenaKumari, SK.AbdulNabi and PuppalPriyanka, "Educational Data Mining and its role inEducational Field", International Journal of Computer Science and Information Technologies, Vol.5(2), 2014, 2458-2461
5. A. Algarni, "Data Mining in Education", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 6, 2016
6. U. K. Sen, "A Brief Review Status of Educational Data Mining", international Journal of Advanced Research in Computer Science & Technology (IJARCST 2015), Vol. 3, Issue 1 (Jan. - Mar. 2015)
7. C. Romero, S. Ventura, "Educational Data Mining: A Review of the State-of-the-Art", IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews, VOL. XX, NO. X, 200X
8. L. C. Liñán, Á. A. J. Pérez, "Educational Data Mining and Learning Analytics: differences, similarities, and time evolution", RUSC. Universities and Knowledge Society Journal, Vol. 12, Núm. 3 (juliol 2015)
9. V. Thanuja, B. Venkateswarlu, G. S. G. N. Anjaneyulu, "Applications of Data Mining in Customer Relationship Management", J. Comp. & Math Sci. Vol.2 (3), 423-433 (2011)
10. Ms. T. K. Anusuya, P. Yasotha, "Estimation of Student Performance Using Artificial Intelligence with LMS Strategies", Journal of Information and Computational Science, Volume 10 Issue 3 – 2020
11. Suhirman, J. M. Zain, H. Chiroma, T. Herawan, "Data Mining for Education Decision Support: A Review", iJET – Volume 9, Issue 6, 2014
12. S. Angra, S. Ahuja, "Analysis of Student's Data using Rapid Miner", Journal on Today's Ideas –Tomorrow's Technologies, Vol. 4, No. 2, December 2016 pp. 109-117
13. M. H. A. Haroun, E. K. Hachem, C. Ziti, M. Bassiri, "How To Motivate Students Through mining Educational Data", International Journal of Engineering & Technology, 7 (4.32) (2018) 75-78
14. Yi Li, XiaoningZhai, "Review and Prospect of Modern Education using Big Data", Procedia Computer Science 129 (2018) 341–347
15. Y. D. Usman, "Educational Resources: An Integral Component for Effective School Administration in Nigeria", Research on Humanities and Social Sciences, ISSN (Paper) 2224-5766 ISSN (Online) 2225-0484 (Online) Vol.6, No.13, 2016
16. Jaya Srivastava, DrAbhay Kumar Srivastava, "Data Mining in Education Sector: A Review", Special Conference Issue: National Conference on Cloud Computing & Big Data
17. H. Deng, X. Wang, Z. Guo, A. Decker, X. Duan, C. Wang, G. A. Ambrose, K. Abbott, "PerformanceVis: Visual analytics of student performance data from an introductory chemistry course", Visual Informatics 3 (2019) 166–176
18. L. D. Hammond, L. Flook, C. C. Harvey, B. Barron, D. Osher, "Implications for educational practice of the science of learning and development", Applied Development Science, 2020, VOL. 24, NO. 2, 97–140

487

19. M. Quadrana, P. Cremonesi, D. Jannach, *"Sequence-Aware Recommender Systems"*, ACM Computing Surveys, Vol. 1, No. 1, Article 1, Publication date: February 2018.

20. E. Gomede, F. H. Gaffo, G. U. Briganó, R. M. d. Barros, L. d. S. Mendes, *"Application of Computational Intelligence to Improve Education in Smart Cities"*, Sensors 2018, 18, 267; doi:10.3390/s18010267

21. K. Chrysafiadi, M. Virvou, *"Student modeling approaches: A literature review for the last decade"*, Expert Systems with Applications, 40, 2013, 4715-4729

22. M. D. Zrakić, A. Marković, Z. Bogdanović, D. Barać, S. Krčo, *"Providing Adaptivity in Moodle LMS Courses"*, Educational Technology & Society, 15 (1), 326–338.

23. F. Martino, A. Spoto, *"Social Network Analysis: A brief theoretical review and further perspectives in the study of Information Technology"*, PsychNology Journal, 2006, Volume 4, Number 1, pp. 53 – 86

24. Tony A Slieman and Troy Camarata, "Case-Based Group Learning Using Concept Maps to Achieve Multiple Educational Objectives and Behavioral Outcomes", Journal of Medical Education and Curricular Development Volume 6: 1–7

25. H. Kaur, *"A Review of Applications of Data Mining in the Field of Education"*, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 4, April 2015

26. S. S. Chouhan, R. Khatri, *"Data Mining based Technique for Natural Event Prediction and Disaster Management"*, International Journal of Computer Applications (0975 – 8887) Volume 139 – No.14, April 2016

27. Data Set https://www.kaggle.com/spscientist/students-performance-in-exams .

AUTHOR PROFILE

Kamal Bunkar M.Tech(I.T.) from SOIT, R.G.P.V. Bhopal, B.E (C.S.) from Govt. Engineering Collage Ujjain. He is working as a Lecturer in Institute of Computer Science, Vikram University, Ujjain-INDIA since last 12 years. His research interest include Networking and Data Mining .

Dr. Sanjay Tanwani, Professor & Head, School of Computer Science, D.A. University, Indore Dean, Engineering Sciences from March 2010 - March 2012 Member, Executiver Council, D A University from Jan 2011 - March 2012.Worked as a programmer in Devi Ahilya University Computer center for four years (1986-1990), then worked as a Lecturer (1990-1993) & Reader (1993-2000) in Computer Science Department of D. A. University, Indore before joining Internet Software division group at Impetus in August 2000 taking 3 years extra ordinary leave from University. Apart from teaching and research, I have done consultancy & development work for package based software like 'Thermal Simulation of buildings' and contributed in the development of various softwares, like examination system, payroll system, and entrance examinations for the university.