

A Review on Educational Data Mining approaches with its uses and applications

Dilshad Kaur

Research Scholar

Guru Gobind Singh College of Modern Technology,

Kharar Punjab

dilshadladdi@yahoo.com

Rajan Sachdeva

Assistant Professor, CSE Department

Guru Gobind Singh College of Modern Technology,

Kharar Punjab

mp13rajan@gmail.com

Abstract— Data mining is the procedure of extracting important and meaningful values from the database. A data warehouse is a type of data storage where data is stored for the purpose of fetching data when needed in the near future. Data warehouses can store all kinds of data. In particular, the type of data depends on the type of industry being used. In most industries, all kinds of data are recorded, but in some companies only that information is stored. This information is useful and meaningful. The data stored in the warehouse are useful for the decision support system. Based on historical data, decisions about future schemes can be made easily or effectively.

This paper organizes a review to the education data mining and the research work that has been done in this domain during last year's.

Keywords—Data Mining, Education Data Mining, Pattern matching.

I. INTRODUCTION

Education data mining is done for evaluating the performance of students of an institute. EDM plays most important role for high quality universities and institutes to set their criteria or level among other universities [1]. The record of the students of a university sets the academic achievement level of its own. Education data mining is also known as student data mining. The term education data mining or student data mining has various meanings according to previous work or literature survey [2]. Student data mining can be performed by measuring the performance of the students and in order to measure the performance of a student there are various ways such as to calculate the learning assessment of the student and calculate the co-curriculum [3]. Most of the previous work focuses on the performance evaluation of graduate students and their success rate. For example in Malaysia, most of the universities and colleges use final grades of the students for the process of student data mining [4]. These final grades of the student is the combination of his performance in various activities like marks obtained in exams, assessment scores, scoring in various co-curriculum activities and Performa of the syllabi or course [5]. The estimation of student performance is an important task to be performed by the universities. The idea behind the evaluation of the student's performance is to keep the record of a student's performance and efficiency of learning process. On the basis of the data collected by analyzing the student's performance various strategies and plans or policies can be created for student's welfare [6]. For evaluating the performance of the students, universities have to maintain their records in various activities. Then these records are used to analyze the student's performance. There are various techniques and methods which are used to analyze the recorded data. Data mining is

one of the techniques [7]. Data mining process in field of education is known as education data mining. Data mining technique is widely used in the field of education for evaluating the student's performance [8]. EDM or student data mining is a process which is used to haul out the useful and meaningful information from large datasets or database of the student's information. Then the extracted data or information is used for predicting the performance of the students [9]. EDM is an application which is an aid to students' data mining process. EDM is an application of data mining. EDM is mainly meant for answer to the problem rises in the education system [10]. EDM tool is used along with the application of data mining which is used for the purpose of efficient of student learning and business products. EDM is an effective tool in order to identify the student data for motivating in studies and other activities and leads to the increased use of technology in various universities and institutes [11].

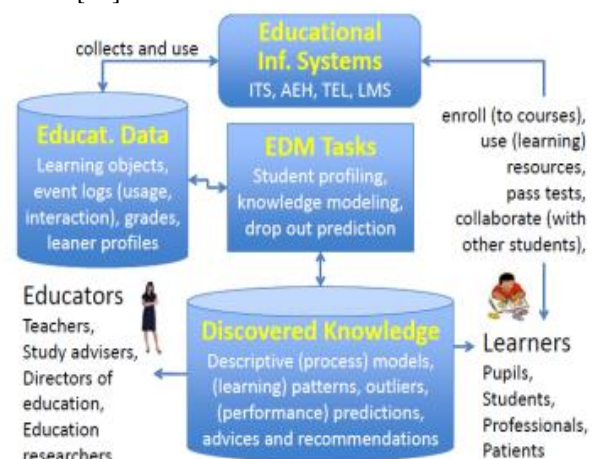


Figure 1 EDM in a nutshell [12]

EDM is the process in which data is stored and collected on the basis of various parameters and various EDM process are implemented on this data for extracting the meaningful information. Then extracted data is stored in a warehouse and on the basis of this data various decisions are taken by the organizations [13].

For an educational institute students are considered as the main assets. The performance of the students of universities leaves great impact on the level of the university and even country [14]. The best performance of the students can lead to the good graduates and post graduates and even such a youth plays an important role for bringing the country towards the success or advancements. The performance of the students plays an important role in setting up the level of the university also [15]. Hence the performance of the students leaves an impact on the social and economic growth of the country. Student's performance also affects the labor market. The criteria of academic achievements are considered at the time of recruitment. Hence the performance of a student plays an important role for making a bright career. Academic performance of student is measured by the CGPA i.e. Cumulative Grade Point Average [16].

II. RELATED WORK

Surjeet Kumar Yadav, [1] In this work author focused on various aspects such as Knowledge Discovery in data mining and various techniques followed for the purpose of data mining in education field. KDD along with the process of data mining was a process in which the meaningful data extraction was performed for the purpose of decision making and future forecasting. The EDM does not use KDD process for feature extraction. Various techniques were applied in the process of classification. Decision tree was one of these classifiers that follow the rule of divide and conquers. In this work there was a brief description of decision tree in education data mining. Decision tree classifiers were applied to the historical data of the student database for predicting the performance of the students. After analyzing the result, the students with more weak points can get more attention of the teachers with proper and right counseling.

Surjeet Kumar Yadav, [2] the amount of data stored in database had been increased day by day. All kind of data is stored in database these days, but it is not necessary that all of the data should be meaningful for each and every process. The database may also contain irrelevant data also. So in order to get the meaningful data, the process of data mining was discovered. Educational data mining is a field of data mining in which the database of the student's information was mined for getting the relevant information for the process of performance prediction. EDM was done for figure out the weak students and help them by providing right counseling to

them. Various techniques can be applied for prediction such as decision tree etc.

Brijesh Kumar Baradwaj, [3] the objective of providing higher education can only be achieved on the basis of performance evaluation of the students. The performance of the student can be evaluated on the basis of various parameters such as CGPA, marks in exams, marks in various co-curriculum activities etc. The informative data was hidden behind this stored data. In order to observe the informative data from student database, various EDM techniques were developed. Most popular techniques for data mining was decision tree because it is helpful in decision making systems. In this work the performance of the student was calculated at the end of the semester by applying classification and decision tree system for analyzing and predicting student's data.

SayaliRajesh Suyal, [4] mining is the process of extracting something meaningful and essential from huge storage or heap. For example mining of gold from sand or rock is known as the process of gold mining instead of rock or sand mining. Hence the process of extracting the meaningful data was known as data mining and also known as knowledge mining. The extracted data or knowledge can be applied for various purposes such as decision making, analysis purpose etc. The education system is such a system which directly affects the future of the country or nation. Hence in order to get bright future the process of education data mining was done so that on the basis of extracted information the help should be provided to the needy students. The data mining contains another process such as classification, regression, association etc. Data mining can be used for various purposes such as extracting the relevant information from large amount of stored data and then making a meaningful use of extracted data such as decision making on the basis of historical data, to improve the system etc.

Paulo Cortez, [5] this paper focused on the education level of Portuguese. This country lies at the end of the tail of Europe because the failure rate of this country is quite high. In this work the focus was on achievements of the students at secondary level with respect to the Business Studies and data mining. In this, the simulation was performed on the basis of real world data. The data was collected from primary sources. Then on the basis of results various parameters were concluded such as the performance of the students was prejudiced by the historical data evaluation. The work developed the most essential student data mining tool and leads to an improvement in the performance of the students and better utilization in the resources of the universities etc.

Amirah Mohamed ShahirI, [6] Education data mining was not an easy task because of the huge and big database availability. The paper had represented a study of Malaysia education system. This work showed that the Malaysian education system had to face a problem which was due to lack of techniques and research in that field. Due to less research

and unknown facts that can affect the system there was non-availability of the mining techniques. Hence this paper focused on the review of the existing work so that the system can be improved by removing the weaknesses of the traditional systems. The main goal of this work was to review or study already done work in this field. This paper also provided the answer to the question that in which way the prediction algorithms can be applied to achieve efficient results. The student's achievement and success level can be enhanced by performing data mining on the historical data.

Alaa Mustafa, [7] Educational data mining was related to the mining process of educational data. Educational data mined for the purpose of evaluating the student's performance. This work focused on the process of analyzing the learning behavior of the students with EDM. In this a case study was reviewed. The data of the case study was collected from the course database. Then pre-processing techniques were applied to the data to make it more refine and usable. Then various data mining techniques were applied for the purpose of classification and clustering and detecting the outlier values. Then the behavior of the student was analyzed and described on the basis of above calculated parameters.

Muluken Alemu Yehuala, [8] in this paper the center of attention was how the data mining can be useful for analyzing the performance of a student on the basis of historical data. Various technologies were available for implement the process of data mining. In this work, CRISP-DM was used for data mining. In order to extract and matching the patterns from the database, Classification and Prediction were performed in this. In this work, the data was gathered from MS-EXCEL and then pre-processing was applied on this data for model building. The models were created for graduated students. In this the process of data mining was implemented with the help of WEKA 3.7 tool or application. On the basis of results and output various decisions were taken by the university authorities to enhance the learning skills of its students.

Jennifer Sabourin, [9] in this work author raised the concept of security in data mining. Here security referred to the security of the student related data from non-interesting organizations and un-authentic personnel. The student's data or academic data of the students was used by various business organizations for the purpose of surveys. Then it enhances the risk of security of the data. In this paper, author represents three criteria's on which the security was dependent. These main criteria were policy, corporate-socio responsibility and public interest or opinion. This paper provides the solution to the problem of security related to these fields.

George Siemens, [10] the process of analysis had been posed a keen interest in field of education for enhancing the university's policies and standards. So in order to fulfill this need, the concept of Education data mining and learning analytics and knowledge is introduced. In this paper the main focus was on these two concepts and its related fields such as

communication between various communities and collaboration of data for the process of data mining.

III. CONCLUSION

Transforming the data set into understandable and meaningful structure after extracting useful information from data is known as data mining. The prime purpose of data mining is to obtain mine cases, affiliations, changes, peculiarities, and quantifiable vital structures from complex data set. K-mean clustering is one data mining technique that is being mostly used for extracting data from a data set because of its simplicity and higher speed.

REFERENCES

- [1] Surjeet Kumar Yadav, "Data Mining Applications: A comparative Study for Predicting Student's performance", *ijitce*, vol 1(12), Pp 13-20, 2012.
- [2] Surjeet Kumar Yadav, "Data Mining: A Prediction for Performance Improvement of Engineering Students using Classification", *WCSIT*, vol 2(2), Pp 51-56, 2012.
- [3] Brijesh Kumar Baradwaj, "Mining Educational Data to Analyze Students' Performance", *ijacsa*, vol 2(6), Pp 63-70, 2011.
- [4] Sayali Rajesh Suyal, "Quality Improvisation Of Student Performance Using Data Mining Techniques", *Vol 4(4)*, Pp 1-4, 2014.
- [5] Paulo Cortez, "Using Data Mining To Predict Secondary School Student Performance", *research gate*, 2018.
- [6] Amirah Mohamed Shahirl, "A Review on Predicting Student's Performance Using Data Mining Techniques", *ELSEVIER*, vol 72, Pp 414-422, 2015.
- [7] Alaa Mustafa, "Mining Student Data to analyze Learning behavior", *Research gate*, 2009.
- [8] Muluken Alemu Yehuala, "Application Of Data Mining Techniques For Student Success And Failure Prediction (The Case Of Debre_Markos University)", *vol 4(4)*, Pp 91-95, 2015.
- [9] Jennifer Sabourin, "Student Privacy and Educational Data Mining: Perspectives from Industry", *IEEE*, Pp 164-170, 2015.
- [10] George Siemens, "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration", Pp 252-254, 2012.
- [11] Cristóbal Romero, "Educational Data Mining: A Review of the State-of-the-Art", *IEEE*, vol 20, Pp 1-19, 2010.
- [12] Luis Talavera, "Mining Student Data To Characterize Similar Behavior Groups In Unstructured Collaboration Spaces", *CSCL*, Pp 17-23, 2004.
- [13] U . K. Pandey, and S. Pal, "Data Mining: A prediction of performer or underperformer using classification", (*IJCSIT*) *Vol. 2(2)*, Pp.686-690, 2011.
- [14] Pimpa Cheewaprabokit, "Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program", *IMECS*, vol 1, Pp 1-5, 2013.

-
- [15] M. R. Beikzadeh and N. Delavari, "A New Analysis Model for Data Mining Processes in Higher Educational Systems,"citesserx, Pp 5-9, 2004.
 - [16] Ajinkya Kunjir," Recommendation of Data Mining Technique in Higher Education", IJCER, vol 5(3), Pp 29-35, 2015.
 - [17] Jayashree M Kudari," Survey on the Factors Influences the Students" Academic Performance", IJERMT, vol 5(6), Pp 30-37, 2016.
 - [18] RYAN S.J.D. BAKER," The State of Educational Data Mining in 2009: A Review and Future Visions", Journal of education data mining, vol 1(1), Pp 1-14, 2009.
 - [19] Dr. Varun Kumar," An Empirical Study of the Applications of Data Mining Techniques in Higher Education", IJACSA, vol 2(3), Pp 80-85, 2011.
 - [20] Neelamadhab Padhy, "The Survey of Data Mining Applications And Feature Scope" International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3,Pp 43-58, 2012.