

The Study of Data Mining Aspects in Web-Based Educational System

Ritika^{1*} Dr. Kalpana Midha²

¹ Research Scholar, OPJS University, Churu, Rajasthan

² Associate Professor, Department of Computer Science, OPJS University, Churu, Rajasthan

Abstract – Data mining (DM) is a tool for finding secret trends and for revealing associations between dataset attributes. Data mining is capable of collecting tacit, previously unknown and potentially useful knowledge from data. The techniques become popular from past years in many application areas such as weather forecasting, engineering, web learning, marketing, medical, education, financial, sport etc. This indicates that the potential of data mining tools to provide novel solutions for decision-makers to solve problems exists in particular areas. The data examined and the information gathered from the field of education utilizing DM techniques is referred to as educational data mining. EDM is able to address the rising need for a comprehensive assessment of the new area of EDM. The EDM is a very new and very small academic area. The main goal of data mining is the detection of hidden connections between objects in a data collection.

Keywords – Data Mining, Educational Data Mining, Knowledge Discovery in Databases, Data Mining Techniques

-----X-----

INTRODUCTION

Data mining (DM) is a tool for finding secret trends and for revealing associations between dataset attributes. Data mining is capable of collecting tacit, previously unknown and potentially useful knowledge from data. Data mining offers opportunities by data analysis. Data mining methods have the ability to uncover concrete trends and laws. The techniques become popular from past years in many application areas such as weather forecasting, engineering, web learning, marketing, medical, education, financial, sport etc. Discovering Information from Giant Databases is regarded as Data Mining. This finds secret information from a variety of data sources in a variety of fields. A variety of strategies can be used in different fields of data mining, together with weather forecasting, oil analysis, industry, scientific, marketing and EDM, etc.[1]. In order to extract and interpret information from educational data sources, a sub-area of data mining has also been developed called EDM. Data mining, analytics & machine learning are used on EDM information to extract information from educational environments. It is increasingly in demand and draws further interest because of the rise in educational data in web based learning and even the development in traditional education. Alarmed by emerging strategies for identifying the distinctive forms of data found in scholastic contexts, it aims to derive useful information in order to develop and understand

learning processes from vast amounts of raw data[2]. Probing standard database documents can provide answers to problems such as "tracking students who have failed the tests," while EDM provides answers to additional problems such as "predicting students who are more likely to pass." Moving to educational institutions, the development of user templates so that student behaviors or results can be forecast well in advance is a key area for EDM implementations. As a result, several academics have begun to explore different data mining techniques to help educators or professors analyze and develop their respective course organizations [3].

DATA MINING

Data mining refers to extracting or mining knowledge from large amounts of data. The term is actually a misnomer. Thus, data mining should have been more appropriately named as knowledge mining which emphasis on mining from large amounts of data. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.[1] Data mining is an interdisciplinary subfield of computer science and statistics with an overall goal to extract information (with intelligent methods) from a data set and transform the information into a comprehensible structure for further use. Data mining is the analysis step of the "knowledge discovery in databases" process or KDD. Aside from the raw analysis step, it also involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating.

The word "data mining" is a misnomer, considering that the goal is to derive trends and knowledge from a ton of data, not merely to collect (mining) data. This is also a common term, because much of the time is expended on some form of large-scale data or data processing (collection, retrieval, storage, study, because statistics) much as any usage of the socially nurturing network of machine preference, like artificial intelligence (e.g., AI) and market awareness. The main challenge of data mining is to self-load or program vast quantities of data to extract previously unknown, interesting trends, e.g. data compilation (bunch analysis), abnormal documents (anomaly detection), and circumstances (association law mining, sequential sequence mining). Usually, this involves the usage of storage methods, such as relational archives. These trends may then be used as a sort of description of the knowledge details which could be used in further research or, for example, in AI which prescient research. For eg, the data mining stage which discern a variety of data sets, which will then be able to be used to achieve increasingly reliable goals by selecting an emotionally supporting network. Neither data analysis, data preparation, nor transmission and reporting findings are part of the data mining phase, but have a position in the general KDD method as additional progresses.

EDUCATIONAL DATA MINING

Educational Data Mining (EDM) is ready to meet the increasing need for a thorough evaluation of the emerging field of EDM. EDM focuses on compilation, chronicling and evaluation of under-study research and appraisal outcomes. EDM is a continuous, confined scholastic field. Main publications on the EDM have emerged in the last two years, although maybe under thirty individuals on the world, who have presented themselves as a part of the EDM. As every new sector, EDM has become out of the current controls and is expanding to reach new ones. Massive quantities of EDM-forming participants come from the Intelligent Tutoring Network (ITS) citizens community, where exposure to large volumes of educational data has been prepared to

render EDM a rational move forward. EDM work imparts a range of common characteristics to the People's Artificial Intelligence in Education (AIED) community. The analysis was done in EDM research, which established approaches of psychometrics and educational statistics. EDM is ready to disrupt or, in any case, to strengthen and expand the observable strategies used in education by bringing to bear the implications of several years of DM and AI work. In the long term, considering the theoretical basis of the majority of EDM scholars, it is fully anticipated that evidence on understudy learning in computer science will be uncovered. It's not shocking, therefore, to discover any scope between the EDM and Computer Science Education (CSE) sectors.

KNOWLEDGE DISCOVERY IN DATABASES (KDD)

Some people treat data mining same as Knowledge discovery while some people view data mining essential step in process of knowledge discovery. The data mining method utilizes AI, measurements, and interpretation methods to locate and display knowledge in a system that is easily understandable. "Information" in the KDD relates to the release of trends isolated from the data being analyzed. A pattern is an articulation that portrays the realities of a subset of data. In this sense, the distinction between KDD and data collection is that KDD deals with the general methodology for extracting information from data when data mining associates with the use of algorithms to dispose of data trends without the basic measures of the KDD approach. (1) Nevertheless, since data mining is an immediate and integral part of the KDD method, most researchers are using the two-pronged method. The incremental interpretation of the KDD method is shown in Figure 1.1.

- To include an overview of the application area, the system's goals And its clients, and the essential earlier foundations and earlier knowledge.
- Choosing a data set or concentrating on a subgroup of factors or data tests That disclosure is to be carried out
- Pre-processing & data cleaning, noise reduction, collection of essential information for modeling, set of techniques for managing missing data fields, time succession accounts and modifications
- Data reduction & projection, seeking appropriate highlights to talk to data, Using dimensional reduction or conversion strategies to minimize quantity

Variables to define symmetric data profiles

- Selection of data mining tasks based on the KDD objective: clustering, grouping, regression, etc.
- Selecting methods and measurements to be used to scan for data trends
- Mining the information: scanning for pattern of intrigue
- Estimating or deciphering the mined patterns, by potential come back to such preceding advances
- Use of this information to advance overall system and address potential conflicts to recent convictions or separate awareness. These are the means by which each KDD & data mining undertaking progresses.

The following diagram shows the process of knowledge discovery process:

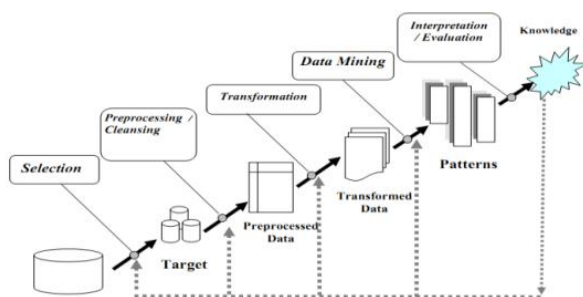


Figure 1.1. Steps of KDD process

DATA MINING TECHNIQUES

Data mining is a computational data processing technique that is widely employed in many fields for the purpose of extracting valuable knowledge from results. Data mining techniques were used to create a model for generating new content for knowledge. Several big data mining strategies have been developed and implemented, including clustering, sorting, clustering, regression, linear trends and decision tree. The following is a description of the main approaches utilized in this field. There is a variety of data mining methods to derive valuable knowledge from the data collection. Some commonly employed data mining techniques for the key categories of data mining problems are:

- Cluster analysis: Cluster analysis divides data into groups or divisions that might be significant or useful. If functional clusters are the goal, the resulting clusters should catch the "actual" data structure. The clustering of similar data forms is called clustering. The cluster analysis integrates results on the basis of the knowledge received. The goal is to insure the objects in a grouping are similar

to each other and distinct from artifacts in other categories. [7].

- Visualization technique: today's need is to handle a large volume of details, and computer programs are helping us do this. Virtual data analytics can help to deal with this influx of information by integrating people into the data analysis system. Visual data mining allows the user to obtain insight into the results, to render massive data similarities visible between various views and to communicate directly with the data[8]. Data visualization is a process by which quantitative data is converted into meaningful pictures.

- Association (correlation and causality): the relationship-trend analysis is a central field of data mining. The laws of association shall decide the interrelationship between the various data objects in the transaction records. The Mining Association Rules consider an fascinating connection between broad repositories. The association rule for mining shall lay down all rules with confidence and trust outside the limits laid down by the client. Thus, the least support and trust requirement has been based on the rules of the mining association at both a basic definition and a multiple stage. The key problem in data mining is the detection of latent correlations between the different attributes of the system. Recent developments in data processing and retrieval technologies have rendered it possible to capture vast volumes of data every day in a number of fields. The partnership pattern analysis is a big field of data mining focused on these findings. Association rules define inter-relationships between specific data items in transaction data.

- Classification: Classification incorporates a number of pre-classified instances to create a form that is used to classify records. Applications for detection of hazards are especially adapted to this form of study. The system of classification includes computational tools, including decision tree, linear learning, neural networks and statistics. The data analysis approach requires understanding and sorting. Training data is validated by a classification algorithm in the learning process. Classification test data are used to approximate the consistency of classification legislation. Where the standard is sufficient, the recommendations can be expanded to incorporate modern data tuples.

- Neural networks: The Neural Network is an extremely new technology that has been generated and is growing fast. There is a weight present in the range of associated input / output modules. Throughout the learning process, weight adjustment stimulates the learning network to evaluate the right class labels of the input tuples. Neural networks have a phenomenal capacity to derive significance from unclear or vague data in order to recognise associations and spot patterns. These became too advanced to be used either by humans or by other machines. They are well designed for continuous quality inputs and outputs. For eg, whether temperature, wind speed, hand-written character reorganization, for teaching a machine to pronounce English text. Neural networks are best adapted for data flows or trends and for modelling or forecasting purposes.

LITERATURE REVIEW

Surjeet Kumar Yadav et al. (2012) 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 sequence to get meaningful data, the process of DM 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.

Sayali Rajesh Suyal et al. (2014) 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.

Bipin Bihari Jayasingh et al. (2016) starts an example study that is taken for a specific foundation, in the specific condition, for the specific group and specific arrangement of understudies. The example data are gathered from a study hall by conveying the poll endeavored by two distinct clumps of understudy having questions relating to Inquiry based and deductive learning. The framework is created and tried twice subsequent to training the substance utilizing inductive strategy and actualized utilizing trait importance, discriminant rules of class separation mining. The outcomes are imagined through bar outlines and shows that the two bunches of students of various years have diverse learning characteristics.

Keno C. Piad, Menchita Dumlao, Melvin A. Ballera, Shaneth C. Ambat et al. (2016) predicts its employability graduates utilizing nine factors. To start with, various arrangement calculations in data mining were tried making strategic relapse with precision of 78.4 is actualized. In view of calculated relapse examination, three scholastic factors legitimately influence; IT_Core, IT_Professional and Gender recognized as huge indicators for employability. The data were gathered dependent on the multi-year profiles of 515 understudies arbitrarily chose at the position office tracer study.

S. M. Merchán et al. (2016) Exhibits and discusses the experience of implementing such data mining techniques and methods to 932 Systems Engineering under-study data from El Bosque University in Bogotá, Colombia; an endeavor attempted to establish a prescient model for under-study academic presentation. As an iterative disclosure and learning process, the experience is broke down as indicated by the outcomes got in every one of the procedure's cycles. Each acquired outcome is assessed in regards to the outcomes that are normal, the data's information and yield portrayal, what hypothesis directs and the congruity of the model got as far as prediction precision. Said congruity is assessed considering specific insights regarding the populace contemplated, and the particular needs showed by the institution.

Paulo Cortez et al. (2018) 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.

Mustafa Agaoglu et al. (2015) in their work designed a technique, which mainly concentrated on the performances of the students. According to this insight, students evaluated the staff's performance by answering the questionnaire as to how they teach them, the efficiency of using teaching aids and the communication skill and so on. Right now, based characterization procedure has used to arrange the exhibition of every staff that took care of the concerned courses. The outcomes demonstrated that C4.5 classifier holds more exactness than contrasted with different procedures utilized in their work. The result of this work uncovered that a large portion of the inquiries in survey for assessing the course were appear to be improper.

Hemaid and El-Halees et al. (2015) did a comparative report to look at the components related with the evaluation of educators' exhibitions. In this investigation, data gathered for instructors from the Ministry of Education and Higher Education in Gaza City. They proposed a model to assess their presentation with strategies of DM like affiliation, arrangement rules (Decision Tree, Rule Induction, k-Nearest Neighbor (KNN), Naïve Bayesian (Kernel)) to decide ways that can assist them with bettering serve the educational procedure and improve their exhibitions ideally and in this manner consider it the exhibitions of instructors in the study hall. In every undertaking, they introduced the separated information and portrayed its significance in the instructor's exhibition domain.

MulukenAlemuYehuala et al. (2015) 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

Brijesh Kumar Baradwaj et al. (2011) 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 cocurriculum 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.

Amirah Mohamed Shahirl et al. (2008) 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 el-Halees et al. (2009) Data Mining can be used in the field of education to enhance our ability of the learning process with a view to characterizing, extricating and examining variables related to the under-study learning process as described. Mining in an educational situation is referred to as educational data mining.

CONCLUSION

Data mining offers opportunities by data analysis. Data mining methods have the ability to uncover concrete trends and laws. The techniques become popular from past years in many application areas such as weather forecasting, engineering, web learning, marketing, medical, education, financial, sport etc.. Data is broken down and information obtained from the educational sector utilizing DM techniques was referred to as educational data mining (EDM). Educational Data Mining (EDM) is prepared to satisfy the increasing need for a comprehensive evaluation of the new area of EDM. Neither data collection, data preparation, nor the interpretation and reporting of results are part of the data mining process, yet have a place with the general KDD process as additional advances.

REFERENCES

Surjeet Kumar Yadav (2012). "Data Mining Applications: A comparative Study for Predicting Student's performance", *ijitce*, vol 1(12), pp. 13-20.

Sayali Rajesh Suyal (2014). "Quality Improvisation Of Student Performance Using Data Mining Techniques", Vol 4(4), pp. 1-4.

Bipin Bihari Jayasingh (2016). *A data mining approach to inquiry based inductive learning practice in engineering education*, in IEEE 6th International Conference on Advanced Computing, pp. 845-850.

Keno C. Piad, Menchita Dumlaio, Melvin A. Ballera, Shaneth C. Ambat (2016). "Predicting IT Employability Using Data Mining Techniques," in third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC).

Hemaid and El-Halees (2015). "Improving teacher performance using data mining", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, No. 2, pp. 407-412.

Muluken Alemu Yehuala (2015). "Application Of Data Mining Techniques For Student Success And Failure Prediction (The Case Of Debre_Markos University)", vol 4(4), pp. 91-95.

S. M. Merchán (2016). "Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic," IEEE Latin America Transactions, vol. 14, no. 6, June 2016.

Paulo Cortez (2018). "Using Data Mining To Predict Secondary School Student Performance", research gate.

Mustafa Agaoglu E. Serra Yurtkoru, and Asli Kucukaslan Ekmekci (2015). 'The effect of ERP implementation CSFs on business performance: an empirical study on users' perception', Procedia-Social and Behavioral Sciences, Vol. 210, pp. 35-42.

Brijesh Kumar Baradwaj (2011). "Mining Educational Data to Analyze Students' Performance", ijacsa, vol 2(6), pp. 63-70.

Amirah Mohamed Shahirl (2015). "A Review on Predicting Student's Performance Using Data Mining Techniques ", ELSEVIER, Vol. 72, pp. 414-422.

Alaa el-Halees (2009). "Mining student's data to analyze e-Learning behavior: A Case Study".

Corresponding Author

Ritika*

Research Scholar, OPJS University, Churu, Rajasthan