

## **Data Mining Technique for Improved Management Information System**

**Dr Sanjeev Kumar Varshney**

**Associate Professor, Faculty of Commerce**

**Veerangana Avantibai Govt. Degree College, Atrauli , Aligarh**

**Abstract** – This article to design a model and use the data mining technique applied to Learning Education System. They allow a all technical institutions to prescribed format in technical education system. This concepts to consider various ways in which the concepts of technical educational data mining are applied to the information extracted from Technical Education System (TES). The data from this system can be evaluated to convert the information collected into useful meaningful information to provide a technical education tailored to the need of each student. This approach seek to improve the effectiveness and efficiency of education by recognizing patterns in student performance.

**Keywords:**Data mining, e-learning, Model.

### **1. Introduction**

Data mining refers to extracting or “mining” knowledge from large amountsof data. The term is actually a misnomer. Remember that the mining of gold from rockor sand is referred to as gold mining rather than rock or sand mining. Thus, data miningshould have been more appropriately named “knowledge mining from data,” which isunfortunately somewhat long. “Knowledge mining,” a shorter term may not reflect theemphasis on mining from large amounts of data. Nevertheless, mining is a vivid termcharacterizing the process that finds a small set of precious nuggets from a great deal ofraw material.

Data mining is as a step in the process of knowledge discovery.Knowledge discovery as a processis depicted in and consists of an iterative sequence of the following steps:

1. Data cleaning (to remove noise and inconsistent data)
2. Data integration (where multiple data sources may be combined)
3. Data selection (where data relevant to the analysis task are retrieved fromthe database)
4. Data transformation (where data are transformed or consolidated into forms appropriatefor mining by performing summary or aggregation operations, for instance)
5. Data mining (an essential process where intelligent methods are applied in order toextract data patterns)
6. Pattern evaluation (to identify the truly interesting patterns representing knowledgebased on some interestingness measures)

7. Knowledge presentation (where visualization and knowledge representation techniques are used to present the mined knowledge to the user)

### **Data Mining Functionalities:**

Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks. In general, data mining tasks can be classified into two categories: descriptive and predictive. Descriptive mining tasks characterize the general properties of the data in the database. Predictive mining tasks perform inference on the current data in order to make predictions.

- Concept/Class Description: Characterization and Discrimination
- Mining Frequent Patterns, Associations, and Correlations
- Classification and Prediction
- Cluster Analysis
- Outlier Analysis
- Evolution Analysis

### **Classification and Prediction:**

The following pre-processing steps may be applied to the data to help improve the accuracy, efficiency, and scalability of the classification or prediction process.

**Data cleaning:** This refers to the pre-processing of data in order to remove or reduce noise (by applying smoothing techniques) and the treatment of *missing values* (e.g., by replacing a missing value with the most commonly occurring value for that attribute, or with the most probable value based on statistics).

**Relevance analysis:** Many of the attributes in the data may be *redundant*. Correlation analysis can be used to identify whether any two given attributes are statistically related.

**Data transformation and reduction:** The data may be transformed by normalization, particularly when neural networks or methods involving distance measurements are used in the learning step. Normalization involves scaling all values for a given attribute. The data can also be transformed by *generalizing* it to higher-level concepts. Concept hierarchies may be used for this purpose. This is particularly useful for continuous valued attributes.

Data can also be reduced by applying many other methods, ranging from wavelet transformation and principle components analysis to discretization techniques, such as binning, histogram analysis, and clustering.

### **Cluster Analysis:**

The process of grouping a set of physical or abstract objects into classes of *similar* objects is called clustering. A cluster is a collection of data objects that are *similar* to one another within the same cluster and are *dissimilar* to the objects in other clusters.

### **Types of Data in Cluster Analysis:**

- Interval Scaled Variables
- Binary Variables
- Categorical, Ordinal, and Ratio-Scaled Variables
- Variables of Mixed Types
- Vector Objects

## **2. Literature Review**

A considerable research work has been done in the field of data mining. In the following the researcher is going to summarize various studies which have been made so far in the literature.

Kumar, Varun, Chadha, Anupama.,<sup>2011</sup>, have addressed that the applications of data mining in educational institution to extract useful information from the huge data sets and providing analytical tool to view and use this information for decision making processes by taking real life examples.

Panda, Subhalaxmi, Pattanaik, P. A., Swarnkar, Tripti.,<sup>2017</sup>, have proposed the approach of Random Forest to predict the career decision for the 12th passing out students. The use of Random Forest has helped the students to take a correct appropriate decision as per their interest and skills and acts a career counselor toolbox.

Luo<sup>1,2</sup>, Qi.,<sup>2008</sup>, has defined the relation between Knowledge, Data Mining, and Knowledge Discovery in Database (KDD) process. They have also proposed Data mining theory, Data mining tasks, Data Mining technology and Data Mining challenges.

Villegas-Ch<sup>1</sup>, W., Mora<sup>2</sup>, S. Luján.,<sup>2017</sup>, have discussed the approach to improve the effectiveness and efficiency of education by recognizing patterns in student performance. In this study they have presented an analysis of the data mining techniques that fit LMS, specifically in terms of a case study applied to the e-learning platform Moodle.

AsifRaheela, MerceronAgathe, Ali Syed Abbas, HaiderNajmiGhani., <sup>2017</sup>, have focused on a small number of courses that are indicators of particularly good or poor performance, it is possible to provide timely warning and support to low achieving students, and advice and opportunities to high performing students.

Baker Ryan S.J.d., Baker R.S.J.d., Baker E. (Eds.), <sup>2010</sup>, have studied that the educational data mining methods often differ from methods from the broader data mining literature, in explicitly exploiting the multiple levels of meaningful hierarchy in educational data. Methods from the psychometrics literature are often integrated with methods from the machine learning and data mining literatures to achieve the goal.

Abad Martínez Fernando, Alicia A. ChaparroCasoLópez, <sup>2016</sup>, have discussed the similarities between the results obtained and those shown in literature, highlighting how simple decision trees allow a greater explanation and interpretation than other models and techniques.

BresfeleanVasile Paul, BresfeleanMihaela, GhisoiuhaveNicolae, <sup>2008</sup>, they have illustrated the experiments in the educational area, based on classification learning and data clustering techniques, made in order to draw up the students' profile for exam failure/success.

Lu Hongjun, Setiono Rudy, Liu Huan, <sup>2017</sup>, have developed algorithms, rules which are similar to, or more concise than those generated by the symbolic methods can be extracted from the neural networks. The data mining process using neural networks with the emphasis on rule extraction is described.

Behrouz Minaei-Bidgoli I , Deborah A. Kashy', GerdKortemeyer', William F. Punch, <sup>2003</sup>, In this study they have presented an approach to classifying students in order to predict their final grade based on features extracted from logged data in an education web-based system. They designed, implemented and evaluated a series of pattern classifiers and compared their performance on an online course dataset.

Singh BhullarManpreet, KaurAmritpal, <sup>2012</sup>, have discussed about data mining and their different phases, advantages and also classified data using Weka data mining tool which helps to understand the data. They have used J48 algorithm to predict the result of the student.

PalSaurabh, ChaurasiaVikas, <sup>2017</sup>, The present research intends to the student achievement in Higher Education using Data Mining techniques. This real-world data was collected by using performance reports and questionnaires. Four Decision Tree algorithms (BFTree,

J48, RepTree and Simple Cart) are applied in this work. The results showed that BFTree algorithm mostly proper to classify and predict student's whose performance is excellent and who's poor during studying the subjects.

Al-Radaideh A. Qasem, M. Emad Al-Shawakfa and I. Mustafa Al-Najjar, <sup>2006</sup>, They have presented an attempt to used the data mining processes, particularly classification, to help in enhancing the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses. The CRISP framework for data mining is used to mining student related academic data.

DelavarilNaeimeh, AyatollahzadehShirazi Mohammad Reza and BeikzadehMohammad Reza, <sup>2004</sup>, in this studythey have presented and justified the capabilities of data mining technology in the context of higher educational system hy'proposing a new model for improving the efficiency and effectiveness of the higher educational process. Higher educational institutes can use this model to identify which part of their processes can be improved by data mining technology and how they can achieve the goal.

AherSunita B, LOBO L.M.R.J., <sup>2011</sup>, surveyed an application of data mining in education system & also presented result analysis using WEKA tool.

BaradwajBrijesh Kumar, Pal Saurabh, <sup>2011</sup>, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here. By this task we extract knowledge that describes students' performance in end semester examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling.

YadavSurjeet Kumar, BharadwajBrijesh Kumar, Pal Saurabh, <sup>2011</sup>, in this study they have discussed use of decision trees in educational data mining. Decision tree algorithms are applied on students' past performance data to generate the model and this model can be used to predict the students' performance.

HALEES EL-ALAA, <sup>2008</sup>,has used educational data mining to analyze learning behavior.In this study, he has collected students' data from DataBase course. After preprocessing the data, he applied data mining techniques to discover association, classification, clustering and outlier detection rules

PandeyUmesh Kumar, Pal S., <sup>2011</sup>, they have used data mining techniques name Byes classification method. Institutions can find those students who are consistently perform

well. This study will help to institution reduce the drop put ratio to a significant level and improve the performance level of the institution.

FDEZ J. ALCALÁ, FERNÁNDEZ A., LUENGO J., DERRAC J., <sup>2010</sup>, have presented three new aspects of KEEL: KEEL (Knowledge Extraction based on Evolutionary Learning) dataset, a data set repository which includes the data set partitions in the KEEL format and shows some results of algorithms in data sets; some guidelines for including new algorithms in KEEL, helping the researchers to make their methods easily accessible to other authors and to compared the results of many approaches already included within the KEEL software.

R.B. Bhise, S.S. Thorat, A.K. Supekar, <sup>2013</sup>, The objective of this study is to introduce educational data mining by describing step by step process using technique of K-means (Clustering Methods). This study will help the teacher to reduced drop-out ratio to a significant level and improve the performance of students.

Gnardellis.T.,Boutsinas .B., <sup>2011</sup>, They have investigated the application of data mining techniques within the education framework. Data mining techniques are used to automatically extract knowledge, in the form of relationships and patterns, from large databases.

Hsia Tai-Chang, Shie An-Jin, Chen Li-Chen, <sup>2008</sup>, they have used data mining techniques to uncover the preferences and future choices of continuing education students of an extension education center at a university in Taiwan.

Pang Bo and Lee Lillian, <sup>2004</sup>, they have proposed a novel machine-learning method that applies text-categorization techniques to just the subjective portions of the document. Extracting these portions can be implemented using efficient techniques for finding minimum cuts in graphs; this greatly facilitates incorporation of cross-sentence contextual constraints.

Siemens George and Baker Ryan S J.d., <sup>2012</sup>, they have argued for increased and formal communication and collaboration between these communities in order to share research, methods, and tools for data mining and analysis in the service of developing both LAK and EDM fields.

Vranic M., Pintar D. and Skocir Z. , <sup>2007</sup>, have presented how data mining algorithms and techniques can be 700 new regular students and approximately. As emphasized, available

data about these used in the academic community to potentially improve students was expected to hold useful information not yet some aspects of education quality.

### **3. Summary of Literature Review**

In the present literature review studied about the data mining based educational technologies allow educators to study how students learn(descriptive studies) and which learning strategies are most effective. Data mining and knowledge discovery techniques can be applied to find interesting relationships between attributes of assessments, and the solution strategies adopted. To purpose a framework for the interesting association rules within a educational system. Taken together and used within the online educational setting, the value of these tasks lies in improving student performance and the effective design of the online courses.

To identify attributes characterizing patterns of performance disparity between various groups of students. We propose a general formulation of contrast rules as well as a framework for finding such patterns. Education systems allow for more accurate assessment and more effective evaluation of the learning process.

### **4. Research Gaps**

After review of literature, the researcher has figure out the following gaps:

Kumar Varun, Chadha Anupama.,<sup>2011</sup>, In the present study, they have discussed the various data mining techniques which can support education system via generating strategic information. Since the application of data mining brings a lot of advantages in higher learning institution, it is recommended to apply these techniques in the areas like optimization of resources, prediction of retainment of faculties in the university, to find the gap between the number of candidates applied for the post, number of applicants responded, number of applicants appeared, selected and finally joined.

Panda Subhalaxmi, Pattanaik, P. A., Swarnkar, Tripti.,<sup>2017</sup>, have used of Random Forest to take a correct appropriate decision as per their interest and skills. The final goal is to give a better insight to design a better Indian education system for Indian students with the effective outcome. This review may extend to larger features to solve complex decision databases in an efficient manner.

Luo<sup>1,2</sup>, Qi.,<sup>2008</sup>, have suggested that the Knowledge Discovery and Data Mining applications play an important role in data that have clearly grown to surpass raw human

processing abilities. The challenges facing advances in this field are formidable. Develop new mining algorithms for classification, clustering, dependency analysis, and change and deviation detection that scale to large databases.

Singh Bhullar Manpreet, Kaur Amritpal, 2012, have suggested that the no efficient method to caution the student about the student about the deficiency in attendance. It doesn't identify the weak student and inform the teacher. Another common problem in larger colleges and universities, some students may feel lost in the crowd. Whether they're struggling to find help with coursework, or having difficulty choosing (or getting into) the courses they need, many students are daunted by the task of working through the collegiate bureaucracy. Since the proposed model identifies the weak students, the teachers can provide academic help for them. It also helps the teacher to act before a student drops or plan for recourse allocation with confidence gained from knowing how many students are likely to pass or fail. Proposed system also shows data graphically according to the need or organization which help them to take important decisions. For future work we also use clustering, with the help of clustering we can see the domain and interest of students in particular field.

Marsh A. Julie, Pane F. John, and Hamilton S. Laura, 2017, they have collective findings from RAND's work on DDDM offer several directions to the broader research community. First, more research is needed on the effects of DDDM on instruction, student achievement, and other outcomes. Research to date has examined effects on instruction to a limited extent and has yet to measure effects on outcomes, although the ongoing ISBA study will be analyzing, among other things, the relationship between data use and student achievement. Future studies linking implementation and impact could shed light on the conditions under which positive effects are most likely to occur.

Al-Radaideh A. Qasem, M. Emad Al-Shawakfa and I. Mustafa Al-Najjar, 2006, have collected a real and large data set from the university student database and apply the model using such data. Moreover, several other classification methods can also be applied to test the most suitable method that suit the structure of the student data and give a better classification accuracy.



## **Conclusion:**

As per our study of literature there are several techniques available to analyze the Data mining is more than running some complex queries on the data to store in the database. They must work with your data, reformat it, or restructure it, regardless of whether using SQL, document-based databases such as Hadoop, or simple flat files. Identifying the format of the information that is need based upon the technique and the analysis that we want to do. After you have the information in the format you need, you can apply the different techniques (individually or together) regardless of the required underlying data structure or data set.

## **References:**

- [1]. Kumar Varun, ChadhaAnupama.,<sup>2011</sup>, “An Empirical Study of the Applications of Data Mining Techniques in Higher Education” (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No.3, March 2011.
- [2]. Panda Subhalaxmi, Pattanaik, P. A., Swarnkar, Tripti.,<sup>2017</sup>, “A Higher Education Predictive Model Using Data Mining Techniques” OE is volume is published and copyrighted by its editors.
- [3]. Luo<sup>1,2</sup>, Qi.,<sup>2008</sup>, “Advancing Knowledge Discovery and Data Mining”, 2008 Workshop on Knowledge Discovery and Data Mining .
- [4]. Villegas-Ch<sup>1</sup>, W., Mora<sup>2</sup>, S. Luján.,<sup>2017</sup>, “Analysis of data mining techniques applied to LMS for personalized education” IEEE.
- [5]. Asif, Raheela, Merceron, Agathe, Ali, Syed Abbas, Haider, NajmiGhani,<sup>2017</sup>, “Analyzing undergraduate students’ performance using educational data mining”
- [6]. Baker Ryan S.J.d., Baker R.S.J.d., Baker, E. (Eds.),<sup>2010</sup>, “Data Mining for Education”, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.
- [7]. Abad Martínez Fernando, Alicia A. ChaparroCasoLópez,<sup>2016</sup>, “Data-mining techniques in detecting factors linked to academic achievement”, bInstitute of Educational Research and Development, Autonomous University of Baja California, Ensenada, B. C., México.
- [8]. BresfeleanVasile Paul, BresfeleanMihaela, GhisoiuhaveNicolae,<sup>2008</sup>, “Determining Students’ Academic Failure Profile Founded on Data Mining Methods”,

Babes-Bolyai University, Faculty of Economics and Business Administration, Cluj-Napoca, Romania.

[9]. Lu Hongjun, Setiono Rudy, Liu Huan, <sup>2017</sup>, “Neuro Rule A Connectionist Approach to Data Mining”, Department of Information Systems and Computer Science National University of Singapore.

[10]. Minaei-Bidgoli Behrouz, A. Kashy Deborah, Kortemeyer Gerd, F. Punch William, <sup>2017</sup>, “Predicting student performance: An application of data Mining methods with an educational web-based system”, 33<sup>rd</sup> ASEE/IEEE Frontiers in Education Conference.

[11]. Bravo Javier, Shafti Leila, Ortigosa Alvaro, <sup>2009</sup>, “Recommendation in Higher Education Using Data Mining Techniques”, Universidad Autonoma de Madrid.

[12]. Singh Bhullar Manpreet, Kaur Amritpal, <sup>2012</sup>, “Use of Data Mining in Education Sector” Proceedings of the World Congress on Engineering and Computer Science Vol I WCECS 2012, San Francisco, USA.

[13]. Pal Saurabh, Chaurasia Vikas, <sup>2017</sup>, “Is Alcohol Affect Higher Education Students Performance: Searching and Predicting pattern using Data Mining Algorithms” International Journal of Innovations & Advancement in Computer Science, (IJIACS) ISSN 2347 – 8616 Volume 6, Issue 4.

[14]. Marsh A. Julie, Pane F. John, and Hamilton S. Laura, <sup>2017</sup>, “Making Sense of Data-Driven Decision Making in Education” The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

[15]. Al-Radaideh A. Qasem, M. Emad Al-Shawakfa and I. Mustafa Al-Najjar, <sup>2006</sup>, “Mining Student Data Using Decision Trees”, International Arab Conference on Information Technology (ACIT'2006), Nov. 2006, Jordan.

[16]. Delavaril Naeimeh, Ayatollahzadeh Shirazi Mohammad Reza and Beikzadeh Mohammad Reza, <sup>2004</sup>, “A New Model for Using Data Mining Technology in Higher Educational Systems”, IEEE.

[17]. Aher Sunita B, LOBO L.M.R.J., <sup>2011</sup>, “Data Mining in Educational System using WEKA” International Conference on Emerging Technology Trends (ICETT).

[18]. Baradwaj Brijesh Kumar, Pal Saurabh, <sup>2011</sup>, “Mining Educational Data to Analyze Students’ Performance”, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6.

- [19]. Yadav Surjeet Kumar, Bharadwaj Brijesh Kumar, Pal Saurabh, <sup>2011</sup>, "Data Mining Applications: A comparative Study for Predicting Student's performance" International Journal of innovative technology & creative engineering (ISSN:2045-711) VOL.1 NO.12.
- [20]. HALEESEL-ALAA, <sup>2008</sup>, "Mining Students data to analyze learning behavior: A case study" are linked to publications on ResearchGate.
- [21]. Pandey Umesh Kumar, Pal S., <sup>2011</sup>, "Data Mining : A prediction of performer or underperformer using classification", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2).
- [22]. FDEZ J. ALCALÁ, FERNÁNDEZ A., LUENGO J., DERRAC J., <sup>2010</sup>, KEEL Data-Mining Software Tool: DataSet Repository, Integration of Algorithms and Experimental Analysis Framework, <http://www.keel.es>.
- [23]. Y. Tang Tiffany and McCalla Gordon, <sup>2010</sup>, "Student Modeling for a Web-based Learning Environment: a Data Mining Approach", American Association for Artificial Intelligence (AAAI).
- [24]. R.B. Bhise, S.S. Thorat, A.K. Supekar, <sup>2013</sup>, "Importance of Data Mining in Higher Education System", underlined in blue are linked to publications on ResearchGate.
- [25]. Gnardellis.T., Boutsinas .B., <sup>2011</sup>, "On Experimenting with Data Mining in Education" Enterprise Programmes for Research and Technology II, General Secretariat for Research & Technology, Hellenic Ministry of Development.
- [26]. Hsia Tai-Chang, Shie An-Jin, Chen Li-Chen, <sup>2008</sup>, "Course planning of extension education to meet market demand by using data mining techniques – an example of Chinkuotechnology university in Taiwan", [tchsia@ctu.edu.tw](mailto:tchsia@ctu.edu.tw) (T.-C. Hsia).
- [27]. Pang Bo and Lee Lillian, <sup>2004</sup>, "A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts",
- [28]. Maqsood Mohd Ali, <sup>2013</sup>, "Role of data mining in Education Sector", IJCSMC.
- [29]. Siemens George and Baker Ryan S J.d., <sup>2012</sup>, "Learning Analytics and Educational Data Mining: Towards Communication and Collaboration",
- [30]. Verykios Vassilios .S., Bertino Elisa, Nai Fovino Igor, <sup>2004</sup>, "State-of-the-art in Privacy Preserving Data Mining" SIGMOD Record, Vol. 33, No. 1.
- [31]. Vranic M., Pintar D. and Skocir Z. , <sup>2007</sup>, "The use of data mining in education environment", 9th International Conference on Telecommunications - ConTEL 2007 ISBN: 978-953-184-111-5.