

January 2015

A Study of the Applications of Data Mining Techniques in Higher Education

SUSHIL VERMA

Department of Computer Applications, SATI Vidisha, sushilverma81@gmail.com

R. S. THAKUR

Department of Computer Applications, MANIT, Bhopal (MP), India, Ramthakur2000@yahoo.com

SHAILESH JALORI

Department of applied Maths & Computer science, SATI, Vidisha, Shilesh_jaloree@rediffmail.com

Follow this and additional works at: <https://www.interscience.in/ijcct>

Recommended Citation

VERMA, SUSHIL; THAKUR, R. S.; and JALORI, SHAILESH (2015) "A Study of the Applications of Data Mining Techniques in Higher Education," *International Journal of Computer and Communication Technology*: Vol. 6 : Iss. 1 , Article 1.

DOI: 10.47893/IJCCT.2015.1263

Available at: <https://www.interscience.in/ijcct/vol6/iss1/1>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer and Communication Technology by an authorized editor of Interscience Research Network. For more information, please contact sritampatnaik@gmail.com.

A Study of the Applications of Data Mining Techniques in Higher Education

¹SUSHIL VERMA, ²R. S. THAKUR & ³SHAILESH JALORI

¹Department of Computer Applications, SATI Vidisha

²Department of Computer Applications, MANIT, Bhopal (MP), India

³Department of applied Maths & Computer science, SATI, Vidisha

E-mail : sushilverma81@gmail.com, Ramthakur2000@yahoo.com, Shilesh_jaloree@rediffmail.com

Abstract - Data mining is used to extract meaningful information and to develop significant relationships among variables stored in large data set. Few years ago, the information flow in education field was relatively simple and the application of technology was limited. However, as we progress into a more integrated world where technology has become an integral part of the business processes, the process of transfer of information has become more complicated. Today, one of the biggest challenges that educational institutions face is the explosive growth of educational data and to use this data to improve the quality of managerial decisions and student's performance. The main objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of Unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance. The paper aims to purpose the use of Data mining techniques to improve the efficiency of higher educational institutions. If data mining techniques such as clustering, decision tree and association can be applied to higher education processes, it can help improve student's performance.

Keywords- *Data Mining, Educational Data Mining, Clustering, Decision tree, Classification, Prediction.*

I. INTRODUCTION

The concept of data mining is the technique of extracting previously unknown information with the widest relevance from databases, in order to use it in the decision-making process. Data Mining is a process of extracting previously unknown, valid, potential useful and hidden patterns from large data sets [1]. As the amount of data stored in Educational database is increasing rapidly. In order to get required benefits from such large data and to find hidden relationships between variables using different data mining techniques developed and used. Clustering and decision tree are most widely used techniques for future prediction. The main goal of clustering is to partition students into homogeneous groups according to their characteristics and abilities [2]. Nowadays; higher educational organizations are placing in a very high competitive environment and are aiming to get more competitive advantages over the other business competitors. Data mining techniques are analysis tool that can be used to extract meaningful knowledge from large data sets [3]. Educational data mining uses many techniques such as decision trees, neural networks, k-nearest neighbor, naive bayes, support vector machines and many others [4]. Decision tree analysis is a popular data mining technique that can be used to explain the

interdependencies among different variables such as attendance ratio and grade ratio. Clustering is one of the basic techniques often used in analyzing data sets. Data mining encompasses different algorithms that are diverse in their methods and aims [5]. Today, data collecting and storing are no longer expensive and difficult task. As a result, datasets are growing explosively. To extract the knowledge and information from these massive datasets has attracted a great deal of scientific attention and has become an important research area. Data mining is a flourishing research field and has become a synonym for the process of extracting hidden and useful information from datasets. Data mining provides many tasks that could be used to study the student performance.

II. DATA MINING TECHNIQUES

Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. While data mining and knowledge discovery in database are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. The sequences of steps identified in extracting knowledge from data are shown in Figure 1. Various algorithms and techniques like Classification, Clustering, Prediction, Association Rule, Decision Trees, Outlier

etc., are used for knowledge discovery from databases. These techniques and methods in data mining need brief mention to have better understanding.

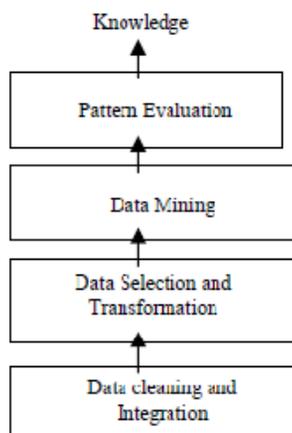


Figure 1: The steps of extracting knowledge from data

A) Clustering: -

Clustering is a technique by which similar records are grouped collectively. Usually this is done to give the end user a high level view of what is going on in the database. Clustering is sometimes used to denote segmentation. Clustering analyzes data objects without consulting a known class label. In general, the class labels are not present in the training data simply because they are not known to begin with [6]. Clustering can be used to generate such labels. The objects are clustered or grouped based on the principle of maximizing the intraclass similarity and minimizing the interclass similarity. That is, clusters of objects are formed so that objects within a cluster have high similarity in comparison to one another, but are very dissimilar to objects in other clusters. Each cluster that is formed can be viewed as a class of objects, from which rules can be derived. Application of clustering in education can help institutes group individual student into classes of similar behavior, Partition the students into clusters, so that students within a cluster (e.g. Average) are similar to each other while dissimilar to students in other clusters (e.g. Intelligent, Weak).

B) Classification and Prediction: -

Classification is the most commonly applied data mining Technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. Classification is the processing of finding a set of models (or functions) which describe and distinguish data classes or concepts, for the purposes of being able to use the model to predict the class of objects whose class label is unknown. The derived model may be represented in various forms, such as classification (IF-THEN) rules, decision trees, mathematical formulae, or neural networks.

Classification can be used for predicting the class label of data objects. However, in many applications, one may like to predict some missing or Unavailable data values rather than class labels. This is usually the case when the predicted values are numerical data, and is often specifically referred to as prediction. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In Classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier [7].

C) Association rule: -

Association analysis is the discovery of association rules showing attribute-value conditions that occur frequently together in a given set of data. Association analysis is widely Used for market basket or transaction data analysis. Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis [3]. Association Rule algorithms need to be able to generate rules with confidence Values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.

D) Decision tree:-

A decision tree is a foretelling model that, as its name implies, can be viewed as a tree. Specifically each branch of the tree is a classification question and the leaves of the tree are partitions of the dataset with their categorization. Decision tree is tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees and Chi Square Automatic Interaction Detection.

E) Outlier: -

A database may contain data objects that do not comply with the general behavior of the data and are called outliers. The analysis of these outliers may help in fraud detection and predicting abnormal values.

III. RELEATED WORK

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes. ("Han j. and Kamber", 2006) Explained

that k-means is a well known clustering algorithm tends to uncover relations among variables already presented in dataset [3]. ("Kifaya", 2009) explained that K-means clustering is a widely used method that is easy and quite simple to understand. Cluster analysis describes the similarity between different cases by calculating the distance. These cases are divided into different clusters due to their similarity [5]. ("Henrik", 2001) concluded that clustering was Effective in finding hidden relationships and associations between different categories of students [8]. ("Galit.et.al ", 2007) Gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams [9]. ("Erdogan and Timor ", 2005) Used educational data mining to identify and enhance educational process which can improve their decision making process [10]. ("Mohammad Reza Beikzadeh", 2004) in order to analyze student's trends and behaviors toward education. Lack of deep and enough knowledge in higher educational system may prevent system management to achieve quality objectives, Data Mining methodology can help bridging this knowledge gaps in higher education system [11]. ("Maclennan.J", 2005) Data mining techniques can be utilized effectively in selecting course, managing student's improving attendance / dropouts providing supplementary classes where necessary, allocating instructors in a better managed way and thus improving overall stature of the institute/university [12]. ("K. H. Rashan", 2011) A Data Mining model can monitor each student's progress by capturing the variables such as previous semester grade, test mark, assignment grade and attendance. The student's performance can also be analyzed based on the features of interpersonal peer groups such as intellectual self confidence, scoring pattern and time spent with peer groups [13]. ("Luan", 2002) described that higher education institutions carry three duties that are Data Mining intensive [14]. They are:

1. Scientific research that relates to the creation of knowledge.
2. Teaching that concerns with the transmission of knowledge.
3. Institutional research that pertains to the use of knowledge for decision making.

(" Brijesh Kumar Baradwaj, ",2011) described that the hidden patterns, associations and anomalies, which are discovered by some Data mining techniques, can be used to improve the effectiveness, efficiency and the speed of the processes [15].

IV. RESEARCH METHODOLOGY

The research methodology adapted is based on the in-depth study of the topic pertaining to the data mining and its application in higher education. The

views of various national and international conferences were taken into consideration while analyzing the data mining applications in the field of higher education. The talks with various academicians, institutions, colleges offering higher education and experts in the field of data mining helped us to find and present the techniques, process and application of data mining in higher education. The main objective of higher education institutions is to provide quality education to its students. One way to achieve highest level of quality in higher education system is by discovering knowledge for prediction regarding enrolment of students in a particular course, alienation of traditional classroom teaching model, detection of unfair means used in online examination, detection of abnormal values in the result sheets of the students, prediction about students' performance. Higher educational system can be viewed from two different angles. One is the outsider environmental motivation and the other is internal educational reason. The outsider environmental motivation can be observed in higher educational system where the higher educational organizations are aiming to be ahead of their business competitors. Therefore they first have to be powered by a proper roadmap and to be demonstrated with an exact guideline of reaching the top-level educational level. The internal educational reason is considered as the proceeding toward improving the educational management system. Higher educational organization is projected with the promises of more speed during the processes, fewer costs, more quality and flexibility. We have identified six main processes in higher educational systems, which are "evaluation, planning, registration, consulting, marketing and examination". Each process can be categorized into some sub-process. As an example, "evaluation", is an educational process. Its main sub-process are "student assessment", "lecturer assessment", "industrial training assessment", "course assessment" and "student registration evaluation". The main idea in our proposed work is improving the current processes to some new and enhanced educational processes, which have got superior advantages over the traditional processes. "Course selection consulting" is a sub-process under the "consulting" educational main process. By applying some of the classification, clustering or association technique on the set of student taken various courses data, the characteristic patterns of previous students who took particular elective subjects or courses, and the association of courses or elective subject by various type of student can be extracted as knowledge and be stored in knowledgebase. The resulting enhanced process is "classification or association students to the most appropriate course and elective subject". The output of the process can be used by the faculty's consultant to present the most suited courses to students and also by educational course planner to have more advanced strategies on student's course planning.

V. CONCLUSION

Among several innovation in recent technology, data mining is making comprehensive changes in the field of higher education. We 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. Data mining techniques capabilities provided effective improving tools for student performance. It showed how useful data mining can be in higher education in particularly to predict the final performance of student.

REFERENCE :

- [1] C. Romero, S. Ventura, E. Garcia (2008), "Data mining in course management Systems: Moodle case study and tutorial", *Computers & Education*, Vol. 51, No. 1, pp. 368-384, 2008.
- [2] Connolly T., C. Begg and A. Strachan (1999), "Database Systems: A Practical Approach to Design, Implementation, and Management" (3rd Ed.). Harlow: Addison-Wesley.687, 1999.
- [3] Han, J. and Kamber, M., (2006), "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems", Jim Gray, Series Editor, 2006.
- [4] Wayne Smith (2005), "Applying Data Mining to Scheduling Courses at a University", *Communications of the Association for Information Systems*, Vol. 16, Article 23, 2005.
- [5] Kifaya (2009), "Mining student evaluation using associative classification and clustering", *Communications of the IBIMA* vol. 11 IISN 1943-7765, 2009.
- [6] C. Romero, S. Ventura (2007), "Educational data mining: A Survey from 1995 to 2005", *Expert Systems with Applications* (33), pp. 135-146, 2007.
- [7] Shaeela Ayesha, Tasleem Mustafa, Ahsan Raza Sattar, M. Inayat Khan (2010), "Data Mining Model for Higher Education System", *Europen Journal of Scientific Research*, Vol.43, No.1, pp.24-29, 2010.
- [8] Henrik (2001), "Clustering as a Data Mining Method in a Web-based System for Thoracic Surgery": © 2001.
- [9] Galit.et.al (2007), "Examining online learning processes based on log files analysis": a case study, *Research, Reflection and Innovations in Integrating ICT in Education*, 2007.
- [10] Erdogan and Timor (2005), "A data mining application in a student database", *Journal of Aeronautic and Space Technologies* July 2005 Volume 2 Number 2 (53-57), 2005.
- [11] Mohammad Reza Beikzadeh, Naeimeh Delavari (2004), "A New Analysis Model for Data Mining Process in Higher Educational Systems" TS3B-2 in M2USIC 2004.
- [12] ZhaoHui, MacIennan.J, (2005), "Data Mining with SQL Server 2005" Wihely Publishing, Inc, 2005.
- [13] K. H. Rahan, Anushka Peiris (2011), "Data Mining Applications in the Education Sector", MSIT, Carnegie Mellon University, retrieved on 28/01/2011.
- [14] Luan, J. (2002), "Data mining and Its Applications in Higher Education" in A.Serban and J.Luan (eds.) *Knowledge Management: Building a Competitive Advantages for Higher Education. New Directions for institutional Research*, No.113.San Francisco, CA: Jossey Bass.
- [15] Brijesh Kumar Baradwaj, Sourabh Pal (2011), "Mining Educational Data to Analyze Student's Performance", *IJACSA* Volume 2, no. 6, 2011.

