



ISSN 2278 – 0211 (Online)

## Use of Data Mining Techniques in Assessing Student and Faculty Needs

**Archna Bajaj**

Ph.D. Scholar, Department of Computer Science and System Studies, Mewar University, Chittorgarh, India

**Dr. S. K. Saxena**

Senior Faculty, Department of Computer Science and Engineering, Delhi Technological University, India

### **Abstract:**

*The aim of this research paper is to use Data Mining Techniques to assess and prioritize the Students and Faculty needs. Students and Faculty form the core part of any Institute or organization and the effectiveness of these institutes depends on meeting their needs. In this study, data is collected from students and faculty from higher education institutes using a closed-ended questionnaire. The focus area of the study is based on providing the services that most likely meet the needs of all the participants. The responses generated help in getting a deep insight into the factors that may lead to enhancing student and faculty satisfaction and decreased student dropout ratio.*

**Keywords:** Data Mining, Educational Data Mining (EDM), Faculty Needs, Student Needs, Student Satisfaction

### **1. Introduction**

According to Giudici, Data Mining is “the process of selection, exploration, modeling of large quantities of data to discover regulations or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of database”[1].

Data Mining or Knowledge Discovery in Databases (KDD) is a collection of exploration techniques based on advanced analytical methods and tools for handling large amount of information. These techniques can find novel patterns that may assist an enterprise in understanding the business better and forecasting [2].

Even though Data Mining has been used in numerous fields which include financial services, Banking sector, retail, healthcare, telecom, food industry to name a few, the application of Data Mining to educational sector is limited. There are increasing research interests in using Data Mining in education. This new emerging discipline called Educational Data Mining (EDM) deals with the developing of methods for exploring data that is originating in an educational context [3].

Quality of higher education institutes implies providing the services, which most likely meet the needs of students, staff and other participants.

“Need” is defined as the gap between current outcomes or outputs and desired (or required) outcomes or outputs [4]. It is important to understand and identify these gaps in terms of ways of providing education so as to help them achieve their goals.

Data Mining acts as an automated assistant that helps in bridging this gap. Emerging Data Mining techniques discover the hidden patterns and associations, changes and anomalies from educational data resulting in improved decision making process in higher educational institutes. This improvement would bring significant advantages like increasing student satisfaction, increasing student learning outcome and improving student-teacher relation.

Clustering, Classification, Association, and Regression etc. are different data mining techniques that can be used in the field of education for discovering different kinds of knowledge. In the proposed work, we will be using these Data Mining Techniques to try and assess student and faculty needs.

### **2. Literature Survey**

Data Mining is the non-trivial extraction of implicit, previously unknown and potential information from data. This encompasses number of technical approaches such as clustering, data summarization, classification, finding dependency Network, analyzing changes and detecting anomalies [5].

Educational Data Mining is an emerging field that focuses on applying Data Mining tools to Education related Data [6].

Researches within Educational Data Mining field have focused on topics like analyzing/predicting student performance, decreasing dropout ratio, student retention, transfer cases, student behaviour analyses etc.

Recent literature on Educational Data Mining is presented below:

Romero and Ventura conducted survey on Educational Data Mining between Years 1995 to 2005 and concluded that Educational Data Mining is a promising area of research [7].

Alaa-El-Haleez gave a case study that used Educational Data Mining to analyze student-learning behavior [8].

B.K.Bhardwaj and Saurabh Pal used classification task on student database to predict the students division on the basis of previous database and reduce fail ratio [9].

Chong Ho Yu used Data Mining techniques to study the predictors affecting University student retention [10].

S.Anupama and Vijaylakshmi used Data Mining techniques to study behavior and performance of students [11].

Ajay Kumar Pal and Saurabh Pal used Data Mining techniques in EDM for predicting performance of students [12].

J.Ranjan and K.Malik used data-mining techniques for exploring the effects of probable changes in processes related to admissions, course delivery and recruitments [13].

### 3. Objectives

1. The proposed work uses Data Mining Techniques to identify the needs of students and teachers and further enhances their skills to help them achieve their career goals.
2. Identify the needs and preferences of students towards appropriate course ware, teaching methodologies, faculty competency and approach and ways of assessment. This can help in finding relation between student satisfaction and student retention/dropout rate.
3. Prioritize the needs of the students. Need patterns can be further studied to identify the focus areas that need to be developed in order to achieve improved learning outcome.
4. To help the courseware coordinators in offering updated courses and specializations.
5. To help the faculty/instructors in developing new, focused/targeted and more effective strategies towards teaching methodologies.
6. Finding similarities and differences between student and teacher needs.

### 4. Research Methodology

A Data Mining process is likely to include the following steps:-

a) Data selection and collection: This step includes finding the best source databases for the data that is required.

b) Cleaning and preparing data: A data store integrates data from a number of databases. When integrating data, preparing the data includes solving problems and dealing with missing data, data conflicts and ambiguities.

c) Data mining exploration and validation: Once data has been collected and cleaned, it is possible to start data mining exploration. Data mining model may be constructed and after taking sample data, number of relevant techniques can be applied.

d) Evaluating and Monitoring: It is important to evaluate the results and choose the best technique. Accuracy and efficiency of these techniques needs to be checked. Further there is need to monitor the performance of these techniques.

#### 4.1. Data Collection Strategy

Data can be collected from educational institutes offering higher education using methods like: - Questionnaire, Surveys, Observations, Interviews and/or Interactions. The data for this study has been collected from Students and Faculty of Undergraduate courses in various Colleges. A sample size of 350 students and 75 faculty has been used in the study. The closed-ended questionnaire comprised of various attributes pertaining to academic needs. These needs were broadly classified under Course ware, Teaching methodologies, Assessment methods and Teachers competency and Approach. The Likert scale based questionnaire was responded to by the participants on the scale of 1 to 5 ranging from Very Important, Fairly Important, Neutral, Not Important and Not at all Important.

#### 4.2. Cleaning and Preparing Data

This collected data needs to be pre-processed and cleaned before Data Mining Techniques can be applied on the same to extract meaningful information. In this step only those fields were selected which were required for data mining. The pre-processing was done in terms of filling up missing values and transforming values from one form to another. The data collected from various sources has been collated into a single dataset for processing.

#### 4.3. Data Mining Exploration

A popular data mining tool, IBM SPSS has been used for initial statistical analysis of the data. The Excel dataset is imported into SPSS. One of the Central tendency measures (Mean, Median and Mode) will be used to assess and prioritize the needs of the participants. This will be followed by classification of the participants on the basis of gender and then clustering the students and faculty for further analysis.

#### 4.4. Data Analysis

Data Mining Techniques that may be used for Data Analysis are as below:

#### 4.4.1. Data Mining Techniques

Data mining is the process of discovering meaningful, new correlations, patterns and trends by sifting through large amounts of data stored in repositories using pattern recognition [14]. Based on the kinds of patterns we are looking for, Data mining tasks can be classified into two categories: Descriptive and Predictive.

A) Descriptive: It describes the data, relationships between variables, identifies the patterns and includes models for overall probability distribution of data. E.g. Clustering, Summarization & Association Rule.

##### i) Clustering:

Clustering is the identification of classes also called clusters or groups for a set of objects whose classes are unknown. The objects are so clustered that the interclass similarities are maximized and interclass similarities are minimized based on some criteria defined on the attributes of objects. Once the clusters are decided, the objects are labeled with their corresponding clusters and common features of the objects in a cluster are summarized to form class description. Different clustering methods are:

Distance based method, Partition method, Hierarchical and Model based method.

##### ii) Summarization:

Summarization is the abstraction or generalization of data. A set of task-relevant data is summarized and abstracted, resulting in smaller set that gives general overview of the data with aggregate information. Summarization maps data into subsets with associated simple descriptions [15].

##### iii) Association:

Association is the discovery of togetherness or connection of objects. An Association rule reveals the associative relationships among objects i.e. appearance of set of objects in a database is strongly related to the appearance of another set of objects. Apriori algorithm can be used for finding association rules. Romero et al. (2008) indicates, "Association rule can be applied to find relation between characteristics of objects" [16].

##### B) Predictive:

This model permits value of one variable to be predicted from known values of another variable. It analyses the current state and past state of attribute and helps in prediction of future: -

Classification, Regression and Trend Analysis.

##### i) Classification:

Classification is the derivation of function or model that determines the class of an object based on its attributes. Classification is a different technique as compared to Clustering. A classification model is constructed by analyzing the relation between the attributes and the classes of objects in the training set. This technique is useful when classes are already known. This technique can be used in predicting the class to which an object is likely to belong. Some of the classification algorithms that can be used are: -

Decision tree Rules and Naive Bayes method.

##### ii) Regression:

It is a Data Mining function that predicts a number. Regression task begins with data set in which target values are known. Regression algorithm estimates value of target as a function of predictors for each case in builds data. These relationships between predictors and target are summarized in a model, which can be applied to different data set in which target values are unknown. Different types of Regression are: -

Linear Regression, Non-Linear Regression, Multi variate Regression.

##### iii) Trend Analysis:

A lot of data available now are time series data that are accumulated over time. Trend Analysis discovers interesting patterns in the evolution history of objects. A model function is constructed to simulate the behavior of object that can be used to predict the future behaviour.

One or more of the above data mining techniques will be used to achieve the objectives of the study.

## **5. Conclusion**

This study is being done with the objective of assessing the needs of Students and Faculty and the results will give an insight into the preferences of the participants. The study focusses on the academic needs of the students and the higher educational Institutes can use the study to streamline their focus areas and improve student satisfaction which is directly proportional to student retention. The comparison between the priorities of students and faculty re-emphasizes the similar and divergent needs of each group.

## 6. Future Scope

This research work is not exhaustive as we the participants are limited to undergraduate students and faculty of colleges. Also, the needs analyzed are academic in nature. More work needs to be done in case of post graduates students and/or an extension of their probable needs with respect to various service areas, geographical locations, type and year of course, career aspirations, socio-economic backgrounds, ethnicity, Age, college environment etc.

## 7. References

- i. P. Giudici, "Introduction," in *Applied Data Mining: Statistical Methods for business and Industries*, 1<sup>st</sup> ed. Sussex, England, John Wiley & Sons, 2003, pp. 2.
- ii. G.K.Gupta, "Introduction," in *Introduction to Data Mining with Case Studies*, 1<sup>st</sup> ed. New Delhi, India, Prentice-Hall of India Pvt. Ltd., 2006, pp. 45.
- iii. C. Romero et al., "Data mining in Course Management system: Moodle Case study and Tutorial", *Computers & Education*, Vol. 51, No.1, 2008, pp. 368-384.
- iv. R. A. Kaufman, "Why Needs Assessment," in *Needs Assessment-Concept and Applications*, 1<sup>st</sup> ed. New Jersey, Educational Technology Publications, 1979, pp. 8.
- v. A. K. Pujari, "Data Mining," in *Data Mining Techniques*, 1<sup>st</sup> ed. Hyderabad, India, Universities Press (India) Pvt Ltd, 2001, pp. 45
- vi. R. S. J. D. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *Journal of Educational Data Mining*, Article 1, Vol. 1, No. 1, pp. 3-17, Fall 2009.
- vii. C. Romero & S. Ventura, "Educational Data mining: A survey from 1995 to 2005," *Expert Systems with Applications*, Vol. 33, No. 1, pp. 135-146, 2007.
- viii. A. El-Halees, "Mining Students Data to analyze learning behavior: A case study," Department of Computer Science, Islamic University of Gaza P.O.Box 108 Gaza, Palestine, 2009.
- ix. B. K. Baradwaj & S. Pal, "Mining Educational Data to Analyze Students," *IJACSA*, Vol. 2, Issue 6, pp. 63-69, Oct. 2011.
- x. Chong Ho Yuet et al., "A Data Mining Approach for Identifying Predictors of Student Retention from Sophomore to Junior Year," *Journal of Data Science*, Vol. 8, pp. 307-325, 2010.
- xi. S. A. Kumar and M. N. Vijayalakshmi, "Relevance of Data Mining Techniques in Edification Sector," *International Journal of Machine Learning and Computing*, Vol. 3, No. 1, pp. 4-6, Feb. 2013.
- xii. A. K. Pal and S. Pal, "Classification Model of Prediction for Placement of Students," *I.J. Modern Education & Computer Science*, Vol. 5 Issue 11, pp. 49-56, Nov. 2013.
- xiii. J. Ranjan & K. Malik, "Effective educational process: A data-mining approach", *VINE*, Vol. 37, Issue 4, pp. 502-515, 2007.
- xiv. A. K. Pujari, "Data Mining," in *Data Mining Techniques*, 1<sup>st</sup> ed. Hyderabad, India, Universities Press (India) Pvt. Ltd, 2001, pp. 46
- xv. M. H. Durham, "Introduction," in *Data Mining- Introductory & Advanced Topics*, 4<sup>th</sup> ed. New Delhi, India, Dorling Kindersley (India) Pvt. Ltd., 2009, pp. 9
- xvi. C. Romero et al., "Data mining in Course Management systems: Moodle case study and tutorial," *Computers & Education*, Vol. 51, No.1. pp. 368-384, May 2007.