

# USAGE OF DATA MINING METHODS FOR IMPROVISATION OF EDUCATION

<sup>1</sup>Bhaves M. Patel, <sup>2</sup>Dr. Ashok R. Patel

<sup>1</sup>Assistant Professor, Dept. of Computer Science, Hemchandracharya North Gujarat University, Gujarat, INDIA

<sup>2</sup>Assistant Professor, Florida Polytechnic University, USA

**Abstract :** Data mining (DM) is the process to discover different patterns, co-relations and anomalies among large datasets to estimate future outcomes. While usage of tools and techniques of data mining to extract knowledge automatically from huge database generated by or related to student's learning and educating activities from educational environment is called Educational data mining (EDM). EDM is very useful in education system particularly when examining students' learning performances. It analyzes education related data to design models to improve student' learning experiences and enhance effectiveness of institution. Hence, EDM can aid educational institutions to offer high quality education for its learners. This research paper focuses on data mining techniques which can be used to improve education.

**Keywords:** Educational Data Mining (EDM); Data Mining (DM); Prediction; Clustering; Classification;

## I. INTRODUCTION

Educational data analysis is not new idea as it was performed by many researchers time by time using traditional analysis methods. But recent advances in computer technology in education including high processing computers and the ability to process fine-grained data from huge datasets, have attracted many researchers to use these technologies to increase interest in designing techniques to analyze the large amounts of data generated in educational environment. Data mining approaches are used widely to mine educational data. The figure 1 illustrates the process of mining educational data.

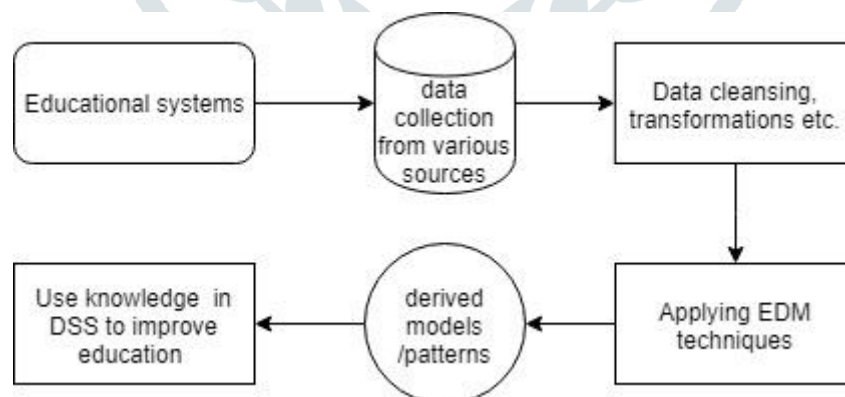


Figure 1 Process flow of data mining in education

As we are data rich, data from different sources (homogenous / heterogenous) of education either online or offline are collected and data cleansing techniques are applied to clean noisy or missing data. Data transformation is also applied to convert all the data into uniform. Once dataset is ready for data mining, different data mining techniques are applied to extract knowledge from it and this knowledge is used by analysts in decision support system(DSS) for improving education.

## II. Algorithms of data mining used in education

Data mining usually is applied on disciplines like classification, categorization, prediction, clustering and visualization. Classification is used to identify associations and clusters, and splits subjects under study. For example, educational systems may use classification broadly to analyze characteristics of students. Categorization uses rule induction algorithms for handling categorical results. Prediction is used to discover data points which are purely based on the description of another related data points. Classification and prediction uses supervised data, while Clustering uses unsupervised data Visualization uses different visualization techniques like interactive graphs; charts etc to display mathematically induced data and scores.

### 1. Classification

Classification is a type of data analysis to extract models which describes important data classes. Such models are used to predict categorical class labels. For example, we can develop a classification model to classify student's performance either poor or good. Such models are used to better understand the data at large. Many researchers have proposed several classification models in machine learning, pattern recognition and statistics. Classification is supervised learning in which all training data are already classified with class labels. Data classification is performed in two stages, consisting of a *learning step* in which classification model is developed and, *classification step* in which the model is used to predict the class label of given data.

### 2. Relationship mining

Relationship mining is used to discover relationships between different variables in a dataset and training them as rules for later use. There are different types of techniques such as association rule mining, sequential pattern mining, correlation mining, and causal data mining. Association rule mining (ARM) is used to identify association between two or more data variables to generate frequent patterns. Sequential pattern mining is used to mine frequently occurring ordered events or subsequences as patterns. For example, sequential patterns are "Students who learn mathematics are likely to learn programming." Correlation mining is one of the most important data mining methods for its capability to identify primary dependencies between objects. Casual data mining is used to find relationship between casual data.

In education, relationship mining can be used for identifying relationships between the student's on-line activities and the final marks and to model learner's problem solving activity sequences. Also it can be utilized into identifying the relationship in patterns of behavior of students, difficulties or mistakes which learners usually encountered with.

### 3. *Clustering*

In data mining, Clustering is the technique of grouping particular set of objects based on their characteristics, aggregating them according to their similarities. It partitions a set of data (or objects) into a set of meaningful sub-classes, called clusters. Clustering is unsupervised learning in which there is no class label in training data. In education, clustering can be used to group students according to their performance and identifying those groups of students which are weak or left the education. And educational institutes can analyze to improve their performance and also reduce the ratio of school leaving.

### 4. *Prediction*

Prediction in data mining is used to identify data points purely on the description of another related data value. It's not essentially associated to upcoming events, but the variables which are used are unknown. Prediction obtains the relationship between a value we know in advance and a value we need to forecast for future reference. For example, prediction models in data mining are used by a education analyst who predict that how much grade a student will obtain in final year, so that student can be prepared for job accordingly. The prediction is also referred to as Numeric Prediction in data mining. Normally regression analysis is used for prediction.

### 5. *Discovery with models*

Of the broad categories of data mining methods mentioned above, classification, relationship mining, clustering and prediction mining are considered universal methods across all types of data mining; however, Discovery with Models is considered one of the more important approaches in educational system.

In this method, a model is designed through prediction, clustering or by human reasoning knowledge engineering and this model is then used as a component in another analysis, namely in prediction and relationship mining. In the prediction, the created model's predictions are used to predict a new variable. For the usage of relationship mining, the designed model allows the analysis between new predictions and additional variables in the study. In many cases, discovery with models make use of authenticated prediction models which have verified generalizability across situations. Key applications of this method include finding relationships among student behaviors, characteristics and contextual variables in the learning environment.

### 6. *Distillation of data for human judgment*

Humans can make conclusions about data that might be beyond the scope in which an automated data mining method provides. For the use of education data mining, data is cleaned for human judgment for mainly two purposes, identification and classification.

For the purpose of identification, data is cleaned to facilitate humans for identifying well known patterns, which might be difficult to interpret. i.e. the learning curve shows the pattern that visibly reflects the relationship among experience and learning over time in educational system.

Data is also cleaned for classifying features of data, which is used to assist the creation of the prediction model. Classification aids to speed up the creation of the prediction model, enormously.

The aim of this method is to display the information in a visual, useful and interactive way in order to understand the huge amounts of educational data and also to assist decision making. In general, this method is useful to educators to understand usage information and efficacy in course activities. Major applications of the distillation of data for human judgment incorporate discovering patterns in student learning, performance etc.

### III. Applications in Educational system

Data mining is widely used in education for improvisation of learner and educator activities.

**Student modeling:** It can be used to find out patterns of characteristics and behaviors of students by mining repositories of student models.

**Student group formation & Detection of undesirable student behaviors:** We can create student groups based on their performance, regularity or any other characteristics. Also we can discover undesirable behavior of student by applying clustering and classification methods.

**Prediction of student performance:** By the use of prediction method, we can predict student performance in future.

**Visualization of analyzed data:** Knowledge can be presented through various visualization techniques like graph, chart etc to better understand it.

**Feedback for educators:** Feedback is the most important part of any system to improve future. By taking feedback from students and other educators, education can be improved by correcting educators in the right direction.

**Curriculum development:** Data mining tools are used to customize learning activities for each student and adapt the pace in which the student completes the course. This is in particularly beneficial for online courses with varying levels of competency.

**Social network analysis:** Social network analysis is a technique to observe the structure of relationships and the effect this social structure has on the attitudes, behavior, and performance of the individual actors or groups.

#### IV. Conclusion

Data mining is a very powerful analytical tool to improve decision making and analyzing new patterns and relationships for educational organizations. Educational data mining contains techniques including data mining, statistics, machine learning. Data mining is required to analyze data collected from teaching and learning, tests learning theories, and policy decision-making etc. There exists a number of opportunities educational data mining, from an analysis at organizational level to the analysis at individual level. Furthermore, educational data mining is widely used and applied by learners, researchers and teachers, even institutions.

#### REFERENCES

1. Peña-Ayala A., Mizoguchi R., Student Modeling by Data Mining, New Challenges for Intelligent Information and Database Systems. Studies in Computational Intelligence, pp 207-216 vol 351. Springer, Berlin, Heidelberg 2011
2. S. L. Tanimoto, "Improving the prospects for educational data mining," in Track on Educational Data Mining, at the Workshop on Data Mining for User Modeling, at the 11th International Conference on User Modeling, pp. 1–6, 2007.
3. S. K. Mohamad and Z. Tasir, "Educational data mining: A review," *Procedia-Social and Behavioral Sciences*, vol. 97, pp. 320–324, 2013.
4. X.chen, M.Vorvoreanu, K.Madhavan, Mining Social Media Data for Understanding Student's Learning Experiences, *Ieeexplore.Ieee.Org*, 7(3), 2014, 246–259
5. N. Upadhyay, V. Katiyar, A Survey on the Classification Techniques in Educational Data Mining, *International Journal of Database Management System (IJDMS)*, 3(11), 2014, 725–728
6. M.A Khan, W.Gharibi and S.K.Pradhan (2014), Data Mining Techniques for Business Intelligence in Educational System: a Case Mining, 978-1-4799-3351-8/14/\$31.00 ©2014 IEEE
7. A. M Morais, J. M. F. R. Araújo and E.B. Costa (2014), Monitoring Student Performance Using Data Clustering and Predictive Modelling, 978-1-4799-3922-0/14/\$31.00 ©2014 IEEE.
8. Baker, Ryan. "Data Mining for Education" , Oxford, UK: Elsevier. 2014
9. B. Patel, B. Prajapati, J. Dharwa, AR Patel, Study and Evaluation of Student Performance Parameters using statistical features of excel, *International Journal of Innovative Research in Computer Science & Technology*, pp 2347-5552