



# INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

## A SURVEY ON THE CLASSIFICATION TECHNIQUES IN EDUCATIONAL DATA MINING

SHRUTI A. JARI<sup>1</sup>, PROF. NIKHIL S. BAND<sup>2</sup>, DR. HEMANT R. DESHMUKH<sup>3</sup>

1. Student of Master of Engineering in (CSE), IBSS college of Engineering and Technology, Amravati, India.
2. Assistant professor Department of (CSE), IBSS College of Engineering and Technology, Amravati, India.
3. Head of the Department of (CSE), IBSS College of Engineering and Technology, Amravati, India.

Accepted Date: 05/03/2015; Published Date: 01/05/2015

**Abstract:** Due to increasing interest in data mining and educational system, educational data mining is the emerging topic for research community. Educational data mining means to extract the hidden knowledge from large repositories of data with the use of technique and tools. Educational data mining develops new methods to discover knowledge from educational database and used for decision making in educational system. The various techniques of data mining like classification. Clustering can be applied to bring out hidden knowledge from the educational data. In this paper, we focus on the educational data mining and classification techniques. In this study we analyze attributes for the prediction of student's behavior and academic performance by using open source data mining tool and various classification methods like decision trees.

**Keywords:** Educational data mining; Classification; Analysis.



PAPER-QR CODE

Corresponding Author: MS. SHRUTI A. JARI

Access Online On:

[www.ijpret.com](http://www.ijpret.com)

How to Cite This Article:

Shruti A. Jari, IJPRET, 2015; Volume 3 (9): 1191-1197

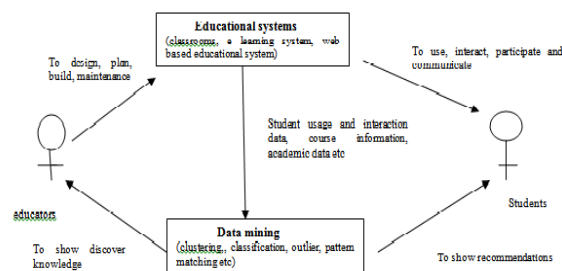
## INTRODUCTION

The examination and study of student's academic performance is not a new exercise but computer based learning environment increases more interest towards student's analysis. The concepts and techniques of data mining can be implemented in education to predict the academic performance of student. On the basis of these kinds of predictions the academic performance of student can be improved. EDM is applied to large amount of data

Accumulated by surveys and various classification techniques are implemented for better performance. The prediction of student's performance has become one of most important needs in order to improve the quality of performance. There is a need of data mining in educational system for the students as well as academics responsible. Educational data mining is an arising regulation that promote the new techniques for extracting the new data that come from educational settings and by using those techniques, a better prediction can be done for student's behavior, academic performance, subject interest etc.

## II. WHAT IS EDUCATIONAL DATA MINING?

Data mining originate a new technique known as educational data mining. In educational data mining, data mining concepts are applied to data that is related to field of education. EDM is the process of transforming the raw data aggregated by education systems. Educational data mining means exploring hidden data that originated from educational settings by using new methods for better interpretation of students and settings they learnt. Educational data mining promote distinct tools and algorithms for analyze the data patterns. In EDM, data is accumulated during learning process and then study can be done with the techniques from statistics, machine learning and other data mining concepts. To extract the hidden knowledge from data came from educational system, the various data mining techniques like classification, clustering, rule mining etc.



In fig we represent the need of educational data mining. The Academics' responsible and educators worked upon the educational system to enhance the performance of students. In this diagram it is shown that educators want to design the educational system then plan to build that system and most important maintain that educational system. Educational systems include traditional classrooms and some innovative learning methods like e learning system, intelligent and adaptive web based educational system etc. The data set can

be extracted from students as students are directly connected with educational system. Now the data is given as input to data mining process and in result it gives recommendations to students and to extract new knowledge to the educators by using various data mining techniques like clustering, classification, pattern matching etc.

#### **i. Goals of Educational data mining:**

Some of the goals of educational data mining are as follows:

1. Prediction of student's learning behavior by building student models that integrate all definite information of students like student's knowledge, behavior, academic information etc.
2. Exploring or upgrading domain models that discriminate the content to be learnt and perfect pedagogical sequences.
3. Analysis of all the effects of various types of instructional support given by learning.
4. Advancing scientific knowledge.

#### **ii. Phases of Educational data mining:**

Educational data mining is concerned with translation of new hidden information from the raw data collected from educational systems. EDM generally consist of four phases:

1. The first phase of educational data mining is to find the relationships between data of educational environment. The aim of establishing these relationships is to utilize these relationships in various data mining techniques like classification, clustering, regression etc.
2. The second phase of educational data mining is validation of discovered relationships between data so that over fitting can be avoided.
3. The third phase is to make predictions for future on the basis of validated relationships in learning environment.

4. The fourth phase is supporting decision making process with the help of predictions.

### iii. Methods of Educational data mining

There are so many promoted methods of educational data mining but all kind of methods lie in one of following specified categories:

1. **Prediction:** given a detail explanation of prediction in his paper. He mentioned that “In prediction, the goal is to develop a model which can infer a single aspect of data (predicted variable) from some combination of other aspects of data (predictor variables).if we study prediction extensively then we get three types of prediction: classification, regression and density estimation. In any category of prediction the input variables will be either categorical or continuous. In case of classification, the categorical or binary variables are used, but in regression continuous input variables are used. Density estimation can be done with the help of various kernel functions.

2. **Clustering:** In clustering technique, the data set is divided in various groups, known as clusters. When data set is already specified, then the clustering is more useful. As per clustering phenomenon, the data point of one cluster and should be more similar to other data points of same cluster and more dissimilar to data points of another cluster. There are two ways of initiation of clustering algorithm. Firstly, start the clustering algorithm with no prior assumption and second is to start clustering algorithm with a prior postulate.

3. **Relationship Mining:** Relationship mining generally refers to contrive new relationships between variables. It can be done on a large data set, having a no of variables. Relationship mining is an attempt to discover the variable which is most closely associated with the specified variable. There are four types of relationship mining: association rule mining, correlation mining, sequential pattern mining and causal data mining. Association data mining is based on if- then rule that is if some particular set of variable value appears then it generally have a specified value. In correlation mining, the linear correlations are discovered between variables. The aim of sequential pattern mining is to extract temporal relationships between variables.

4. **Discovery with Models:** it includes the designing of model based on some concepts like prediction, clustering and knowledge engineering etc. This newly created model’s predictions are used to discover a new predicted variable.

5. **Distillation of Data for Human Judgment:** There are two objectives for human judgment for which distillation of data can be done: Identification and

Classification. As per phenomenon of identification, data is represented in a way that human can easily recognize the well specified patterns.

### **III. LITERATURE SURVEY**

#### **i. Efficiency of decision trees in predicting student's academic performance**

It suggested an approach for predicting the student's performance in examination. In data collection, a slight modification has been done in defining the nominal values for the analysis of accuracy. As per need of system, data is preprocessed, and integer values are converted into nominal values. In this paper, the implementation of decision trees rules can be done by dividing the data into two groups. In decision tree the attribute for splitting at each node of tree is normalized information gain. The attribute having highest normalized information gain is chosen to make decision. This paper analyzes the accuracy of algorithm in two ways, the first is by comparing the result of tree with the original marks obtained by student in terms of efficiency.

#### **ii. Classification model of prediction for placement of students**

In new approach of classification to predict the placement of students. This approach provides the relations between academic records and placement of students. In this analysis, various classification algorithms are employed by using data mining tools for study of student's academic records. In this approach the training algorithm uses a set of predefined attributes. The most widely used classification algorithms are, naïve Bayesian classification algorithm, multilayer perceptron. For the high dimensional inputs the naïve Bayesian classification is best technique. Multilayer perceptron is most suitable for vector attribute values for more than one class. For testing, the 10 fold cross validation is selected as this evaluation approach. Here, a no of tests are regulated for estimation of input variables: chi square test, information gain test and gain ratio test. Each of the tests makes the concernment of variable in another way. According to this analysis, among three selected best algorithms, the best algorithm is Naïve Bayes classification.

#### **iii. Study of factors analysis affecting academic achievement of undergraduate students in international program**

In this paper analysis to identify the weak students so that the academic performance of those weak students can be improved. In this study, open source data mining tool is used to estimate aspects for predicting the student's academic performance. In this study, data set to characterize classifier (decision tree, neural network). To predict the accuracy, a cross validation

with 10 folds is used. In this study, to explore the proposal, two classification algorithms have been accepted and distinguished: The investigation process consists of three main steps: data preprocessing, attribute filtering and classification rules. According to this analysis, it is suggested the decision tree model is more accurate than the neural network model. It can be concluded that the decision tree technique has better efficiency data classification for this data set.

#### IV. CONCLUSION:

This paper described about the Educational data mining, goals of educational data mining and phases of educational data mining and existing classification techniques. Various classification techniques can be implemented on the data set but which classification technique will be applied on the data to improve the academic performance of students, it is important. In this paper, we made a comparison analysis on different existing approaches and methods of classification of data sets. We also analyzed the advantages and shortcomings of each algorithm applied to data set. So we can say that this paper will provide a beneficial glance of existing solution for classification with their advantages and shortcomings.

#### V. REFERENCES

1. Kumar S. Anupama and N. vijaylaxmi M.2011 Efficiency of Decision trees in predicting Student's Acedemic performance.
2. L. Ramanathan, Dhanda S. and D. S. kumar 2013 Predicting Student's Performance using Modified ID3 Algorithm
3. Pal A. kumar and Pal S. 2013 Classification Model of Prediction for Placement of Students
4. Cheewaparakobkit P.2013 Study of Factors Analysis Affecting Academic Achievement of Undergraduate Students in International Program
5. Bresfelean V. Paul 2007 Analysis and Predictions on Student's Behaviour using Decision Trees in Weka Environment Babes Bolyai University
6. Baker Ryan S.J.d. Data mining for education Carnegie Mellon University.
7. [http://docs.oracle.com/html/E10312\\_01/dm\\_concepts.htm](http://docs.oracle.com/html/E10312_01/dm_concepts.htm) Data Mart Concepts
8. <http://datawarehouse4u.info/OLTP-vs-OLAP.html> OLTP vs OLAP
9. <http://olap.com/category/bi-solutions/predictive-analytics> Predictive Analysis

10. Patil, Preeti S.; Srikantha Rao; Suryakant B. Patil (2011). "Optimization of Data Warehousing System: Simplification in Reporting and Analysis". IJCA Proceedings on International Conference and workshop on Emerging Trends in Technology (ICWET) (Foundation of Computer Science) 9 (6): 33–37.