Scientific Journal of Impact Factor (SJIF): 3.134
ISSN (Online): 2348-4470
ISSN (Print) : 2348-6406

International Journal of Advance Engineering and Research Development

Volume 2, Issue 2, February -2015

A NOVEL EFFICIENT MODEL FOR PREDICTING STUDENTS PERFORMANCE USING CLASSIFICATION METHOD

Deepika D Patil[1], Ramesh Medar[2]

[1]Computer Science and Engineering , GITCET
2Computer Science and Engineering, GITCET

Abstract: *Student performance in university courses is of great concern to the higher education where several factors may affect the performance. Currently the huge amount of data is stored in educational database these database contains the useful information for predicting student performance. In this paper , the classification rule generation process is based on the decision tree as a classification method where the generated rules are studied and evaluated. A system that facilitates the use of the generated rules is built which allows students to predict the final grade in a course under study.*

Keyword-*Data Mining, Classification, Decision tree, Knowledge Discovery in database (KDD), ID3 Algorithm*

I. INTRODUCTION

The inventory of information technology in various fields has lead the large volumes of data storage in various formats like records, files, documents, images, sound, videos, scientific data and many new data formats. The data collected from different applications require proper method of extracting knowledge from large repositories for better decision making. The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data.

Using these techniques many kinds of knowledge will be discovered like association rules, classifications and clustering. The discovered Knowledge can be used for prediction regarding enrollment of students in a particular course, prediction about student's performance and so on. The main objective of this paper is to use data mining methodologies to study student's performance in end. Examination plays a vital role in any student's life. The marks obtained by the student in the examination decide his future. Therefore it becomes essential to predict whether the student will pass or fail in the examination. If the prediction says that a student tends to fail in the examination prior to the examination then extra efforts can be taken to improve his studies and help him to pass the examination.

II. RELATED WORK

Brijesh Kumar Baradwaj and saurabh pal [5] describes that the quality 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 particular course, detection of abnormal values in the result sheets of the students, prediction about students' performance and so on, the classification method is used to evaluated student's performance and as there are many approaches that are used for data classification, the decision tree method is used here.

Han and Kamber [4] describes data mining software that allow the user to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process.

Mohammed M.Abu Tair and Alaa M. EI-Halees [6] applied the educational data mining concerns with developing methods for discovering knowledge from the data that came from educational domain. After preprocessing the data, we applied data mining techniques to discover association, classification, clustering and outlier detection rules. In each of these four tasks, we present the extracted knowledge and describes its importance in educational domain.

Monika Goyal and Rajan Vihra [3] applied the data mining techniques to achieve improvement in the efficiency of higher education institution. If the data mining techniques such as clustering, decision tree and association are applied for higher education processing, it would help to improve students' performance, to measure their retention, their lifecycle management and the grant fund management of an institution.

III. DATA MINING TECHNIQUES

Data mining refers to learning in a practical, nontheoretical sense. Data Mining defines extracting or "mining" knowledge from large amount of data. Data Mining Techniques are used to operate on huge amount of data to obtain hidden patterns and also for decision making [5].

Different Techniques used for Data Mining like Classification, Clustering, Regression, Artificial Intelligence, Decision trees, Genetic Algorithm, Nearest Neighbor method are used for knowledge discovery from database. The Following are the steps for extracting data from database.

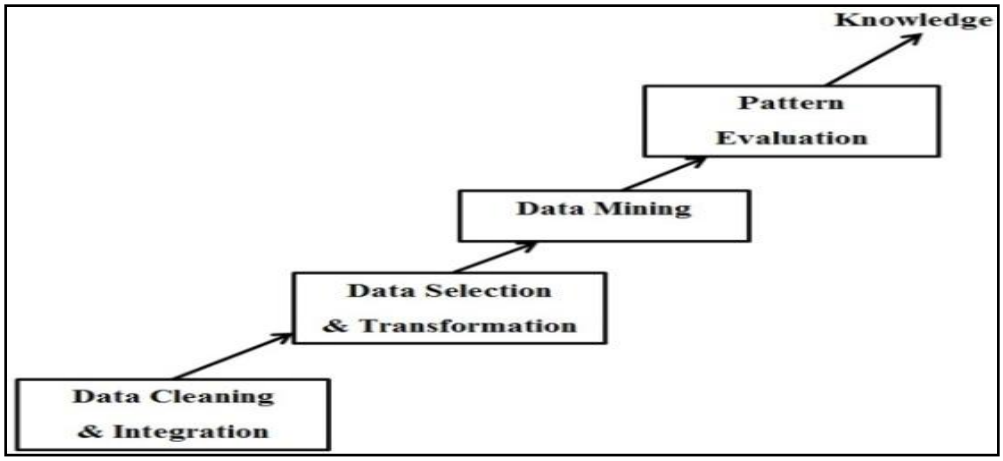

Figure1. Steps of Extracting Knowledge from Data

3.1. Classification

The Classification method of data mining technique, it is a method of finding a set of models which describes and distinguishes data models, for a purpose of being able to use model to predict the class of objects whose class label is unknown. The derived model may be represented in many ways such as neural networks, mathematical formulas, decision tree etc. The data classification process involves learning and classification .In classification test data are utilized to calculate the accuracy of the classification rules. If the accuracy is such that it is excepted then the same rule is applied to different data types. Hence we use ID3 algorithm to represent logical rules of student final grade.

3.2. Clustering Analysis

Clustering is a method of finding group of object which belong to same domain or category. Clustering analyses data objects without consulting known class label. In educational data mining, clustering is a process of grouping students according to their own behavior. According to clustering, clusters used to distinguish student's performance according to their activities and behavior. In this they have been clustered into different groups like, punctuality, exams, and academics and so on.

Clustering Application in the engineering domain can help institution to group the students into different clusters depending upon the performance of individuals. Partitioning the students into clusters represents that the students having the similar behavior to each other.

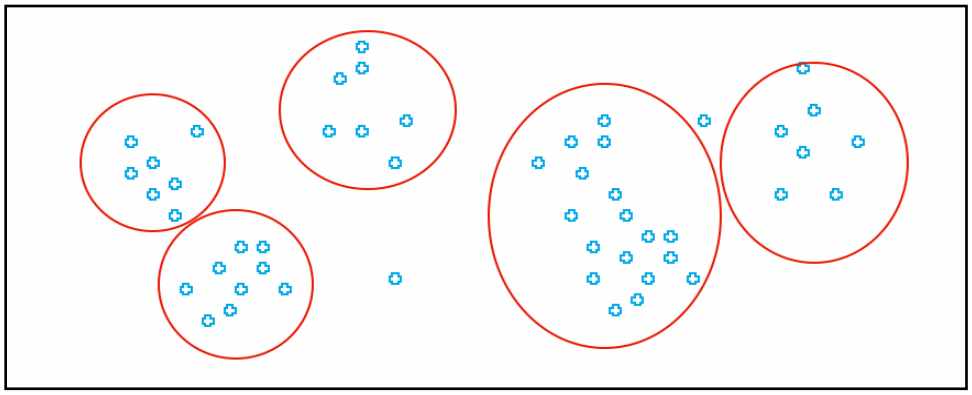

Figure 2. Picture showing clusters with similar behavior

3.3. Association rule applying

Association rule mining is a conventional and well researched method for determining interesting correlation relationship between attributes in large dataset. Association analysis is the discovery of association rules showing attribute-value condition that occurs frequently together in a given data set. A typical example of data mining is market basket analysis. Rule Support and Confidence are two measures of interestingness.

3.4. Decision Tree method

Decision tree are popularized structure for supervised learning. Decision trees are mostly used in mainly operational research especially in decision analysis, to identify the attributes to reach the goal [3].There are number of decision making algorithm used and ID3 algorithm is used here.

IV. DATA MINING PROCESS

4.1. Data Preparation and Analysis

The data from the database is used and the data is collected from one of the institution for preparing the data set of students form an Information System Department on the sampling method for the session 2010 to 2015. The size of data records is in terms of thousands. For the analysis purpose the data from the different department is joined to make a single data.

4.2. Data selection and transformation

In Selection method the data field which are required to analyze the result of students, only that data is selected. For the prediction of the student's performance the required data from the database is extracted and is used for the analysis or transformation. Some of the variables used for selected are given in following table [11].

Variables	Description	Possible value
Dep	Department of students	(Technical, Scientifical, Commercial, Industrial)
HSD	Higher School Degree of students	(Excellent, Very good, Good, Acceptance)
Midterm	Midterm marks of students	Excellent>=85%, Very good=85%> &<75%, Good=65%> & <75%, Average 55%> & <65% , Poor<55%.
LG	Lab grade	Poor, Average, Good
SEM	Seminar	Poor, Average, Good
ASS	Assignment	Yes, No
FG	Final grade	Excellent>=85%, Very good=85%> &<75%, Good=65%> & <75%, Average 55%> & <65% , Poor<55%.

Figure3. Student Related Variables

4.3. Decision Tree

Decision tree is a flow-chart like structure, where each internal node is represented by rectangles and the leaf nodes are denoted by ovals. All internal nodes have two or more child nodes. Its most commonly used algorithm because of its ease of implementation and easier to understand compared to other classification algorithm [10**].**

4.4. ID3 algorithm

The steps of ID3 algorithm is as follows:

1. Let P be the set of training instance
2. Choose an attribute that best differentiates the instances
3. Create a tree node that value is chosen as an attribute. Create child node from this link where each link represents represents a unique value for the chosen attribute. If the subclass does not satisfy the predefined criteria and there is at least one attribute to further subdivide the path of the tree.

An entropy and Gain are calculated. The index used to measure degree of impurity is Entropy [7].
The Entropy is calculated as follows

$$\textbf{Entropy}=\Sigma_j - P_j \log_2 P_j$$

Splitting criteria used for splitting of nodes of the tree is its Gain. To obtain the best attribute for a particular node in a tree can be obtained by calculating Information gain. The Information gain can be calculated as follows:

$$\textbf{Gain(S,A)=Entropy(S)-}\Sigma_{v \in Values(A)} |Sv|/|S|*\textbf{Entropy (Sv)}$$

V. OUTCOME DETECTION

The dataset of the students has been obtained from the Engineering institute for the academic year 2010-2015.This paper with the help and gain tries to predict the behavior or performance of the students based on the attributes provided in the database.

VI. CONCLUSION

In this paper Data Mining techniques were applied to predict the performance of the students. Mainly the ID3 algorithm was used on the bases of student's database. This prediction will help to improve the result of the students in the upcoming year. From this the students who need special attention is predicted and that student is provided with that attention to reduce failing ration and taking appropriate action at right time.

VII. REFERENCE

[1] Abeer Badr El Din Ahmed, Ibrahim Sayed Elaraby," A prediction for Student's Performance Using classification Method, World Journal of Computer Application and Technology, 2014
[2] Anju Rathee, Robin Prakash Mathur, "Survey on Decision Tree Classification algorithms for the Evaluation of Student Performance", International Journal of Computers & Technology, 2013
[3] Hongjie Sun, Research," Student Learning Result System based on Data Mining", 2010.
[4] Monika Goyal, Rajan Vohra2, "Applications of Data Mining in Higher Education", 2012
[5] Brijesh Kumar Baradwaj, Saurabh Pal, "Data mining: machine learning, statistics, and databases", 1996.
[6] Mohammed M. Abu Tair, Alaa M. El-Halees, "Mining Educational Data to Improve Students", 2012
[7] Nikhil Rajadhyax, Rudresh Shirwaikar," Data Mining on Educational Domain", 2012.
[8] K.Shanmuga Priya, A.V.Senthil Kumar, "Improving the Student's Performance Using Educational Data Mining"'2013
[9] Bhise R.B, Thorat S.S, Supekar A.K, "Importance of Data Mining in Higher Education System", 2013.
[10] Surjeet Kumar Yadav, Brijesh Bharadwaj, and Saurabh Pal, Mining Education Data to Predict Student's Retention: A comparative Study", 2012.
[11] Varun Kumar, Anupama Chadha, "Mining Association Rules in Student's Assessment Data", 2012.

