

# Analysis of Student Performance by using Data Mining Concept

Miss .Pooja M .Dhekankar

Student of Master of engineering (CSE)

G.H. Raisoni college of Engineering and Technology  
Amravati, India

Poojadhekankar151@gmail.com  
9503293558

Dinesh S. Datar

Dinesh.datar@raisoni.net

G.H. Raisoni college of Engineering and Technology  
Amravati, India

**Abstract:** Data mining methodology has a tremendous contribution for researchers to extract the hidden knowledge and information which have been inherited in the data used by researchers. Its related with developing and improving methods for discovering knowledge from data that extract from educational domain. This paper represents to categorize the students into grade order in all their education studies and it helps to improve 'students' academic performance. In this format of paper we used educational data mining techniques to improve 'students' performance, and detects the problem of low academic performance of students. For evaluation of better performance of students we use some interesting techniques like association rules, clustering, classification and outlier detection.

**Keywords:** Association Rule, Classification, Outlier Detection Clustering.

\*\*\*\*\*

## I. INTRODUCTION

There are many tremendous improvement research interests in using data mining in educational sector. This modern emerging sector, called educational data mining, concerned with improved methods that extract knowledge from data come from the educational sector[5]. Data mining is a technique of sorting which is actually used to extract hidden patterns from huge databases. This concepts and methods can be applied in various fields like marketing, medicine, real estate, customer relationship management, engineering, web mining, etc. Educational data mining is a new emerging or advanced technique of data mining that can be applied on the data related to the field of education. The data can be collected from past used data and operational data reside in the databases of educational institutes. The data of students can be personal information or academic performance. Furthermore it can be achieve from e-learning database systems which have a huge amount of data and information used by most institutes[5][6]. It uses many techniques for proper implementation of data mining concepts such as decision trees, neural networks, naive bayes, K-Nearest neighbour and many others. Using these techniques different kinds of knowledge can be discovered using association rules, classification and clustering[5][6]. By using this we extract knowledge that describes students performance in examination and all their detail information. From This huge amounts of data, the first task is to sort them out, cluster analysis is to classify the raw data in a reasonable way. clustering is a bunch of physical or abstract objects, as per the degree of similarity between them, divided into several groups, and makes the same data objects within a groups of high similarity and different groups of data objects which are not similar[1]

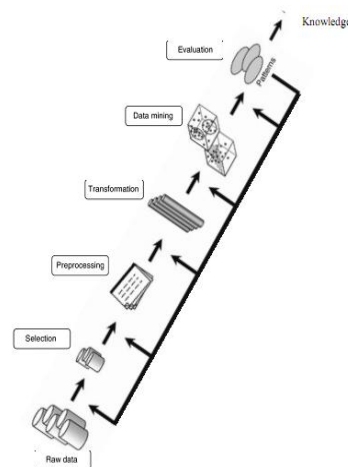


Figure 1: The steps of extracting knowledge from data

## II. RELATED WORK

Although, using concept of data mining in the higher education is a recent research sector, there are many research techniques in this field. That is because of its improvement in educational institutes. In this paper that use students data to analyze their learning behavior to predict their performance[2]. Using discovered association rules, rules are sorted using lift metric method. Then we apply two way classification methods which are rule induction and naive Bayesian classifier to predict the grade and performance of students. We clustered the students into groups using k-mean clustering algorithm. Finally, We apply outlier detection to find out all outliers present in the data. Two outlier methods are used distance-based approach and density-based approach. Each one of these tasks goes hand in hand to improve the performance of students. Our aim is to study and show how data mining concept can be implemented in higher education to improve student' performance. He used students' data capture from database table and collected all available data including personal

detail and academic detail information of students, course records and data from e-learning system particularly classification to help in improving the quality of the higher educational system by evaluating student data to study the main attributes that may affect the student performance in courses. The extracted classification rules are based on the decision tree as a classification method, the extracted classification rules are studied and evaluated. Then we apply association rule mining analysis based on students' failed courses to identifies students' failure patterns. The aim of their study is to find hidden relationship between the failed courses and suggests relevant causes of the failure to improve the poor results of students' performances.

In this area; we describe the improvement results of applying the data mining methods to the data, for each of the four data mining tasks; classification Association rule, , clustering and outlier detection, and how we get the benefit from the discovered and applied knowledge.

### A. Association Rule

Association rule is a advanced and well researched technique for discovering interesting relations between variables in large databases. Association rules are usually apply to satisfy a user-specified minimum support and auser-specified minimum confidence at same level. Association rule generation is usually divided into two individual steps:

First, The minimum support is applied to find all frequent item sets in a database. Second, these frequent item sets and the minimum confidence constraint are required to form rules. Finding all frequent item sets in a database is difficult since it involves searching all possible item sets (item combinations). The set of possible item sets is the power set over I and has size  $2^n - 1$ . Figure 2 shows a sample of association rules discovered from data students with average grade, with their support, confidence [3].  
 [Lower\_class\_grade=Poor, Higher\_class\_grade=Good]-> [Grade=Average]  
 (Support: 0.19, Confidence: 0.757)  
 [Lower\_class\_grade=Good, Higher\_class\_grade=Poor]-> [Grade=Average]  
 (Support: 0.105, Confidence: 0.731)

### B. Classification

Classification is the way of searching a model that distinguishes and describes data models, classes or concepts, for the reason of being able to apply the model to identify the class of objects whose class label is not known to the user. The derived model is based on the analysis of a set of training data. It is very essential to know that classification rules are different than rules generated or derived from association technique. The concept of association rules are characteristic rules, but classification rules are prediction rules[2].

If lower\_class\_grade=good and Higher\_class\_grade=good then Topper  
 If Lower\_class\_grade=poor and Higher Class\_grade=good then Average  
 If Lower\_class\_grade=poor and Higher\_class\_grade=poor then Below Average.

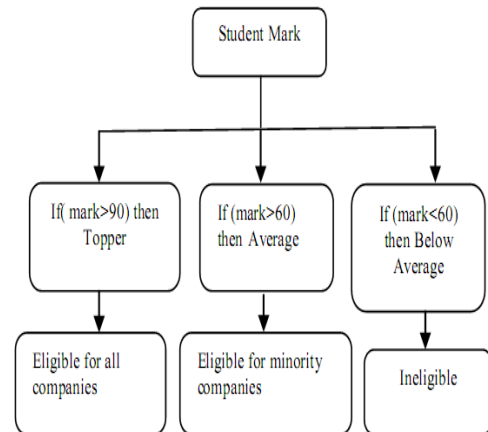


Fig2. Student Classification

### D. Outlier Detection

In this paper, we apply outlier analysis to identify outliers present in the student database collection. Distance-based approach find out the count of outliers in the present database depends on the distance to their k- nearest neighbors , and the result of this method is to flag the records either to be outlier or not, with true or false value [4].Density-based approach depicts the local densities of particular area and declares instances in low density regions as potential outliers.

### C. Clustering

Clustering technique is the task of segmenting and arranging a diverse group into a number of identical subgroups or clusters. Clustering of data is a technique in which we form clusters of object that are same in function and characteristics. The criterion for detecting the similarity is implementation dependent. Clustering techniques is often confused with classification, but there are some variations. In classification the objects are assigned to predefined classes, where as in clustering techniques the classes are also to be declared. Clustering methods may be divided into two sub classes based on the clustering structure and functions which produce hierarchical cluster and partitioning cluster.

## III. LITRATURE REVIEW

The building blocks of data mining is the evolution of a field with the confluences of various disciplines, which includes database management systems(DBMS), Statistics, Artificial Intelligence(AI), and Machine Learning. The era of data mining applications was conceived in the year1980 primarily by research-driven tools focused on single tasks [8]. In this paper we are actually Implement some standard techniques for better out obtaining from the applied concept of data mining .The applied techniques in this paper are:

1. Association Rule
2. Classification
3. Outlier Detection
4. Clustering

In this way the above techniques are implemented for analysis of student performance in educational field by using the technique of data mining .

#### IV. CONCLUSION

In this paper data mining techniques are efficiently used to categorize the level of students. One of the data mining techniques that is classification, accurately classifies the data for categorizing student based on the levels. Particularly we describe the association rules and we clarify the rules using lift metric. Then we used two classification techniques which are Rule Induction and Naïve Bayesian classifier to detect the performance of the graduate student. As one important function of data mining, clustering analysis either as a separate tool to discover data sources distribution of information, as well as other data mining algorithm as a preprocessing step, the cluster analysis has been into the field of data mining is an important research topic. Clustering is used to the group the students according to their grade and proficiency. This goes a long way to help how define the recruitment process in a easier manner.

#### REFERENCE

- [1] Yujie.zhang.(2012),“Clustering Methods in Data Mining with its Applications in High Education”, International conference on Education Technology and Computer (ICETC 2012) IPCSIT vol.43.
- [2] Mohammed M. Abutair , Alaa M.El-Halees.(2012),“Mining Educational data to improve students” performance: A case study”, International Journal of Information and Communication Technology Research, Volume 2 no.2
- [3] El-Hales-A.(2008),“Mining Students Data to Analyze Learning Behavior: A Case Study”, The 2008 International Arab Conference of Information Technology(ACIT2008)-Conference Proceedings, University of Sfax,Tunisia,Dec 15-18.
- [4] Han.J and Kamber.M.(2006),“Data Mining: Concepts and Techniques,2ndedition. The Morgan Kaufmann series in Data Management Systems, Jim Gray, series editor.
- [5] Romero, C. , Ventura, S. and Garcia, E. (2008) ‘Data mining in course management systems: Moodle case study and tutorial’, Computers & Education, vol. 51, no. 1, pp. 368-384.
- [6] Romero, C. and Ventura, S. (2007) ‘Educational data Mining: A Survey from 1995 to 2005’, Expert Systems with Applications (33), pp. 135-146.
- [7] Al-Radaideh, Q., Al-Shawakfa, E. and Al-Najjar, M. (2006) ‘Mining Student Data Using Decision Trees’, The 2006 International Arab Conference on Information Technology(ACIT'2006) – Conference Proceedings.
- [8] Piatetsky-Shapiro, Gregory. 2000. The Data-Mining Industry Coming of Age. IEEE Intelligent Systems.
- [9] Amandeep Kaur Mann, Navneet Kaur, Survey Paper on Clustering Techniques, IJSETR,2278 – 779.
- [10] Pradeep Rai, Shubha Singh, A Survey of Clustering Techniques, International Journal of Computer Applications, 2010.