Volume 3, Issue 8, August 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Study on Student Data Analysis Using Data Mining Techniques

Umamaheswari. K[*]
Department of Computer Science, NGM College
India

S. Niraimathi
Department of Computer Applications, NGM College
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. It is a processing procedure of extracting credible, novel, effective and understandable patterns from database. This paper is used to categorize the students into grade order in all their education studies and it helps in interview situation. This study explores the socio-demographic variables (age, gender, name, lower class grade, higher class grade, degree proficiency and extra knowledge or skill, etc). It examines to what extent these factors helps to categorize students in rank order to arrange for the recruitment process. Due to this, all students get benefitted and it also reduces the short listings. Here, clustering, association rules, classification and outlier detection has been used to evaluate the students performance.***

Keywords* — *Data mining, Clustering, Classification, Association rule, Outlier detection, Preprocessing.

I. Introduction

Data mining is a type of sorting technique which is actually used to extract hidden patterns from large databases. Data mining 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 technique of data mining that can be applied on the data related to the field of education. It uses many techniques 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. By using this we extract knowledge that describes students' performance in the end of the semester examination and all their details. In the face of huge amounts of data, the first task is to sort them out, cluster analysis is to classify the raw data in a reasonable way. The so called clustering is a group of physical or abstract objects, according to the degree of similarity between them, divided into several groups, [1] and makes the same data objects within a groups of high similarity and different groups of data objects which are not similar.

II. RELATED WORK

Data mining in higher education is a recent research field and this area of research is gaining popularity because of its potentials to educational institutes. In this paper that use students data to analyze their learning behavior to predict the results.Mohammed M.AbuTair, Alaa M.EI-Halees [2], had a survey on educational data mining [1993-2007] they collected graduate students information and applied data mining techniques to discover knowledge. Using discovered association rules, they sorted the rules using lift metric. Then they used two way classification methods which are rule induction and naive Bayesian classifier to predict the grade of graduate students. They also clustered the students into groups using k-mean clustering algorithm. Finally, they used outlier detection to detect all outliers in the data. Two outlier methods which are distance-based approach and density-based approach were used. Each one of these tasks goes hand in hand to improve the performance of graduate students. Romero and Ventura[5],did on a survey on educational data mining between 1995 and 2005.They concluded that educational data mining is a promising area of research and it has specific requirements not presented in other domains.Thus,work should be oriented towards educational domain of data mining.Bharadwaj and pal[6] conducted study on the students performance based on selecting 300 students from 5 different degree college conducting BCA(Bachelor of Computer Application) course of Dr.R.M.L.Awadhuniversity,Faizabad,India.Based on Bayesian classification method using 17 attributes, it was found that the factors like students grade in senior secondary exam, living location, medium of teaching, mother's qualification, students other habit, family annual income and student's family status were highly correlated with the student academic performance.

III. Graduate Students Dataset and Recruitment Preprocessing

The dataset is a collection of final year student's information. To group the students data using clustering technique, it may hold the academic, residence and personal record of the student. It includes students whole study details from its beginning. When data can be taken from the educational field then this new emerging field is termed as "Educational Data Mining" [3]. The main objective of higher educational institutions is to provide proper placement facilities to the student. For this reason they categorize the student based on their skill level. Skill level is ranked in the form of CGPA grade taking into account end semester marks and also based skill test. These are used evaluate to evaluate the results. According to the student level, they categorize them into groups. The institution allows the eligible students to attend, the recruitment process based on the companies criteria. The eligible students were categorized using clustering technique. This makes easier to select and reduce time as well. Due to this process recruited candidates are high and filtering is low. And there is more chance for all level of candidates to be recruited. Before applying the data mining techniques on the data set, there should be a methodology that governs our work. In this section we apply the data mining techniques to the data.

A. Association Rule

Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. Association rule generation is usually split up into two separate steps:

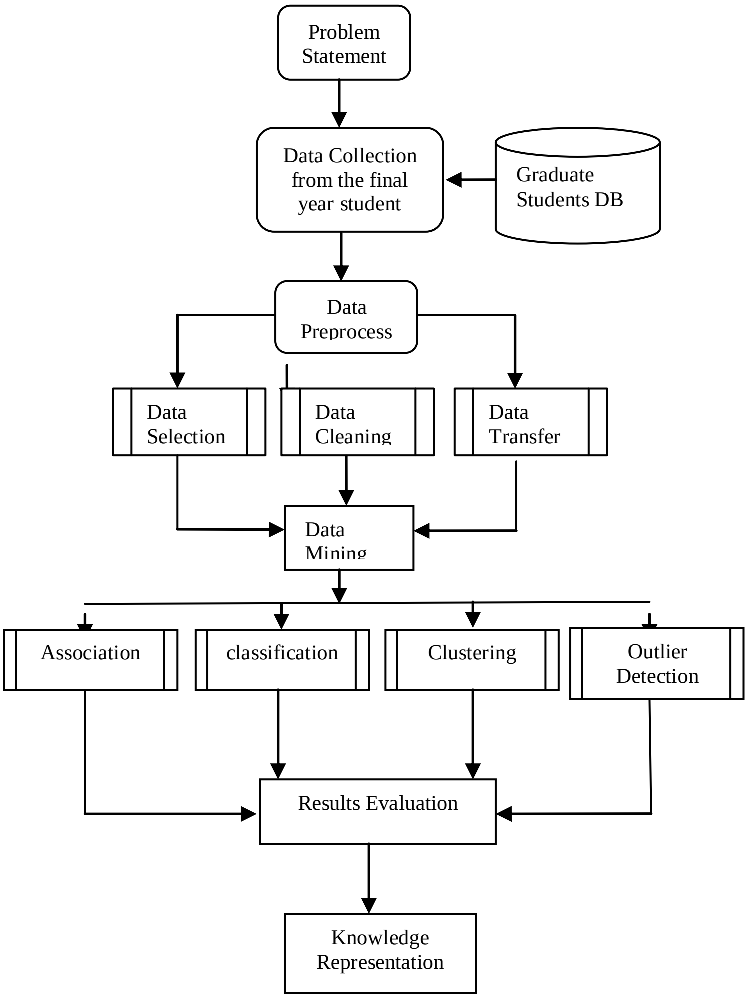

Fig 1: Data mining Methodology

First, minimum support is applied to find all frequent item sets in a database. Second, these frequent item sets and the minimum confidence constraint are used 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 $\boldsymbol{2^n-1}$(excluding the empty set which is not a valid item set).Figure 2 depicts a sample of association rules discovered from data students with average grade, with their support, confidence [2].

[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 process of finding a model that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. The derived model is based on the analysis of a set of training data. It is important to know that classification rules are different than rules generated from association. Association rules are characteristic rules, but classification rules are prediction rules [5].

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.

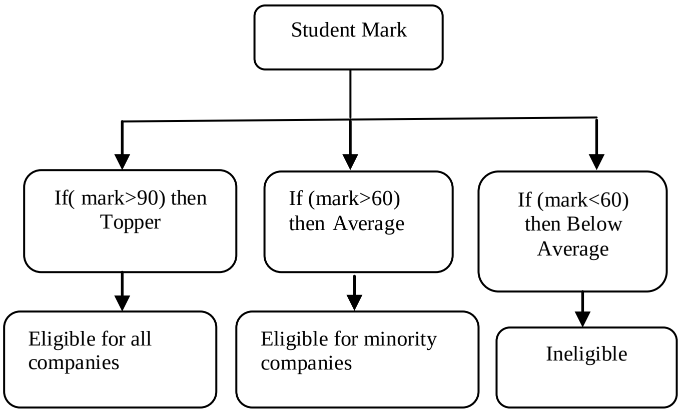
Student Mark
If(mark>90) then Topper
If (mark>60) then Average
If (mark<60) then Below Average
Eligible for all companies
Eligible for minority companies
Ineligible

Fig 2: Student Classifications

C. Clustering

Data Clustering is a method in which we make cluster of objects that are somehow similar in characteristics. The criterion for checking the similarity is implementation dependent. Clustering is often confused with classification, but there are differences. In classification the objects are assigned to predefined classes, where as in clustering the classes are also to be defined. Clustering methods may be divided into two categories based on the cluster structure which they produce hierarchical cluster and partitioning cluster.

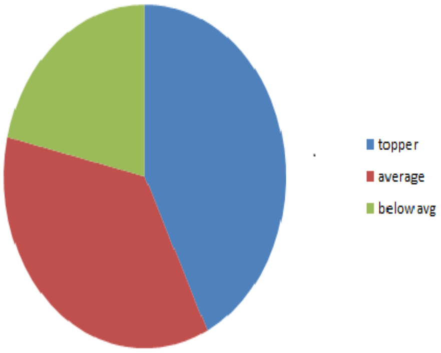

Fig 3: Clustering Students Based On Performance

D.Outlier Detection

In this paper, we outlier analysis to detect outliers in the student dataset. Distance-based approach identifies the number of outliers in the given data set based on the distance to their k nearest neighbours, and the result of applying this method is to flag the records either to be outlier or not, with true or false value [4].Density-based approach computes local densities of particular regions and declares instances in low density regions as potential outliers.

IV. CONCLUSIONS

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. 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.

REFERENCES

[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] Er.Rimmy Chuchra.(2012),'Use of Data Mining Techniques for the Evaluation of Student Performance: A case study', International Journal Of Computer Science and Management Research,vol 1.

[4] Han.J and Kamber.M.(2006),'Data Mining: Concepts and Techniques,2nd edition. The Morgan Kaufmann series in Data Management Systems, Jim Gray, series editor.

[5] 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.

[6] B.k.Bharadwaj and s-pal(2011),'Data Mining: A Prediction for Performance Improvement using Classification',(IJCSIT) International Journal of Computer Science and Information Security (IJCSIS) vol.9,no.4,pp-136-140.

[7] M.I. López and J.M Luna,' Classification via clustering for predicting final marks based on student participation in forums'.

[8] R. Rabbany, M. Takaffoli and O. Zaïane, 'Analyzing participation of students in online courses using social network analysis techniques', Proceedings of Educational Data Mining, 21-30, 2011.

[9] S. K. Yadav, B.K. Bharadwaj and S. Pal, 'Data Mining Applications: A comparative study for predecting students performance', International Journal of Innovative Technology and Creative Engineering, vol 1, No. 12, ISSN:2045-8711, 2011.

[10] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, 'Mining student data using decision trees', International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006.

[11] Minaei-Bidgoli B.; Kashy D.; Kortemeyer G.; 2003. 'Predicting student performance: an application of data mining methods with an educational web-based system'. In Proc. of IEEE Frontiers in Education. Colorado, USA, 13–18.

[12] Available [online] http://en.wikipedia.org/wiki/Data_mining