

Survey on Educational data mining –A Perspective approach in School and College level

Ms.Jose Arockia Rani
Research Scholar, VISTAS
arockiarani89@gmail.com

Mr.R.Balamurugan
Asst.Prof, VISTAS
bmurugan.scs@velsuniv.ac.in

ABSTRACT:

Assessment as a dynamic process produces data that reasonable conclusions are derived by stakeholders for decision making that expectedly impact on student learning outcomes. Data mining is very useful in educational field to find important pattern from the data. The prediction will help the teachers to identify the weak students and help them to improve their performance.

Keywords: Data mining; Education; Students performance; stake holders; Decision Trees.

1.INTRODUCTION:

Performance monitoring involves assessments which serve a vital role in providing information that is geared to help students, teachers, administrators, and policy makers take decisions .Students performance is an essential part in higher learning institutions. This is

because one of the criteria for a high quality university is based on its excellent record of academic achievements [2].

One of the important facts in institution in the rapid growth of educational data main goal of any educational institution is to improve education quality prediction of student's performance in education institution is one way to reach the good quality in education system [3]. It is most of higher learning institutions in Malaysia used the final grades to evaluate student performance. Final grades are based on course structure, assessment mark, final exam score and also extracurricular activities [4].

The proposed systematically review is to support the objective of the study which is the following,

1. To study and identify the gaps in exiting perdition methods.

2. To study and identify the variable used in analyzing student performance.
3. To student existing prediction methods for predicting student performance.

The main objective of this paper is to provide an overview on the data mining techniques that can be used to predict student performance. In this paper also focuses on the prediction algorithm can be used to identify the most important attributes in a student data. We could like this improve student achievement and success more effectively in an efficient way using educational data mining techniques. It could bring the benefits and impacts to students educators and academic institutions. It is provide many techniques being proposed to evaluate student performance. Data mining is one of the most popular techniques to analyze student performance.

DM provides various methods for analysis process which include classification, clustering, and association rule. Classification, which is one of the prediction types classifies data (constructs a pattern) based on the training set and uses the pattern to classify a new data (testing set). Clustering is the process of grouping records in classes that are

similar, and dissimilar to records in other classes. In relationship mining, the goal is to discover the relationship exist between parameters [2, 3]

The next section will focused on the methodology of survey in predicting student performance.

2.Application of educational data mining:

Educational Data Mining (EDM) is an emerging discipline concerned with developing methods for exploring the unique types of data that come from educational settings and using those methods are developing the student. [5]

2.1.Student performance prediction:

A simple student performance and monitoring system based on various data mining techniques was developed with the predictor attributes including students demographic details course average score in 1st to 5th semester overall gain performance.[7]

Predicting student academic performance those student are on risk can be taken to improve their performance . Due to this result of the college can be improve. The Educational Data Mining (EDM) can be used to get the feedback for the

teachers so that teacher can be improve teaching method. There are many data mining techniques can be used for this like classification, clustering, regression, association, rule mining .

2.2 Classification of students:

Student performance gets the knowledge from large data and to discover the parameter the contributed to the student 'success'

The classification techniques are applied to the students data. The study also conducts a three classifications techniques

DT(Decision Tree)NB(Naïve Bayes)and RB(Rule Based) using weka tool. predicted student for three types of categories 1.Good 2.average 3.poor.

The teacher taken to early actions to help and assist the poor and average category students to improve their results.

2.3.Trends in admission:

This study investigated the changing achievement relationships among students based on ages at entry into first grade and ages relative to their classmates. Relative age, class age, sex, parental education, home environment, and type of community were entered in stepwise multiple regression analyses. Data mining provides powerful techniques for

various fields including educational field is rapidly increasing due to the massive amount of student data which can be used to discover valuable pattern pertaining student's learning behavior.

3. Algorithm using in Educational Data Mining

3.1.ALGORIHM OF C4.5:

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan.^[8] C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5is often referred to as a statistical classifier.C4.5 decision tree methods for classifying anomalous and normal activities in a computer network. The k-Means clustering method is first used to partition the training instances into k clusters using Euclidean distance similarity. On each cluster, representing a density region of normal or anomaly instances, we build decision trees using C4.5 decision tree algorithm. The decision tree on each cluster refines the decision boundaries by learning the subgroups within the cluster. To obtain a final conclusion we exploit the results derived from the decision tree on each cluster.

3.2 APRIORI ALGORITHM:

Apriori algorithm is a classical algorithm for mining association rule that considers the time, number of database scans, memory consumption, and the interestingness of the rules. Discover a FIS data mining association algorithm that removes the disadvantages of APRIORI algorithm and is efficient in terms of number of database scan and time. Frequent Item set Mining (FIM) is an important data mining problem which detects frequent item sets scin a transaction database. It plays a fundamental role in many data mining tasks that attempt to find interesting patterns from databases, such as association rules, correlations, sequences, episodes, classifiers, clusters, etc. Many algorithms have been proposed to solve the problem.

3.3. FUZZY LOGIC :

Fuzzy set theory is used in application involving educational assessment and performance as it is regarded as efficient and effective in uncertain situations involving performance assessment. It is known that export Fuzzy scoring systems noted[9]. First, cluster heads are typically selected among all sensor nodes within the network, and

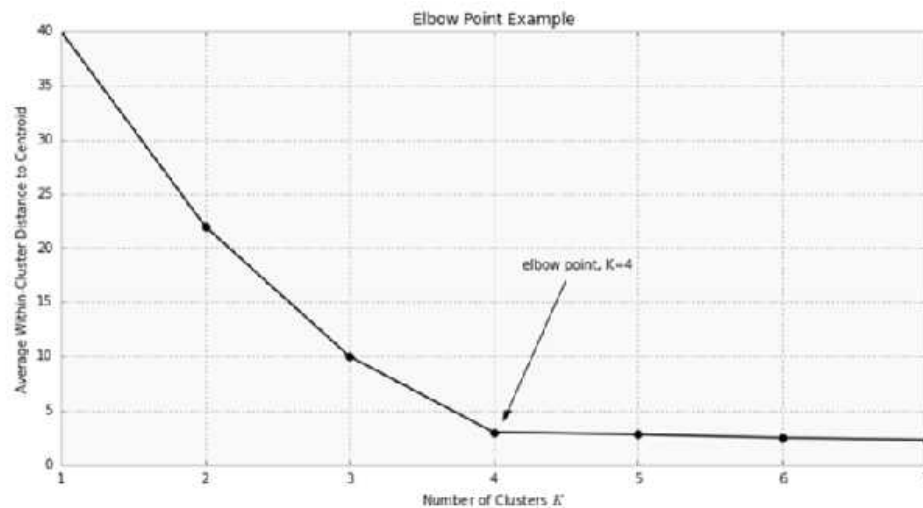
consequently, unbalanced clusters may be generated. Second, the controllable parameters are defined manually. Third, the protocol is not adjusted and tuned based on application specifications

3.4. K-MEANS ALGORITHM:

Clustering is the process of partitioning a group of data points into a small number of clusters. K-means remains the most popular clustering algorithm because of its simplicity. k-means clustering is a method of vector quantization, originally from signal processing, that is popular for cluster analysis in data mining. In general, we have n data points $x_i, i=1...n$ that have to be partitioned in k clusters. The goal is to assign a cluster to each data point. K-means is a clustering method that aims to find the positions $\mu_i, i=1...k$ of the clusters that minimize the *distance* from the data points to the cluster. K-means clustering solves

$$\operatorname{argmin}_c \sum_{i=1}^k \sum_{x \in c_i} d(x, \mu_i) = \operatorname{argmin}_c \sum_{i=1}^k \sum_{x \in c_i} \|x - \mu_i\|^2$$

where c_i is the set of points that belong to cluster i . The K-means clustering uses the square of the Euclidean distance $d(x, \mu_i) = \|x - \mu_i\|^2$.



4.CONCLUSION:

The main objective of educational data mining is to improve teaching-learning process. Predicting major application of Educational Data Mining(EDM) is a student performance such a assist student in improving their academic performance. The result of the study shows that we can predict the graduation performance in computer science student in University program. So using decision tree students' performance can be predicted. The student whose performance is poor. can be warned. The management can take necessary action to improve their performance by giving more attention, taking extra lectures etc. Due to such measures

student performance can be improved. The number of failures can be reduced. Ultimately college results also get improved.

5.FUTURE SCOPE:

In this paper, regular students from schools and colleges taken for surveying purpose. Apart from this, lot of papers available in the area of distance education and online courses are not taken for consideration. The methodology used for that different from that of traditional one. In future ,lot of space available for surveying purpose.

Reference:

1.Council N."Knowing what student knows. The science and design of educational assessment " National

- Academic press- Washington D.C.2001.
2. M. of Education Malaysia, National higher education strategic plan (2015).URL<http://www.moe.gov.my/v/pelan-pembangunan-pendidikan-malaysia-2013-2025>
3. Carlos Márquez Vera, Cristóbal Romero Morales and Sebastián Ventura Soto, "Predicting of school failure and dropout by using data mining techniques", The IEEE Journal of Latin-American Learning Technologies (IEEE-RITA) , Vol. 8, No. 1, pp 7-14, Feb 2013.
4. U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.
5. Nolan J "A Prototype Application of Fuzzy Logic and Expert Systems in Education Assessment". AAAI/IAAI Proceedings pp 1134-1139. 1998
6. Ogor EN. Student academic performance monitoring and evaluation using data mining techniques. Electronics, Robotics and Automotive Mechanics Conference, 2007 Sep. p. 354–9. Crossref
- 7.Quinlan, J. R. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, 1993.
- 8.Jump up^Ian H. Witten; Eibe Frank; Mark A. Hall (2011). "Data Mining: Practical machine learning tools and techniques, 3rd Edition". Morgan Kaufmann, San Francisco.
- 9.A.Dinesh Kumar ,Dr.V.Radhika —A Survey on Predicting Student Performancell, A.Dinesh Kumar et al, (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (5) , 2014, 6147-6149

