Academic Performance Analysis System

Rachana G
Department of Computer Science and Engineering
The National Institute of Engineering
Mysuru, India
rachanagalebhe3010@gmail.com

Rohan Karnik S R
Department of Computer Science and Engineering
The National Institute of Engineering
Mysuru, India
sr.rohanro@gmail.com

Siddhi Singh
Department of Computer Science and Engineering
The National Institute of Engineering
Mysuru, India
siddhi.singh28@gmail.com

Sunidhi
Department of Computer Science and Engineering
The National Institute of Engineering
Mysuru, India
lal.sunidhi21@gmail.com

S Lokesh
Associate Professor, Department of Computer Science and Engineering
The National Institute of Engineering
Mysuru, India
lokesh.sl29@gmail.com

Abstract— Educational Data Mining (EDM) is used to develop methods for exploring the unique types of data emerging from educational institutions and other academic areas, and using these methods to analyze and understand the performance of students in their academic curriculum. It is very important and challenging to analyze the performance of students. Educational Data Mining is used to develop new algorithms and tools in order to analyse the patterns in the academic data . EDM, using these methods, applies techniques from statistics, machine learning, and data mining to analyze data collected from educational institutions. Apriori algorithm and decision tree algorithm can be used to analyze students' performance in different courses and faculty performance in course conduction based on the performance of students. This project is an approach to exploit data mining techniques to determine real time patterns in students' academic data to analyze students' performance as well as teachers' performance. Along with that the performance of teacher's can be improved by conducting various workshops and meetings. In this project we will be analyzing how students' performances vary with respect to different blooms knowledge level mapped questions.

Keywords—Educational Data Mining, Decision Tree, Analytics, Apriori Algorithm

I. INTRODUCTION

Students are the main possessions of various universities in this world. Universities make an effort to develop and deploy extraordinary techniques and different analysis programs to analyse the performance of students. It is equally important to analyse and measure the performance of teachers in-order to complete the development of students in various domains. The proposed work mainly focuses on the development of each student whose analysis is done on daily basis.

Measuring the academic performance of students becomes a critical task since each   students' performance can differ depending upon the factors like their internal skills, social backgrounds and environmental constraints. A teacher can analyse students' performance using data mining models and based on that remedial actions can be taken to improve students' performance. A teacher can identify the risk involved while teaching the course by validating the data mining model by adding a proper strategy which does not yield a good result since all students are made to learn the same strategy applied. In-order to overcome this disadvantage this proposed model is developed and deployed based on real time student's performance data.

Real time data is collected by day to day results answered by students using Clicker, where the questions are added by respective faculty of that subject of the under respective branch of undergoing semester. Further, the mark lists are collected from examination section during the internals and at the end of each semester, to take remedial actions for future batch students for the same course. This proposed model is more helpful for identifying better course conduction strategy for a semester of a particular branch. Each course is based on the students' performance question-wise which are mapped to different blooms knowledge levels which include cognitive domain (knowledge-based), affective domain (emotion-based) and psychomotor domain (action-based). We proposed a model to mine the academic performance of students using association rule learning (apriori algorithm). Mainly we use apriori algorithm in which potential rules are made based on support, confidence and lift. For the analysis and display of results of each student, group of students and indirect analysis of teacher, one of the classification method called decision tree algorithm is used. Our idea is to identify existing patterns of data in previous years' students' performance and maintain database for it.

This paper is divided into five sections. Various related literature survey is presented in the second section. Third section describes methodology implemented using data mining technique and classification technique.

Fourth section display results of proposed model. Fifth section concludes the paper.

II. RELATED WORK

Yiwei Wu, Jiajin Le and Xia Xiang [1] proposed a system to assess student achievement using Factor and Cluster Analysis. The main idea is to determine the fields in which a student is good and the fields in which he is not good. Factor analysis is used to reduce the number of variables in several common factors. The common factor is obtained when a large amount of data is simplified and summarized in information; You can also reduce the number of variables while reproducing the intrinsic relationship between the variables. After this, the group analysis is used to divide the students into different groups according to the common factor score of each student. According to the results of the factorial analysis and the grouped analysis, the students are divided into several categories according to the common factor that can help the students of each category to better understand the concepts in which they are not good and the teachers can use different educational programs for different categories of students.

Ishwank Singh, Abhay Bansal and A Sai Sabitha [2] proposed an analysis of student performance using a grouping algorithm. Data mining techniques are used to analyze the general performance of students based on various parameters such as attendance, grades, projects, internships and other extracurricular activities. The research works use the K-means grouping algorithm to classify students into different groups. It will also help students and teachers focus on improvement strategies by monitoring student performance. The analysis is useful during university admissions and business placements. The K-means clustering algorithm is used because it is easy to implement due to its computational efficiency.

SenolZafer Erdogan eat al used data mining to determine the correlation between the performance of the entrance exam and the success of the students. They made use of cluster analysis and the k-means algorithm. The objects are grouped according to the centroids of the group. The objects are grouped to their nearest centroid. After all the objects are assigned, the calculations are made once more. Using this method, the grouping was performed to determine the success rate of the students [3].A.Parkavi and K. Lakshmi proposed to identify educational environment variable and find a relationship between them to determine the facts that improve the outcome of education provided by the academic institution [4].

Sagardeep Roy and Anchal Garg used the analysing of educational data, learning and the decision tree algorithm are used to understand to student learn, predict academic performance of students and how to improve student learning [5] .

Ms. Tismy Devasia, Ms. Vinushree T P and Mr. Vinayak Hegde used a web application that uses the ingenious Bayesian extraction technique to extract useful information. The result illustrates that the ingenuous Bayesian algorithm provides more precision over other methods such as Regression, Decision Tree, Neural Networks, etc. for comparison and prediction [6].

The main aim of our research work is to identify and analyze the performance of teachers and students in the educational institute. The Apriori algorithm and the decision tree algorithm are well suited and effective, so we have chosen it in our work.

III. METHODOLOGY

The various data mining techniques used in this paper are Apriori Algorithm and Decision Tree Algorithm.

1) Apriori Algorithm: It operates on transactional and relational databases. The frequency of each individual item in the database is identified and extended to a larger item sets seeing those item sets appear adequately often in the database. It uses a "bottom up" technique, using BFS (breadth-first search) and Hash tree structure. Here, frequently occurring subsets are extended one item at a time. This is called candidate generation, and groups of candidates are checked against the data present. The algorithm is terminated after no other successful extensions are discovered.

According to downward closure lemma, the candidate set contains only frequent-length item sets, others are pruned out.

For example, consider the following database:

TABLE 1. STUDENT DATABASE

Name	C	C++	Engg. Math.
Alvin	4	5	2
Lobo	4	4	5
Bob	4	5	2
Alice	4	4	5

The conclusions that can be determined from this database are as follows:

1. All (100%) of students who do well in C also do well in C++.

2. 50% of students who do well in C, C++ also do well in Engg. Math.

3.   50% of students who do not do well in C, C++ also do not do well in Engg. Math.

The main limitation with Apriori algorithm is that during the process of candidate generation, a large number of subsets are generated. The databases are scanned excessively, due to which performance is reduced.

2) Decision Tree Algorithm- This algorithm is in accordance with the supervised learning algorithms, which can also be applied in determining regression or classification/grouping problems.

The central motive behind using Decision Tree is to generate a training model which can be used to predict class or value of target variables by acquiring decision rules understood from provided databases and the former data (training data).

Decision Trees algorithm is very easy to understand, in comparison with other classification algorithms. It solves the problem in the form of trees. The internal nodes of the tree correspond to a property, and the leaf nodes correspond to class labels.

Decision Tree Algorithm Pseudocode:

1.   From the dataset, the best attribute is chosen, and placed as the root of the tree.

2.   The training set is broken down into subsets. The subsets will be constructed in a way such that each subsets will have data with the duplicate value for an attribute.

3.   The above two steps are repeated on every subset till the leaf node is found for all the branches of that tree.
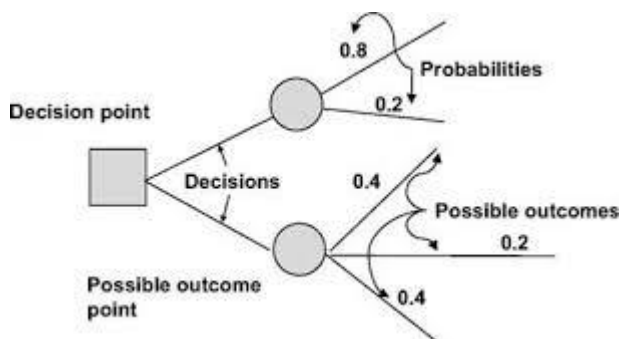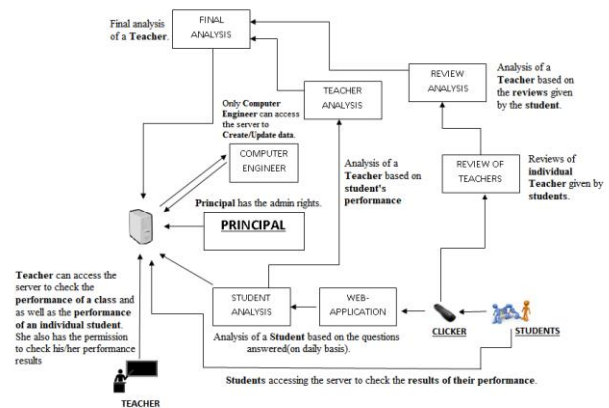

Figure 1. Decision Tree

IV. SYSTEM DETAILS

[1] System Architecture

The above figure conveys that principal who is the admin has authority to access local server for students' performance and teachers' performance. Analysis of students(on daily basis) is done through clicker inputs. Students and teachers have their individual profiles which can be accessed from local server by them. The students are analysed based on their performance in class, for individual subjects/topics. The teachers are analysed in two stages: first on the basis of the reviews given by students, and then on the basis of the performance of his students. Only the computer engineer is given the authority to modify or update the records of teachers and students, after getting a consent from the principal.

[2] Context Flow Diagram
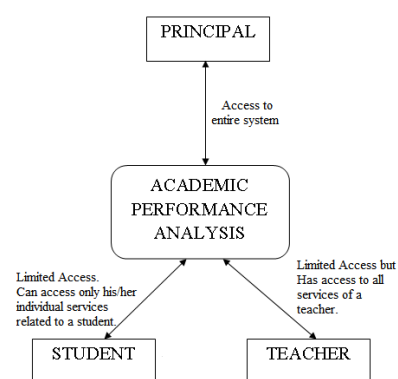
This figure represents the high-level view of system architecture. The principal will have access to all the data of teachers and students but won't be able to alter anything. Teacher will have access to his own profile and his students' profile, but has no authority to change anything. Student can view his profile for marks, attendance, feedback etc.

V. CONCLUSION

Now we have thoroughly understood the drawbacks from literature survey and this paper proposes the designed model and required measures and solutions have been found for them. The student's performance in different blooms level questions can be measured by extracting the required pattern of each student's performance which are frequently extracted and occurred on daily basis from student's performance data-set. By seeking the patterns from student's analysis and performance required measures can be taken by mapping each question to blooms knowledge and thus indirectly, lecturers' analysis can be done through combination of reviews from each student and from the performance of each class.

VI. REFERENCES

[1] Yiwei Wu, Xiaoling Xia and Jiajin Le, "Evaluation of Students' Achievements Based on Factor Analysis and Cluster Analysis", Computer Science and Technology, Donghua University Shanghai, China. IEEE (2017)

[2] Ishwank Singh, A SaiSabitha, Abhay Bansal, "Student Performance Analysis Using Clustering Algorithm", 2016 6th International Conference - Cloud System and Big Data Engineering (Confluence). IEEE

[3] A. M. Shahiri, W. Husain and N. A. Rashid, "A Review on Predicting Student's Performance using Data Mining Techniques," in The Third Information Systems International Conference.

[4] A. Parkavi and K. Lakshmi, "Pattern Analysis of Blooms Knowledge Level Students Performance using Association Rule Mining", in 2017 IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM)

[5] Sagardeep Roy and Anchal Garg, "Analyzing Performance of Students by Using Data Mining Techniques", in 2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)

[6] Ms.Tismy Devasia, Ms.Vinushree T P and Mr.Vinayak Hegde, "Prediction of Students Performance using Educational Data Mining", in 2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)

[7] J.K.Jothi and K.Venkatalakshmi, "Intellectual performance analysis of students by using data mining techniques", International Journal of Innovative Research in Science, Engineering and Technology, vol 3, Special iss 3, March 2014.

[8] G. Dekker, M. Pechenizkiy and J. Vleeshouwers. "Predicting students drop out: A case study." In Educational Data Mining 2009.

[9] V.Ramesh, P.Parkavi and K.Ramar, "Predicting Student Performance:A Statistical and Data Mining Approach," International Journal of Computer Applications, vol. 63, no. 8, 2013