

Predicting the Course Knowledge Level of Students using Data Mining Techniques

Thapaswini P S

Assistant Professor, Computer Science and Engineering Department, VCET, Puttur, India

Abstract: *Data mining techniques can be used to analyze the pattern of data in different fields. Based on the analyses' results recommendations can be made to decision making authorities. The data mining techniques can be used in educational domain to improve the outcome of the educational sectors. In this paper the authors have carried out research study by devising an algorithm and tool to determine students' course knowledge level using data mining techniques. This helps the faculty and students to take necessary remedial actions to improve performance in courses.*

Keywords: Data mining, educational data analysis, association mining

1. Introduction

The large and increasing amount of data enhance the challenge for using data analysis tools to discover regular and irregular pattern of data [1]. Data mining field is growing up as a discipline in which tools can be used for analyzing the data, revealing the undiscovered knowledge and provisions of automated decisions in different domains [2]. One of the domain in which data mining is used is Education domain which is called as Educational data mining [3]. The main purpose of educational organizations is to improve the quality of teaching and learning so the outcome of the education is effective and stake holders are satisfied [4].

In educational domain, the need for analysis and prediction of students' performance is increasing. As Institution is moving with the main goal of outcome based education, it has a plan to improve outcome based education [2]. This can be fulfilled using recommendation system, which can help the students, and faculty by providing valuable suggestions to attain outcome based education. This determination is done using Association rule mining. We study to find whether the determination techniques in data mining help the educational institutions to determine the students' course knowledge level in the courses. This helps to identify the risks of students' performance in course internal and external examinations. We study to find whether the multiple linear regression algorithm and association rule mining helps us to determine the cluster of course knowledge level to which the student is more relevant to.

2. Related Works

Data Mining applications are enormously getting used in Education domain to improve the performance of the students. Researchers and educationalists have worked in different applications of Educational data mining (EDM). Researchers have scientifically investigated data of Educational domain using data mining, thus the outcome of the education will be improved [3] [4].

Data mining is useful in providing students (learners) with recommendations to improve their learning effectively and efficiently [5]. The learners' behavioral pattern can be

mined using data mining and different levels of recommendations can be produced using recommender systems. The recommendation system is designed to provide individualized recommendations to students to improve their learning effectiveness [3]. Students' performance can be predicted based on their past performance. Based on the student activity carried out during learning, their performance in final exams can be predicted. Learning analytics and Educational data mining process can give recommendations to learners to intervene and optimize the outcome of course learning [2] [3]. User knowledge modeling can be used to infer the skills and procedural knowledge level of students. This model has been used to provide recommendations to tutors in giving problems to students. The recommendations are based on the performance of students by attempting to skills performed for solving problems [3]. Using trend analysis, students' performance in exams can be analyzed and the effects of new policies in education can be identified [4].

When multi disciplinary students are enrolling in to a course, it is a difficult for the instructor to determine the students' capacity and their background knowledge. Object oriented model integrated with data mining techniques can be used for predicting the compositions of students background knowledge based on the prerequisites of the course and prior course knowledge of the students. Accordingly, the instructor can conduct the course by providing remedial sessions and supplementary materials of course to improve the students learning [5].

3. Literature Survey

Suresh Kalathur [1] - When students of various backgrounds enroll in a class, it is hard for the instructor to infer the composition of the class and the background level of all the students. This paper discusses an object - oriented model coupled with data mining techniques that will predict the class configuration based on the course prerequisites and the prior courses taken by the students. A Course object captures the static part of a course. Each Semester, there will be Course Offerings capturing the dynamic content of the Course that varies from one offering to another. The Course Syllabus, as usually described in the Course Catalog, is associated with its respective Course. Use case diagram is

Volume 10 Issue 8, August 2021

www.ijsr.net

Licensed Under Creative Commons Attribution CC BY

being developed. The primary use case of the system is to build the student profile for a particular course offering and present the results to the instructor teaching the course. Now that the students' profiles have been determined, the model should provide the feedback to the instructor.

A M Shahiri [2] - In predicting the student performance there is lack in existing system to analyze and monitor the student progress. First, the study on prediction methods are insufficient. Second is due to lack of investigation on the factors affecting students' achievement in particular courses. This section intensively discusses the important factor on predicting the student performance. They are attribute and prediction method. The attributes are widely used in CGPA and internal assessment. This technique is feasible value for future education. In order to build the predictive modeling several tasks which are used are regression and categorization. It will help the educational system to monitor the students' performances in a systematic way. On predicting the performances provides early feedbacks to the students that help them to achieve better in the further tests.

4. Problem Statement

We want to predict the students' knowledge blooming level based on their performance in the continuous internal tests. This prediction is done using association rule mining and regression analysis. For this analysis, java programming tools are used. Based on the students' attempts and scores in questions according to blooms taxonomy level they are mapped, students' performance in future tests can be predicted. If they are performing low in particular blooms knowledge level in a course, recommendations can be given, so students' performance can be improved in future tests and examinations.

We want to determine the overall knowledge level of student by studying how he has performed in different blooms' level mapped questions in different courses. Based on this prediction can be done for future courses about his performance and recommendations can be done so his performance can be improved in different blooms taxonomy knowledge level. To perform the active learning tasks in class, students of all knowledge levels of a course can be grouped. So they can do their collaborative activities better.

5. Proposed Method

Students' performance data in different knowledge level of courses is used to predict students' performances in future exams. This system uses 3 modules.

a) Admin module.

- Registers the faculties to the system.
- Import marks from excel file.
- Processes the student marks and performs the analysis.

b) Analysis module

- Marks from the excel file is taken for processing.
- Analysis to determine the students' level is done.

c) Faculty module

- Faculties will view the analysis of student.
- Recommendations are provided through mails.

Based on that remedial classes can be conducted to improve the knowledge level of students so the overall performance of students in courses also improved.

6. Implementation

Students' performances in Continuous internal examinations of courses are collected.

The collected data contains:

- 1) Each question mapped to blooms taxonomy level
- 2) Maximum marks for each question
- 3) Question wise marks scored by each student

The overall activities are classified as following:

- Collection of Data,
- Preparation of data,
- preprocessing the data,
- Processing the data,
- Results and Analysis.

We have collected student performance data set from our college database. The data set contains the question numbers and their blooms taxonomy level mapped, maximum marks, roll number of students and the marks scored by individual students question wise. The data is collected from various sources, so they have cleaned properly, by filling the missing values. After that the cleaned data is transformed into table form to ensemble it for data mining process. Regression algorithm helps in predicting the academic blooming level performance of students in courses.

Algorithm to Find Best Knowledge Level of Student

- 1) Read each row from the excel file.
- 2) Then count how many questions are from each blooms taxonomy level and maintain in tot_tax [] array.
- 3) Blooms' Taxonomy level for each question is stored in tax [] array.
- 4) Then read the maximum marks for each question, store it in max [] array.
- 5) For each student perform the following
 - a) Initialize skills [] array to zero.
 - b) Read marks scored by student in each question one by one from each column
 - c) Do the following for each mark from a cell in excel sheet of row rx and column cx
 - Check among the 3 revised blooms taxonomy, which taxonomy the question belong to i is the blooms level, where $1 \leq i \leq 3$
 - Find out the percentage of mark scored in each question as following and add it to the skills [i].
skills [i]=skills [i]+mark *100/max [cx]
 - d) After calculating skills of each student based on marks scored in each question, find out the average skill of student in each knowledge level as following.
 - For each revised blooms level, k from 1 to 3 do the following for each student:
 1. Skills [k]=skills [k]/total_tax [k]
 - ii. To identify best_skill of each student, do the following:
 - iii. Initialize top_percentage as 0.

1. For each blooms level k from 1 to 3, do the following for each student:
 - a. If $top_percentage \leq skills[k]$
 - i. $Top_percentage = skills[k]$
 - ii. $Best_skill = k$
 - e) Display/store the students' best skill identified as the following:
 - If $best_Skill == 1$, then student is good in remembering skill
 - If $best_skill == 2$, then student is good in understanding skill
 - If $best_skill == 3$, then student is good in applying skill.

Once the levels in various internals are known, the levels are being compared to check if there is improvement in the further tests.

7. Conclusion

In this paper we determine the course knowledge level of students using associative rule mining to perform the collaborative active learning tasks. This helps to identify the risks of students' performance in course internal and external examinations by identifying in which knowledge level of course students are performing poor, based on that remedial action can be taken to determine the cluster of course knowledge level to which the student is more relevant to.

References

- [1] S. Kalathur, "An Object - Oriented Framework for Predicting Student Competency, " in Proceedings of SERP'06 Las Vegas, 2006.
- [2] 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.
- [3] R. A. Huebner, "A survey of educational data mining research, " Research in Higher Education Journal.
- [4] K. Elissa, "Title of paper if known, " unpublished.
- [5] R. Nicole, "Title of paper with only first word capitalized, " J. Name Stand. Abbrev., in press.
- [6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto - optical media and plastic substrate interface, " IEEE Transl. J. Magn. Japan, vol.2, pp.740 - 741, August 1987 [Digests 9th Annual Conf. Magnetism Japan, p.301, 1982].
- [7] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.