

Mining Association Rules in Student Assessment Data

Charanjit Bamrah¹, Minakshi Bhandari², Nirali Maniar³, Prof. Vandana Munde⁴

IT Department, Mumbai University, Thakur College, Kandivali (East), India

Abstract: The objective of the educational institution that is producing good results in their academic exams can be achieved by using the data mining techniques which can be applied to analyze the performance of the students and to impart the quality of education in the educational institutions. Data mining is used to extract meaningful information and to develop relationships among variables stored in large data set. In this case, Apriori algorithm is used which extracts the set of rules, specific to each class and analyzes the given data to classify the student based on their performance in academics. Students are classified based on their involvement in doing assignment, internal assessment tests, attendance etc., which helps to analyze the performance of the student based on the pattern extracted from the educational database.

Keywords: Education, Data Mining, Association Rules

I. INTRODUCTION

Education is an essential element for the betterment and progress of a country. It enables the people of a country civilized and well mannered. Today the important challenge that education faces, is reaching a stage to facilitate the colleges in having more efficient, effective and accurate educational processes. To date, educational organizations are placed in a very high competitive environment and are aiming to get more competitive advantages over the other competitors. To remain competitiveness among educational field, these organizations need deep and enough knowledge for a better assessment, evaluation, planning, and decision-making. The required knowledge cannot be gained from the tailor made software used now a days.

Data mining incorporates a multitude of techniques from a variety of fields including databases, statistics, data visualization, machine learning and others. The data mining technology can discover the hidden patterns, associations, and anomalies from educational data. This knowledge can improve the decision making processes in higher educational systems. Data mining is considered as the most suited technology appropriate in giving additional insight into the lecturer, student, alumni, manager, and other educational staff behavior and acting as an active automated assistant in helping them for making better decisions on their educational activities.

The data mining techniques can help the institutes in extracting patterns like students having similar characteristics, Association of students' attitude with performance, what factors will attract meritorious students and so on. The past several decades have witnessed a rapid growth in the use of data and knowledge mining as a means by which academic institutions extract useful hidden information in the student result repositories in order to improve students' learning processes.

The main objective of this paper is to use data mining methodologies to study students' performance in their

courses. In this research, we will be using Association rules discovery techniques.

II. LITERATURE SURVEY

2.1 Data Mining

Data mining has attracted a great deal of attention in the information industry and in society as a whole in recent years, due to the wide availability of huge amounts of data and the imminent need for turning such data into useful information and knowledge. Data mining can be viewed as a result of the natural evolution of information technology. Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both.

It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining is a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions.

The first and simplest analytical step in data mining is to describe the data — summarize its statistical attributes (such as means and standard deviations), visually review it using charts and graphs, and Look for potentially meaningful links among variables (such as values that often occur together). But data description alone cannot provide an action plan.

You must build a predictive model based on patterns determined from known results, then test that model on results outside the original sample. The diagram given below depicts data mining process as an essential step in knowledge discovery and the role it plays in the entire knowledge discovery process.

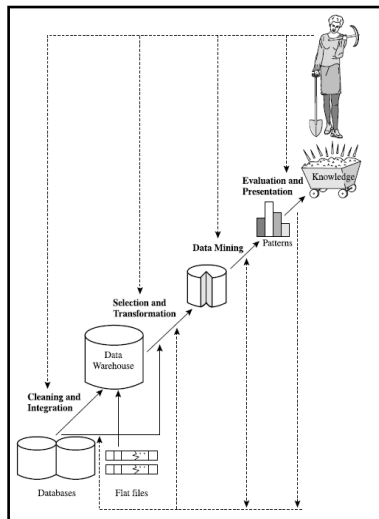


Figure 1

2.2 Association Rules

In data mining, association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness. It aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories.

Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem is usually decomposed into two sub problems. One is to find those item sets whose occurrences exceed a predefined threshold in the database; those item sets are called frequent or large item sets. The second problem is to generate association rules from those large item sets with the constraints of minimal confidence.

Association Rules will allow to find out rules of the type: If A then B where A and B can be particular items, values, words, etc. An association rule is composed of two item sets:

1. Antecedent or Left-Hand Side (LHS)
2. Consequent or Right-Hand Side (RHS)

Many algorithms for generating association rules were presented over time. Some of the popular known algorithms are Apriori, Eclat and FP-Growth which is used to mine frequent item sets. The mining exploits infrequent data and high lowest support and high lowest confidence values.

2.3 Apriori Algorithm

Apriori is a classic algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight

general trends in the database: this has applications in domains such as market basket analysis.

The rules generated by Apriori algorithm makes it easier for the user to understand and further apply the result. Employed the association rule method specifically Apriori algorithm to identifying novel, unpredicted and exciting samples in hospital infection control. Another study by employed Apriori algorithm to generate the frequent item sets and designed the model for economic forecasting, presented their methods on modeling and inferring user's intention via data. Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time.

III. SCOPE OF THE PROJECT

Higher education, throughout the world is delivered through universities, colleges affiliated to various universities and some other recognized academic institutes.

Today one of the biggest challenges, the educational institutions face, is the explosive growth of educational data and to use this data to improve the quality of managerial decisions to deliver quality education. This system hopes to improve the quality of education by analyzing the data and discover the factors that affect the academic result so as to increase success chances of students. In this perspective we use association rules discovery techniques.

IV. PROBLEM DEFINITION

In our existing system, a group of students are allotted to one professor to observe the performance of those students. This process takes place in the form of keeping records manually. However, this system takes up a lot of time and efforts to maintain and analyse the records. Therefore to overcome this problem, a new system is to be developed which not only maintains the data in the database but also helps in analyzing the performance of students based on the records. The system must allow the professors to access the data of the students, retrieve it and predict the performance of the students.

Level 0 DFD :

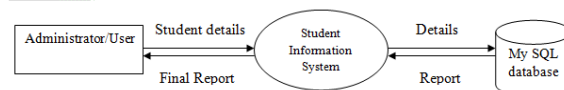


Figure 2: Data Flow Diagram Level 0

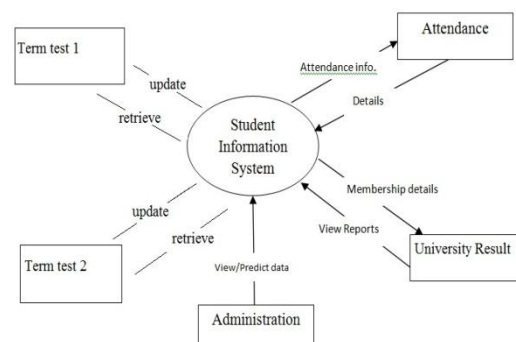


Figure 3: Data Flow Diagram Level 1

V. IMPLEMENTATION

In the college the overall performance of a student is determined by internal assessment as well as external assessment. Internal assessment is made on the bases of a student's term test marks, attendance. While at the same time external assessment of a student based on marks scored in university exam. The proposed model helps to predict the students about poor, average and excellent based on class performance as well as class attendance from the generated rules.

5.1 Data Mining Process

The steps of data mining process are as follows:

5.1.1 Data Selection

The data have been generated by different reports of Term Tests, Attendance, and University Results. The initial data contains the details gathered from a number of students with 5 listed attributes which include Rollno, Term Test1 marks, TermTest2 marks, Attendance, University Results. The data contains numeric value. The target is represented as analysis report. The analysis report was grouped according to three categories (Excellent, Average and Poor). The selected attributes are as mentioned above. The data were then processed for generating rules.

5.1.2 Data Transformation

In order to use this data for generating rules, the data had to be transformed. Another set of records with same attributes are used but in this case the numeric values are converted into string values (poor, average, excellent) according to the given criteria

- ≤ 50 : Poor
- 51 to 75: Average
- 76 to 100 : Excellent

The data was then ready to be mined using association rules.

5.2.3. Rule Generation

Rule Generation consists of two parts:

5.2.3.1 Generating frequent item sets

The frequent item sets generated for the sample data are

1. [Excellent, Excellent, Average, Excellent]
2. [Average, Excellent, Average, Average]

5.2.3.2 Generating Association rules

The association rules are generated using Apriori algorithm.

```

L1 = {frequent items};
for (k= 2; Lk-1 !=∅; k++) do begin
  Ck = candidates generated from Lk-1
  for each transaction t in database do
    The count that are enclosed in t of all candidates in
    Ck is to be incremented
  Lk = candidates in Ck with min_sup
end
return  $\cup_k L_k$ ;

```

Figure 4

The Association rules generated are:

- Average^Excellent^Average => Average
- Excellent^Excellent^Average => Excellent

VI. CONCLUSION

The paper analyzed the potential use of one of the data mining technique called association rule mining in enhancing the quality of students performances. The extracted rules helps to predict the performance of the students and it identify poor, good and excellent students. The performance report of the student also helps to improve the result of the student. This system also helps to identify those students which need special attention to reduce fail ratio and taking appropriate action for the next semester examination.

REFERENCES

- [1] D. Magdalene Delighta Angelina, "Association Rule Generation for Student Performance Analysis using Apriori Algorithm", The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 1, No. 1, March-April 2013, p12-16
- [2] Brijesh Kumar Baradwaj & Saurabh Pal, "Mining Educational Data to analyse Student's Performance", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011, p 63-69
- [3] Varun Kumar, Anupama Chadha, "Mining Association Rules in Student Assessment Data", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September 2012, p211-215
- [4] <http://www.onlinephpstudy.com>
- [5] <http://www.w3schools.com/php/>
- [6] <http://www.mysql.com/>
- [7] <http://in3.php.net/>
- [8] <http://xampp.en.softonic.com/>
- [9]