



International Journal of Control Theory and Applications

ISSN : 0974-5572

© International Science Press

Volume 10 • Number 18 • 2017

Data Mining for Student's Trends Analysis Using Apriori Algorithm

Atta-ur-Rahman¹ and Sujata Dash²

¹ CCSIT, University of Dammam Saudi Arabia, Email: aaurrehman@uod.edu.sa

² North Orissa University Baripada, Odisha, India, Email: sujata238dash@gmail.com

Abstract: Main objective of the educational institutions is to give high quality education to its students and provide competent IT professionals to the market. One way to improve quality in higher education system is by discovering patterns for prediction about the enrollment of students in a particular program, prediction about students' performance and so on. On the other hand it is also important to choose appropriate field for student at enrolment time. Advertise in the area where the education about particular program not well known. Knowledge Discovery in databases (KDD) and Data Mining techniques can be used for extracting useful knowledge from data. In this research, the association rule mining technique is used to find hidden patterns and evaluate students' performance and trends. Apriori algorithm is used for finding associations among attributes. The proposed decision support system has been tested by embedding it in Course Management System (CMS). The system efficiency has been tested upon the data of 3500 students.

Keywords: Association Rule Mining; Apriori Algorithm; Trend Analysis

1. INTRODUCTION

In education sector, trend analysis provides useful help to such requirements. The trend analysis regarding past education of the student, can help them to explore a better field. If the student has had poor grades in past, then what would be the better option for next? Whether he be suggested to adopt computer science or not? Through data mining techniques, the multiple trends of students towards education and the study can be analyzed.

Data Mining is the major process of discovering knowledge. It focusing upon different Methodology for extracting useful knowledge from data and there are several useful KDD tools to extracting the knowledge. This knowledge can be used to increase the quality of education. Data mining can be used for decision making in educational system and for student's trend mining. [1]

Association rules mining is one of the data mining technique which is expected to be very useful in applications. It describes the relationship between support, confidence and interestingness. The support and confidence are usually referred as interestingness measures of an association rule. Association rule mining is the process of finding all the association rules with the condition of minimum support and minimum confidence.

Initially, the support and confidence values are computed for all the rules. Association rules mining was proposed by Agarwal. Many algorithms for generating association rules were presented over time. Some of the popular known algorithms are Apriori and FP-Growth which is used to mine frequent item sets. [2]

The main objective of this paper is to use data mining methodologies to study student's performance in the courses and analysis of the trends. Whether the student's location affects the performance or not? The data analysis exhibits that from rural area quite less number of the student enrolls in computer science degree; hence their grades are very good, mostly passed at grades B or C. Therefore, proper advertisement and counseling be provided to the students of the rural area. On the basis of above mentioned facts, the student's educational history and performance in computer science field. The system can be a touchstone and a proper career guide for the students interested in computer sciences.

Rest of the paper is organized as follows, section 2 is about the related work, section 3 narrates the material and methods, section 4 explains the research methodology, and section 5 is about the implementation and results while section 6 concludes the paper.

2. RELATED WORK

There are increasing research interests in using data mining in education. This new emerging field, called Educational Data Mining, concerns with developing methods that discover knowledge from data originating from educational environments. To be specific, the discussion focuses on data mining, data mining process, how information retrieves from different sources for mining process, review of different data mining techniques, work done in different problems using different mining techniques and finding the problem that not yet discuss.

Asif, M. and J. Ahmad concluded in the light of the results that Fp-Tree algorithm is better than the Apriori for large data sets for finding associations. But on the other hand, we cannot fully disregard the Apriori because for small datasets it works efficiently. These two algorithms have shown great results in the domain like software engineering. A decision support system based on Apriori and FP-Tree will boost the efficiency of the decision making process in software engineering. [3]

As described by Alaa el-Halees, Data Mining is a technique that is useful in educational sector, it can enlarge user's understanding the process of learning. Educational data mining has been used to improve performance of graduate students, and solve the issue of lower grades of the students. Data is preprocessed and then mining technique were applied to find out various association, clustering, association and outlier detection rules. Rule Induction and Naive Bayesian classifier methods have been used to predict the grade of the student. Clustering has been done using k-mean algorithm. Final results were found quite useful for the institute.[4]

Bhardwaj and Pal studied performance of student of different colleges and universities. 300 students were selected for this purpose. After applying classification method of Bayesian classification, different factors such as grade of student in senior secondary exam, teaching medium, location of student, qualification of mother, family income and some other factors highly effect performance of student. It was concluded that mother's low education affects the student's performance. Otherwise students have good grades. Performance of students also effects on living area and family factor as well. Home tuition is also an important factor considered in student's good performance. [5]

Pandey and Pal collected data from different colleges and universities and carried out study on performance bases of almost six hundreds plus students. By using Classification techniques on language, background qualification and category, useful results were found. Background qualification badly affects the upcoming student's degree performance. Student having no computer study background may or may not be affect student performance. We can easily take the decision on the mining results. If a student is done its past education like inter or graduate in bad grade it also may influence the next field result you are going to choose. Language is

also a factor on the basis we take some decision because it happened that large number of student face difficulty with language. [6]

Hijazi and Naqvi performed a study on performance of students on a data of almost three hundreds plus students including two hundreds and twenty-five plus males and seventy five plus female from different colleges of Punjab University. It was stated that "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed. Using regression analysis techniques, results concluded that family income of student and education of mother highly effects student's performance in academics. [7]

As stated by Al-Radaideh et al. data mining techniques have been applied, classifications techniques particularly to improve higher education system quality. Rules extracted using classification methods specially decision tree. Algorithm helps students to find out grades of courses. Mining techniques has been utilized to provide proper classification approach to students to support their basic planning. System has been validated by proper methods to ensure accuracy in the model. It was concluded that accuracy of the proposed model was 87.9%. Two hundred and eighteen students were successfully classified out of two hundred and forty eight students. [8]

According Baker, R. S. J. D., and K. Yacef following goals are set by researchers for educational data mining.

- To predict future learning patterns of student.
- Improving domain models and discovers models that determine learning contents.
- Study different effects of educational support that can be achieved by software.

As basic Apriori Algorithm can't be used for education data mining due previously mentioned shortcoming, various authors have suggested variant in Apriori algorithm to address the mentioned shortcomings. [9]

Algorithm of Association Mining Rule by Xin-hua Zhu, Ya qiong Deng and Qing-ling Zeng stated that they have taken computer cultural foundation courses grades as an example. The proposed algorithm determines grades of examinations held for a specific course. Compared with Apriori algorithm, associations have been found between scores and chapters, chapters and colleges at same final values. [10]

According to Romero, C, Ventura, S. and Garcia, E data mining in course management systems, data can be collected form historical and operational data reside in the databases of educational institutes. The student data can be personal or academic. Also it can be collected from e-learning systems which have a large amount of information used by most institutes. Educational data mining uses many techniques such as decision tree, rule induction, neural networks, k-nearest neighbor, naïve Bayesian and many others. By using these techniques, many kinds of knowledge can be discovered such as association rules, classifications and clustering. [11]

In [12] and [13] authors proposed a semantic network based intelligent search in the religious repository. Atta-ur-Rahman (2013) [14] proposed a teacher assessment and profiling system (TAPS) using a fuzzy rule based system (FRBS) and Apriori Algorithm. In this paper, FRBS was used to assess the user while Apriori Algorithm was used as profiling technique. Significance of the proposed scheme was shown through examples. System can help organization to sort out weaknesses of faculty members and to increase the quality of education provided to the students. Association and relationship among data items has been found using Apriori algorithm. Atta-ur-Rahman (2013) [15] proposed a fuzzy rule based system (FRBS) for network user's behavior classification. The proposed scheme makes use of the users' various logs like web, machine and network logs to classify him/her behaviorally. Simulation results are presented to signify the effectiveness of proposed scheme. Web, databases, hardware and other application logs provide information of user to classify. System helps in securing privacy of system/user and also can secure network from un-authorized and criminal mind users. Proposes work will help organization to make precautionary measures to secure the system.

3. MATERIAL AND METHODS

3.1. Data mining process

The workflow of the system consists of three phases: Data Integration Phase, Mining Phase and Prediction and analysis Phase. Dataset will be formulated for this purpose.

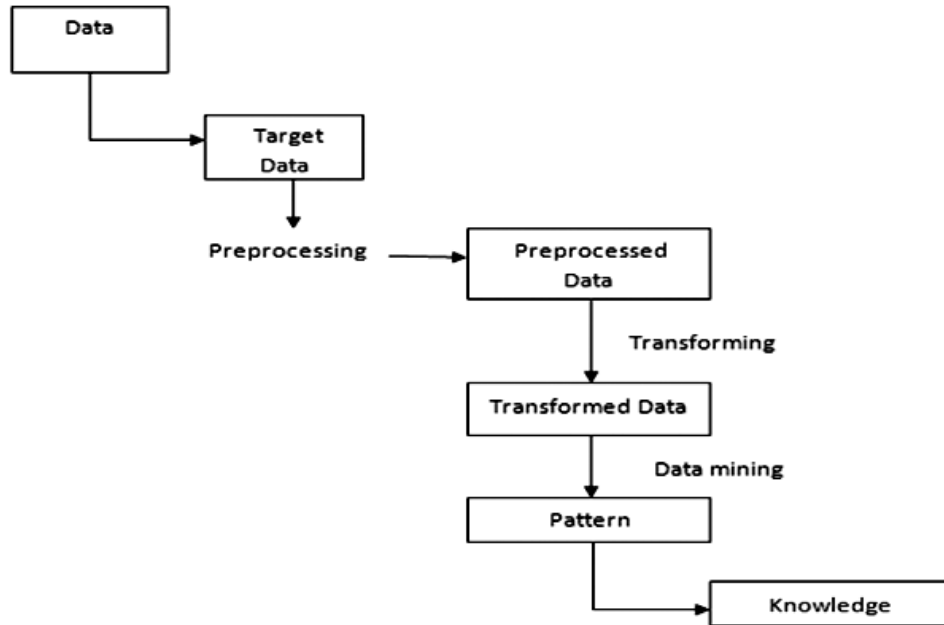


Figure 1: Data Mining Process

3.2. Data Extraction

In educational system, the information about students’ past education subjects, marks, board or university, grade, location and gender is taken on the admission time. Data is saved on CMS. Student’s result data according to the course is maintained semester wise. Assignments marks, quiz marks as well as practical score are also entered with respect to subjects. CGPA is calculated from quality points and credit hours. Some of the Meta information of CMS is given below.

3.3. Data collection and preparation

In our study, we have considered student’s data of Barani institute of Information Technology (BIIT) of Rawalpindi, Pakistan. On the basis of the data collected some attributes have been considered to analysis student’s trends. The attributes used for this purpose:

Table 1
Attributes and Description

Attribute	Description	Values
Discipline	Discipline of students enrolled in institute	BCS,BIT,MCS,MIT
Past Education	Education history of students enrolled in institute	Arts, Science, Commerce, Computer
Past Education Grade	Past education performance	A/B/C or 1 st ,2 nd div.
Degree Req.	CGP req. to complete the degree	Min 2.5
Address	Where students belongs	String containing full address

3.4. Data preprocess

One of the important steps of Data Mining process is data pre-processing. Data Pre-processing is used to identify the missing values, noisy data and irrelevant and redundant information from dataset. We use the data in percentage or in characters to describe grades for the above mentioned attributes.

Table 2
Attributes and Values

Attributes	Values
Past Education	Arts, Science, Commerce, Computer
Past Education Grade	A/B/C
Discipline	BCS,BIT,MCS,MIT
Degree Grade	A/ B/C/I/F
Location	Urban/Rural

- Past education means the education that student completed before the degree started. That is academic record.
- Past education grade consider as A/B/C calculated from the marks obtained or grade given by board.
- Current degree program in which student is enrolled.
- Address in the form of full address string existing in database. It is transformed in Urban/Rural category.
- Degree grade of student is considered by the GPA of student. Grade is like A/B/I/F is transformed.

4. METHODOLOGY

In this paper, tool named Weka is used for apply on dataset. The data used for analysis purpose in the form of CSV file. Data format problem is not accruing in Weka tool for result generation. There are many data mining

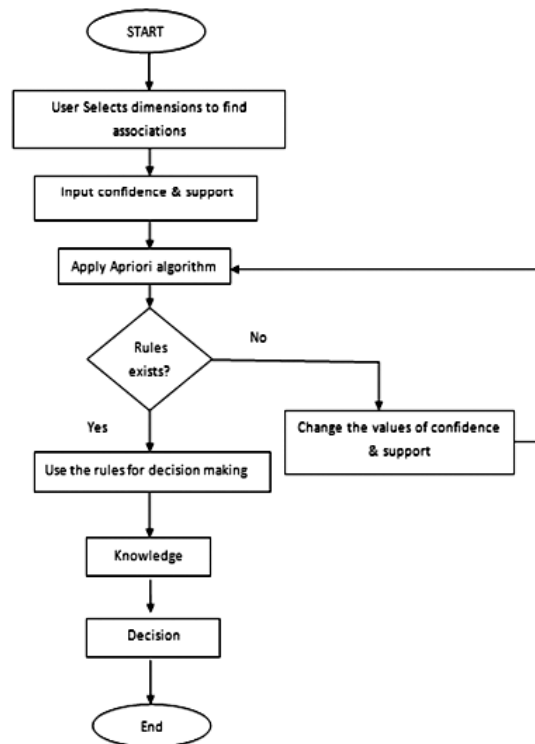


Figure 2: Flow Chart

algorithms for data pre-processing and analysis are exist in this tool. It is very rich tool that is applied for finding interesting relation and hidden pattern in dataset. This tool provides all support for mining algorithm. Apriori algorithm with min support and confidence value with lift value can easily used to achieve mining results.

4.1. Apriori algorithm

In 1994 Agarwal introduced apriori algorithm. The Apriori Algorithm is used for mining frequent itemsets for association rules. Apriori uses a “bottom up” approach, where frequent subsets are extended one item at a time. A step is known as candidate generation and groups of candidates as well. [11].

4.2. Association rule mining

Association rule mining is usually split up into two separate steps:

- First, minimum support is applied to find all frequent itemsets in a database.
- Second, these frequent item sets and the minimum confidence constraint are used to form rules.
- Want to find a group of items that tend to occur together frequently.
- The association rules are often written as $X \rightarrow Y$ meaning that whenever X appears Y also tends to appear. X and Y may be single items or sets of items.

Support indicates the frequency of the pattern. A minimum support is necessary if an association is going to be of some business value.

$$\text{Support}(X) = \text{no. of transactions which contain the itemset } X / \text{total no. of transactions}$$

Confidence denotes the strength of the association between X and Y.

$$\text{Confidence}(X \rightarrow Y) = \text{Support}(XUY) / \text{Support}(X)$$

5. RESULT AND DISCUSSION

The dataset of students has been obtained from BIIT department of Arid Agriculture University, Rawalpindi. In this dataset we find various association rules among attributes. Dataset of 3500 students has been used to analyze final degree grades.

5.1. Final degree grade analysis

5.1.1. According to past education and discipline

Data of past education of students has been collected on the admission time with the division or grade. Past education is normally in commerce, arts, science or computer science area. Data values exited in CGPA have been transformed into grades: A/B/C and F.

Table 3
Attributes and Description

<i>Attribute</i>	<i>Description</i>
Past Education	Like science, commerce, arts etc
Discipline	Final degree program
Degree_ Grade	Grade in final degree

These attributes has been used to show the student's tendency associated with university result. Number of association rules is shown below:

Minimum support: 5% (0.05)

Minimum confidence: 30% (0.3)

PAST_EDU=SCIENCE DEGREE_GRADE=B

==> DISCIPLINE=MCS conf:(0.84)

> PAST_EDU=COMPUTER DEGREE_GRADE=B ==> DISCIPLINE=MCS conf:(0.72)

> PAST_EDU=COMPUTER DEGREE_GRADE=C ==> DISCIPLINE=MCS conf:(0.71)

> PAST_EDU=COMMERCE DEGREE_GRADE=B ==> DISCIPLINE=MIT conf:(0.51)

> DISCIPLINE=MIT DEGREE_GRADE=B

==> PAST_EDU=COMMERCE conf:(0.45)

> PAST_EDU=ARTS

==> DEGREE_GRADE=C conf:(0.37)

> DISCIPLINE=BCS DEGREE_GRADE=B ==>PAST_EDU=SCIENCE conf:(0.9)

> DISCIPLINE=BCS DEGREE_GRADE=A

==> PAST_EDU=SCIENCE conf:(0.9)

The above findings (rules) have been taken from the past education of the students their grades in computer science subjects. The rules express that the students, who already had learned computer in the early stages (like ICS) have won brilliant grades. Whereas, the students had commerce background, mostly enrolled in MIT and passed the degrees in B or C grades. Students, who had science background, have gained brilliant grades in almost every discipline. Furthermore, on the basis of findings students with commerce background can be encouraged to enroll in MIT discipline.

5.1.2. According to location and discipline

Location is extracted from the address string of student that is a combination of several values. Two values, rural and urban are considered to be analyzed. Degree grade is transforming into A/B/C/F/I categories. Discipline means the degree program in which student is currently enrolled.

Table 4
Attributes and Description

<i>Attribute</i>	<i>Description</i>
Location	Urban/Rural
Degree_Grade	Grade in final degree
Discipline	Students' final Degree

These attributes show the student's trend with number of association rules that we found:

Minimum support: 10% (0.1)

Minimum confidence:50% (0.5)

> DISCIPLINE=BCS==>Location=RURAL
conf:(1)

> DEGREE_GRADE=B==>Location=RURAL conf:(1)

- DEGREE_GRADE=C==>Location=RURAL conf:(1)
- DISCIPLINE=MCS==>Location=RURAL
conf:(1)
- DISCIPLINE=BIT==> Location=RURAL
conf:(1)
- DEGREE_GRADE=A==>Location=RURAL conf:(1)
- DISCIPLINE=MIT==>Location=RURAL conf:(1)
- DISCIPLINE=BCS DEGREE_GRADE=B ==>
Location=RURAL 55 conf:(1)

The above results expose that the students coming from rural areas enrolled in almost every discipline and passed the degrees in good grades. Upon these evidences, the computer science should be promoted in rural areas of the country and the students are persuaded through proper guidance and effective advertisement.

5.1.3. Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

6. CONCLUSION

This paper presents the application of data mining techniques for students' trend analysis in different aspects. The data being mined was the enrollment data of the students along with the students' grades. This research actually connects the students' background with the degree and grades outcomes of the students. Trend analysis in student's educational history with respect to their performance in current degree. Students having non-computer science educational history passed their computer science in B or C grade. This trend is helpful for administration to guide new students. The advertisement should be made in the rural areas to promote computer science on the basis of analysis.

ACKNOWLEDGEMENT

My sincere thanks to all the people who have contributed in any way in carrying out this work.

REFERENCES

- [1] R. Agrawal, Christos Faloutsos & Arun N. Swami (1994), "Efficient Similarity Search in Sequence Databases", Proceedings of the 4th International Conference of Foundations of Data Organization and Algorithms, Pp. 69–84.
- [2] Magdalene, D (2013) Association Rule Generation for Student Performance Analysis using Apriori Algorithm The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), Vol. 1, No. 1
- [3] Asif, M. and J. Ahmad (2015) Data Analysis of Effectiveness of Apriori and Frequent Pattern Tree Algorithm in Software Engineering Data Mining *International Conference on Intelligent Systems, Modelling and Simulation, Vol 1, IEEE.*
- [4] Abu, T. and M. El-Halees (2012) Mining Educational Data to Improve Students' Performance: A Case Study International Journal of Information and Communication Technology Research, vol. 2, no. 2.
- [5] Bharadwaj, B.K. and S. Pal (2011a) Mining Educational Data to Analyze Students' Performance International Journal of Advance Computer Science and Applications (IJACSA), vol. 2.
- [6] Kumar, S., U. Pandey and S. Pal (2011b) Data Mining : A prediction of performer or underperformer using classification (IJCSIT) International Journal of Computer Science and Information Technologies

- [7] Hijaz, T.S. and S.R. Naqvi (2006) Factors Affecting Students' Performance: A Case of Private Colleges Bangladesh e-Journal of Sociology. vol. 3
- [8] Al-Radaideh, Q.A., A. Al-Ananbeh and E.M. Al-Shawakfa (2011) A Classification Model For Predicting The Suitable Study Track For School Students (IJRRAS)
- [9] Baker, R. S. J. D., and K. Yacef (2009) The State of Educational Data Mining in 2009: A Review and Future Visions Journal of Educational Data Mining, Article 1, Vol 1, No 1
- [10] Xin-hua Zhu, Ya qiong Deng, Qing-ling Zeng (2009) Application of Mixed-Weighted-Association-Rules-Based Data Mining Technology in College Examination grades Analysis College of Computer Science and Information Technology Guangxi Normal University, Guilin, Guangxi 541004, China
- [11] Romero, C, Ventura, S. and Garcia, E. (2008) Data mining in course management systems: Moodle case study and tutorial Department of Computer Sciences and Numerical Analysis, University of Córdoba, 14071 Córdoba, Spain
- [12] Shahzadi, N., Atta-ur-Rahman and A. Shaheen, "Semantic Network based Semantic Search of Religious Repository", *International Journal of Computer Applications (IJCA)*, Vol. 36, No. 9, pp. 1-5, December, 2011.
- [13] Shahzadi, N., Atta-ur-rahman and M.J. Sawar, "Semantic Network based Classifier of Holy Quran", *International Journal of Computer Applications (IJCA)*, Vol. 39, No. 5: pp. 43-47, February, 2012.
- [14] Atta-ur-rahman, "Teacher Assessment and Profiling using Fuzzy Rule based System and Apriori Algorithm", *International Journal of Computer Applications (IJCA)*, Vol. 65, No. 5, pp. 22-28, 2013.
- [15] Atta-ur-Rahman, D.N. Zaidi, M.H. Salam and S. Jamil, "User Behavior Classification using Fuzzy Rule Based System", *13th International Conference on Hybrid Intelligent Systems (HIS'13)*, pp. 118-123, December 04-06, 2013, Tunisia.