Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Rule-Based Classification of Student Performance using Apriori Method

Sushil Kumar Verma[1], Dr. Dinesh Kumar Sahu[2], Dr. Varsha Namdeo[3]

[1]Research Scholar- M.Tech., Dept. of IT, SRK University, M.P, India
[2]Associate Professor, Dept. of IT, SRK University, M.P, India
[3]Professor, Dept. of IT, SRK University, M.P, India

Abstract

Researchers have increasingly utilized data mining techniques to uncover hidden patterns and extract meaningful insights from raw data. In recent years, data mining has been effectively applied in various domains, including education, to enhance the quality of learning and institutional performance. Within the educational context, data mining facilitates the categorization of student performance by analysing academic data stored in institutional databases. Student performance is categorized based on multiple academic parameters. The categorization of student performance enables educators to identify academically weak students early in the semester.

Keywords: Data mining, Educational data mining, Association rule mining, Apriori algorithm, Data Transform.

1. Introduction

Data mining encompasses a variety of algorithms and techniques, including clustering, classification, regression, prediction, and association rule mining—each serving distinct analytical purposes. Among these, **association rule mining** plays a crucial role in uncovering hidden patterns and relationships within datasets that can support informed decision-making [1]. Association rules, typically expressed in an intuitive *if–then* format, offer an easily interpretable form of knowledge representation.

The **Apriori algorithm** is one of the most widely used methods for mining association rules and has been successfully applied across numerous domains [1, 2, 3]. However, **educational data mining (EDM)** differs from traditional data mining disciplines in both methodology and outcomes, as it focuses on extracting insights that are directly relevant to academic environments. Mining association rules from educational data has emerged as a vital research area, given that the resulting patterns and rules can significantly influence institutional effectiveness. Numerous studies have demonstrated that applying association rule mining within education can assist instructors and students in discovering meaningful knowledge, enhancing learning outcomes, and reducing failure rates.

Association rule mining is user-centric because its objective is find out the interesting rules from which knowledge can be derived [4, 5, 15]. Interestingness of rules means that they are novel, externally significant, unexpected, nontrivial, and actionable. An association mining aids the process in order to facilitate the process, filter and exist the rules for further interpretation.

2. Related work

- Developed a data mining model aimed at identifying similar patterns within educational data and predicting student performance.
- Analysed multiple factors influencing student performance, such as demographic, academic, psychological, and socio-economic variables. Using association rule mining**,** they constructed an analytical model to identify key performance determinants.
- Proposed a data collection and storage model based on student responses to specific questions. They applied association mining to these datasets to uncover individual learning patterns among students.
- investigated complex educational systems and demonstrated how machine learning techniques**,** particularly association rule mining, could identify sensitive learning patterns. Their work explored relationships between tutors and students based on demographic and course data.
- Introduced a framework to enhance educational processes through data mining techniques by uncovering hidden patterns and trends, thereby enabling more accurate and sophisticated predictive analyses of student performance.
- Developed a classification-based predictive model to estimate students' overall performance using internal assessment data across different course modules offered within a semester.

3. Association Rule Mining Process

The process of mining association rules involves of two main parts. First, we have to recognize all the itemsets contained in the data that are sufficient for mining association rules. These combinations have to illustrate at least a certain frequency to be worth mining and are thus called frequent itemsets. The second step will produce rules out of the discovered frequent itemsets.

Let $I = \{i_1, i_2, ..., i_m\}$ is a set of items, $T = \{t_1, t_2, ..., t_n\}$ is a set of transactions, each of which contains items of the itemset *I*. Thus, each transaction t_i is a set of items such that $t_i \subset I$. An association rule is an implication of the form: $X \to Y$, where $X \subset I, Y \subset I$ and $X \cap Y = \phi$. X (or Y) is a set of items, called itemset [1,6].

Support: The support of a rule is represented by the formula

$$supp(X \to Y) = \frac{|X \cap Y|}{n}$$

Where $|X \cap Y|$ is the number of transactions that contain all the items of the rule and n is the total number of transactions [4,7].

Confidence: The confidence of a rule shows the percentage of transactions containing X which also contain Y [4].

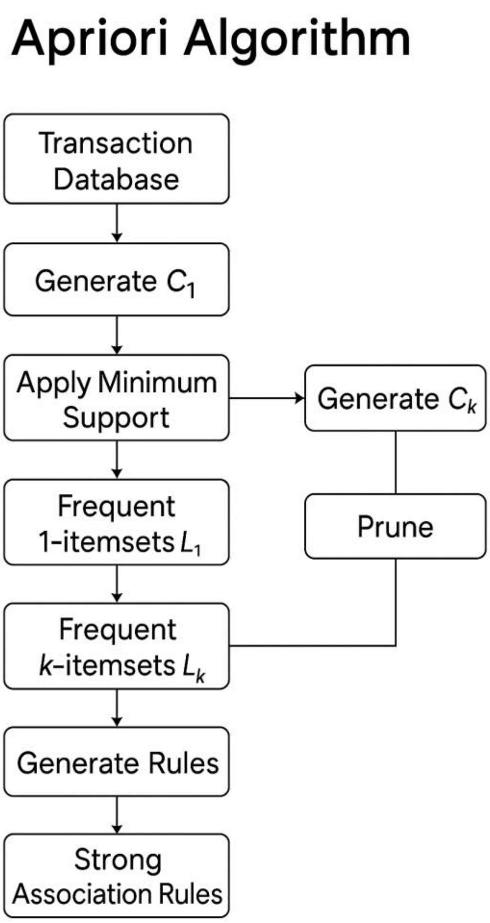
$$conf\,(X \rightarrow Y) = \frac{|X \cap Y|}{|X|}$$

Mining frequent patterns from a given dataset is not a insignificant task. All sets of items that occur at least as frequently as a user-specified minimum support have to be identified at this step. Diverse algorithms attempt to allow efficient finding of frequent patterns. After having generated all patterns that meet the minimum support requirements, rules can be generated out of them

Apriori algorithm is the simplest and most popular algorithm for discovering association rules from a large database [10]. Apriori algorithm works using two main steps.

- Frequent itemsets are called (L_K), these itemsets are greater than or equal to *minsupp*.

Candidate itemsets are called (C_K), and generated from (L_{K-1}). It is required to scan the original database for each itemset in the candidate generation step, in order to calculate the support value.



3.1 Experimental Result

S.No.	PAR	PSM	ATT	MSM	ESM
1	'first'	'first'	'good'	'good'	'first'
2	'first'	'first'	'good'	'good'	'first'
3	'first'	'first'	'average'	'average'	'first'
4	'second'	'first'	'good'	'good'	'first'
5	'second'	'first'	'good'	'good'	'first'
6	first'	'second'	'poor'	'average'	'first'
7	first'	'second'	'good'	'average'	'first'
8	'second'	'second'	'average'	'average'	'second'
9	'second'	'second'	'poor'	'average'	'second'
10	'second'	'second'	'poor'	'poor'	'third'

The analysis using association analysis is being done with the help of WEKA tool. WEKA, formally called Waikato Environment for Knowledge Analysis. Result Generated by WEKA software is presented using Table 2 shows the rule generated by WEKA.

Minimum support: 0.25 (9 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 19

Generated sets of large itemsets:

Size of set of large itemsets L (1): 17

Size of set of large itemsets L (2): 28

Size of set of large itemsets L (3): 11

Size of set of large itemsets L (4): 3

Table 2 - Rule Generated by WEKA software

S.No.	Rules Generated by WEKA	Conf	Lift	Lev:	Conv:
1	*PSM=first ==> ESM=first*	1	4.57	0.17	9.64
2	*PAR=second ESM=second ==> PSM=second*	1	1.5`	0.12	5
3	*PAR=first PSM=first ==> ESM=first*	1	3.54	0.13	6.45
4	*MSM=good ESM=first==> PSM=first*	1	4.16	0.12	6.02
5	*PSM=first MSM=good==> ESM=first*	1	3.54	0.11	5.46
6	*ATT=poor ESM=fail ==> PAR=third*	1	3.51	0.1	5.01
7	*PSM=first ATT=good==> MSM=good*	1	3.59	0.1	5.07
8	*PSM=first ATT=good ==> ESM=first*	1	3.59	0.1	5.04
9	*PAR=second MSM=average ESM=second ==> PSM=second*	1	1.5	0.08	3.2
10	*ATT=good MSM=good ESM=first ==> PSM=first*	1	3.17	0.11	4.32

The interpretation of the above association rules for different confidence values depicts that the students' performance will be first in ESM if he/she got the first division in PSM or first in PSM and PAR or First in PSM and ATT is good in this semester or first in PSM and MSM was good.

Conclusion and Future Work

The reputation of an academic institution are strongly influenced by the performance of its students. This study presents a methodology based on the association analysis algorithm to identify groups of academically weak students in the current semester. The approach involves comparing student performance across previous semesters using parameters such as undergraduate marks, previous semester results, current semester attendance, and mid-semester scores.

The proposed methodology aims to support academic planners and administrators in taking early corrective measures to enhance overall institutional outcomes in subsequent years. The Apriori algorithm has been utilized in this study for association rule generation. As a part of future work, the model will be extended by applying the Apriori algorithm to another association mining and classification and clustering algorithm to achieve improved prediction accuracy and more refined performance analysis.

References

1. Han, J. and Kamber, "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Series in Data Management Systems", Jim Gray, Series Editor, 2006.
2. Brijesh Kumar Baradwaj, Sourabh Pal, "Mining Educational Data to Analyze Student’s Performance", IJACSA, Volume 2, Nov - 2011.
3. Agrawal, Rakesh, Imielinski, Tomasz, Swami, Arun,“Mining Association Rules between Sets of Items in Large Databases”, Proceedings of the ACM SIGMOD International Conference on Management of Data, 1993.
4. J. Han and M. Kamber, " Data Mining: Concepts and Techniques", Morgan Kaufmann, 2000.
5. Shyamala K. and Rajagopalan S. P., “Data Mining Model for a better Higher Educational System", Information Technology Journal, Vol. 5, No. 3, pp. 560-564, 2006.
6. Ranjan J. and Malik K., "Effective Educational Process: A Data Mining Approach", Vol. 37, Issue 4, pp 502-515, 2007.
7. J.Mamcenko, Jelena, Irna Sileikiene, Jurgita Lieponiene and Regina Kulvietiene, "Analysis of E-Exam data using data mining techniques", proceeding of 17th international conference on information and software technologies, pp. 215-219, 2011.
8. M.Ramaawami and R.Bhaskaran, "A CHAID based performance prediction model in educational data mining", 2010.
9. T.Hadzilacos, D.Kalles, C.Pierrakeas and M.Xenos, "On small data sets revealing big differences", Advance in artificial intelligence, pp.512-515, 2006.
10. Azhar Rauf, Sheeba, “Enhanced K-Mean Clustering Algorithm to Reduce Number of Iterations and Time Complexity”, Middle-East Journal of Scientific Research, Vol. 12, Pp. 959-963, 2012.
11. Jaideep Vaidya, "Privacy Preserving K-Means Clustering over Vertically Partitioned Data", proceeding of SIGKDD, Washington, DC, USA, August 24-27, 2003.
12. S.A.Kumar and M.N.Vijayalakshmi, "Efficency of decision tree in predicting student’s academic performance", first international conference on computer science, Engineering and applications, Vol. 2, pp.335-343, 2011.
13. Ceglar, Aaron, Roddick, John F., "Association mining: ACM Computing Surveys ", Volume 38 , Issue 2, 2006.

14. Liu, Bing, "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data", Springer, 2007.
15. Agrawal, Rakesh, Srikant, Ramakrishnan, “Fast Algorithms for Mining Association Rules”, Proceedings 20th Int. Conf. on Very Large Data Bases, VLDB, 1994.