Journal of Contemporary Educational Research, 2024, Volume 8, Issue 2
http://ojs.bbwpublisher.com/index.php/JCER
ISSN Online: 2208-8474
ISSN Print: 2208-8466

Application Analysis of Nursing Students' Grades in Course Relevance Based on Association Rule Mining Algorithm Apriori

Xuemei Li[1,2], Edward C. Jimenez[1]*

[1]Angeles University Foundation, PhD Educ-EM, MacArthur Hwy, Angeles, 2009 Pampanga, Philippines
[2]Shandong College of Traditional Chinese Medicine, Educational Technology Center, Yantai 264199, Shandong Province, China

***Corresponding author:** Edward C. Jimenez, jimenez.edward@auf.edu.ph

Abstract: By analyzing the correlation between courses in students' grades, we can provide a decision-making basis for the revision of courses and syllabi, rationally optimize courses, and further improve teaching effects. With the help of IBM SPSS Modeler data mining software, this paper uses Apriori algorithm for association rule mining to conduct an in-depth analysis of the grades of nursing students in Shandong College of Traditional Chinese Medicine, and to explore the correlation between professional basic courses and professional core courses. Lastly, according to the detailed analysis of the mining results, valuable curriculum information will be found from the actual teaching data.

Keywords: Grade analysis; Apriori algorithm; Course relevance; Data mining

Online publication: February 25, 2024

1. Introduction

With the increasing popularity of informatization in colleges and universities, the student performance data in the school's educational administration management system has also increased. In recent years, when faced with these large amounts of data, scholars have begun to study how to make better use of them and dig out valuable information from these data, instead of just focusing on simple operations on the data. Using association rules to find the correlation between courses provides a great guiding significance for schools to set courses, formulate teaching plans, revise talent training plans, and better carry out teaching activities.

The follow-up content in this paper includes the following parts. Firstly, we summarize the relevant research on the use of Apriori association rule algorithm in teaching management. Secondly, we introduce the Apriori association rule algorithm. Thirdly, we expound on the performance analysis of nursing students based on the association rule algorithm, including experimental design, data preparation, and model building, as well as experimental results and analysis.

2. Related research

This chapter mainly summarizes the relevant literature on the use of Apriori association rule algorithm in the teaching management of colleges and universities.

Zhu *et al.* [1] studied the characteristics of the data in the university management system, used the Apriori algorithm to mine the student achievement data, and analyzed the factors that have an impact on student achievement. This has a very positive effect on the operation of the university management system.

Huang [2] analyzed the relevance of art courses in colleges and universities in her article. This research is also based on data mining technology, using the Apriori association rule algorithm. According to the results of data mining, valuable curriculum information is obtained, the curriculum can be rationally optimized, and a reasonable basis can be provided for revising the syllabus.

Wu *et al.* [3] conducted data mining on the massive data in the student management system, found the correlation between computer science and technical courses, and analyzed the reasons for the uneven course performance. The article also uses the Apriori algorithm. This study provides a reference for improving the teaching quality of computer science and technology courses.

The formulation of talent training programs plays a key role in the cultivation of talents in colleges and universities. In the article by Huang *et al.* [4], using the Apriori data mining algorithm, the degree of correlation between courses is clarified. Their analysis results can provide important references for talent training programs.

In addition, the Apriori algorithm also has related applications in music education. Li [5] revealed the connection between music courses in his article. The internal connection between courses provides more professional guidance for students' learning.

Association rule mining using Apriori algorithm is an important research direction of data mining, which is applied in different fields. The direction mentioned in this article is only for college teaching. Most of these associations analyzed are based on disciplines, which are highly logical and referable.

3. Apriori association rule algorithm

The association rule algorithm is to find the association of itemsets from a large amount of data, and discover the valuable relationship hidden in a large amount of data. All kinds of affairs in life will generate value, and mining the value of these affairs is also a form of data mining. Association rules reflect the interdependence and association between a thing and other things. If there is a certain correlation between two or more transactions, one of them can be predicted by other transactions.

3.1. Important indicators

In this study, we focus on three indicators of association rule algorithms, including support, confidence, and lift.

3.1.1. Support[3]

Support indicates how often a data item appears in a transaction. The support degree of association rule X=>Y is equal to the support degree of itemset $X \cup Y$, the formula is as follows [3].

$$\text{Support}(X=>Y) = \text{Support}(X \cup Y) = \frac{\text{Count}(X \cup Y)}{\text{Count}(D)}$$

3.1.2. Confidence[3]

Confidence, the confidence of the rule X=>Y indicates how likely it is that transactions containing X in D also contain Y. Confidence represents the strength of the certainty of this rule, denoted as confidence (X=>Y). Usually, we will specify the minimum confidence threshold according to our own mining needs, denoted as minconf [3]. The formula is as follows.

$$\text{Confidence}(X \cup Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

3.1.3. Lift[3]

Lift refers to the frequency with which item X and item Y appear together, but at the same time, the frequency of each of these two items must be considered. Lift reflects the correlation of X and Y in association rules. Lift greater than 1 and higher indicates a higher positive correlation. Lift less than 1 and lower indicates a higher negative correlation. Lift equal to 1 indicates no correlation[3]. The formula is as follows.

$$\text{Lift}(X,Y) = \frac{P(X \cup Y)}{P(X)P(Y)}$$

3.2. Apriori algorithm

The Apriori algorithm process is divided into two steps:

Through iteration, all frequent itemsets in the transaction database are retrieved, that is, itemsets whose support degree is not lower than the threshold set by the user.

Frequent itemsets are used to construct association rules that satisfy the user's minimum confidence.

The implementation code of Apriori algorithm is given in **Table 1**.

Table 1. Apriori algorithm [6]

```
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 is enclosed in t of all candidates in Ck is to be incremented
Lk = candidates in Ck with min_sup
end
return ⋃k Lk;
```

4. Analysis of nursing students' grades based on association rules algorithm

This paper uses the Apriori association rule algorithm to mine the correlation between students' grades. The complete analysis process includes four parts: experimental design, data preparation, model building, and experimental results and conclusions.

4.1. Experimental design

The research data in this article were derived from the educational administration system of Shandong College of Traditional Chinese Medicine. After screening, the data used for the experiment were the course scores of the 2019 nursing students in the Department of Nursing for the past three years. These courses included 12

public compulsory courses, 7 professional basic courses, and 7 professional core courses. The data source type exported by the school's educational administration system was an Excel worksheet. With the help of IBM SPSS Modeler 18.0 data mining tools, the Apriori association rule algorithm was selected to mine the correlation between professional basic courses and professional core courses.

4.2. Data preparation

4.2.1. Data selection

The research object selected in this paper was the final examination results of 1105 students from 18 administrative classes of 2019 nursing major in Shandong College of Traditional Chinese Medicine. In order to ensure the validity and integrity of data, all data were exported from the educational administration management system.

4.2.2. Data cleaning

(1) Deleting irrelevant data: The exported raw data included information such as student number, name, course name, and course credits, some of which were not needed in this study, such as course credits. All the fields that were not the objects of this study were deleted.

(2) Deleting irrelevant course grades: The information mined in this paper was the correlation between professional basic courses and professional core courses, so we deleted the scores of unrelated courses. The deleted courses included 12 public compulsory courses. After deletion, only 7 professional basic courses and 7 professional core courses remained. There were also students' names and student numbers, which were used to identify the uniqueness of students.

(3) Data cleaning: In the original data, some students were marked as "absent," "postponed," "fraudulent," and other special circumstances. According to the school policy, the scores of these students were replaced with 60 points.

(4) The course name was simply marked as class, followed by the serial number. The conversion name is shown in **Table 2**.

Table 2. Course conversion name

Course name	Course attributes	Mark
Human Structure	Professional basic course	class1
Fundamentals of TCM Nursing	Professional basic course	class2
Introduction to Nursing	Professional basic course	class3
Immunology and Pathogenic Biology	Professional basic course	class4
Pathology	Professional basic course	class5
Physiology	Professional basic course	class6
Pharmacology	Professional basic course	class7
Health Assessment	Professional core course	class8
Fundamentals of Nursing	Professional core course	class9
Pediatric Nursing	Professional core course	class10
Medical Nursing	Professional core course	class11
Critical Care Nursing	Professional core course	class12
Surgical Nursing	Professional core course	class13
Obstetrics and Gynecology Nursing	Professional core course	class14

4.2.3. Data conversion

Apriori association rule algorithm is a data mining algorithm based on the Boolean Association Rule, which does not support continuous data. Therefore, we need to convert the original data to Boolean-type data. Firstly, we converted the original data into numerical data. Secondly, we converted numerical data into discrete Boolean data. The significance of this operation was to reduce the complexity of the score data.

As different teachers have subjective factors when evaluating different classes' results, we followed certain rules when we converted data in order to reduce the error caused by this reason. This rule was to first calculate the average score of each course. When the score of the course was higher than the average score, the score was converted to 1. When the score of the course was lower than the average score, the score was converted to 0.

4.3. Model building

With the help of IBM SPSS Modeler 18.0 software and Apriori algorithm, an association analysis model was established, as shown in **Figure 1**. The source file of the model was filtered, type set, and finally connected to the Apriori association rule model. In order to achieve the diversity of results, a web diagram was added.

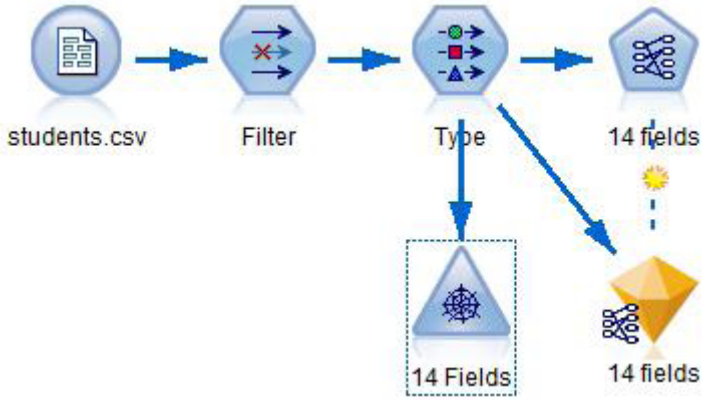


Figure 1. Algorithm mining model of Apriori association rules

In the process of model establishment, the specific operation process and main parameter settings of IBM SPSS Modeler 18.0 were divided into the following points.

(1) Importing data: Open the software, add the "Var. file" node, and import the cleaned data source files. Set "Lines to scan for column and type" to 1200 to read the complete dataset, as shown in **Figure 2**. The student name belongs to private information and has been hidden in the figure. It should be noted that the student number is not a record beginning with "2019," but a student who resumed his studies from 2017 or 2018 to 2019.

(2) Adding "Filter" node: The setting at this node is to filter out the "name" field, because this node is not related to the mining results, as shown in **Figure 3**.

(3) Adding "Type" node (**Figure 4**): First of all, we must click "Read Values" to read the data to the "Values" field to prevent error prompts during operation. Secondly, we need to change the "Measurement" of the ID attribute to "Continuous" and the "Role" to "Record ID." Thirdly, we also need to change the "Measurement" of the class1 to class14 attributes to "Flag" and the "Role" to "Both," which means that these attributes can be used as input data or output data.

(4) Adding "Apriori" node: It should be noted that the parameters of this model node need to be adjusted repeatedly to obtain more accurate results, as shown in **Figure 5**. Double-click the "Apriori" node with the mouse. On the "Model" tab, set the "Minimum antecedent support" to 30, the "Minimum rule

confidence" to 92, and the "Maximum number of antecedents" to 5. According to the settings, we get 10 rules.

(5) Adding "Web" node: The web diagram can more intuitively see the relationship between statistical results. Add a "Web" node after the "Type" node. Add all courses in the Plot tab. In the "Options" tab, set the "Maximum number of links to display" to 25, as shown in **Figure 6**.

(6) After setting all nodes, save the settings, click the "Run" button, run the mining model, and view and analyze the results.

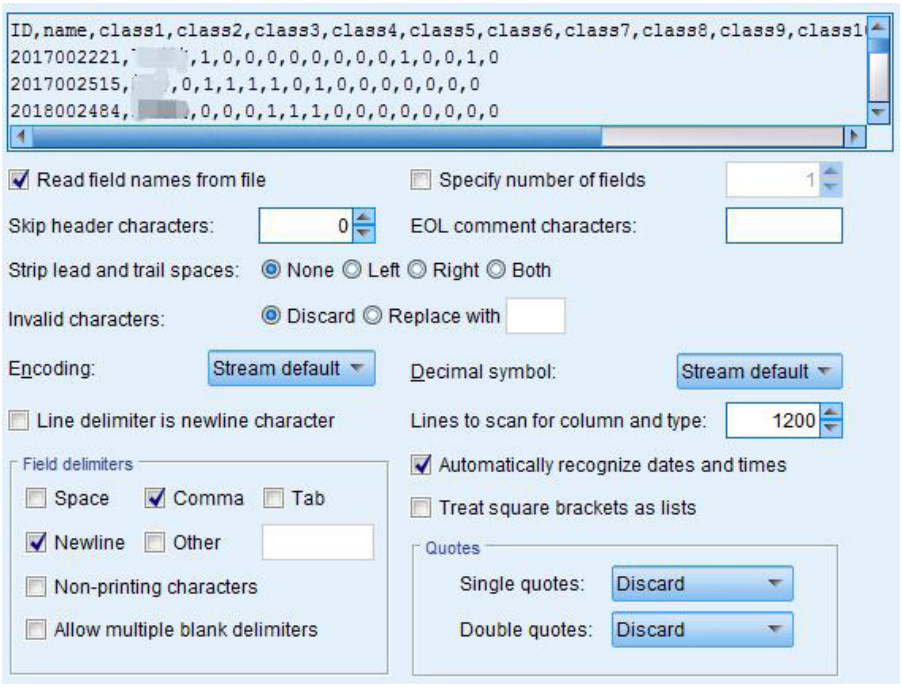


Figure 2. "Var. file" node parameter settings

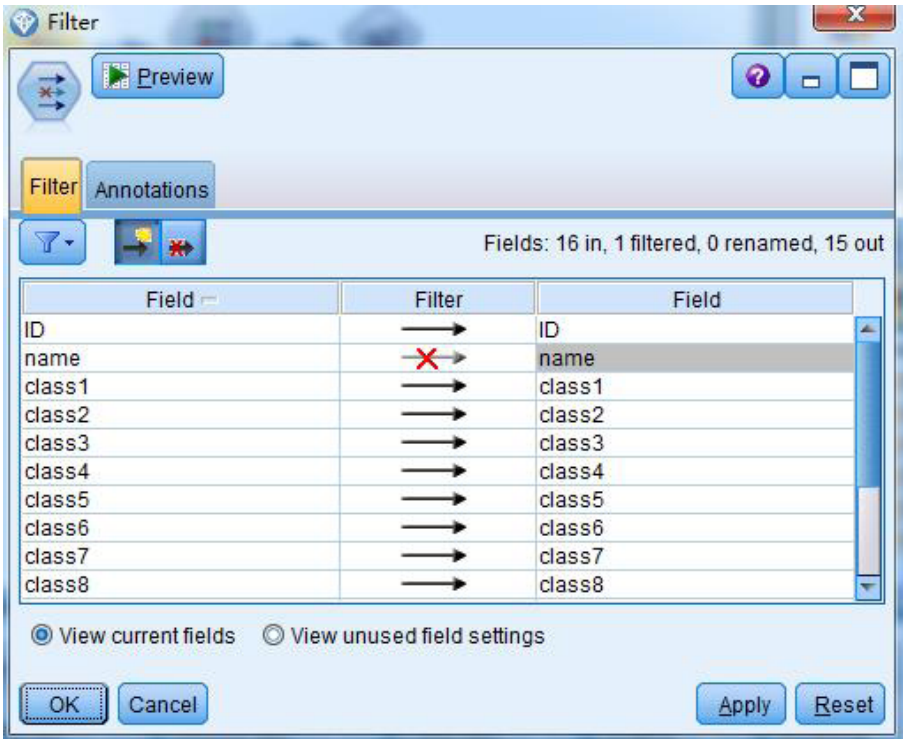


Figure 3. "Filter" node parameter settings

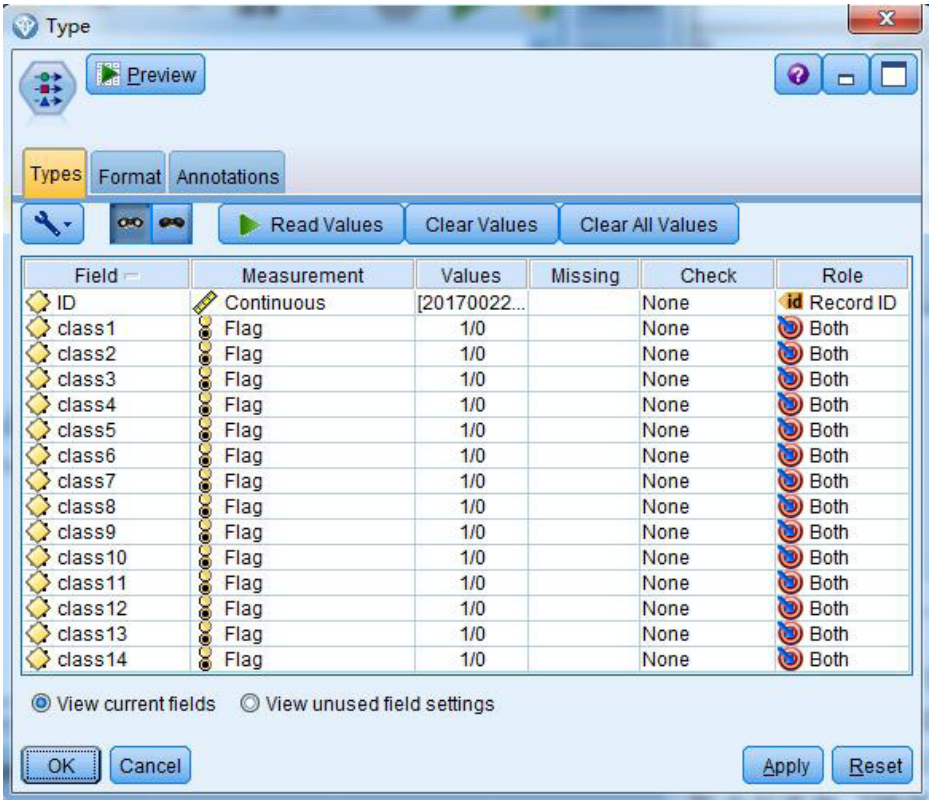


Figure 4. "Type" node parameter settings

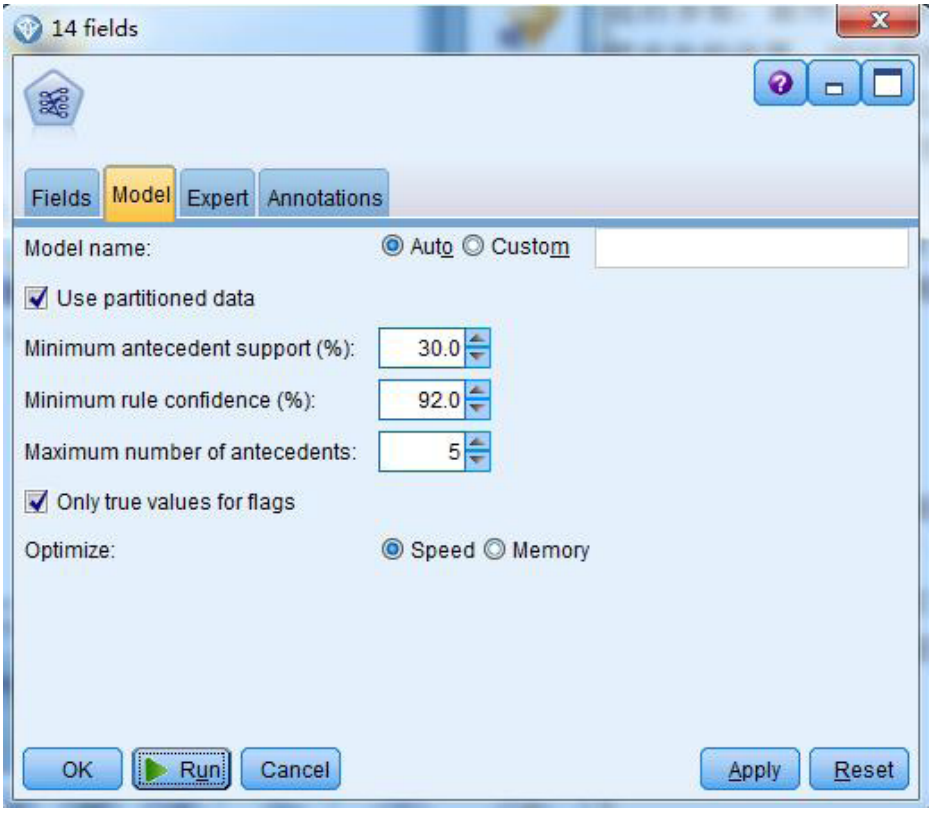


Figure 5. "Apriori" node parameter settings

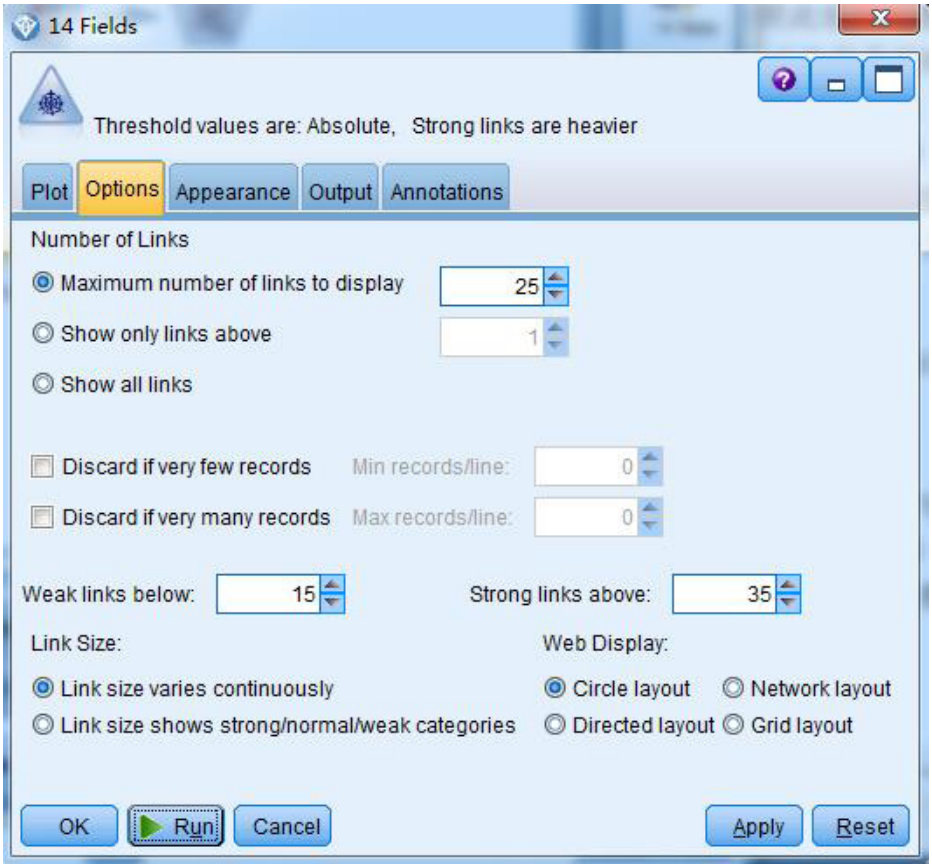


Figure 6. "Web" node parameter settings

4.4. Experimental results

This section analyzes the results obtained from the mining model to provide a reasonable basis for schools to set courses or modify teaching plans.

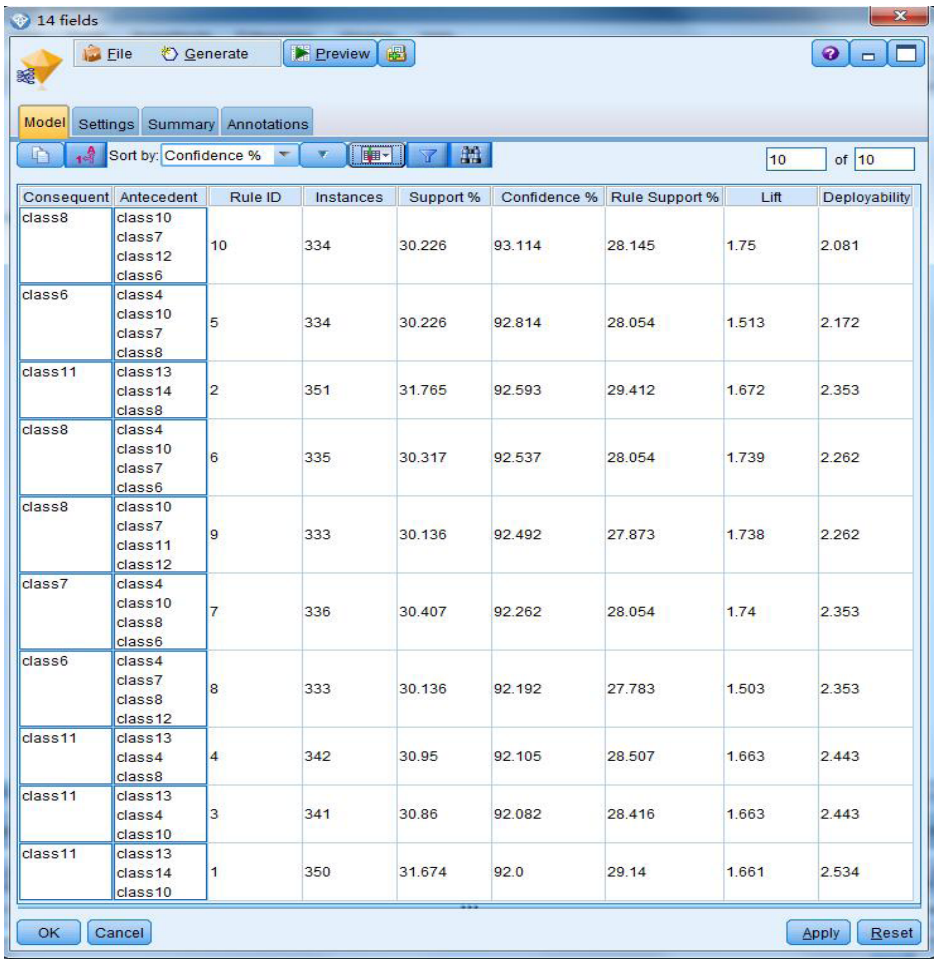
4.4.1. Excavation results

According to the parameters set in the experiment, the association rule results are shown in **Figure 7**, and a total of 10 results are obtained. At the same time, we also get a web diagram, as shown in **Figure 8**.

4.4.2. Results analysis

According to the network diagram in **Figure 8**, we can see that the connecting lines between courses are thick and thin. The thicker the connecting line, the greater the influence between courses. The more connecting lines, the closer the connection with other courses. According to the analysis of association rules and web graph, we can draw the following conclusions.

(1) Among the 10 association rules, it can be seen that 3 of them appear as "Consequent" for professional basic courses, and there are two courses in total. These two courses are Physiology and Pharmacology respectively. The "Consequent" of the other 7 rules are all professional core courses. This shows that the courses set up by the school's academic affairs office are basically all right. This abnormal result may be due to some differences in the standards of some teachers for reviewing test questions, or the method used when converting student performance data is not suitable.



Consequent	Antecedent	Rule ID	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
class8	class10 class7 class12 class6	10	334	30.226	93.114	28.145	1.75	2.081
class6	class4 class10 class7 class8	5	334	30.226	92.814	28.054	1.513	2.172
class11	class13 class14 class8	2	351	31.765	92.593	29.412	1.672	2.353
class8	class4 class10 class7 class6	6	335	30.317	92.537	28.054	1.739	2.262
class8	class10 class7 class11 class12	9	333	30.136	92.492	27.873	1.738	2.262
class7	class4 class10 class8 class6	7	336	30.407	92.262	28.054	1.74	2.353
class6	class4 class7 class8 class12	8	333	30.136	92.192	27.783	1.503	2.353
class11	class13 class4 class8	4	342	30.95	92.105	28.507	1.663	2.443
class11	class13 class4 class10	3	341	30.86	92.082	28.416	1.663	2.443
class11	class13 class14 class10	1	350	31.674	92.0	29.14	1.661	2.534

Figure 7. Association rule results

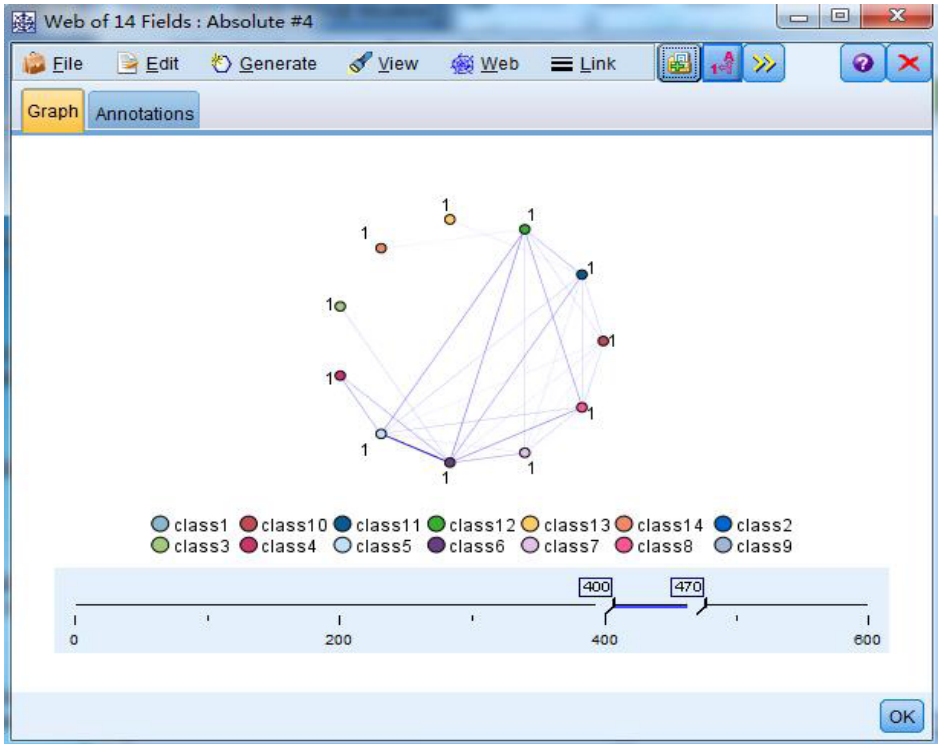


Figure 8. Web results

(2) From the network diagram in **Figure 8**, it can be seen that the courses connected at both ends of the thickest line are Pathology and Physiology. This result indicates that the two courses are highly related, but there is no direct proof in the association rules, and the relationship between the two courses has not been excavated. We speculate that it may be because these two courses do not meet the "Minimum antecedent support" and "Minimum rule confidence."

(3) Another conclusion can be seen in the web diagram, there are two courses that are most connected with other courses. These two courses are Physiology and Medical Nursing. Physiology is a professional basic course, and it is also strongly related to the other 7 courses, and the lines are relatively thick. Medical Nursing is a professional core course. Compared with Physiology, the connecting line of this course is relatively thinner. Medical Nursing is not regarded as a professional basic course, which only proves that it is related to other courses. Considering the content of this course, it really should serve as a professional core course.

(4) From the 10 association rules, there are five courses that do not appear in the rules. These courses are Human Structure, Fundamentals of TCM Nursing, Introduction to Nursing, Pathology, and Fundamentals of Nursing. Only Fundamentals of Nursing is a professional core course, and the others are professional basic courses. Such a situation in the results should be because the professional core courses set up by the academic affairs office are all oriented nursing courses.

(5) The rule "Introduction to Nursing => Fundamentals of Nursing" was not found in the model. These two courses have a sequential relationship, but judging from the results, the two courses are not closely related.

Through the analysis of the above results, we can identify the correlation between the courses, and these results can be used as the basis for schools to formulate teaching plans, and revise talent training programs and curriculum settings.

5. Conclusion

In this paper, with the help of IBM SPSS Modeler data mining software, the Apriori association rule algorithm and web graph are used to analyze the students' grades. We find the correlation between courses, calculate the support and confidence between courses, and dig out valuable information hidden in student performance. We also show the degree of influence between courses through a web diagram. There may be certain errors in this result or it does not conform to our daily rules, but we can refer to the results appropriately. As teachers, we must continue to strengthen analysis and formulate new courses based on the actual situation of students. Teaching management is a dynamic process that requires us to make constant adjustments to cultivate better talents for society.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Zhu X, Zong Y, 2022, Construction and Application Analysis of University Management System Based on Association Rule Mining Algorithm Apriori. Scientific Programming, 2022(2022):9 pages. https://doi.org/10.1155/2022/4367267

[2] Huang C, 2022, Research on the Relevance of Art Courses in Colleges and Universities Based on Data Mining. Scientific Programming, 2022(2022): 7 pages. https://doi.org/10.1155/2022/6896816

[3] Wu X, Zeng Y, 2019, Using Apriori Algorithm on Students' Performance Data for Association Rules Mining, 2nd International Seminar on Education Research and Social Science (ISERSS 2019). Atlantis Press, Malaysia, 205–208.

[4] Huang C, Pu S, 2020, Correlation Analysis on the Courses of Civil Engineering Based on Association Rules Mining, 2020 International Conference on Social Science, Economics and Education Research (SSEER 2020). Atlantis Press, Singapore, 28–32.

[5] Li T, 2021, Application of APRIORI Correlation Algorithm on Music Education Curriculum Association Rules, Journal of Physics: Conference Series (Vol. 1955, No. 1, p. 012067), IOP Publishing.

[6] Angeline DMD, 2013, Association Rule Generation for Student Performance Analysis Using Apriori Algorithm. The SIJ Transactions on Computer Science Engineering & its Applications (CSEA), 1(1): 12–16.