International Journal of Grid and Distributed Computing
Vol. 14, No. 1, (2021), pp. 313-324

Association Rules for Taking Students' Subject Using Apriori Algorithm on Higher Education

Vivi Nila Sari[1], Sarjon Defit[2]
[1,2] Universitas Putra Indonesia YPTK, Padang, Indonesia.
E-mail: [1]vivinilasari@upiyptk.ac.id, [2]sarjond@yahoo.co.uk

Abstract

Information technology at University of Putra Indonesia Yptk Padang has advanced and has been following technological developments in the current 4.0 Revolution era. Almost all lecturers follow and master the development of information technology to support its quality in the process of teaching, research and community service in accordance with the Tri Dharma of Higher Education which must be worked by a lecturer. In the learning process, a lot of data that will be stored without exception is the student grade data. One of the highlights is the application of data mining to determine the pattern of each value with another value. Data mining aims to find relationships between one item and another in a data set. In this research, data mining that will be used is using the Association Rule Apriori algorithm. The Apriori method will generate Support, List, and Confident values. The values generated by Apriori will determine the formation of a pattern. This study resulted in 12 new rules that meet the minimum support value (25%) and minimum confidence (50%). Based on the data on the value data of the UPI Yptk Padang management students, the pattern found is the relationship between the value of one course and the value of other courses.

Keywords: *Association Rules Algorithm; Data Mining; Student Value Data Department of Management;The Apriori Method*

INTRODUCTION

At University of Putra Indonesia Yptk Padang there is a Management department that is quite attractive from year to year. But there are still many students who find it difficult to attend lectures because they do not like some courses. The most important thing in determining the success or failure of a student is their score at the end of the semester. From the results of this semester's grades, we can see what concentration the student will choose in semester 5. The research this time was to determine the pattern of a student's value data related to the pre-requisite score for taking the next course. Please note that in semester 3 there are courses that must meet the prerequisite courses to be taken. Introduction to Business, Algorithm Logic and Introduction to Accounting 1, 2 & 3 courses, are prerequisite courses when management students will continue to semester 3. So that researchers want to analyze the relationship between these courses and prerequisite courses with the application of data mining. Data mining is a process of digging up more value in a database by looking at patterns from data so as to produce useful information that cannot be found manually [1]. Data Mining is also known as Knowledge Discovery in Database (KDD). [2] define data mining as the process of extracting

ISSN: 2005-4262 IJGDC

interesting patterns (implicit, unknown, and potentially exploited) from large data [3] . As the amount and type of data increases, so does the challenge to process them.

In this case, data mining has a big role in processing and extracting data. Data mining is divided into several tasks, including: association, classification, clustering, and sequence patterns [4]. Association is a data mining task that has long been used to find consumer behavior from transaction databases. The benefit of association is to find the relationships between the elements in the database. There are several association discovery methods that are often used, including the Apriori algorithm [5]. Here the research uses data on student grades for semester 1, 2 & 3 of the 2018/2019 school year in the Management class (M5) of academic year 2018 which is the raw data. The total number of students is 27 students. For this reason, in this article data mining will be used to extract unknown information from the academic year 2018 Management (M5) student score database. Information is related to the association between the value of one course and the value of another course.

Problem Formulation and Purpose

- How to analyze the value of student data in finding associative rules between a combination of items and forming patterns of combinations of itemsets with a priori algorithms?
- How to implement data mining with association rule algorithms to analyze student grade data?
- How can these supporting factors be used as information in improving students' learning abilities?

LITERATURE REVIEW

Data Mining

Data mining is the activity of extracting data from very large data sets to find information that has its own use as needed [6]. Data mining can also be called the process of looking for added value which contains information that has not been known from some existing data [7]. Data Mining Stages There are 7 (seven) data mining stages, namely:

1. Data cleaning
   Data cleaning is a process to eliminate irrelevant data. The data that is discarded is sometimes compared beforehand with the hypothesis that has been made. So that in the next process you can easily find the desired results
2. Data integration (data integration)
   Data integration is the process of combining data from several databases into one new database. Not a small amount of the data needed is taken from various databases or text files.
3 Data selection (data selection). Not all of the data that is already in the database are needed, therefore it is necessary to select data for data that is really needed in the next process.
4 Data transformation (data transformation)
   Data is combined or modified according to the processes used in data mining. Because some data mining formats require special data formats for processing.
5 Mining process Is the process of extracting data from a database or collection of data to obtain information that is hidden from the processed data
6 Pattern Evaluation (pattern evaluation)

ISSN: 2005-4262 IJGDC

In this process are the results of data mining techniques in the form of patterns that will be tested on previously created hypotheses. So that it will get conclusions that are close to the results or hypotheses for the next process.

7 Knowledge Presentation

(Knowlegde Presentation) This is included in the final step of data mining, in this stage it is time to present the results that have been done by implementing the analysis obtained. So that it will get a real conclusion.

Apriori Algorithm

A priori algorithm is one type of algorithm that exists in data mining that uses association rules [8]. The use of the a priori algorithm itself is to find the frequency and association of an itemset with other itemsets from the processed data set, which have determined the minimum requirements for the value of support and the minimum requirements for confidence values first [9].

In its use to find patterns of linkage of items with one another, the a priori algorithm is widely used by supermarkets to explore information that has not been known before, for example a supermarket has a lot of transaction data, a supermarket manager can find out the purchasing patterns of its consumers by using an algorithm a priori [10]. The information obtained in the a priori algorithm is in the form of "if - then" as for example if a consumer buys items X and Y, then 50% of consumers might buy item Z, the pattern information is obtained from processing transactions so far.

Whether or not the association rule is useful or not can be seen by looking at the support value of the combination of an item and the confidence value of the relationship between one item and another item contained in the association rule. [11] and to find support and confidence from an item you can use the formula:

$$Support = \frac{\Sigma (X \cup Y)}{N} x\ 100\%$$

Information:

S = $Support$

$\Sigma (X \cup Y)$ = Number of transactions containing X and Y

N = Number of Transactions

$$Confidence = \frac{Support\ (X \cup Y)}{Support\ X} x\ 100\%$$

Completion Technique

1. Determine Minimum Support and Confidence. At this stage, the minimum support and the most accurate minimum confidence from the transaction data is determined so that it will produce the most accurate information as well. At this stage, the author will give a minimum support value limit of 25% and a confidence value of 50%.

ISSN: 2005-4262 IJGDC

2. Analyzing High Frequency Patterns. After the data is collected and the minimum support value has been determined, this stage looks for all frequencies for each itemset, that is, the itemset that has a minimum support of 0.25 or equal to 25% that has been previously determined.

RESEARCH METHODS

In conducting research conducted in this data mining final project, there are several steps that will be carried out, namely:

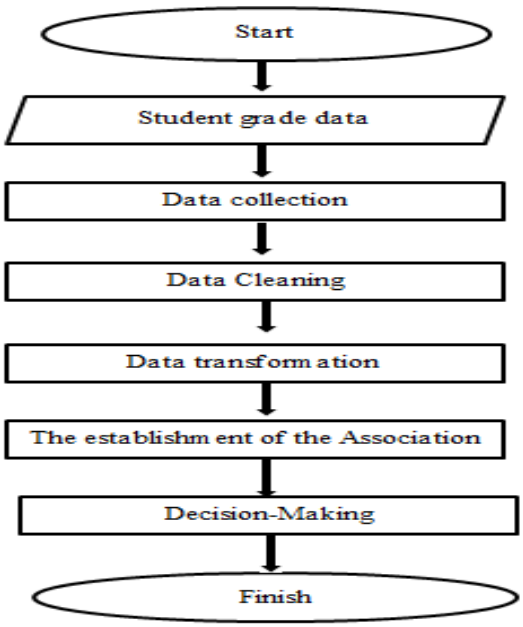


Figure 1. Data Mining Process Flowchart

The description on the flowchart will be explained in the discussion chapter. And it will be explained each sub chapter. From the whole research methods are correlated with one another. It will form a rule using the a priori method is.

RESEARCH RESULTS AND DISCUSSION

The researcher performed a manual calculation analysis of student grade data, especially the grades of Introduction to Business, Algorithm Logic and Introduction to Accounting 1 & 2, to find association rules between these courses by applying the Knowledge Discovery in Data steps. In this analysis the authors use Microsoft Excel 2010 as a tool for calculation and data sorting. And the program of WEKA as data testing [12]. The following is a discussion that will be carried out through analysis of the current system.

Table 1. Course Data

No.	Subject

ISSN: 2005-4262 IJGDC

1	Introduction to Business (MK-1)
2	Logic and Algorithms (MK-2)
3	Introduction to Accounting 1 (MK3)
4	Introduction to Accounting 2 (MK-4)
5	Financial Management (MK-5)
6	Marketing Management (MK-6)
7	Business Ethics (MK-7)
8	Data Base Management (MK-8)
9	Cost Accounting (MK-9)
10	International Business (MK-10)

Cource: Management Students Faculty of Economics and Business 2020

Of the 10 Subjects, a list of course combinations was obtained from 27 academic year 2018 Management Students as shown in Table 2 below:

Table 2. Data for Student's Subjects

College Student	Subject Itemset
1	MK1, MK5, MK6, MK7, MK10
2	MK2, MK8
3	MK3, MK4, MK9
4	MK1, MK5
5	MK1, MK5, MK6
6	MK3, MK4, MK9,
7	MK3, MK4
8	MK1, MK2, MK3, MK4
9	MK1, MK7
10	MK1, MK6, MK7
11	MK1, MK2, MK3, MK4, MK5, MK6. MK9
12	MK1, MK10
13	MK2
14	MK2, MK3, MK4, MK8, MK9
15	MK1, MK3, MK4, MK6, MK9
16	MK1
17	MK1, MK2, MK7, MK8
18	MK1, MK7, MK10
19	MK1, MK6
20	MK1, MK2, MK6, MK8
21	MK3
22	MK4

ISSN: 2005-4262 IJGDC

23	MK1, MK6, MK10
24	MK1, MK2, MK3, MK4,MK5, MK6, MK9, MK10
25	MK2,MK3, MK4, MK8, MK9
26	MK3, MK4, MK9
27	MK1, MK6

Source: Source: Faculty of Economics and Business Management Students 2020

The Process of Using the Association Rule Technique with the Apriori Algorithm.

This stage is the data mining stage, where at this stage the application of association rules with a priori algorithms will be discussed to find patterns of training for village midwives. Starting the association rules with a priori algorithm first determining the minimum support and minimum confidence. Researchers determine the minimum support is 25% and the minimum confidence is 50%. After that find all the combinations of items. Then enter the problem point of the association rule which consists of two processes:

1. Finding frequent itemsets,
   Namely itemset that has support that is greater than minimum support.
   a. The subset of frequent itemset must be a frequent itemset.
   b. Find frequent itemset iteratively from 1-item to k-item.
2. Use a frequent itemset (starting from 2-itemset) to generate association rules

A. Formation of Itemset

The process of forming C1 or what is called 1-Itemset with a minimum amount of support = 25%, in this process an accumulative count will be made of the number of courses taken by a student for every 1 item of courses conducted by 27 students, with the help of formulations. formula as follows:

$$Support\ A = \frac{\sum \text{Transactions contain A}}{\sum \text{Number of Transactions}} \times 100\ \%$$

Then after that the results of Suport A are obtained as a training item which is shown in Table 3 below:

Tabel 3. ***Support 1-Itemset***

No.	Itemset	Frequency	Support
1	Introduction to Business (MK-1)	17	62,96 %
2	Logic and Algorithms (MK-2)	9	33,33 %
3	Introduction to Accounting 1 (MK3)	11	40,74 %
4	Introduction to Accounting 2 (MK-4)	11	40,74 %
5	Financial Management (MK-5)	5	18,52 %
6	Marketing Management (MK-6)	10	37,03 %
7	Business Ethics (MK-7)	5	18,52 %

ISSN: 2005-4262 IJGDC

8	Data Base Management (MK-8)	5	18,52 %
9	Cost Accounting (MK-9)	8	29,62 %
10	International Business (MK-10)	5	18,52 %

Source: Manual Processed Data, 2020

Of the 10 courses attended by students, it turns out that there are only 7 subjects that meet the 1-Itemet (Frequent) criteria, namely {MK-1,MK-2,MK-3,MK-4,MK-6 and MK-9 }then the analysis continues by looking for Support B (2-Itemset).

B. Combination of 2 Itemset

The process of forming C2 or what is called the 2-Itemset with a minimum number of support = 25% produces 21 combination groups with the results shown in Table 4 below:

Table 4. 2-Itemset Combinations

No.	Combination	Frequency	Support
1.	MK1, MK2	5	18,52 %
2	MK1.MK3	4	14,81 %
3.	MK1,MK4	4	14,81 %
4	MK1,MK6	10	37,03 %
5	MK1,MK9	3	11,11 %
6	MK2,MK3	5	18,52 %
7	MK2,MK4	5	18,52 %
8	MK2,MK6	3	11,11 %
9	MK2,MK9	4	14,81 %
10	MK3,MK4	10	37,03 %
11	MK3,MK6	3	11,11 %
12	MK3,MK9	8	29,63 %
13	MK4,MK6	3	11,11 %
14	MK4,MK9	8	29,63 %
15.	MK6,MK9	3	11,11 %

Source: Manual Processed Data, 2020

The 2-Itemset combination data above that is frequent for a minimum support of 25% is 4 combinations, namely {MK-1, MK-6} and {MK-3, MK-4}, with the itemset being {MK-1, MK3, MK -4, MK-6 and MK-9}. The next analysis is to look at the 3-Itemset combination. The results are as in Table 5 with the following 4 combination groups:

Table 5. 3-Itemset Combinations

No.	Combination	Frequency	Support
1	MK-1, MK-3, MK-4	4	14,81 %
2	MK-1, MK-3, MK-6	3	11,11 %
3	MK-1, MK-3, MK-9	3	11,11 %

ISSN: 2005-4262 IJGDC

4	MK-1, MK-4, MK-6	3	11,11 %
5	MK-1, MK-4, MK-9	3	11,11 %
6	MK-1, MK-6, MK-9	3	11,11 %
7	MK-3, MK-4, MK-6	3	11,11 %
8	MK-3, MK-4, MK-9	8	29,63 %
9	MK-3, MK-6, MK-9	3	11,11 %
10	MK-4, MK-6, MK-9	3	11,11 %

Source: Manual Processed Data, 2020

From the calculation of the support value for the 3-Itemset presented in Table 5, it turns out that there is one group that still has a minimum support value of 25%, namely {MK-3, MK-4, MK-9} thus the calculation is stopped until the 3-Itemet combination and continues. Calculate the Confidence value for each Frequent 2-Itemet and 3-Itemet combination.

Establishment of Association Rules

At this stage, look for the confidence value in the results of the last combination which has met the minimum requirements for the support value, namely in stage 2 of the combination so that it can be continued to the next stage of the process by determining the value of confidence using the following formula:

$$\textit{Confidence}\ A \rightarrow B = \frac{\textit{Support}\ (AUB)}{\textit{Support}\ A} \times 100\ \%$$

Table 6. Support and Confidence Results

No.	Combination	Support	*Confidence*
1.	MK-3 → MK-4	37,03 %	90,90 %
2.	MK-4 → MK-3	37,03 %	90,90 %
3	MK-3 → MK-9	29,63 %	72,72 %
4	MK-9 → MK-3	29,63 %	100%
5	MK-4 → MK-9	29,63 %	72,72 %
6	MK-9 → MK-4	29,63 %	100%
7	MK-3 → MK-4, MK-9	29,63 %	72,72 %
8	MK-4, MK-9 → MK-3	29,63 %	100%
9	MK-4 → MK-3, MK-9	29,63 %	72,72 %
10	MK-3, MK-9 → MK-4	29,63 %	100%
11	MK-9 → MK-3, MK-4	29,63 %	100%
12	MK-3, MK-4 → MK-9	29,63 %	80%

Source: Manual Processed Data, 2020

All of the above values have a confidence value ≥ the minimum confidence value, namely 50%.

ISSN: 2005-4262 IJGDC

Knowledge Presentation

From the search process using a priori algorithm with a minimum value of support of 25% and a minimum of 50% confidence, the results of the association rule that appear are 12 rules, namely:

1. MK-3 → MK-4 { S = 37,03 % ; C = 90,90 %}
   This means: if a student has taken an introductory accounting 1 course then the student will take an Introduction to Accounting 2 course in the next semester as much as 37.03%. The level of truth of students choosing introductory accounting 1 and introductory accounting 2 courses is 90.90% meaning that not all of them take these courses.

2. MK-4 → MK-3 { S = 37,03 % ; C = 90,90 %}
   This means: if a student takes an introductory accounting course 2 then the student has taken the Introduction to Accounting 1 course in the previous semester as much as 37.03%. The level of truth of students choosing introductory accounting 2 and introductory accounting 1 courses is 90.90%, meaning that not all of them take these courses.

3. MK-3 → MK-9 { S = 29,63 % ; C = 72,72 %}
   This means: if a student has taken an introductory accounting course 1 then the student will take a Cost Accounting course in the next semester as much as 29.63%. The level of correctness of students choosing introductory accounting 1 and Cost Accounting courses is 72.72% meaning that not all of them take these courses.

4. MK-9 → MK-3 { S = 29,63 % ; C = 100 %}
   This means: if a student takes a cost accounting course, the student has taken the Introduction to Accounting 1 course in the previous semester as much as 29.63%. The level of correctness of students choosing cost accounting courses and introductory accounting 1 is 100% meaning that all those who take cost accounting courses must have taken Introductory Accounting 1 courses.

5. MK-4 → MK-9 { S = 29,63 % ; C = 72,72 %}
   This means: if a student has taken an introductory accounting course 2 then the student will take a Cost Accounting course in the next semester as much as 29.63% The level of correctness of students choosing introductory accounting 2 and Cost Accounting courses is 72.72%, meaning that not all of them take these courses.

6. MK-9 → MK-4 { S = 29,63 % ; C = 100 %}
   This means: if a student takes a cost accounting course, 100% of the student has taken Introduction to Accounting 2 in the previous semester. The level of correctness of students choosing cost accounting courses and introductory accounting 2 is 100%, meaning that all those who take cost accounting courses must have taken Introduction to Accounting 2 courses.

7. MK-3 → MK-4, MK-9 { S = 29,63 % ; C = 72,72 %}

ISSN: 2005-4262 IJGDC

This means: if a student has taken introductory accounting 1 then the student will take introductory accounting 2 and Cost Accounting in the next semester as much as 29.63% The level of correctness of students choosing introductory accounting 1, introductory accounting 2 and Cost Accounting courses is 72.72% meaning not all of them take these courses.

8. MK-4, MK-9 → MK-3 { S = 29,63 % ; C = 100 %}
   This means: if a student takes courses, introductory accounting 2 and Cost Accounting, then the student has taken introductory accounting 1 in the previous semester as much as 29.63%
   The level of correctness of students choosing courses Introduction to Accounting 2, Cost Accounting and introduction to accounting 1 is 100% meaning that all those who take introductory courses in accounting 2 and Cost Accounting must have taken introductory accounting 1 courses.

9. MK-4 → MK-3, MK-9 { S = 29,63 % ; C = 72,72 %}
   This means: if a student has taken introductory accounting 2 then the student will take introductory accounting 1 and Cost Accounting in the next semester as much as 29.63% The level of correctness of students choosing introductory accounting 2, introductory accounting 1 and Cost Accounting courses is 72.72% which means that not all of them take these courses.

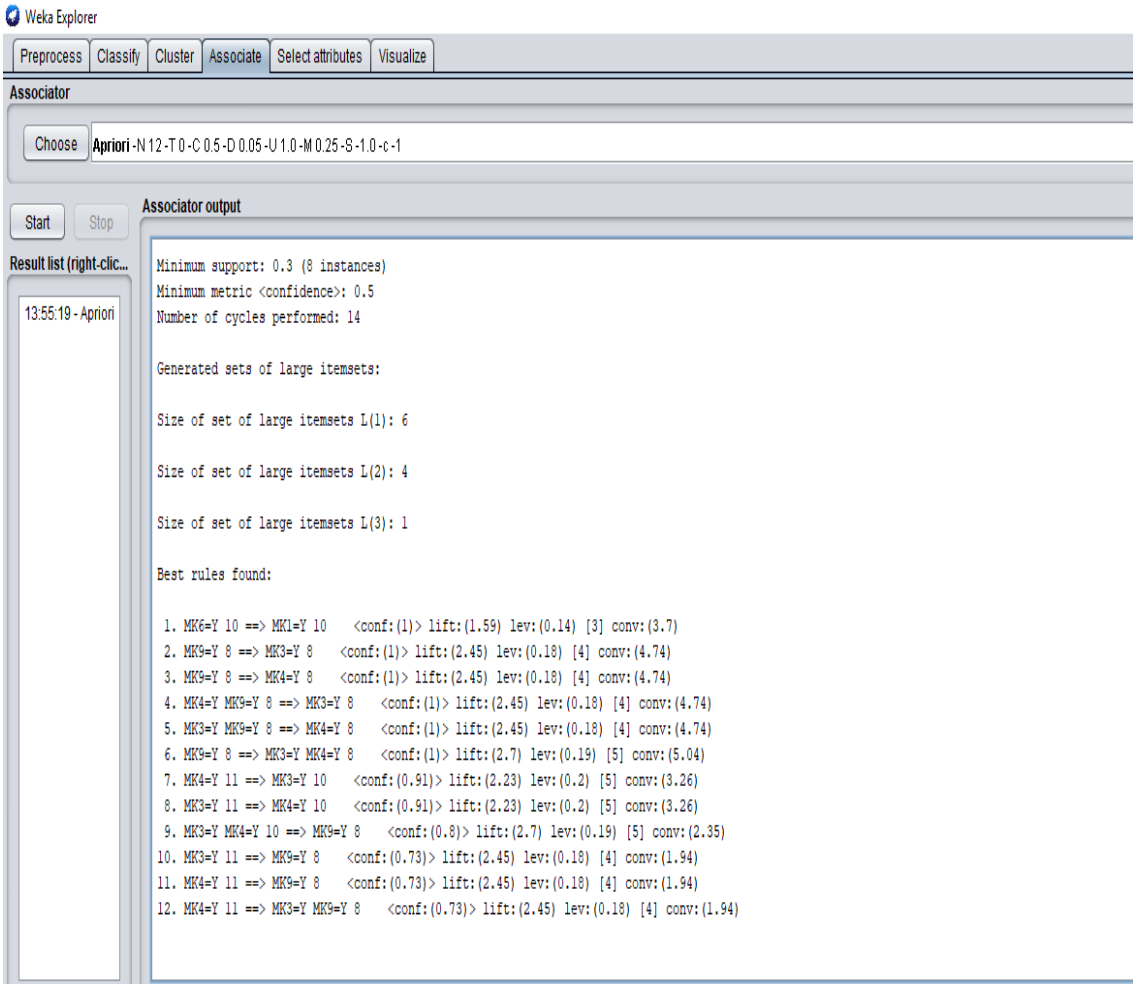
10. MK-3, MK-9 → MK-4 { S = 29,63 % ; C = 100 %}
    This means: if a student takes Introductory Accounting 1 and Cost Accounting courses then the student has taken introductory accounting 2 courses as much as 29.63%. The level of correctness of students choosing Introductory Accounting 1, Cost Accounting and introductory accounting 2 courses is 100% meaning that all those who take introductory accounting 1 and Cost Accounting courses must have taken introductory accounting 2 courses.

11. MK-9 → MK-3, MK-4 { S = 29,63 % ; C = 100 %}
    This means: if a student takes a course in Cost Accounting, the student has taken introductory accounting 1 and introductory accounting 2 in the next semester as much as 29.63%.
    The level of correctness of students taking Cost Accounting courses, Introduction to Accounting 1 and introduction to accounting 2 is 100% meaning that all those who take Cost Accounting courses must have taken Introduction to Accounting and Introduction to Accounting 2.

12. MK-3, MK-4 → MK-9 { S = 29,63 % ; C = 80 %}
    This means: if a student has taken Introduction to Accounting 1 and introductory accounting 2 then the student will take Cost Accounting as much as 29.63%. The level of correctness of students choosing introductory accounting 1, introductory accounting 2 and Cost Accounting courses is 80% meaning that not all of them take these courses.

ISSN: 2005-4262 IJGDC

This is also proven by processing using the WEKA 3.8 application with the following results:



From the process of looking for data information on 27 student scores by limiting the minimum value of support by 25% or equal to 0.25 and a minimum value of confidence of 50%.

CONCLUSION

Based on the research and discussion carried out, several things can be concluded as follows:

Based on data processing that is done both manually and by using the WEKA 3.8 application, it can be concluded that: there are 12 knowledge taking patterns for students majoring in management and the items that appear are Introduction to Accounting 1 (MK-3), Introduction to Accounting 2 (MK-4).) and Cost Accounting (MK-9). Therefore, the Academic Supervisor can give direction to new students that: if they have taken the Introduction to Accounting 1 (MK-3) course then it is advisable to take the Introduction to Accounting 2 (MK-4) and / or Cost Accounting (MK-3) courses. 9) and vice versa. This knowledge is very useful for improving student learning so that the prerequisite courses for taking courses in the next semester are more focused.

ISSN: 2005-4262 IJGDC

BIBLIOGRAPHY

[1] S. H. Xian, "Book Review: 'Flood Damage Survey and Assessment: New Insights from Research and Practice,'" *Water Econ. Policy*, 2019.
[2] A. Ahlemeyer-Stubbe and S. Coleman, "Data Mining Definition," in *A Practical Guide to Data Mining for Business and Industry*, 2014.
[3] J. Apostolakis, "An introduction to data mining," *Struct. Bond.*, 2010.
[4] G. Köksal, I. Batmaz, and M. C. Testik, "A review of data mining applications for quality improvement in manufacturing industry," *Expert Systems with Applications*. 2011.
[5] H. Toivonen, "Apriori Algorithm," in *Encyclopedia of Machine Learning and Data Mining*, 2017.
[6] B. Liu and L. Zhang, "A survey of opinion mining and sentiment analysis," in *Mining Text Data*, 2012.
[7] C. Dawson and C. Dawson, "Educational data mining," in *A–Z of Digital Research Methods*, 2019.
[8] G. A. Syaripudin and E. Faizal, "IMPLEMENTASI ALGORITMA APRIORI DALAM MENENTUKAN PERSEDIAAN OBAT," *JIKO (Jurnal Inform. dan Komputer)*, 2017.
[9] K. Ohara and H. Motoda, "Apriori," 2009.
[10] Dr. Nevine Makram Labib and Mohamed Sayed Badawy, "A Proposed Data Mining Model for the Associated Factors of Alzheimer's Disease," *Int'l Conf. Data Min.* , 2014.
[11] N. Adha, L. T. Sianturi, and E. R. Siagian, "IMPLEMENTASI DATA MINING PENJUALAN SABUN DENGAN MENGGUNAKAN METODE APRIORI (Studi Kasus : PT. Unilever)," *Maj. Ilm. INTI*, 2017.
[12] A. Naik and L. Samant, "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime," in *Procedia Computer Science*, 2016.

ISSN: 2005-4262 IJGDC