Evaluation of student academic performance using e-learning with the association rules method and the importance of performance analysis

by Ramadiani Ramadiani

Submission date: 22-Jul-2020 10:20AM (UTC+0700)
Submission ID: 1360653539
File name: Ramadiani_2020_J._Phys.__Conf._Ser._1524_012107.pdf (839.55K)
Word count: 4340
Character count: 22494

Evaluation of student academic performance using e-learning with the association rules method and the importance of performance analysis

Evaluation of student academic performance using e-learning with the association rules method and the importance of performance analysis

R Ramadiani[1], A Kurniawan[1], Z Arifin[1], M L Jundillah[1], R Alex[1], A Azainil[2], and A N Hidayanto[3]

[1]Faculty of Computer Science and Information Technology Mulawarman University, Samarinda, East Kalimantan, Indonesia
[2]Faculty of Teacher Training and Education Mulawarman University, Samarinda, East Kalimantan, Indonesia[3]Faculty of Computer Science Indonesia University, Depok, Jakarta, Indonesia

Corresponding Author: ilkom.ramadiani@gmail.com

Abstract. In managing the student study process, there is a pattern that occurs repeatedly every year. The recurring data will produce information in the form of student profiles, entry paths, student academic results, study period, average graduate-level and much other information as long as students take the process of teaching and learning activities. This research was conducted to predict student performance and e-learning satisfaction at the Computer Science department in Mulawarman University using the Association Rules method and Importance Performance Analysis. The sample in this study was 389 data of computer science graduate students. Based on the results of the research that has been done, it can be concluded that the graduation rate of computer science students of the Faculty of Computer Science and Information Technology, as follows; students who have graduated most have a GPA interval of 2.76 - 3.50 with male gender and take a study period of more than 6 years which has a support value of 0.321 and a confidence value of 0.628. Their perception of e-learning according to IPA coordinates community, collaboration, materials, social media, knowledge, synthesis, application, understanding, multimedia, evaluation, video, and news in quadrant II has a high level of importance with a relatively high level of performance and must be maintained.

1. Introduction

At present, every university is required to have a competitive advantage by utilizing all the resources they have. In addition to facilities, infrastructure and human resources, information systems are one of the resources that can be used to explore and enhance competitive advantage. Information systems can be used to obtain process and disseminate information, to support daily operational activities, while supporting strategic decision-making activities. There is some survey towards educational data mining and its scopes, such as student's satisfaction, faculty, classroom and student evaluation, students' dropout, course registration, new student enrollment, and collaborative activities [1]. The student's performance level can be improved by identifying the unit test, attendance, assignment and graduation, and giving additional guidance to improve the university result [2]. In this study, we evaluated students' performance using e-learning with the association rules method and the importance of performance analysis.

The data provided in an information system is used to support decision-making activities. These decision-making activities are not enough to rely solely on operational data, data analysis is needed to explore the potential of existing information. Decision-makers try to utilize the data warehouse that they already have to extract useful information, to help decision making. This led to the emergence of a new branch of science to overcome the problem of extracting information, or an important or interesting pattern of large amounts of data, called data mining. The use of data mining techniques is expected to provide knowledge that was previously hidden in the data warehouse so that it becomes valuable information.

The problem in this study is how to implement data mining with the association algorithm rule to analyze the academic performance of computer science students in the Faculty of Computer Science and Information Technology? How to evaluate student perception uses e-learning by using importance-performance analysis? And how can these supporting factors be used as information to improve the quality of e-learning performance and to predict the length of study of each student?

1.1 Data mining

Data mining as knowledge discovery in databases (KDD) is an activity that includes gathering, using historical data to find order, patterns or relationships in large data. This data mining output can be used to make future decisions. The development of KDD causes the use of pattern recognition to diminish because it has become a part of data mining [3]. The main reason why data mining is needed is because of the large amount of data that can be used to produce useful information and knowledge[3].

Data Mining is one of the fields that is growing rapidly because of the huge need for added value from large-scale databases that are increasingly accumulated in line with the growth of information technology. Data mining is divided into several groups based on tasks that can be done, namely [3]: 1) *Description*. Sometimes researchers and analysts simply want to try to find ways to describe the patterns and trends contained in the data. 2) *Estimation*. The estimation is almost the same as the classification, except that the estimated target variable is more numerical than the category. The model is built using a complete record that provides the variable value of the target as a predictive value, then in the next review, the estimated value of the target variable is made based on the value of the predictive variable. 3) *Prediction*. Predictions are almost the same as classifications and estimates, except that in predicting the value of the results there will be in the future. 4) *Classification*. In classification, there is a target category, for example, income classification can be separated into three categories, namely high, medium and low income. 5) *Clustering*. Clustering is grouping records, observing or paying attention to and forming classes of objects that have similarities. Clusters are collections of records that have similarities to one another and have similarities with records in other clusters. 6) *Association*. Task associations in data mining are finding attributes that appear at one time. In the business world, it is called a shopping cart analysis.

1.2 Pattern recognition, machine-learning, e-learning, and IPA

Pattern recognition is a discipline that studies ways to classify objects into classes or categories and recognize data trends. Depending on the application, these objects can be in the form of patients, students, credit applicants, images or signals or other measurements that need to be classified or searched for regression function [3-4].

Data mining, often also called knowledge discovery in the database, is an activity that includes collecting, using historical data to find order, patterns or relationships in large data sets. The output of data mining can be used to improve future decision making. So the term pattern recognition is rarely used because it includes part of data mining [3-4].

Machine Learning is an area in artificial intelligence or artificial intelligence that is related to the development of techniques that can be programmed and learned from past data. Pattern recognition, data mining and machine learning are often used to refer to the same thing. This field intersects with the science of probability and statistics sometimes also optimization. Machine learning is an analytical tool in data mining.

Association analysis or association rule mining is a data mining technique to find associative rules between combinations of items. Association rules in the form of "if ... then ..." or "if ... then ..." are knowledge generated from the function of association rules [5-6]. Association rule is one method that aims to find patterns that often appear between many transactions, where each transaction consists of several items [5-6].

Association rule is one method that aims to find patterns that often arise between many transactions, where each transaction consists of several items so that this method will support the system of recommendation through the discovery of patterns between items in transactions that occur. Especially one of the stages of association analysis called high-frequency pattern analysis attracts many researchers to produce efficient algorithms. The importance of an associative rule can be known by two parameters, support, namely the percentage of the combination of items. Association analysis is defined as a process to find all associative rules that meet the minimum requirements for support and minimum requirements for confidence [6].

The support value of an item is obtained by the formula [9]:

$$S = \frac{\Sigma(Ta+Tc)}{\Sigma(T)} \quad (1)$$

where

S = *Support*

$\Sigma(Ta+Tc)$ = Number of transactions containing antecedent and consequence

Σ(T) = Number of transactions

The confidence value of rule A -> B is obtained from the formula [7]:

$$C = \frac{\Sigma(Ta+Tc)}{\Sigma(Ta)} \quad (2)$$

where

C = *Confidence*

$\Sigma(Ta+Tc)$ = Number of transactions containing antecedent and consequence

$\Sigma(Ta)$ = Number of transactions containing antecedent.

1.3 Cross industry-standard process for data mining

Cross-Industry Standard Process for Data Mining (CRISP-DM) is a consortium of companies established by the European Commission in 1996 and has been established as a standard process in data mining that can be applied in various industrial sectors. The data mining process based on CRISP-DM consists of 6 phases; Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment.

1.4 R Programing

R is a relatively complete statistical analysis system, as a result of various statistical research collaborations in all parts of the world. The current R can be said to be the lingua franca for modern statistical computing purposes [7]. The syntax of the R language is generally equivalent to commercial S + statistical software. For most statistical analysis purposes, programming with R is almost identical to programming with S + [7].

1.5 e-Learning

E-Learning process on the Internet can be done using a variety of platforms or applications. It is expected to facilitate sharing the knowledge to the user. It does not mean to replace the existing class; it serves to enrich and be an alternative at the right time and needed [8-12]. Mobile learning is related to smartphones or tablets, portability, and ready for accessibility. The challenge for mobile learning is how to create lessons, which can download in an instant and run without stalling. The user experience must be intuitive and very fast. Many factors influence a person to choose and use the application.

1.6 Importance performance analysis (IPA)

The Importance Performance Analysis method was invented by [13-14] was used in 1977 to carry out an analysis of the level of customer satisfaction on organizational services or products. In a long period of time Importance, Performance Analysis has been utilized in responding to and developing marketing strategies, because this method helps companies understand customer wants and needs. User satisfaction is the user's perception; this is related to performance, service and product quality. The IPA technique combines measures of performance and perceived interests of users into a two-dimensional plot to facilitate the interpretation of data [13-14]. This plot group attributes into four categories or quadrants to specify in allocating limited resources. Four quadrants are usually identified as the main priority (Quadrant I), Maintain Performance (Quadrant II), Low Priority (Quadrant III), and Excessive (Quadrant IV).

2. Research method

2.1 Data collection

Data collection in this study was carried out with four types of methods, literature study, interview, questioners, and observation.

2.1.1 Data

The data in this study are the academic data of the students of the Computer Science Study Program, Faculty of Computer Science and Information Technology, Mulawarman University. The total data is 389 respondents.

2.1.2 Variable. The independent variables in this study are gender, entry process, GPA and study period, e-learning performance, product quality, and service. The dependent variable in this study is the timeliness of student graduation and student satisfaction. The data mining proses consists of 6 phases, namely [4]: business, understanding, data understanding, data preparation, modeling, evaluation and deployment

Table 1. Data on student graduation

Description	Attribute
Student ID	Number Student ID is the identity code owned by students
Student Name	Name of the student concerned
Place, Date of birth	Place Date of birth of the student concerned
Gender	the Sex of the student
Religion	a religion adopted by students
Address of the student's	Address of the student's origin address
Entrance	The type of entry into the college that students take
Origin of School	The origin of the high school from the student
Study Program	Study Program of the student concerned
Graduation Year	Years of graduating from students
Graduation date	Graduation date of the student concerned
Cumulative Achievement Index (GPA)	Cumulative Achievement Index (GPA) is a measure of a student's ability at a certain time measured on an SKS
Length of Study	The length of study taken from first entering college
Parent's name	The name of the parent of the student concerned
Parent's address	The address of the parent of the student concerned
Thesis title	Thesis title of the student concerned
Graduation Period	Graduation period followed by students concerned
e-Learning evaluation	Student perception in e-learning performance and satisfaction

2.2. Data cleaning and preparation

From the data obtained then the selection of attributes and part of the data in the existing attributes will be converted to facilitate the data mining process because the data will be processed with data mining tools. The training data used in this study is the academic data of students of computer science study program in the year of graduation from 2012 to 2018. There are 389 data respondents were collected.

2.2.1. Data integration and transformation. To improve accuracy and efficiency, algorithms are deemed necessary to transform data and attributes that have been collected.

Table 2. Research data attributes after integration and transformation

Description	Scale	Attribute
Gender	Nominal	Male, female
Entrance	Nominal	SNMPTN, SBMPTN
Cumulative Achievement Index (GPA)	Ordinal	A1 (3.51 to 4.00.) A3 (2.76 to 3.50.) A3 (2.00 to 2.75.)
Study Period	Ordinal	B1 (from 3 to 4 years.) B2 (from 4 to 6 years.) B3 (Study period is over 6 years.)
Graduation Year	Nominal	2012, 2013, 2014, 2015, 2016, 2017, 2018
e-Learning satisfied	Ordinal	Very satisfied (6-7) Quite satisfied (4-5) Very dissatisfied (1-3)
e-Learning performance	Ordinal	Very important(6-7) quite important (4-5) very not important (1-3)

2.2.2 Modeling and evaluation. In modeling staged data Mining techniques will be used with the Association Rule method using the Apriori algorithm with RStudio tools, which will produce associative rules or graduation transaction patterns. In this stage evaluation of the effectiveness and quality of the model is used, whether the Rule Association Method with a priori algorithm has achieved the objectives set at the initial stage.

2.2.3 Deployment. The results of this study in the form of an analysis of the percentage of student academic performance using the Association Rules method which is expected to be used by subsequent researchers as well as the parties concerned as a material in determining the student's academic performance with the attributes described. In addition, the results of this analysis can also be used as a basis for designing a decision support system to identify students' academic performance evaluations.

2.3. Results testing and analysis techniques

This study uses inferential statistical analysis techniques to display student graduation rates using the association rules method. Association rule mining is a procedure for finding relationships between items in a specified data set. Association rule includes two stages: look for the most common combination of itemset and defining the Condition and Result.

3. Results and Discussion

3.1. Data calculation

This research data was obtained from the academic section of the Faculty of Computer Science and Information Technology at Mulawarman University. Based on the data collection process, data that has been collected can be entered at the transformation stage, in accordance with the scale that has been set by the following rules:

(1) For each male sex will be symbolized by "L" and "P" for each female sex.
(2) For each SBMPTN entry point, it will be symbolized with "1" and "5" for the SNMPTN
(3) Each GPA is adapted from the weight rating based on the 2015 Mulawarman University academic rules with the following range:
    a. For each semester achievement index, from 3.51 to 4.00 becomes A1.
    b. For each semester achievement index starting from 2.76 to 3.50 becomes A2.
    c. For each semester achievement index starting from 2.00 to 2.75 becomes A3.
(4) For the duration of the study period, the adaptation is as follows:
    a. The study period 3 to 4 years becomes B1.
    b. The length of the study period from 4 to 6 years becomes B2.
    c. The study period above 6 years becomes B3
(5) E-Learning evaluation;
    a. Performance
    b. Product quality
    c. Service

Based on the rules that have been established, 389 data are obtained as a result of the transformation of research attributes as seen in the student academic data resulting from the transformation.

3.2. Apriori algorithm

A priori algorithm is the most famous algorithm for finding high-frequency patterns. High-frequency patterns are item patterns in a database that has a frequency or support above a certain threshold which is called a minimum term of support or threshold. A priori algorithm is divided into several stages called iterations. Each iteration produces a high-frequency pattern with the same length starting from the first pass which produces a high-frequency pattern with a length of one.

3.2.1. First candidate. The first step is to scan the database to find support for each item. After the support of each item is obtained, items that have support from the minimum support are selected as high-frequency patterns with a length of 1 or often abbreviated as 1-set items as in Table 3.

Table 3. First candidate

Itemset	Count
A1	109
A2	279
A3	1
B1	20
B2	180
B3	189
L	254
P	135

3.2.2. Second candidate. The second iteration produces 2-item sets that each set has two items. First, a 2-item candidate set is made from a combination of all 1-item sets. Then for each candidate 2-item set, the support is calculated by scanning the database, such as Table 4.

Table 4. Second candidate

Itemset	Count
A1,B1	14
A1,B2	75
A1,B3	20
A2,B1	6
A2,B2	104
A2,B3	169

Itemset	Count
A3,B2	1
A1,L	54
A1,P	55
A2,L	199
A2,P	80
A3,L	1
B1,L	16
B1,P	4
B2,L	101
B2,P	79
B3,L	137
B3,P	52

3.2.3. Third candidate. In this iteration produces 3 items which have year items in it.

Table 5. Third candidate

Itemset	Count
A1,B1,P	3
A1,B1,L	11
A1,B2,P	44
A1,B2,L	31
A1,B3,P	8
A1,B3,L	12
A2,B1,P	1
A2,B1,L	5
A2,B2,L	69
A2,B3,P	44
A2,B3,L	125
A3,B2,L	1
A2,B2,P	35

3.3. Data analysis and calculation

Data analysis and calculations are carried out on 389 existing data, in accordance with the candidates that have been obtained. The calculation process was carried out in this study, to obtain information on student graduation rates, with a calculation model based on association rules method. The calculation method of association rules is done using the following formula

1. Finding *support value*

$$\text{A1,B1,P} = \frac{3}{389} = 0.008$$
$$\text{A1,B1,L} = \frac{11}{389} = 0.028$$
$$\text{A1,B2,P} = \frac{44}{389} = 0.113$$
$$\text{A1,B2,L} = \frac{31}{389} = 0.08$$
$$\text{A1,B3,P} = \frac{8}{389} = 0.021$$
$$\text{A1,B3,L} = \frac{12}{389} = 0.031$$
$$\text{A2,B1,P} = \frac{1}{389} = 0.002$$
$$\text{A2,B1,L} = \frac{5}{389} = 0.013$$
$$\text{A2,B2,P} = \frac{35}{389} = 0.09$$
$$\text{A2,B2,L} = \frac{69}{389} = 0.177$$

A2,B3,P	=	$\frac{44}{389}$	=	0.113
A2,B3,L	=	$\frac{125}{389}$	=	0.321
A3,B2,L	=	$\frac{1}{389}$	=	0.002

2. Finding *Confidence value*

A1,B1,P	=	$\frac{3}{14}$	=	0.214
A1,B1,L	=	$\frac{11}{14}$	=	0.786
A1,B2,P	=	$\frac{44}{75}$	=	0.587
A1,B2,L	=	$\frac{31}{75}$	=	0.413
A1,B3,P	=	$\frac{8}{20}$	=	0.4
A1,B3,L	=	$\frac{12}{20}$	=	0.6
A2,B1,P	=	$\frac{1}{6}$	=	0.167
A2,B1,L	=	$\frac{5}{6}$	=	0.83
A2,B2,P	=	$\frac{35}{104}$	=	0.337
A2,B2,L	=	$\frac{69}{104}$	=	0.663
A2,B3,P	=	$\frac{44}{169}$	=	0.26
A2,B3,L	=	$\frac{125}{169}$	=	0.74
A3,B2,L	=	$\frac{1}{1}$	=	1

3.4. Graduation chart

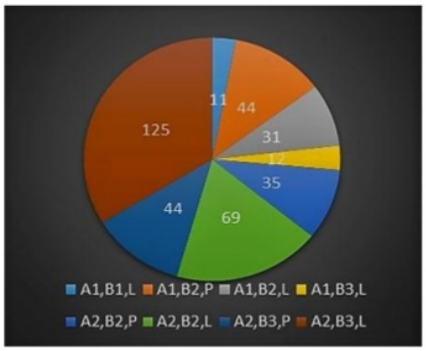

Figure 1. Graduation chart

Based on the Figure 1Graduation chart, the students who graduated most had a GPA interval of 2.76 - 3.50 with the male gender and studied more than 6 years which had a support value of 0.321 and a confidence value of 0.628. Figure 2 describes the coordinates of each category. The next process is to make a chart of IPA quadrants based on these coordinate points.

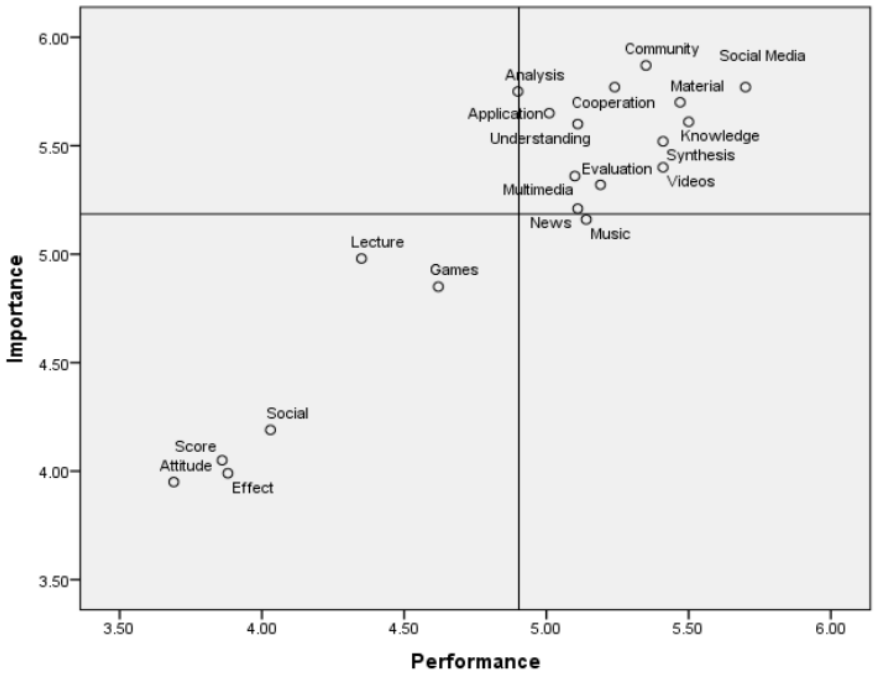

Figure 2. Distibution of IPA quadrants

Based on the results of the four-quadrant IPA graphs in the Figure, It can be concluded the criteria that fall into each quadrant are: Quadrant I has a relatively high level of importance but has lacking performance so it needs to be improved in improving e-learning performance. The category included in this quadrant is Analysis. Quadrant II has a relatively high level of importance with a relatively high level of performance so that it is in accordance with the wishes of the user and must be maintained. The categories included in this quadrant are Community, Collaboration, Materials, Social Media, Knowledge, Synthesis, Application, Understanding, Multimedia, Evaluation, Video, and News. Quadrant III has a relatively low level of importance with a relatively low level of performance so that it does not need to make improvements because it is not a top priority of the user. Categories included in this quadrant are Lecturers, Games, Social, Values, Effects, and Attitudes. Quadrant IV has a relatively low level of importance but has a relatively high level of performance so the university needs to relocate these excess performance resources to other quadrants that need performance improvement. The category in this quadrant is Music.

4. Conclusion

Based on the results of the research that has been submitted, it can be concluded that the highest level of student graduation has a GPA interval of 2.76 - 3.50 with male sex and study period of more than 6 years, which has a support value of 0.321 and a confidence value of 0.628. Based on the results of the analysis that has been done, it can help predict student graduation rates. Students can pay attention to the course of their studies and the Study Program to be able to take preventive actions related to the length of study of each student. According to IPA coordinates, in quadrant II has a high level of importance with a relatively high level of performance and must be maintained. Whereas to produce a better level of accuracy, the variables studied can be added to other data related to students' academic activities during their studies.

Acknowledgments
The research was funded by the Directorate of Research and Community Service of the Directorate General of Research and Development of the Ministry of Research, Technology and Higher Education of the Republic of Indonesia in accordance with the Research Contract No.069 /SP2H/LT/DRPM/2019

References

[1] Pooja T, Anil M and Manisha 2015 *International Journal of Computer Applications* **110** (**15**) 60-68
[2] Suchita B and Rajeswari K 2013 *International Journal of Computer Science and Information Technology* **2** (**7**) 273 – 279
[3] Fadlina 2014 *InformasidanTeknologiIlmiah* (INTI) **3** (**I**)144-154
[4] Sadiq H, Dahan N A, Mutaher Ba-Alwib F, Najoua R 2018 *Indonesian Journal of Electrical Engineering and Computer Science* **9** (**2**) 447~459
[5] Jananto A 2012 *JurnalTeknologiInformasiDinamik* 17(2) Program Studi Sistem Informasi, Universitas Stikubank
[6] Syaifullah A M 2010 *NaskahPublikasi* AMIKOM Yogyakarta
[7] Siti A, Marisi A, Evy S 2015 *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)* **04**(**2**) 127 – 134
[8] Ramadiani et al. 2017 *Procedia Computer Science* (**116**) 373-380
[9] Ramadiani et al. 2016 *International Conference on Engineering and Technology for Sustainable Development (ICET4SD)* Yogyakarta IOP Conf. Series Materials Science and Engineering 012016 105 1-9
[10] Ramadiani et al. 2013 *JSI UI* **9**(**1**) 6-12
[11] Ramadiani R Atan, Selamat M H, Abdullah R, Pa N C and Azainil 2017 *2nd International Conference on Science in Information Technology (ICSITech)* Balikpapan IEEE Xplore (7852626) 158-162
[12] Ramadiani R, Azainil A, Frisca F, Hidayanto A N, Herkules H 2019 *International Journal of Innovation and Learning (IJIL)* **26** (**1**) 1-26
[13] Wong M S, Hideki N, George P 2011 *Journal of Theoretical and Applied Electronic Commerce Research* (**6**)17–30
[14] Silva F, Fernandes P O 2012 *Management* (**6**) 1609-1614

Evaluation of student academic performance using e-learning with the association rules method and the importance of performance analysis

ORIGINALITY REPORT

17%	9%	14%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	digilib.unimed.ac.id Internet Source	2%
2	Wilson Guachamin-Acero, Zhiyu Jiang, Lin Li. "Numerical study of a concept for major repair and replacement of offshore wind turbine blades", Wind Energy, 2020 Publication	2%
3	garuda.ristekbrin.go.id Internet Source	2%
4	iranarze.ir Internet Source	2%
5	Arif Dwi Laksito, Kusrini. "Apriori Algorithm Optimization using Temporary Table", 2019 International Conference on Information and Communications Technology (ICOIACT), 2019 Publication	2%
6	Yennimar Yennimar, Reyhan Achmad Rizal. "Comparison of Machine Learning Classification	1%

Algorithms in Sentiment Analysis Product Review of North Padang Lawas Regency", SinkrOn, 2019
Publication

7	Submitted to Kensington College of Business - Brunei Student Paper	1%
8	compscienceedu.blogspot.com Internet Source	1%
9	Submitted to Universiti Tenaga Nasional Student Paper	1%
10	N.A. Ramadiani, N.A. Azainil, N.A. Frisca, Achmad Nizar Hidayanto, N.A. Herkules. "An integrated model of e-learning continuance intention in Indonesia", International Journal of Innovation and Learning, 2019 Publication	1%
11	Submitted to Binus University International Student Paper	1%
12	Taufan Samudra Akbar, Maman Abdurohman, Aji Gautama Putrada. "Smart Bag Prototype with Apriori Algorithm", 2019 International Symposium on Electronics and Smart Devices (ISESD), 2019 Publication	1%
13	Submitted to Universitas Prima Indonesia	

Student Paper

1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		