

DOI: <http://dx.doi.org/10.21123/bsj.2021.18.3.0554>

Measuring Positive and Negative Association of Apriori Algorithm with Cosine Correlation Analysis

Dewi W Wardani

Informatics Department, Universitas Sebelas Maret, Indonesia.

E-mail: dww_ok@uns.ac.id

ORCID ID: <https://orcid.org/0000-0002-3178-2741>

Received 29/2/2020, Accepted 10/11/2020, Published Online First 21/2/2021, Published 1/9/2021



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

Abstract:

This work aims to see the positive association rules and negative association rules in the Apriori algorithm by using cosine correlation analysis. The default and the modified Association Rule Mining algorithm are implemented against the mushroom database to find out the difference of the results. The experimental results showed that the modified Association Rule Mining algorithm could generate negative association rules. The addition of cosine correlation analysis returns a smaller amount of association rules than the amounts of the default Association Rule Mining algorithm. From the top ten association rules, it can be seen that there are different rules between the default and the modified Apriori algorithm. The difference of the obtained rules from positive association rules and negative association rules strengthens to each other with a pretty good confidence score.

Key words: Apriori, Association-rule-mining, Cosine-correlation-analysis, Data-mining.

Introduction:

Association rule analysis is used to obtain association rules which often appear in dataset (1). The default Association Rule Mining (ARM) finds association rules between items which existed in a transaction which is called positive association rules (PAR). Whereas, the negative association rules (NAR) which shows the association between low-frequent-itemset, is also essential to be analyzed. Its argued that with the proper analysis, NAR will strengthen the positive association rules. Therefore, it will add more advantages to ARM.

To measure the certainty and usability level of a rule, support and confidence score are used. Those measurements are not good enough because confidence is only a conditional probability prediction value of two or more item sets (2). They cannot measure the correlation value between two item sets. It may cause misleading association analysis results and then it can cause more significant problems later on the extended applications.

This weakness can be handled by adding a correlation analysis on ARM (3). Correlation analysis shows the relationship between two or more variables (4). The combining of the correlation analysis will increase the quality of

ARM by producing rules with better correlation item sets.

Previous studies related to this research only concern positive association rules. Some of them are followed by cosine correlation analysis as well (5). There were also some studies which produced PAR and NAR without any correlation analysis (6, 7). The reason for the using correlation analysis is that cosine correlation analysis has a null-invariant characteristic which is useful in analyzing vast data (8). Another work indicated that negative correlation was found but was ignored in mining personality of students (9). In earlier work, the use of negative correlation in process mining was useful to detect fraud (10).

Therefore, in this work will study the measuring PAR and NAR by using Apriori algorithm on the support-confidence framework and cosine correlation analysis. Apriori algorithm is chosen to be modified as the oldest and well known association rule algorithm. This algorithm is simple yet it has been proven to be customized in solving problems. Recent work has used it to help semantic maps to illustrate the imbalance of implications between functions (11). Apriori is also helped to handle rules which are defined by using the

examination of only two τ -dependent tables by implementing nondeterministic information systems Apriori based (12). Road's accident in Dubai can be analyzed successfully by Apriori (13). The investigation pattern of contact can be used to provide optimal node selection in IoT networks's (14). Therefore, if the modification of Apriori algorithm can return significant results, it can be assumed that the modification is a promising approach. The advantages of this research are:

- (i). Enriching data mining methods, especially ARM method with the Apriori algorithm.
- (ii). Improving association rule quality by producing PAR and NAR followed by cosine correlation analysis.

Related Work:

Negative ARM was usually ignored and has got the interest recently (15). The idea, to mining the negative correlation in association rule, aims to reduce the not interesting obtained rules and to achieve more interesting rules (16). One work (17) proposed the approach to reduce the set of PAR and NAR with a low computational cost. One reason is to reduce the computational cost, especially in a vast dataset, multi-objective evolutionary algorithm. The concern to mining negative association is also shown by (9). In the market-basket analysis domain, negative association rules are also needed to identify the pattern of products, to identify whether any conflict among each other or any products compliment to each other's. They also propose an approach which reduces database scans. (18) used the extension of Apriori algorithm in the spatial domain. The spatial mining generates positive and negative frequent itemset which will be applied at a temporal bar at a specific situation. The results surpass the ones which do not find the negative correlation. The other work provides the mining of positive and negative as a service (19). It works on XML data and the other work (20), concerned in mining the negative correlation in infrequent itemset.

In health domain, negative association which is combined with regression is useful in determining the depressive symptoms in middle aged and elderly (21). Negative correlation is also needed in existing biclustering algorithms for microarrays by proposing NBic-ARM (22). The constraint is added to discovering the negative association rule in several domains (23). For all the related work, the measure of negative correlation has not been investigated. Therefore, this work also propose the cosine correlation to measure the negative correlation.

In theoretical aspect, some approaches have

shown as well that negative correlation has improved the performance of a specific method. A method combines negative correlation learning (NCL) with convolution neural network (1-dim CNN) to solve insufficient data problem (24). Negative correlations also helped to relax the variance property in the Mixed Generalized Ordered Response (MGOR) models (25). Modified Neural Network algorithm with adaptive negative correlation (NEA_ANCL) has shown evolutionary result (26). Similar work has shown as well than negative correlation learning neural network which combines with Particle-swarm optimization which can solve denoising problem in wavelet analysis technique (WAT) (27).

The Approach:

A. Positive Association Rules (PAR) and Negative Association Rules (NAR)

PAR ($A \square B$) refers to the association rule between items which exist in a transaction. For example, the association rule that shows which items will be bought together by a customer. Here are some characteristics of PAR (20):

1. A
2. $\text{support}(A \rightarrow B) = \text{support}(A, B) \geq \text{minsup}$
3. $\text{confidence}(A \rightarrow B) = \frac{\text{support}(A, B)}{\text{support}(A)} \geq \text{minconf}$

NAR refers to the association rule between existing items and non-existing items in a transaction. For example, the association rule that shows a customer will not buy both items A and B. This type of association rule is presented as ($\neg A \square B$) or ($A \square \neg B$). Support, confidence, and cosine values are also used to find NAR. $\text{Support}(\neg A) = 1 - \text{support}(A)$. Based on it, $\text{support}(A \square \neg B)$ can be counted as $\text{support}(A \square \neg B) = \text{support}(A, \neg B) = \text{support}(A) - \text{support}(A, B)$

Here are the characteristics of NAR (18):

1. A
2. $\text{support}(A \rightarrow B) = \text{support}(A, B) \geq \text{Minsup}, \text{support}(B) \geq \text{Minsup}$ and $\text{support}(A, B) < \text{Minsup}$
3. $\text{support}(A \rightarrow \neg B) = \text{support}(A, \neg B) \geq \text{Minsup}$
4. $\text{confidence}(A \rightarrow \neg B) = \frac{\text{support}(A, \neg B)}{\text{support}(A)} \geq \text{Minconf}$

Note: minimum support (Minsup), minimum confidence (Minconf)

B. Correlation

Correlation is a statistical measurement of relationship between two or more variables (21). Correlation analysis result shows how strong the correlation between two or more variables is. Correlation rule is valid if its correlation value fulfills the minimum correlation value required. There are some correlation measurements. One of them is cosine correlation. Given two itemsets (A and B), cosine value can be known as follows:

$$\text{cosine}(A,B) = \frac{P(A,B)}{\sqrt{P(A) * P(B)}} \\ = \frac{\text{support}(A,B)}{\sqrt{\text{support}(A) * \text{support}(B)}}$$

Cosine value ranges from 0-1. The bigger its value, the more transactions that consist both itemsets (A and B). The smaller its value, less transaction which consists of both itemsets (A and B) (22). Cosine value is a good correlation indicator because it is null-invariant characteristic. It means that null-transaction will not affect correlation measurement. Null-transaction, a transaction which does not contain any of the tested itemsets, is an important property to measure correlation in a huge transactional database (9).

C. Apriori Algorithm enriched with negative correlation

In this work, Apriori algorithm is used as main method which will be modified. Apriori algorithm will be enriched by negative association rule and cosine correlation analysis for each itemsets whose support values are eligible.

This approach will be done as follows:

1. Generate k-item-sets derived from (k-1)-itemsets.
2. Counting support value of each itemset
 - a. Itemsets which have support \geq Minsup will be counting its cosine value.
 - i. If cosine \geq Mincos, then it will generate 2k-2 distinctive association patterns (antecedent \square consequent).
 - ii. Each rule which has confidence \geq Minconf will be included in PAR.
 - iii. PAR is a positive associative rule which has the entire requirement in support, confidence, and cosine.
 - b. Itemsets which have support $<$ Minsup will be generated into distinctive 2k-2 association patterns.

- i. Each pattern will be used to generate A \square B (Antecedent Negative Rule/ANR) and A \square \neg B (Consequent Negative Rule/CNR).
- ii. Every form which has support \geq Minsup will be counted for its cosine.
- iii. Every form with cosine \geq Mincos will be counted for its confidence.
- iv. If its confidence \geq Minconf then it will be included in NAR (which consists of ANR and CNR). NAR is a negative association rule which has the entire requirement in support, confidence, and cosine.

Note: minimum cosine (Mincos)

As below is the pseudo code of the modified Apriori algorithm:

Input:

D, database of transaction

T, transactions in *D*

Minsup, minimum support value threshold

Mincos, minimum cosine value threshold

Minconf, minimum confidence value threshold

Output:

PAR, positive rules

NAR, negative rules

Step:

```

find 1-frequent itemsets
for (k = 2, Lk-1  $\neq$   $\emptyset$ , k++) do {
    generate candidate Ck which has no infrequent
    itemsets
    for each Ck do {
        if support  $\geq$  Minsup then {
            Lk = Lk + Ck
        }
        generate rules as in pattern
        for each rule do {
            if (cosine  $\geq$  Mincos) then {
                if confidence  $\geq$  Minconf then {
                    PAR = PAR + rule
                }
            }
        }
    }
} else {
    generate rules_pattern
    generate rules {ANR( $\neg$ A  $\square$  B), CNR(A  $\square$   $\neg$ B)}
    for each rule do {
        if (support  $\geq$  Minsup) then {
            if (cosine  $\geq$  Mincos) then {
                if (confidence  $\geq$  Minconf) then {
                    NAR = NAR + rule
                }
            }
        }
    }
}
return PAR and NAR
    
```

The addition on the default ARM using Apriori algorithm for frequent itemsets mining on the support-confidence framework is briefly explained below:

```
1. The enriching with cosine correlation
analysis on positive ARM .
if support  $\geq$  Minsup then {
Lk = Lk + Ck
generate rules as in pattern
for each rule do {
if (cosine  $\geq$  Mincos) then {*
if confidence  $\geq$  Minconf then {
PAR = PAR + rule
} } } }
```

The default Apriori algorithm only considers the value of support and the score of confidence. In this work, the proposed method will enrich with the cosine correlation analysis, after calculating the value of support for itemsets that has support values \geq Minsup (as written in the pseudo-code line with mark “*”). Any itemset that satisfies the provisions of the Mincos value will be for its confidence value. Association rules whose confidence values \geq Minconf is a valid positive association rules.

2. The enriching on NAR mining also includes Cosine correlation analysis.

The default Apriori algorithm does not produce negative association rules. Therefore, this research uses itemsets that does not comply with the minimum support to develop a negative association rules.

```
generate rules_pattern
//Each itemsets are arranged in the pattern of
antecedent - consequent (A, B).
generate rules {ANR( $\neg A \square B$ ), CNR( $A \square \neg B$ )}
//Each pattern is used to generate association of
ANR ( $\neg A \square B$ ) and CNR ( $A \square \neg B$ ).
if (support  $\geq$  Minsup) then
//The next step is similar to the positive association
rules mining, which is to calculate its
support value.
if (cosine  $\geq$  Mincos) then
//Association rules that satisfy the minimum
support provisions will be calculated its
Cosine value.
if conf  $\geq$  Minconf then {
NAR = NAR + rule
}
//If it meets the minimum cosine value, the next step
is to calculate its confidence value. Each
association rules also meet the minimum
cosine value, then it is a valid negative association
rules.
```

Materials and Methods:

Dataset and the experimental settings

As a new idea in adding the negative correlation, the experiments will be more

investigated for real data instead of theoretical aspects. The main aim of experiments is to make sure that the negative correlation can make the obtained rules are stronger compare to without it. Therefore, the experiment simply performed the variety of normal experiments. In the future work the experiments can move to the more comprehensive experiments which consider more aspects.

The experiments are conducted by extracting PAR and NAR in the support-confidence framework with Cosine correlation analysis using Apriori algorithm for frequent itemsets mining. This research will also perform PAR mining using the default ARM algorithm with Apriori algorithm for frequent itemsets mining on a support-confidence framework. Roberto Bayardo from the UCI dataset prepares the data. The domain is mushroom data that can be accessed in this repository (<http://fimi.ua.ac.be/data/>). There are 8124 records with 23 fields in each file. The whole album is comprised of over 119 different items.

The scenario of experiments is by performing the variation of a minimum score of support, confidence and the cosine. The Minsup is 30%, 40% and 50%. The Minconf is 50%, 60%, 70% and 80%. The Mincos is 50%, 60%, 70% and 80%. The 1st scenario uses Minsup 30% and Minconf 50%, the 2nd scenario uses Minsup 30% and Minconf 60%, and the next scenarios are the combination of those three scores respectively.

Results

In this part, the details of the experiment's results are explained. The test will show the top ten obtained rules for each scenario (the variation of. Each scenario covers of PAR, ANR and CNR. In case that the results in a plot are not much different compared to the other scenario, the specific outcome of PAR, ANR and CNR will not be shown in detail. In the end, the analysis of results will be discussed in the next subsection. The minimum number of cosine will be used only for ANR and CNR.

From the 1st scenario in Tables 1- 4, the PAR obtained in the three tops, 86 \square 85, 34 \square 85 and 34, 86 \square 85. The obtained consequent of ANR and CNR are enriching the overall results, such as $\neg 58 \square 86$, 56 $\square \neg 94$. There are no results of ANR and CNR, which is contradictive comparing to the outcome of PAR. There is also a rule that pretty interesting from CNR, along with the increasing score of Mincos. It obtained 1 $\square \neg 28$ (if it is poisonous, it has aroma) and 28 $\square \neg 1$ (if it doesn't have aroma, it isn't poisonous). It seems CNR supports the confidence of the results. The obtained

results of PAR, ANR and CNR are not much different comparing to the 1st scenario.

Table 1. The Number of Rules with Minsup = 30% and Minconf = 50%

ARM Algorithm		Number of Association Rule			
Default		20.194			
Modified	Positive	Mincos=50% 20.194	Mincos=60% 17.292	Mincos=70% 11.637	Mincos=80% 3.020
	Negative	385	240	122	58

Table 2. Top Ten PAR (A □ B) With Minsup = 30% And Minconf = 50%

Default Apriori		Modified Apriori							
		Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%	
A	B	A	B	A	B	A	B	A	B
86	85	86	85	86	85	85	85	85	85
34	85	34	85	34	85	34	85	34	85
34,86	85	34,86	85	34, 86	85	34,86	85	34,86	85
90	85	24,53	86,90,94	24,53	86,90,94	24,53,24,3	86,90,94	24,53,24,34,5	86,90,94
						4,53		3	
34,90	85	24,34,53	86,90,94	24,34,53	86,90,94	24,53	86,90,94	24,53	86,90,94
86,90	34	24,53	34,86,90,94	24,53,85	34,86,90,94	24,53,85	34,86,90,94	24,53,85	34,86,90,94
86,90	85	24,53,85	86,90,94	24,53,	86,90,94	24,53	86,90,94	24,53	86,90,94
34,	85	24,53	86,90,94	24,34,53,	86,90,94	24,34,53,8	86,90,94	24,34,53,85	86,90,94
86,90				85		5			
85,	34	24,34,53,	85,86,90,	24,34,53	85,86,90,9	24,34,53	85,86,90,9	24,34,53,85	85,86,90,94
86,90		85	94		4		4		94
86,90	34,	24,34,53	85,86,90,	24,34,53	85,86,90,9	24,34,53	85,86,90,9	24,34,53	85,86,90,94
	85		94		4		4		94

The obtained results of the 2nd (Table 5), 3rd (Table 6) and 4th (Table 7) scenarios are similar comparing to the results of the 1st scenario. The order of the obtained results is not different either.

Table 3. Top Ten ANR (¬A □ B) With Minsup= 30% And Minconf = 50%

Modified Apriori									
Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%			
A	B	A	B	A	B	A	B	A	B
58	86	58	86	58	86	58	86	58	86
28,59,90	86	28,59,90	86	28,59,90	86	28,59,90	86	28,59,90	86
9	86	9	86	9	86	9	86	9	86
28,59,63	86	28,59,63	86	28,59,63	86	28,59,63	86	28,59,63	86
2,63,90,93	86	2,63,90,93	86	2,63,90,93	86	2,63,90,93	86	2,63,90,93	86
58	34	58	34	58	34	58	34	58	34
28,59,90	34	28,59,90	34	28,59,90	34	28,59,90	34	28,59,90	34
9	34	9	34	9	34	9	34	9	34
28,59,63	34	28,59,63	34	28,59,63	34	28,59,63	34	28,59,63	34
2,63,90,93	34	2,63,90,93	34	2,63,90,93	34	2,63,90,93	34	2,63,90,93	34

Table 4. Top Ten CNR (A □ ¬B) With Minsup= 30% And Minconf = 50%

Modified Apriori									
Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%			
A	B	A	B	A	B	A	B	A	B
56	94	56	94	56	94	56	94	56	94
94	56	94	56	94	56	94	56	1	28
1	28	1	28	1	28	1	28	28	1
28	1	28	1	28	1	28	1	24	93
23	94	23	94	23	94	23	94	93	24
23	58	23	58	23	58	23	58	90	58
94	23	94	23	94	23	94	23	86	8
52	94	52	94	52	94	52	94	86	28,59,90
94	52	94	52	93	58	93	58	34	58
93	58	93	58	56	9	56	9	34	28,59,90

Table 5. The Number of Rules with Minsup = 30% and Minconf = 60%

ARM Algorithm		Number of Association Rule			
Default		16.986			
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	16.986	14.606	10.217	3.020
		241	205	117	58

Table 6. The Number of Rules with Minsup = 30% and Minconf = 70%

ARM Algorithm		Number of Association Rule			
Default		13.513			
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	13.513	11.253	8.193	2.891
		111	104	83	44

Table 7. The Number of Rules with Minsup = 30% and Minconf = 80%

ARM Algorithm		Number of Association Rule			
Default		13.513			
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	10.036	7.803	5.568	2.150
		59	59	56	33

The 5th scenario which is shown in Tables 8 - 11 return pretty similar results as well comparing to the 1st scenario, that ANR and CNR support the confidence of an overall obtained rules. The rule 1 □

¬28 (if it is poisonous, it has aroma) and 28 □ ¬1 (if it doesn't have aroma, it isn't poisonous) is still obtained as well.

Table 8. The Number of Rules with Minsup = 40% and Minconf = 50%

ARM Algorithm		Number of Association Rule			
Default		2.211			
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	2.211	2.211	1.425	104
		74	74	70	6

Table 9. Top Ten PAR (A □ B) With Minsup = 40% And Minconf = 50%

Default Apriori				Modified Apriori					
		Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%	
A	B	A	B	A	B	A	B	A	B
86	85	86	85	86	85	86	85	86	85
34	85	34	85	34	85	34	85	34	85
34,86	85	34,86	85	34,86	85	34,86	85	34,86	85
90	85	90	34	90	34	90	34	90	34
34,90	85	86,90	34	86,90	34	86,90	34	86,90	34
86,90	34	85,86,90	34,85	85,86,90	34,85	85,86,90	34,85	85,86,90	34,85
86,90	85	86,90	85	86,90	85	86,90	85	86,90	85
34, 86,90	85	34,90	85	34,90	85	34,90	85	34,90	85
85, 86,90	34	86,90	85	86,90	85	86,90	85	86,90	85
86,90	34,85	34,86,90	85	34,86,90	85	34,86,90	85	34,86,90	85

Table 10. Top Ten ANR (¬A □ B) With Minsup= 40% And Minconf = 60%

Modified Apriori									
Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%			
A	B	A	B	A	B	A	B		
52	90	52	90	52	90	93	24		
2,59,90	86	2,59,90	86	2,59,90	86	28	1		
59,63,93	86	59,63,93	86	59,63,93	86	1	28		
2,63	86	2,63	86	2,63	86				
2,28	86	2,28	86	2,28	86				
59,90,93	86	59,90,93	86	59,90,93	86				
2,59,90	34	2,59,90	34	2,59,90	34				
59,63,93	34	59,63,93	34	59,63,93	34				
2,63	34	2,63	34	2,63	34				
2,28	34	2,28	34	2,28	34				

The obtained results of the 6th, (Table 12) 7th (Table 13) and 8th (Table 14) scenarios are similar comparing to the results of the 5th scenario.

The order of the obtained results is not different, either

Table 11. Top Ten CNR (A \square -B) With Minsup= 40% And Minconf = 50%

Modified Apriori							
Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%	
A	B	A	B	A	B	A	B
1	28	1	28	1	28	1	28
28	1	28	1	28	1	28	1
1	23	1	23	1	23	24	93
24	93	24	93	24	93		
59	1	59	1	59	1		
24	2	24	2	24	2		
63	1	63	1	63	1		
24	28	24	28	24	28		
90	52	90	52	90	52		
86	2,59,90	86	2,59,90	86	2,59,90		

Table 12. The Number of Rules with Minsup = 40% and Minconf = 60%

ARM Algorithm		Number of Association Rule			
Default		1.820			
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	1.820	1.820	1.124	104
		53	53	50	6

Table 13. The Number of Rules with Minsup = 40% and Minconf = 70%

ARM Algorithm		Number of Association Rule			
Default		1.569			
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	1.569	1.569	897	98
		41	41	39	6

Table 14. The Number of Rules with Minsup = 40% and Minconf = 80%

ARM Algorithm		Number of Association Rule			
Default		1.341			
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	1.341	1.341	693	68
		39	39	38	6

The result of the 9th scenario is shown in Tables 15 - 18. The similar results are obtained in the 10th (Table 19), 11th and 12th scenarios. The exact the same results as shown in Table 20 came out in 11th and 12th scenarios included the obtained

number of rules. Therefore, the received amount of combination (50%, 80%) the same as the combination (50%, 70%). In overall, the results are consistent

Table 15. The Number of Rules with Minsup = 50% and Minconf = 50%

ARM Algorithm		Number of Association Rule			
Default		656			
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	656	656	656	56
		0	0	0	0

Table 16. Top Ten PAR (A □ B) With Minsup = 50% And Minconf = 50%

Default Apriori		Modified Apriori							
		Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%	
A	B	A	B	A	B	A	B	A	B
86	85	86	85	86	85	86	85	86	85
34	85	34	85	34	85	34	85	34	85
34,86	85	34,86	85	34,86	85	34,86	85	34,86	85
90	85	90	34	90	34	90	34	90	34
34,90	85	86,90	34	86,90	34	86,90	34	86,90	34
86,90	34	85,86,90	34,85	85,86,90	34,85	85,86,90	34,85	85,86,90	34,85
86,90	85	86,90	85	86,90	85	86,90	85	86,90	85
34, 86,90	85	34,90	85	34,90	85	34,90	85	34,90	85
85, 86,90	34	86,90	85	86,90	85	86,90	85	86,90	85
86,90	34,85	34,86,90	85	34,86,90	85	34,86,90	85	34,86,90	85

Table 17. Top Ten ANR (¬A □ B) With Minsup= 50% And Minconf = 50%

Modified Apriori							
Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%	
A	B	A	B	A	B	A	B
-	-	-	-	-	-	-	-

Table 18. Top Ten CNR (A □ ¬B) With Minsup= 50% And Minconf = 50%

Modified Apriori							
Mincos=50%		Mincos=60%		Mincos=70%		Mincos=80%	
A	B	A	B	A	B	A	B
-	-	-	-	-	-	-	-

Table 19. The Number of Rules with Minsup = 50% and Minconf = 60%

ARM Algorithm	Number of Association Rule				
Default	436				
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	436	436	436	56
		0	0	0	0

Since PAR is the main goal in Apriori, it also resumes the difference of the PAR obtained results, shown in Table 21. Implicitly, it is shown

that the modified Apriori does not ruin the default one

Table 20. The Number of Rules with Minsup = 50% and Minconf = 80%

ARM Algorithm	Number of Association Rule				
Default	356				
Modified	Positive	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
	Negative	356	356	356	56
		0	0	0	0

Table 21. The Number of Rules with Minsup = 50% and Minconf = 80%

Default Apriori		Modified Apriori			
Minsup	Minconf	Mincos=50%	Mincos=60%	Mincos=70%	Mincos=80%
30	50	0	14.370	42.373	85.045
	60	0	14.0115	39.850	82.220
	70	0	16.724	39.3694	78.605
	80	0	22.249	44.519	78.577
40	50	0	0	35.549	95.296
	60	0	0	38.241	94.285
	70	0	0	42.829	93.753
	80	0	0	48.322	93.586
50	50	0	0	0	91.463
	60	0	0	0	87.155
	70	0	0	0	84.269
	80	0	0	0	84.269

A. Evaluation and Discussion

This evaluation is carried out to analyze the results of experiments that have been done. The experimental results will be used as the analysis of the association rules generated, the number of association rules made, and by looking at the top ten results produced by each algorithm. The results of this analysis will show the obtained effects by the embedding of a NAR mining and Cosine correlation analysis.

There are differences between the results of association rules in the default Apriori and the modified Apriori algorithm. Several association rules generated by the default Apriori algorithm are not included in the results of the modified Apriori for not meeting the minimum cosine terms. The larger the minimum cosine value is given, the more different results have occurred. The more significant the minimum cosine value is given, the more significant the difference between the numbers of PAR generated by the default Apriori and the number of PAR made by the modified Apriori. The number of NAR only produced by the modified Apriori algorithm continues to decrease for the increasing of the minimum cosine value at various support values (30%, 40%, 50%) and confidence (50%, 60%, 70%, 80%).

The experimental results provide PAR and NAR which can be used to enrich the knowledge of an item. They also indicate a mutually reinforcing relationship between association rules which are equally included in the top ten rules of association of each type. By implementing with various minimum support, cosine, and confidence, there are some association rules which always been at a top 3 on the results of positive association rules mining. They are:

- $86 \square 85$ (if the veil is white, then the veil type is partial),
- $34 \square 85$ (if the attachment gills are free, then the veil type is partial), and
- $34,86 \square 85$ (if the veil is white and attachments gills are free, then the veil type is partial).

The minimum support = 30% and minimum confidence = 50%, 60%, 70%, 80% produce NAR of ANR $\neg 58 \square$ (if there is rootstalk, then the veil is white). While with the minimum support = 30% and minimum confidence = 50%, 60%, 70% found NAR in the form of CNR $86 \square \neg 58$ (if the veil is white, then there is rootstalk). Both association rules are included in the top ten results from each type. The relationship between the two items (86 and 58) further confirms that the relationship between them and additional knowledge about item

86, which was included in the top ten results of PAR mining.

NAR in the form of CNR at minimum support = 30% and 40% resulted in two mutually reinforcing association rules, namely $1 \square \neg 28$ (if the mushroom is poisonous, then it has no aroma) and $28 \square \neg 1$ (if mushroom has no aroma, then it is not poisonous). In addition, the NAR in the form of ANR on minimum support = 40% yield $\neg 1 \square 28$ (if mushroom is not poisonous, then it has no aroma) and $\neg 28 \square 1$ (if it has aroma, then it is poisonous), which further strengthens the relationship between item 1 and 28.

There are some association rules which fulfill support requirement, but they do not satisfy the cosine requirement. It caused the different sequence of association rules produced by the default and the modified Apriori algorithm. The series in the association rules of the negligence Apriori and the modified Apriori on various minimum cosine values resulting from the association rules that do not surpass the minimum cosine value.

Conclusions:

This work have enriched the defaults Apriori algorithm with negative ARM. The aim is to enhance the results of the obtained rules. Therefore, it will strengthen the analysis. This work use Cosine correlation analysis to extend the result with negative association rule. The results of the series of experiments against mushroom database show that this work obtains different results of association rules mining. The difference is both in terms of quantity and sequence between the default and the modified ARM algorithm. The obtained results, both PAR and NAR, strengthen to each other with a pretty good confidence score. The near future work is to study the performance of ARM algorithm in more massive datasets.

Author's declaration:

- Conflicts of Interest: None.
- I hereby confirm that all the Figures and Tables in the manuscript are mine. Besides, the Figures and images, which are not mine, have been given the permission for re-publication attached with the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in Universitas Sebelas Maret.

References:

1. Kaur M, Kang S. Market Basket Analysis: Identify the changing trends of market data using association rule mining. *Procedia Comput Sci.* 2016;85:78–85.

2. Ohri M, Thakur K. An Enhanced Apriori and Improved Algorithm for Association Rules. 2016.
3. Saito H, Monden A, Yücel Z. Extended association rule mining with correlation functions. In: 2018 IEEE International Conference on Big Data, Cloud Computing, Data Science & Engineering (BCD). IEEE; 2018. p. 79–84.
4. Gogtay N, Thatte U. Principles of correlation analysis. *J Assoc Physicians India*. 2017;65(3):78–81.
5. Said AM, Dominic P, Zailani S. A new scheme for extracting association rules: market basket analysis case study. *Int J Bus Innov Res*. 2011;6(1):28–46.
6. Bagui S, Dhar PC. Mining positive and negative association rules in Hadoop's MapReduce environment. In: Proceedings of the ACMSE 2018 Conference. 2018. p. 1–1.
7. Li C, Hao F, Zhao L, Song L, Dong X. Analysis of medical and healthcare data based on positive and negative association rules. In: 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). IEEE; 2017. p. 1559–1564.
8. Han J, Pei J, Kamber M. Data mining: concepts and techniques. Elsevier; 2011.
9. Gara GPP, Padoa FRF. Mining Association Rules on Students Profiles and Personality Types. In: Proceedings of the International Multiconference of Engineers and Computer Scientists. 2015.
10. Sarno R, Dewandono RD, Ahmad T, Naufal MF, Sinaga F. Hybrid Association Rule Learning and Process Mining for Fraud Detection. *IAENG Int J Comput Sci*. 2015;42(2).
11. Zeng N, Xiao H. Inferring implications in semantic maps via the Apriori algorithm. *Lingua*. 2020 Feb 1:102808.
12. Sakai H, Nakata M, Watada J. NIS-Apriori-based rule generation with three-way decisions and its application system in SQL. *Inf Sci*. 2020;507:755–771.
13. John M, Shaiba H. Apriori-Based Algorithm for Dubai Road Accident Analysis. *Procedia Comput Sci*. 2019;163:218–227.
14. Redhu S, Hegde RM. Optimal relay node selection in time-varying IoT networks using apriori contact pattern information. *Ad Hoc Netw*. 2020;98:102065.
15. Kadir ASA, Bakar AA, Hamdan AR. Frequent absence and presence itemset for negative association rule mining. In: Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on. IEEE; 2011. p. 965–970.
16. Piao X, Wang Z, Liu G. Research on mining positive and negative association rules based on dual confidence. In: Internet Computing for Science and Engineering (ICICSE), 2010 Fifth International Conference on. IEEE; 2010. p. 102–105.
17. Martin D, Rosete A, Alcalá-Fdez J, Herrera F. A new multiobjective evolutionary algorithm for mining a reduced set of interesting positive and negative quantitative association rules. *IEEE Trans Evol Comput*. 2014;18(1):54–69.
18. Shaheen M, Shahbaz M, Guergachi A. Context based positive and negative spatio-temporal association rule mining. *Knowl-Based Syst*. 2013;37:261–273.
19. Çokpınar S, Gündem Tİ. Positive and negative association rule mining on XML data streams in database as a service concept. *Expert Syst Appl*. 2012;39(8):7503–7511.
20. Wu X, Zhang C, Zhang S. Efficient mining of both positive and negative association rules. *ACM Trans Inf Syst TOIS*. 2004;22(3):381–405.
21. Zhou Q, Wang T, Basu K. Negative association between BMI and depressive symptoms in middle aged and elderly Chinese: Results from a national household survey. *Psychiatry Res*. 2018;269:571–578.
22. Houari A, Ayadi W, Yahia SB. Mining negative correlation biclusters from gene expression data using generic association rules. *Procedia Comput Sci*. 2017;112:278–287.
23. Jabbour S, El Mazouri FE, Sais L. Mining Negatives Association Rules Using Constraints. *Procedia Comput Sci*. 2018;127:481–488.
24. Xu L, Zhu D, Chen X, Li L, Huang G, Yuan L. Combination of one-dimensional convolutional neural network and negative correlation learning on spectral calibration. *Chemom Intell Lab Syst*. 2020 Apr 15;199:103954.
25. Balusu SK, Pinjari AR, Mannering FL, Eluru N. Non-decreasing threshold variances in mixed generalized ordered response models: A negative correlations approach to variance reduction. *Anal Methods Accid Res*. 2018;20:46–67.
26. Sheng W, Shan P, Chen S, Liu Y, Alsaadi FE. A niching evolutionary algorithm with adaptive negative correlation learning for neural network ensemble. *Neurocomputing*. 2017;247:173–182.
27. Ma T, Wang C, Wang J, Cheng J, Chen X. Particle-swarm optimization of ensemble neural networks with negative correlation learning for forecasting short-term wind speed of wind farms in western China. *Inf Sci*. 2019;505:157–182.

قياس الارتباط الايجابي والسلبي بخوارزمية Apriori مع تحليل ارتباط جيب التمام

ديوي ورداني

قسم المعلوماتية، جامعة الاحد عشر من مارس، اندنوسيا.

الخلاصة:

يهدف هذا العمل إلى معرفة قواعد الارتباط الإيجابية وقواعد الارتباط السلبية في خوارزمية (Apriori) باستخدام تحليل ارتباط جيب التمام. يتم تطبيق الخوارزمية الافتراضية وخوارزمية استخراج قواعد الارتباط المعدلة على قاعدة بيانات الفطر لمعرفة الفرق في النتائج. أظهرت النتائج التجريبية أن خوارزمية استخراج قواعد الارتباط المعدلة يمكن أن تولد قواعد ارتباط سلبية. وتعطي إضافة تحليل ارتباط جيب التمام قدرًا أصغر من قواعد الارتباط عما هو من كميات خوارزمية استخراج قواعد الارتباط الافتراضية. من خلال قواعد الارتباط العشرة الأولى، يمكن ملاحظة وجود قواعد مختلفة بين الخوارزمية الافتراضية وخوارزمية Apriori المعدلة. إن اختلاف القواعد التي تم الحصول عليها من قواعد الارتباط الإيجابية وقواعد الارتباط السلبية يقوي بعضها البعض بدرجة جيدة جدًا.

الكلمات المفتاحية: Apriori، استخراج قواعد الارتباط، استخراج البيانات، تحليل ارتباط جيب التمام.