EasyChair Preprint

№ 11140

Mamdani Fuzzy on Improving the Minimum Confidence Final Value Accuracy in the Apriori Algorithm

Erman Sibarani, Erna Budhiarti Nababan and Syahril Efendi

October 23, 2023

Mamdani Fuzzy on Improving the Minimum Confidence Final Value Accuracy in the Apriori Algorithm

1nd Erman Sibarani
Faculty of Computer Science and Technology Information
University of Sumatera Utara
Medan, Indonesisa
ermansibarani@students.usu.ac.id

2nd Erna Budhiari Nababan
Faculty of Computer Science and Technology Information
University of Sumatera Utara
Medan, Indonesisa
ernabrn@usu.ac.id *

3nd Syahril Efendi
Faculty of Computer Science and Technology Information
University of Sumatera Utara
Medan, Indonesisa
syahril1@usu.ac.id

***Abstract*— Finding rules from data is an active area of research in artificial intelligence. Active research is carried out based on regular exploration of item sets, eventually leading to the creation of rules. Many algorithms provide suggestions to fix different problems. Many optimization methods aim to minimize frequent itemset generation and analysis time for data reduction. The apriori algorithm is a data mining algorithm used to analyze databases based on their frequency, based on an association rule learning system. In this paper, we modified the apriori algorithm in such with fuzzy Mamdani for generate package items within a minimum support value. The generated item sets can also help the decision maker to forming new packages for the customers. At the same time, the improved Apriori algorithm has a great improvement in operating efficiency when it conducts association rule analysis on the data set with a large amount of item indexes. Fuzzy Mamdani is here as a supporting method that can increase the accuracy of confidence values and produce a model with the hope of being able to eliminate itemset based on the a priori algorithm. The proposed procedure can reduce the time consumption of Apriori based association rule mining by up to 50% while still maintaining substantial similarity in output, contingent upon user input.**

Keywords—Apriori Algorithm, Confidence Value, Fuzzy Mamdani

I. INTRODUCTION

The apriori algorithm is a well-known classic algorithm for mining association rules [1]. Apriori is the most well-known association rule algorithm, but according to [5], The conventional Apriori algorithm consumes too much time and space for the 2nd iteration of the frequent itemset generation. For example, given 104 single frequent items, the candidate set for the 2nd iteration is almost 108. Checking the occurrence of such a large candidate set for each transaction is impossible. The model or shape produced by this algorithm has a very important influence on the determination of items that will have a relationship with other items, namely the confidence value. Therefore we must ensure the accuracy of the confidence value produced, by increasing the accuracy of the confidence value, the final results will be more accurate, and with that, the machine will study models that have very accurate results.

Mamdani Fuzzy is a method that is easy to implement and has been widely used in various fields such as decision support systems, artificial intelligence, and so on [7]. This method is quite effective in classifying. In addition, this method is also equipped with the rule-based generation which can be optimized to increase the speed and accuracy of results.

The purpose of the research conducted was to increase the accuracy value of the Apriori method by classifying it using the Fuzzy Mamdani method for the resulting confidence value and analyzing the results of a comparison between the Apriori combination method and the conventional Apriori method. Producing a system that can improve the accuracy of the final results of the Apriori algorithm with the help of the Mamdani Fuzzy method. Generate a model that can be learned by a computer system against sales associate rules.

II. PROPOSED FEATURE: MAMDANI FUZZY ON IMPROVING THE APRIORI ALGORITHM

Review of some articles related to Mamdani's a priori algorithm and fuzzy algorithm. Implemented prefix element set and compression matrix to optimize MapReduce based Apriori algorithm on Hadoop. This study explains how to optimize MapReduce-based Apriori algorithm on Hadoop by applying prefix element set and compression matrix. In this case, the researcher uses the prefix - itemset to solve the problem at the connection stage, then the researcher provides an item transaction matrix to store transaction data and transfer it to internal memory., then combine that item's properties. transaction matrix with frequent itemset property to compress the item transaction matrix. With the implementation of the algorithm, the size of the matrix will gradually decrease. Next, the researcher uses the intersection strategy to quickly calculate the support level for the commodity. Finally, the researchers refined the MapReduce-based optimized Apriori algorithm to be able to handle very large data [9].

The study is titled Analysis of Methicillin-Resistant Staphylococcus Aureus Using Apriori, Dbscan and K-Means Algorithms. This study uses three data mining techniques: Apriori algorithm, DBSCAN algorithm and K-means algorithm. Based on this research, the a priori algorithm shows that leucine is dominant in the genome, traits such as pathogens that cause malaria, dengue fever, and lung cancer. [8].

Fuzzy logic is an extension of traditional reasoning, in which x is either a member of set A or not, or x can be a member of set A with a certain degree of membership. The ability of fuzzy models to map fuzzy values is why in many

cases we use fuzzy inference models that use fuzzy values to produce an explicit or definite result. determined. [3].

Fuzzy Mamdani is one branch of artificial intelligence developed to help optimal decision-making from several alternatives. Fuzzy is used to resolve situations where there is uncertainty about whether something is right or wrong. This situation is often a situation that occurs in the real world [4].

The Mamdani method is often called the Max-Min method. This method was introduced by Ebrahim Mamdani in 1975. To get the result, it is necessary to perform 4 steps:

1. Forming the fuzzy set, in the Mamdani method, the input variables and output variables are divided into one or more fuzzy sets.
2. Implicit application function, in the Mamdani method, the implication function used is min.
3. The arrangement of rules, unlike monotonic inference, if the system is made up of many rules, draws inferences from the set and correlation between the rules. There are 3 methods used to achieve the fuzzy inference system: max, additive, and probabilistic.
4. Defuzzification is a fuzzy set derived from the combination of fuzzy rules, while the results obtained are in the form of numbers in the domain of the fuzzy set. So, if we give it a fuzzy set within a certain range, it will be able to get a certain crisp value at output. [4].

III. RESEARCH METHODOLOGY

Research methodology is structured concepts about rules, activities, and procedures used by researchers in telling how the research process is carried out. The research method owned by the researcher is about the steps and processes experienced during the research. To explain the process step by step in the sub-chapters of this research. In general, it can be observed in Figure 1 as follows:

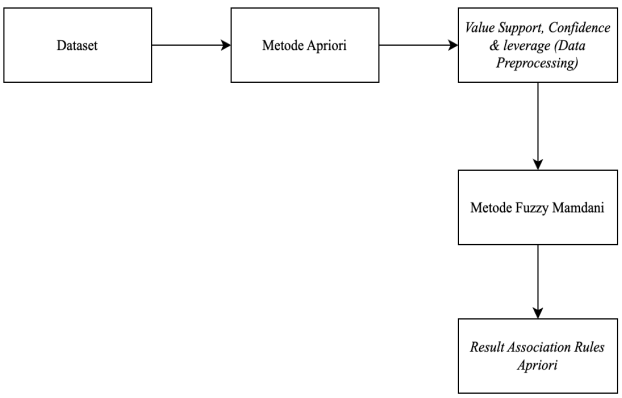

Figure 1 Research Design

In the picture above, the process of the research stages is as follows:

1. Adjust the dataset according to the needs, namely the product sales dataset.
2. Then the dataset will be processed using the Apriori method by dividing into 2 (two) data groups, namely Training Data and Testing Data.
3. The resulting Confidence value will be classified using the Fuzzy Mamdani method aimed at increasing the accuracy of the results produced.

IV. RESULT AND DISCUSSION

The process steps of the Apriori algorithm using the Association method can be described in the form of a flowchart diagram as shown in Figure 3.2 below:

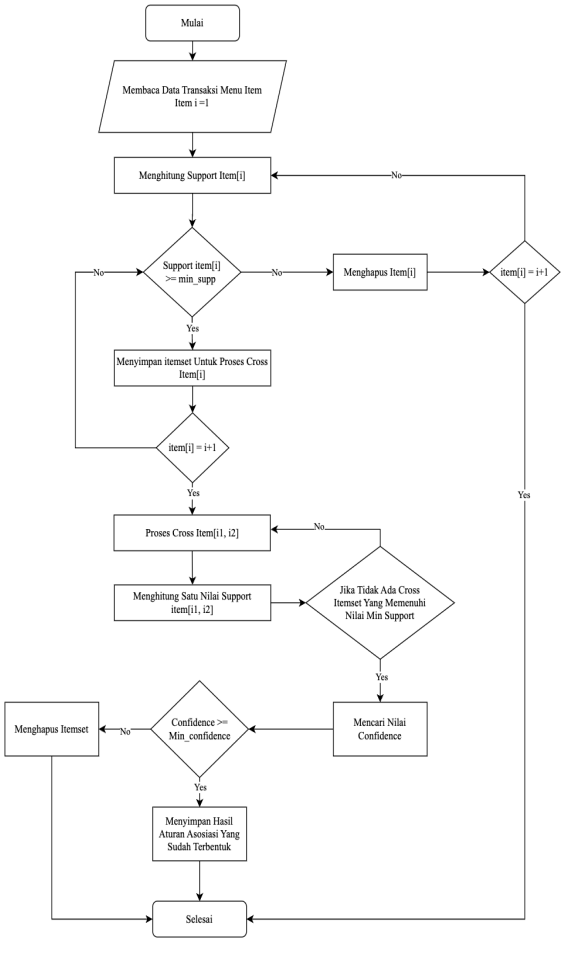

Figure 2 Apriori Algorithm

The data mining technique used in this study is the association technique. System modeling aims to find association rules where the resulting association rules will be used as a reference in determining export service sales promotion activities. At this stage, to find out the performance of the proposed method, a dataset is used, namely transaction data sourced from *https://www.kaggle.com/code/ozlemilgun/market-basket-analysis-with-a priori algorithm/data.*

The dataset used has 541,921 records and 8 attributes. In this case the author provides 5 examples of data records in tabular form. To find out the form of data used can be observed in Table 1: As an example, the process of extracting association rules will be carried out with the assumption that the minimum support is 0.01% and the minimum confidence is 0.05%.

TABLE 1 READ DATA TRANSACTION

No	Description	Quantity	...	Customer ID

1	Wrap Red Apples	25	...	17850
2	Save The Planet Cotton Tote Bag	12	...	17850
3	Coffee Mug Cat + Bird Design	8	...	17850
4	Save The Planet Cotton Tote Bag	6	...	17850
5	Office Mug Warmer Choc Blue	6	...	17850
...	...	...	...	...
541920	Poppy's Playhouse Bedroom	3	...	12680

Iteration 1 starts by training candidate 1 – itemset (C1) from the transaction data and calculating the support amount. The support calculation is the number of occurrences of the item in the transaction divided by the total number of transactions. The minimum support level specified is 0.01%, after which entries with a support value less than 0.01% will be removed. The resulting large itemset 1 (L1) can be seen in Table 3 below:

Support (Wrap Red Apples) = (Number of Transactions Contains Wrap Red Apples / Total Transactions) * 100

Support (Wrap Red Apples) = (68 / 541920) x 100 = 0,0125

TABLE 2 CANDIDATE 1 – ITEMSET (C1)

No	Item Set	Total	Support %
1	Wrap Red Apples	680	0,0125 %
2	Save The Planet Cotton Tote Bag	700	0,0129 %
3	Coffee Mug Cat + Bird Design	400	0,0074 %
4	Save The Planet Cotton Tote Bag	350	0,0065 %
5	Office Mug Warmer Choc Blue	600	0,0111 %

The specified minimum support is 0.01%, then items that have a support value of less than 0.01% are removed. The resulting large-itemset 1 (L1) can be observed in table 3 below:

TABLE 3 LARGE – ITEMSET (L1)

No	Item Set	Total	Support %
1	Wrap Red Apples	680	0,0125 %
2	Save The Planet Cotton Tote Bag	700	0,0129 %
5	Office Mug Warmer Choc Blue	600	0,0111 %

In the second iteration, cross element L1 to form C2 (candidate element set containing 2 elements) and calculate its support. For contestants with the same item one counts, for example when item set {Wrap Red Apples} is combined with item set {Wrap Red Apples} the result is only {Wrap Red Apples} and not Wrap Red Apples, Wrap Red Apples.}. Then the next iteration with the same steps as iteration 1:

1. Calculate Support Value

$$Support(X \rightarrow Y) = \frac{Number(X,Y)}{Num(All\ Samples)}$$

$$Support(Wrap\ Red\ Apples, Save\ The\ Planet\ Cotton\ Tote\ Bag)$$

$$= \frac{Transaction\ Count \rightarrow\ Support}{Total\ Transaction}\ x\ 100\%$$

Support (Wrap Red Apples, Save The Planet Cotton Tote Bag)

= (65/ 541920*) x 100 % = 0,0129 %*

2. Calculate Confidence Value

$$Confidence(X \rightarrow Y) = \frac{Support(X \rightarrow Y)}{P(X)}$$

$$Confidence\begin{pmatrix} Wrap\ Red\ Apples, \\ Save\ The\ Planet\ Cotton\ Tote\ Bag \end{pmatrix}$$

$$= \frac{Transaction\ Count \rightarrow\ Support}{Count\ of\ Related\ Transactions}\ x\ 100\%$$

Support (Wrap Red Apples, Save The Planet Cotton Tote Bag)

= 70/ 680 *= 0,102*

3. Calculate Lift Value

$$Lift(X \rightarrow Y) = \frac{Confidence(X \rightarrow Y)}{P(X)}$$

$$Lift(X \rightarrow Y) = \frac{0{,}102}{680}\ x\ 100\% = 0{,}015$$

TABLE 4 CANDIDATE 2 – ITEMSET (C2)

No	Item Set	Total Transaction	Support %	Confidence	Lift
1	Wrap Red Apples, Save The Planet Cotton Tote Bag	70	0,0129	0,102	0,015
2	Wrap Red Apples, Office Mug Warmer Choc+Blue	50	0,0092	0,073	0,01
3	Save The Planet Cotton Tote Bag, Office Mug Warmer Choc+Blue	48	0,0089	0,07	0,011

The minimum support level specified is 0.01%, after which entries with a support value less than 0.01% will be removed. The resulting large itemset 2 (L2) can be seen in Table 5 below:

TABLE 5 LARGE – ITEMSET (L2)

No	Item Set	Total Transaction	Support %	Confidence	Lift
1	Wrap Red Apples, Save The Planet Cotton Tote Bag	70	0,0129	0,102	0,015

Then, the application of the Mamdani fuzzy method is done in background processing, where the method will be executed after obtaining the support, confidence and lift values using the method calculation. Apriori method. The steps to take are as follows:

1) Fuzifikasi

The fuzification stage will form membership based on predetermined parameters, spesifically support, confidence, and lift value,

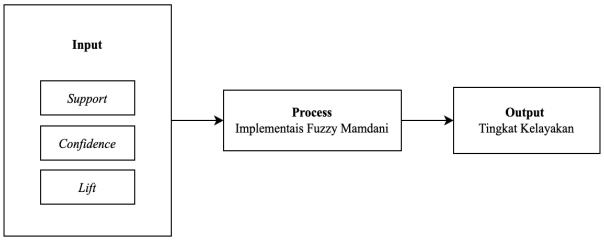

Figure 3 Mamdani Fuzzy Process Analysis

From the image of the Mamdani fuzzy analysis process above, it can be observed in Table 6 below:
Parameters are obtained from the calculation results of the a priori algorithm.

Support Value = 0,0129
Confidence Value = 0,102
Lift Value = 0.015

TABLE 6 Linguistic Value Variables

Function	Nama Variabel	Nilai
Input	*Support*	0,0129
	Confidence	0,102
	Lift	0,015
Output	Tingkat Kelayakan	0,94625

The next fuzzification process is the process of determining the fuzzy set of variables that have been previously set, both input and output variables.

TABLE 7 FUZZY SET

Function	Variable Name	Domain
Support	Small	[0.001;0.002;0.005;0.13]
	Medium	[0,011;0,013;0,014;0,015]
	High	[0.009;0.011;0.012;0.01]
Confidence	Small	[0.01;0.04;0.06;0.1]
	Medium	[0.14;0.16;0.18;0.2]
	High	[0.4;0.6;0.8;1]
Lift	Small	[0.001;0.002;0.005;0.13]
	Medium	[0,011;0,013;0,014;0,015]
	High	[0.009;0.011;0.012;0.01]

2) Formation of the Fuzzy Association

The first step of the Mamdani fuzzy method is to form fuzzy sets. Based on Table 8, perform a curve representation of each variable to know the membership function of each set. Curvilinear representation of each variable can be observed in Figures 4 to 6.

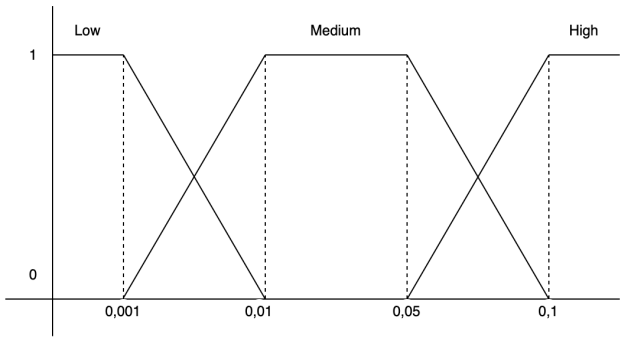

Figure 4 Increase in Support Sets.

In Figure 4 above, the lowest parameter value is 0.001 and the highest is > 0.1.

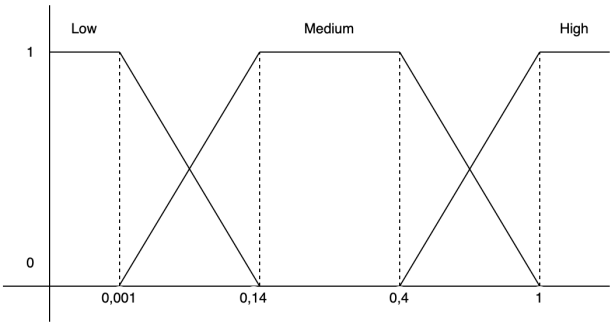

Figure 5 Increase in Confidence Sets

In Figure 5 above, the lowest parameter value is 0.01 and the highest is > 0.1.

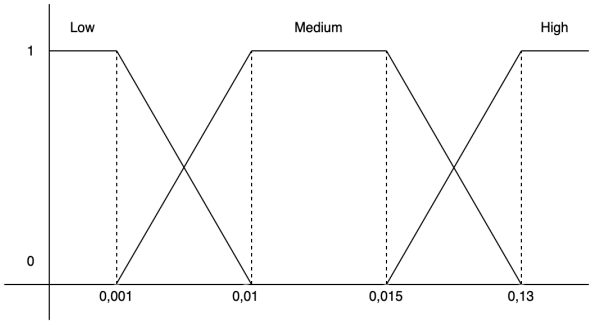

Figure 6 Increase in Lift Sets

In Figure 6 above, the lowest parameter value is 0.001 and the highest is > 0.1.

3) Fuzzy Inference

The method used in the implication function of the Mamdani fuzzy method is the min method. The final rules of the fuzzy inference system in this study are determined by combining existing sets, then the author analyzes which rules occur based on the collected field data. Below are the final rules set forth by the author, which can be seen in table 8:

TABLE 8 FUZZY INFERENCE SYSTEM

No	IF Support	AND Confidence	AND Lift	THEN Predictions
R1	Low	Low	Low	Small

R2	Low	Low	Medium	Small
R3	Low	Low	High	Small
R4	Low	Medium	Low	Small
R5	Low	Medium	Medium	Small
R6	Low	Medium	High	Small
R7	Low	High	Low	Big
R8	Low	High	Medium	Big
R9	Low	High	High	Big
R10	Medium	Low	Low	Small
R11	Medium	Low	Medium	Small
R12	Medium	Low	High	Small
R13	Medium	Medium	Low	Big
R14	Medium	Medium	Medium	Small
R15	Medium	Medium	High	Big
R16	Medium	High	Low	Big
R17	Medium	High	Medium	Big
R18	Medium	High	High	Big
R19	High	Low	Low	Small
R20	High	Low	Medium	Small
R21	High	Low	High	Small
R22	High	Medium	Low	Big
R23	High	Medium	Medium	Big
R24	High	Medium	High	Big
R25	High	High	Low	Big
R26	High	High	Medium	Big
R27	High	High	High	Big

Defuzzification is carried out at the final stage, in determining the result of defuzzification will convert every result from the inference engine expressed in fuzzy set form. At this stage, the output generated at the fuzzification stage is {'small': 0, 'big': 0.94625}, where tempmsmall = 225, small= 0, tetmpbig = 290, big = 0.94625, sumbu x = 5, sumbu y = 4.

Formula ((tempsmall * small) + (tempbig * big)) / ((small * x) + (big * y)) = ((225 * 0) + (290 * 0.94625)) / ((0 * 5) + 0.94625 * 4)) = 72.50.

Defuzzification is performed in the final step, to determine the defuzzification result, it converts each result of the inference engine represented as a fuzzy set. The yield produced at the defuzzing stage is 72.50 with a large range.

Based on the explanation of the results of the above algorithm implementation, Fuzzy Mamdani achieved quite good results in helping to increase the accuracy value of the a priori algorithm by forming three membership functions, which are the support value, trust, and support. To see the results, you can observe pictures 2 and 3:

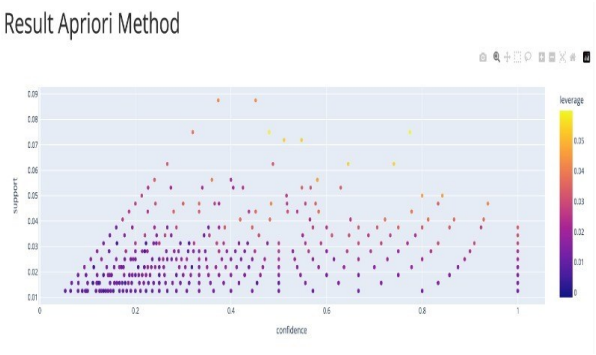

Figure 7 Result Apriori Method

Figure 3 shows the difference in the results which are very far and very striking. It is evident that the performance of the Mamdani fuzzy algorithm for increasing the accuracy of the final value on the a priori algorithm is very good and very feasible to use to improve more accurate results on the a priori algorithm.

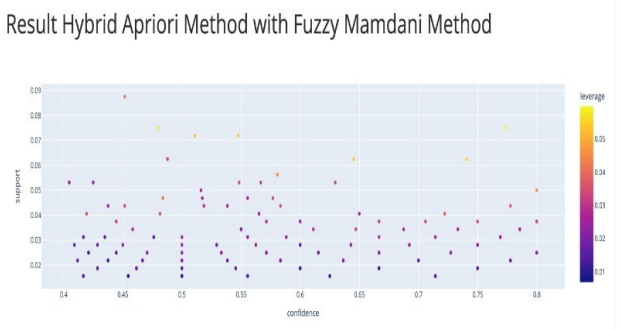

Figure 8 Result Hybrid Apriori Method with Fuzzy Mamdani Method

V. CONCLUSION

Based on the results of the tests and discussions that have been carried out, the authors produce the following

1. The minimum support and confidence values greatly affect the performance of the a priori algorithm, the smaller the specified support and confidence values, the more candidate items that become item recommendations for the processed transaction dataset.
2. The success of the Mamdani fuzzy algorithm in increasing accuracy. The a priori algorithm can provide good accuracy values capable of producing > 20□tter than the results of the classical algorithm. This can be seen in Figure 8; however, the successful performance of the fuzzing task is also determined by the minimum defuzzification value generated.
3. The performance of the Mamdani fuzzy algorithm provides absolute accuracy because the algorithm can only operate based on the established environment.

VI. REFERENCES

[1] Akas, Minhajul Fahim, A. G.M. Zaman, and Abed Khan. 2020. "Combined Item Sets Generation Using Modified Apriori Algorithm." PervasiveHealth: Pervasive Computing Technologies for Healthcare: 4–6.

[2] Al-Maolegi, Mohammed, and Bassam Arkok. 2014. "An Improved Apriori Algorithm for Association Rules." International Journal on Natural Language Computing 3(1): 21–29.

[3] Ameh Ardi, Yuan, Efendi, Syahril, and Nababan, Erna Budhiarti. 2020. "Mamdani and Sugeno Fuzzy Performance Analysis on Rainfall Prediction." Randwick International of Social Science: 161–1

[4] Yogachi, E F, V M Nasution, and G Prakarsa. 2021. “Design and Development of Fuzzy Logic Application Mamdani Method in Predicting the Number of Covid-19 Positive Cases in West Java.” *IOP Conference Series: Materials Science and Engineering* 1115(1): 012031.

[5] Rathee, Sanjay, Manohar Kaul, and Arti Kashyap. 2015. “R-Apriori: An Efficient Apriori Based Algorithm on Spark.” PIKM 2015 - Proceedings of the 8th Ph.D. Workshop in Information and Knowledge Management, co-located with: CIKM 2015: 27–34.

[6] Balaji, Mahesh, and G. Subrahmanya Vrk Rao. 2013. “An Adaptive Implementation Case Study of Apriori Algorithm for a Retail Scenario in a Cloud Environment.” Proceedings - 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing, CCGrid 2013: 625–29.

[7] W. B. Zulfikar, ed., Implementation of Mamdani Fuzzy Method in Employee Promotion System, Materials Science and Engineering No. 288, IOP Publishing (Annual Applied Science and Engineering Conference, 2017).

[8] Lee, Min Young, and Yoon, Taeseon. 2016. “Analysis of Methicillin-Resistant Staphylococcus Aureus Using Apriori, DBSCAN, and K-Means Algorithms.” *ACM International Conference Proceeding Series*: 23–28.

[9] Sun, Rui, and Yuqiang Li. 2020. “Applying Prefixed-Itemset and Compression Matrix to Optimize the MapReduce-Based Apriori Algorithm on Hadoop.” *PervasiveHealth: Pervasive Computing Technologies for Healthcare*: 89–93