Clustering of uninhabitable houses using the optimized apriori algorithm

Al-Khowarizmi[1], Marah Doly Nasution[2], Yoshida Sary[3], Bela[3]
[1]Department of Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
[2]Department of Mathematics Studies, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
[3]Department of Information System, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Received Jan 4, 2024
Revised Jan 30, 2024
Accepted Mar 4, 2024

Keywords:

Algorithm
Apriori
Clustering
Uninhabitable houses
Unsupervised learning

ABSTRACT

Clustering is one of the roles in data mining which is very popularly used for data problems in solving everyday problems. Various algorithms and methods can support clustering such as Apriori. The Apriori algorithm is an algorithm that applies unsupervised learning in completing association and clustering tasks so that the Apriori algorithm is able to complete clustering analysis in Uninhabitable Houses and gain new knowledge about associations. Where the results show that the combination of 2 itemsets with a tendency value for Gas Stove fuel of 3 kg and the installed power meter for the attribute item criteria results in a minimum support value of 77% and a minimum confidence value of 87%. This proves that a priori is capable of clustering Uninhabitable Houses to help government work programs.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Al-Khowarizmi
Department of Information Technology, Faculty of CS & IT, Universitas Muhammadiyah Sumatera Utara
Jl. Kapt. Mukhtar Basri No 3, Medan 20238, Indonesia
Email: alkhowarizmi@umsu.ac.id

1. INTRODUCTION

Data mining is a technique that is very necessary to support the success of artificial intelligence and data science principles [1], [2]. Data mining has 5 basic roles such as association, clustering, classification, forecasting, and prediction [3], [4]. Each war must be based on a dataset and a learning model from the data. The model for learning from data is of course observed based on supervised leaning or unsupervised leaning [5], [6]. In supervised leaning the role of data mining that can be processed is classification, forecasting and prediction, while in unsupervised leaning the role of data mining that can be completed is association and clustering [7].

Focusing on clustering, clustering is a technique for grouping data based on basic similarities and differences in the dataset [8]. The purpose of clustering is to divide data sets into group data sets that have similar and different characteristics [9]. Clustering does not require training data on data objects [10], so many applications use clustering, as in [11] Optimizing business data to increase the effectiveness and accuracy of business data by utilizing clustering techniques in business data analysis services that are smarter and show maximum grouping above 80%. Meanwhile in research [12] used the fuzzy clustering algorithm to group student success results and influencing factors in the dataset so that the results of the clustering research were a student work ratio of 96.7%, a student engagement ratio of 97.5% and a behavior ratio of 95.1%.

Clustering can also be said to be data in forming data patterns so that they can be utilized by other methods [13]. There are many algorithms that can solve clustering problems, one of which is Apriori [14]. The Apriori is an algorithm with unsupervised leaning that is able to solve association and clustering problems. On research [14] carried out clustering using the a priori algorithm on 609 medical records on digestive diseases where the research aimed to explore drug use rules where the results of clustering using the A priori algorithm showed confidence in the analysis results to be greater than 0.91 with a level of support greater than 20% of the information without applying the concept of data mining like clustering. Meanwhile on [15] optimizing the performance of the Apriori algorithm in conducting Clusters on Hadoop where the research results show that the Apriori algorithm is superior by implementing MapReduce-Based compared to Apriori in general.

Various problems in everyday life can of course be solved with a priori algorithms [16], so these algorithms need to be analyzed and optimized for their performance in performing clustering to produce new knowledge which can be called associations [17]. However, the optimization process must be tested on the dataset. A dataset that really supports everyday problems is Uninhabitable Houses [18], [19]. Uninhabitable Houses are owned by the community and are not intended for habitation, so Uninhabitable Houses need guidance from the government in order to provide assistance to become Inhabitable Houses. So, it is necessary to cluster Uninhabitable Houses using the Apriori algorithm which is optimized based on the final result, namely new knowledge based on associations.

2. MATERIAL AND METHOD

2.1. Dataset

The dataset in this research is Uninhabitable Houses data in a village. The process of supporting the government's work program in self-procuring housing has displaced many people. However, the problem that arises from building houses independently by people who have limited resources is the lack of planning and technical knowledge regarding inhabitable houses. This causes the condition of houses built in the long term to become Uninhabitable Houses [20], [21].

2.2. Apriori optimization

The Apriori algorithm is a data mining method for detecting patterns in the dataset to be studied [22]. The application of association rules in data mining aims to detect information from items that are connected to each other in the form of association rules. Association rules are obtained from the results of calculations which consist of 2 measures, namely [11]:

a. The support value is determined according to (1). Support values for two items are used in (1). The parameter T A states the number of transactions containing A, T A∩B transactions containing A and B, T A∩B∩C transactions containing A, B and C, and Ttotal the total number of transactions [23].

$$Support\ (A) = \frac{T_A}{T_{Total}} x\ 100\% \quad (1)$$

$$Support\ (A,B) = \frac{T_{A\cap B}}{T_{Total}} x\ 100\% \quad (2)$$

$$Support\ (A,B,C) = \frac{T_{A\cap B\cap C}}{T_{Total}} x\ 100\% \quad (3)$$

Where,
T_A is a state the number of transactions containing A,
$T_{A\cap B}$ is a transaction containing A and B,
$T_{A\cap B\cap C}$ is a transaction containing A, B and C,
T_{Total} is the total transaction amount.

b. In calculating confidence, itemset exchange is carried out. For example, a combination of 2 itemsets, namely A → B, then reversed to become B → A. Likewise with a combination of 3 itemsets, namely A, B → C, then reversed to become A, C → B and B, C → A. Each itemset support value maybe it will remain the same, but it will likely have a different confidence value. This is to find out which confidence value is the largest for each itemset. The confidence calculation for a combination of 2 itemsets is stated in (4). The confidence calculation for a combination of 3 itemsets is stated in (5) [24].

$$Confidence\ (A,B) = \frac{T_{A\cap B}}{T_A} \quad (4)$$

$$Confidence\ (A,B,C) = \frac{T_{A\cap B\cap C}}{T_{A\cap B}} \quad (5)$$

The Apriori algorithm is defined as a data mining algorithm that is often used in the association rule method [25]. A priori algorithms play a role in finding high frequency patterns. High frequency patterns are

patterns of items whose frequency is above a certain threshold in a database. The stages of a priori include the following [23]:
- Formation of candidate itemsets. The combination of (k-1)- itemsets obtained from the previous iteration can form a candidate itemset [26].
- Calculation of support for each k-itemset candidate. To measure the number of transactions that have items, support is needed from each candidate which is obtained by examining the database that will be used. How to find support can be done using calculations in (1) and (2).
- High frequency pattern analysis. High frequency patterns are determined from k-itemset candidates that exceed the minimum support value.
- If the high frequency pattern is no longer obtained, the entire process will stop.

2.3. General architecture

In presenting this paper, of course a rule with a general architecture is formed. Where the general architecture describes this series of research so that it matches the expected results. The general architecture can be seen in Figure 1.

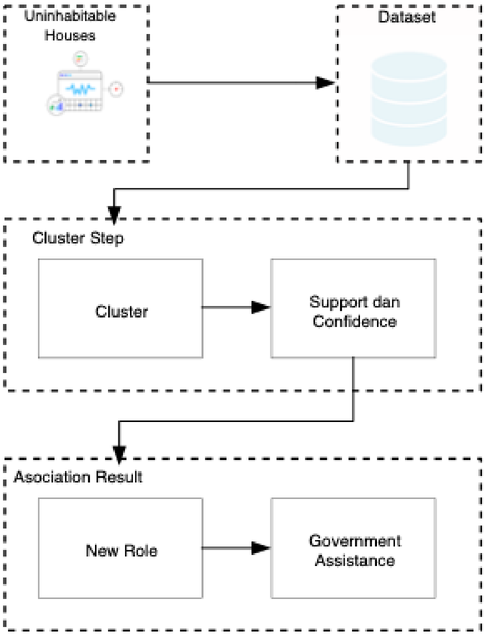

Figure 1. General architecture

In Figure 1 the steps are explained as follows.
- List data in a village where there are Uninhabitable Houses.
- Store it in the data warehouse and become a dataset.
- Enter the clustering process and generate support and confidence.
- After step 3, we enter new knowledge, namely roles, so that priority data can support the government assistance process.

3. RESULT AND DISCUSSION

In this paper, the dataset of attribute items used is from variables determined from population data, which consists of the results of direct observations of the community or residents who are entitled to receive house renovation assistance in Fisherman Indah Village. The population data was taken from 2016-2019. Then we will use data samples from the selection, pre-processing/cleaning data, and transformation stages.

Where the data that has been processed is tested in the Rapidminer tool with 19 attributes that have been obtained from 9 population data itemsets as transaction items and then tested using the association method with a maximum of 3 items with a minimum support value of 0.4 or 40% and a minimum confidence value of 0.8 or 80%. In this paper we will discuss mainly the results of manual calculations carried out to determine support and confidence value calculations. The results of using the RapidMiner tool to obtain algorithm testing results in item data collection, forming itemset candidates, calculating support for each k-itemset candidate, calculating confidence value and finally the formation of associations.

According to manual calculations using the a priori algorithm, it can be seen in Table 1 which is a representation of input data in the form of input data when goods transactions occur. The process of the Al Priori algorithm on this data is based on various formulas. The following is an example of input data used for the data mining process in the form of a transaction item table for selecting home renovation assistance as follows in Table 1.

Table 1. Transaction item house data

No	Variable	Attribute items
1	X1	One's own, Permanent, Rooftile, Wall, Ceramics, Municipal Waterworks, Electricity Meter, Gas Stove stove, Self-employed
2	X2	Rent, Semi Permanent, Rooftile, Woven bamboo, Cement, Municipal Waterworks, Electricity Meter, Gas Stove stove, Self-employed
3	X3	One's own, Permanent, Zinc, Wall, Dirt Floor, Well water, Electricity Meter, Gas Stove stove, Self-employed
4	X4	One's own, Permanent, Zinc, Woven bamboo, Dirt Floor, Municipal Waterworks, Non Electricity Meter, Gas Stove, Fisherman
5	X5	Rent, Semi Permanent, Zinc, Wall, Dirt Floor, Well water, Electricity Meter, Gas Stove stove, Self-employed
6	X6	Rent, Permanent, Rooftile, Wall, Ceramics, Well water, Electricity Meter., Gas Stove, Self-employed
7	X7	One's own, Semi Permanent, Zinc, Wall, Cement, Well water, Electricity Meter, Gas Stove Stove, Self-employed
8	X8	One's own, Semi Permanent, Zinc, Woven bamboo, Cement, Municipal Waterworks, Electricity Meter, Gas Stove, Fisherman
9	X9	One's own, Semi Permanent, Rooftile, Woven bamboo, Dirt Floor, Well water, Electricity Meter, Kerosene Stove, Fisherman

From the results of the representation in Table 1, it can be seen that the frequency pattern is carried out based on the support value in analyzing data on potential recipients of house renovation assistance. Where table 1 also shows home ownership items. However, from the detailed patterns seen in Table 1, the next process is to carry out calculations using the a priori algorithm. and testing through several itemset schemes which are detailed as follows:

3.1. Support value 1 itemset

The results of the calculation of the support value are obtained using a calculation sample with a minimum support of ≥ 40% as follows.

$$Support\ (\text{A}) = \frac{\text{Number of Transactions Containing A}}{\text{Total Transactions}} x\ 100\%$$

So,

$$Support\ (Gas\ Stove\ 3\ Kg) = \frac{8}{9} x\ 100\% = 88\%$$

$$Support\ (Electricity\ Meter) = \frac{8}{9} x\ 100\% = 88\%$$

$$Support\ (One's\ own) = \frac{6}{9} x\ 100\% = 66\%$$

$$Support\ (Self - Employed) = \frac{6}{9} x\ 100\% = 66\%$$

$$Support\ (S.Permanent) = \frac{5}{9} x\ 100\% = 55\%$$

As for the following Table 2, the formation of support values from 1 itemset of data is as follows:

3.2. Support value 2 itemset

The results of the calculation on the Support value are obtained using a calculation sample with a minimum Support ≥ 40% as follows.

$$Support\ (\text{A}) = \frac{\text{Number of Transactions Contains A and B}}{\text{Total Transactions}} x\ 100\%$$

So,

$$Support\ (Gas\ Stove\ 3\ Kg, Electricity\ Meter) = \frac{7}{9} x\ 100\% = 77\%$$

$$Support\ (Gas\ Stove\ 3\ Kg, One's\ own) = \frac{5}{9} x\ 100\% = 55\%$$

$$Support\ (Gas\ Stove\ 3\ Kg, Self - Employed) = \frac{6}{9} x\ 100\% = 66\%$$

$$Support\ (Gas\ Stove\ 3\ Kg, S.Permanent) = \frac{4}{9} x\ 100\% = 44\%$$

$$Support\ (Gas\ Stove\ 3\ Kg, Electricity\ Meter) = \frac{7}{9} x\ 100\% = 77\%$$

The following is Table 3. The formation of support values from 2 data itemsets is as follows:

Table 2. Results of support values with 1 itemset

No	*Itemset* 1	Amount	*Support*	Percent
1	Gas Stove 3 Kg	8	0.88	88%
2	Electricity Meter.	8	0.88	88%
3	One's own	6	0.66	66%
4	Self-employed	6	0.66	66%
5	S.Permanent	5	0.55	55%
6	Zinc	5	0.55	55%
7	Well water	5	0.55	55%
8	Wall	5	0.55	55%
9	Woven bamboo	4	0.44	44%
10	Rooftile	4	0.44	44%
11	Municipal Waterworks	4	0.44	44%
12	Permanent	4	0.44	44%
13	Dirt Floor	4	0.44	44%
14	Fisherman	3	0.33	33%
15	Cement	3	0.33	33%
16	Rent	3	0.33	33%

Table 3. Results of support values with 2 itemsets

No	*Itemset* 1	*Itemset* 2	Ammount	*Support*	Percent
1	Gas Stove 3 Kg	Electricity Meter.	7	0.77	77%
2	Gas Stove 3 Kg	One's own	5	0.55	55%
3	Gas Stove 3 Kg	Self-employed	6	0.66	66%
4	Gas Stove 3 Kg	S.Permanent	4	0.44	44%
5	Gas Stove 3 Kg	Zinc	5	0.55	55%
6	Gas Stove 3 Kg	Well water	4	0.44	44%
7	Gas Stove 3 Kg	Wall	5	0.55	55%
8	Gas Stove 3 Kg	Woven bamboo	3	0.33	33%
9	Gas Stove 3 Kg	Rooftile	3	0.33	33%
10	Gas Stove 3 Kg	Municipal Waterworks	4	0.44	44%
11	Gas Stove 3 Kg	Permanent	4	0.44	44%
12	Gas Stove 3 Kg	Dirt Floor	3	0.33	33%
13	Gas Stove 3 Kg	Cement	3	0.33	33%
14	Gas Stove 3 Kg	Rent	3	0.33	33%
15	Electricity Meter	One's own	5	0.55	55%
16	Electricity Meter	Self-employed	6	0.66	66%
17	Electricity Meter	S.Permanent	5	0.55	55%
18	Electricity Meter	Zinc	4	0.444	44%
19	Electricity Meter	Well water	5	0.55	55%
20	Electricity Meter	Wall	5	0.55	55%
21	Electricity Meter	Woven bamboo	3	0.33	33%
22	Electricity Meter	Rooftile	4	0.44	44%
23	Electricity Meter	Municipal Waterworks	3	0.33	33%
24	Electricity Meter	Permanent	3	0.33	33%
25	Electricity Meter	Dirt Floor	3	0.33	33%
26	Electricity Meter	Cement	3	0.33	33%
27	Electricity Meter	Rent	3	0.33	33%
28	One’s own	Self-employed	3	0.33	33%
29	One’s own	S.Permanent	3	0.33	33%
30	One’s own	Zinc	4	0.44	44%

3.3. Support value 3 itemset

The results of the calculation on the Support value are obtained using a calculation sample with a minimum Support ≥ 40% as follows.

$$Support = \frac{\text{Number of Transactions Contains A, B and C}}{\text{Total Transactions}} x\ 100\%$$

So,

$$Support(Gas\ Stove\ 3\ Kg, Electricity\ Meter, One's\ own) = \frac{4}{9} x\ 100\% = 44\ \%$$

$$Support(Gas\ Stove\ 3\ Kg, Electricity\ Meter, Self-Employed) = \frac{6}{9} x\ 100\% = 66\ \%$$

$$Support(Gas\ Stove\ 3\ Kg, Electricity\ Meter, S.Permanent) = \frac{4}{9} x\ 100\% = 44\ \%$$

$$Support(Gas\ Stove\ 3\ Kg, Electricity\ Meter, Zinc) = \frac{4}{9} x\ 100\% = 44\ \%$$

The following is Table 4 for the formation of support values from 3 data itemsets as follows:

Table 4. Results of support values with 3 Itemsets

No	*Itemset* 1	*Itemset* 2	*Itemset* 3	Ammount	Support	Percent
1	Gas Stove 3 Kg	Electricity Meter.	One's owni	4	0.44	44%
2	Gas Stove 3 Kg	Electricity Meter.	Self-employed	6	0.66	66%
3	Gas Stove 3 Kg	Electricity Meter.	S.Permanent	4	0.44	44%
4	Gas Stove 3 Kg	Electricity Meter.	Zinc	4	0.44	44%
5	Gas Stove 3 Kg	Electricity Meter.	Well water	4	0.44	44%
6	Gas Stove 3 Kg	Electricity Meter.	Wall	5	0.55	55%
7	Gas Stove 3 Kg	Electricity Meter.	Rooftile	3	0.33	33%
8	Gas Stove 3 Kg	Electricity Meter.	Municipal Waterworks	3	0.33	33%
9	Gas Stove 3 Kg	Electricity Meter.	Permanent	3	0.33	33%
10	Gas Stove 3 Kg	Electricity Meter.	Cement	3	0.33	33%
11	Gas Stove 3 Kg	Electricity Meter.	Rent	3	0.33	33%
12	Gas Stove 3 Kg	One's own	Self-employed	3	0.33	33%
13	Gas Stove 3 Kg	One's own	Zinc	4	0.44	44%
14	Gas Stove 3 Kg	One's own	Wall	3	0.33	33%
15	Gas Stove 3 Kg	One's own	Municipal Waterworks	3	0.33	33%
16	Gas Stove 3 Kg	One's own	Permanent	3	0.33	33%
17	Gas Stove 3 Kg	Self-employed	S.Permanent	3	0.33	33%
18	Gas Stove 3 Kg	Self-employed	Zinc	3	0.33	33%
19	Gas Stove 3 Kg	Self-employed	Well water	4	0.44	44%
20	Gas Stove 3 Kg	Self-employed	Wall	5	0.55	55%
21	Gas Stove 3 Kg	Self-employed	Rooftile	3	0.33	33%
22	Gas Stove 3 Kg	Self-employed	Permanent	3	0.33	33%
23	Gas Stove 3 Kg	Self-employed	Rent	3	0.33	33%
24	Gas Stove 3 Kg	S.Permanent	Zinc	3	0.33	33%
25	Gas Stove 3 Kg	S.Permanent	Cement	3	0.33	33%
26	Gas Stove 3 Kg	Zinc	Well water	3	0.33	33%
27	Gas Stove 3 Kg	Zinc	Wall	3	0.33	33%
28	Gas Stove 3 Kg	Zinc	Dirt Floor	3	0.33	33%
29	Gas Stove 3 Kg	Well water	Wall	4	0.44	44%
30	Gas Stove 3 Kg	Wall	Permanent	3	0.33	33%

Confidance Value (Cf), the association rule search is formed after obtaining a high frequency pattern to calculate the confidence value where the minimum confidence value that has been determined is 0.8 or 80%.

$$Confidence = P(B \mid A)$$

$$Confidance\ P(X, Y) = \frac{\text{Transaction Amount Contains X and Y}}{\text{Transactions Containing X}} x\ 100\%$$

$$Confidance\ (One's\ own, Gas\ Stove\ 3\ Kg = \frac{5}{6} x\ 100\% = 83\%$$

$$Confidance\ (One's\ own, Electricity\ Meter = \frac{5}{6} x\ 100\% = 83\%$$

$$Confidance\ (Self - Employed, Wall = \frac{5}{6} x\ 100\% = 83\%$$

The following is the process of forming association rules using pattern analysis as shown in Table 5.

Table 5. Formation of association rules

No	*Premises*	*Conclusion*	*Support*	*Confidance*
1	One's own	Gas Stove 3 Kg	0.55	0.83
2	One's own	Electricity Meter.	0.55	0.83
3	Self-employed	Wall	0.55	0.83
4	Self-employed	Gas Stove 3 Kg, Wall	0.55	0.83
5	Gas Stove 3 Kg, Self-employed	Wall	0.55	0.83
6	Self-employed	Electricity Meter., Wall	0.55	0.83
7	Electricity Meter., Self-employed	Wall	0.55	0.83
8	Gas Stove 3 Kg, Electricity Meter.	Self-employed	0.66	0.85
9	Gas Stove 3 Kg	Electricity Meter.	0.77	0.87
10	Electricity Meter.	Gas Stove 3 Kg	0.77	0.87
11	Self-employed	Gas Stove 3 Kg	0.66	1.00
12	Zinc	Gas Stove 3 Kg	0.55	1.00
13	Wall	Gas Stove 3 Kg	0.55	1.00
14	Municipal Waterworks	Gas Stove 3 Kg	0.44	1.00
15	Permanent	Gas Stove 3 Kg	0.44	1.00
16	Self-employed	Electricity Meter.	0.66	1.00
17	S.Permanent	Electricity Meter.	0.55	1.00
18	Well water	Electricity Meter.	0.55	1.00
19	Wall	Electricity Meter.	0.55	1.00
20	Rooftile	Electricity Meter.	0.44	1.00
21	Self-employed	Gas Stove 3 Kg, Electricity Meter.	0.66	1.00
22	Gas Stove 3 Kg, Self-employed	Electricity Meter.	0.66	1.00
23	Electricity Meter., Self-employed	Gas Stove 3 Kg	0.66	1.00
24	Gas Stove 3 Kg, S.Permanent	Electricity Meter.	0.44	1.00
25	Electricity Meter., Zinc	Gas Stove 3 Kg	0.44	1.00
26	Gas Stove 3 Kg, Well water	Electricity Meter.	0.44	1.00
27	Wall	Gas Stove 3 Kg, Electricity Meter.	0.55	1.00
28	Gas Stove 3 Kg, Wall	Electricity Meter.	0.55	1.00
29	Electricity Meter., Wall	Gas Stove 3 Kg	0.55	1.00

The association rule search is formed after obtaining a high frequency pattern that has been obtained in a combination of 2 items. Use an equation formula to calculate the confidence value where the minimum confidence value determined by the user is 80%. To find association rules, only use the values of 2 itemsets by setting a minimum confidence of 80%. So, the clusters that form the rule association can be seen in Table 6.

Table 6. Association rules

No	Role	Support	Confidance	
1	If the job is Self-employed then it has a Wall house	77%	5/6	0.83
2	If the Gas Stove fuel is 3 kg then it has an Electricity Meter	77%	7/8	0.87
3	If the power is installed Electricity Meter then it has a 3 Kg Gas Stove	66%	7/8	0.87
4	If the work is self-employed then it has a 3 Kg Gas Stove	55%	6/6	1
5	If the roof is Zinc then it has a 3 kg Gas Stove	55%	6/6	1
6	If the wall is a wall then it has a 3 kg Gas Stove	44%	6/6	1
7	If the water source is Municipal Waterworks then it has a 3 Kg Gas Stove	44%	6/6	1
8	If the building is permanent then it has a 3 kg gas stove	33%	6/6	1
9	If the work is self-employed then it has an electricity meter.	33%	6/6	1
10	If the building is semi-permanent then it has an Electricity Meter.	66%	6/6	1
11	If the water source is Well water then it has an Electricity Meter.	55$	6/6	1
12	If it's a wall then it has an Electricity Meter.	55%	6/6	1

In this paper, of course, optimization is carried out using an a priori algorithm to find association rules for itemset data patterns with a minimum support of 40% and a minimum confidence of 80%. On the analysis process page display, the next step that will be displayed is the modeling step of the analysis process carried out by the system using attribute item data. The model process display using the a priori method is carried out to determine the data items resulting from the analysis which is seen in Figure 2.

Figure 2 shows a visualization of itemset association rules with high frequency values produced by rapidminer testing. Display of the results of the itemset association rule values with the confidence values produced by Rapidminer testing. from Figure 2 also forms an association rule which is tested on the Uninhabitable Houses data shown in Figure 3.

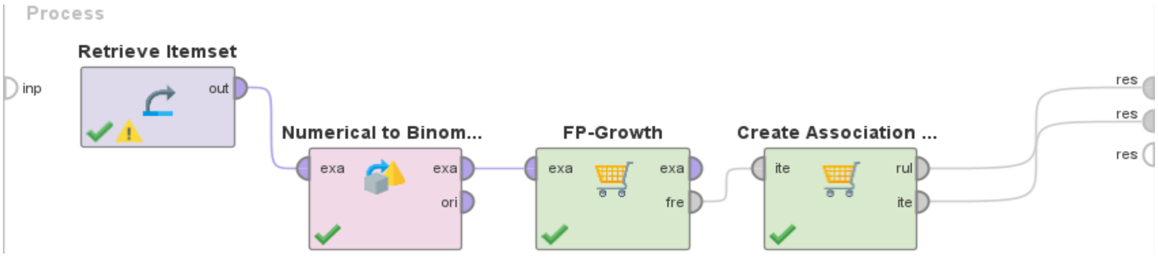

Figure 2. Apriori process in clustering

Figure 3 shows the association rule search is formed after obtaining a high frequency pattern that has been obtained in a combination of 2 items. Use an equation formula to calculate the confidence value where the minimum confidence value determined by the user is 80%. To find association rules, only use the values of 2 itemsets by setting a minimum confidence of 80%.

AssociationRules

```
Association Rules
[S.Permanen] --> [Gas 3 Kg] (confidence: 0.800)
[Sumur] --> [Gas 3 Kg] (confidence: 0.800)
[Seng] --> [Meteran] (confidence: 0.800)
[Seng] --> [Milik Sendiri] (confidence: 0.800)
[Sumur] --> [Wiraswasta] (confidence: 0.800)
[Sumur] --> [Tembok] (confidence: 0.800)
[Tembok] --> [Sumur] (confidence: 0.800)
[Gas 3 Kg, Milik Sendiri] --> [Meteran] (confidence: 0.800)
[Meteran, Milik Sendiri] --> [Gas 3 Kg] (confidence: 0.800)
[S.Permanen] --> [Gas 3 Kg, Meteran] (confidence: 0.800)
[Meteran, S.Permanen] --> [Gas 3 Kg] (confidence: 0.800)
[Seng] --> [Gas 3 Kg, Meteran] (confidence: 0.800)
[Gas 3 Kg, Seng] --> [Meteran] (confidence: 0.800)
[Sumur] --> [Gas 3 Kg, Meteran] (confidence: 0.800)
[Meteran, Sumur] --> [Gas 3 Kg] (confidence: 0.800)
[Gas 3 Kg, Milik Sendiri] --> [Seng] (confidence: 0.800)
[Seng] --> [Gas 3 Kg, Milik Sendiri] (confidence: 0.800)
[Gas 3 Kg, Seng] --> [Milik Sendiri] (confidence: 0.800)
[Sumur] --> [Gas 3 Kg, Wiraswasta] (confidence: 0.800)
[Sumur] --> [Gas 3 Kg, Tembok] (confidence: 0.800)
[Tembok] --> [Gas 3 Kg, Sumur] (confidence: 0.800)
[Gas 3 Kg, Tembok] --> [Sumur] (confidence: 0.800)
[Sumur] --> [Meteran, Wiraswasta] (confidence: 0.800)
[Meteran, Sumur] --> [Wiraswasta] (confidence: 0.800)
```

Figure 3. Results of association rules

4. CONCLUSSION

Based on the results of the analysis of the calculation pattern using the a priori algorithm method, it can be seen that it is based on a combination of 2 itemsets with a tendency value for Gas Stove fuel of 3 kg and the installed power meter for the attribute item criteria with the result being a minimum support value of 77% and a minimum confidence value of 87%. In the data mining testing system in selecting and clustering Uninhabitable Houses, several forms are displayed to process the input attribute item data. However, testing in this paper shows clustering in Uninhabitable Houses with an a priori algorithm that is optimized by adding new knowledge in the form of associations in house renovation assistance with the help of rapidminer testing tools.

REFERENCES

[1] Al-Khowarizmi and Suherman, “Classification of skin cancer images by applying simple evolving connectionist system,” *IAES International Journal of Artificial Intelligence*, vol. 10, no. 2, pp. 421–429, Jun. 2021, doi: 10.11591/IJAI.V10.I2.PP421-429.

[2] M. E. Al Khowarizmi, Rahmad Syah, Mahyuddin K. M. Nasution, “Sensitivity of MAPE using detection rate for big data forecasting crude palm oil on k-nearest neighbor,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 3, pp. 2696–2703, 2021, doi: 10.11591/ijece.v11i3.pp2696-2703.

[3] A. Dogan and D. Birant, “Machine learning and data mining in manufacturing,” *Expert Systems with Applications*, vol. 166, p. 114060, 2021, doi: https://doi.org/10.1016/j.eswa.2020.114060.
[4] N. Maleki, Y. Zeinali, and S. T. A. Niaki, “A k-NN method for lung cancer prognosis with the use of a genetic algorithm for feature selection,” *Expert Systems with Applications*, vol. 164, p. 113981, 2021, doi: https://doi.org/10.1016/j.eswa.2020.113981.
[5] K. K. Hiran, R. K. Jain, K. Lakhwani, and R. Doshi, *Machine Learning: Master Supervised and Unsupervised Learning Algorithms with Real Examples*. BPB Publications, 2021.
[6] S. Bashath, N. Perera, S. Tripathi, K. Manjang, M. Dehmer, and F. E. Streib, “A data-centric review of deep transfer learning with applications to text data,” *Information Sciences (Ny).*, vol. 585, pp. 498–528, 2022.
[7] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, “A systematic review on supervised and unsupervised machine learning algorithms for data science,” *Supervised and Unsupervised Learning for Data Science*, pp. 3–21, 2020.
[8] T. M. Ghazal, “Performances of K-means clustering algorithm with different distance metrics,” *Intelligent Automation & Soft Computing*, vol. 30, no. 2, pp. 735–742, 2021.
[9] K. Bandara, C. Bergmeir, and S. Smyl, “Forecasting across time series databases using recurrent neural networks on groups of similar series: A clustering approach,” *Expert Systems with Applications*, vol. 140, p. 112896, 2020.
[10] Y. Zhang, C. Song, and D. Zhang, “Deep learning-based object detection improvement for tomato disease,” *IEEE access*, vol. 8, pp. 56607–56614, 2020.
[11] N. Wang and N. Wang, “Design of an intelligent processing system for business data design of an intelligent processing clustering system for business data analysis based on improved algorithm analysis based on improved clustering algorithm,” *Procedia Computer Science*, vol. 228, pp. 1215–1224, 2023, doi: 10.1016/j.procs.2023.11.105.
[12] H. Han, “Fuzzy clustering algorithm for university students’ psychological fitness and performance detection,” *Heliyon*, vol. 9, no. 8, p. e18550, 2023, doi: 10.1016/j.heliyon.2023.e18550.
[13] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, “Chapter 4 - Algorithms: The basic methods,” I. H. Witten, E. Frank, M. A. Hall, and C. J. B. T.-D. M. (Fourth E. Pal, Eds. Morgan Kaufmann, 2017, pp. 91–160.
[14] J. Wu *et al.*, “A study of TCM master Yan Zhenghua’s medication rule in prescriptions for digestive system diseases based on Apriori and complex system entropy cluster,” *Journal of Traditional Chinese Medical Sciences*, vol. 2, no. 4, pp. 241–247, 2015, doi: 10.1016/j.jtcms.2016.02.007.
[15] S. Singh, R. Garg, and P. K. Mishra, “Performance optimization of MapReduce-based Apriori algorithm on Hadoop cluster,” *Computers & Electrical Engineering*, vol. 67, pp. 348–364, 2018, doi: 10.1016/j.compeleceng.2017.10.008.
[16] E. Kaya, B. Gorkemli, B. Akay, and D. Karaboga, “A review on the studies employing artificial bee colony algorithm to solve combinatorial optimization problems,” *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105311, 2022.
[17] H. Luo *et al.*, “Associations of β-Fibrinogen Polymorphisms with the Risk of Ischemic Stroke: A Meta-analysis.,” *Journal of Stroke and Cerebrovascular Diseases*, vol. 28, no. 2, pp. 243–250, Feb. 2019, doi: 10.1016/j.jstrokecerebrovasdis.2018.09.007.
[18] N. Shinohara, K. Hashimoto, H. Kim, and H. Yoshida-Ohuchi, “Fungi, mites/ticks, allergens, and endotoxins in different size fractions of house dust from long-term uninhabited houses and inhabited houses,” *Building and Environment*, vol. 229, p. 109918, 2023.
[19] Y. Liu, F. Yu, J. Xu, and P. Xin, “Identification of dangerous rural houses using oblique photogrammetry and photo recognition technology,” in *2023 International Conference on Pattern Recognition, Machine Vision and Intelligent Algorithms (PRMVIA)*, 2023, pp. 70–75.
[20] S. M. Berliana, A. W. Augustia, P. D. Rachmawati, R. Pradanie, F. Efendi, and G. E. Aurizki, “Factors associated with child neglect in Indonesia: Findings from National Socio-Economic Survey,” *Children and Youth Services Review*, vol. 106, no. September, p. 104487, 2019, doi: 10.1016/j.childyouth.2019.104487.
[21] Y. Abe, K. Yamada, R. Tanaka, K. Ando, and M. Ueno, “Dynamic living space: toward a society where people can live anywhere in 2050,” *Food Bioprod. Process.*, p. 105151, 2023, doi: 10.1016/j.futures.2024.103363.
[22] M. Sornalakshmi *et al.*, “Hybrid method for mining rules based on enhanced Apriori algorithm with sequential minimal optimization in healthcare industry,” *Neural Computing and Applications*, pp. 1–14, 2020.
[23] R. Papi, S. Attarchi, A. Darvishi Boloorani, and N. Neysani Samany, “Knowledge discovery of Middle East dust sources using Apriori spatial data mining algorithm,” *Ecological Informatics*, vol. 72, no. July, p. 101867, 2022, doi: 10.1016/j.ecoinf.2022.101867.
[24] X. Zhang and J. Zhang, “Analysis and research on library user behavior based on apriori algorithm,” *Measurement: Sensors*, vol. 27, no. April, p. 100802, 2023, doi: 10.1016/j.measen.2023.100802.
[25] E. V. Altay and B. Alatas, “Intelligent optimization algorithms for the problem of mining numerical association rules,” *Physica A: Statistical Mechanics and its Applications*, vol. 540, p. 123142, 2020.
[26] C. Wang and X. Zheng, “Application of improved time series Apriori algorithm by frequent itemsets in association rule data mining based on temporal constraint,” *Evolutionary Intelligence*, vol. 13, no. 1, pp. 39–49, 2020.

BIOGRAPHIES OF AUTHORS

Dr. Al-Khowarizmi was born in Medan, Indonesia, in 1992. He is a Dean in Faculty of Computer Science and Information Technology at Universitas Muhammadiyah Sumatera Utara (UMSU). He got Doctoral Degree from Universitas Sumatera Utara in 2023. His main research interest is data science, big data, machine learning, neural network, artificial intelligence and business intelligence. He can be contacted at email: alkhowarizmi@umsu.ac.id.

Dr. Marah Doly Nasution is a researcher and lecturer in Universitas Muhammadiyah Sumatera Utara (UMSU). Currently, He finished his Doctoral degree recently in Mathematics Computation at 2020 in Universitas Sumatera Utara. His research interest is in modelling and simulation, internet of things and data mining. He can be contacted at email: marahdoly@umsu.ac.id.

Yoshida Sary is a researcher and lecturer in Universitas Muhammadiyah Sumatera Utara (UMSU). Currently, she finished master degree recently in information technology in Universitas Putra Indonesia 'YPTK' Padang. Her research interest is in algorithm and programming, design information system, and data mining. She can be contacted at email: yoshidasary@umsu.ac.id.

Bela is a researcher and student in Universitas Muhammadiyah Sumatera Utara (UMSU). Currently. Her research interest is in algorithm and programming, design information system, and data mining. She can be contacted at email: bela@students.umsu.ac.id.