



Occupational Correlation to the Level of Community Welfare Using The Apriori Algorithm (Case Study: Mangga Village)

Aulia Adlin Revaldi^{1*}, Novriyenni², Lina Arliana Nur Kadim³

^{1,2,3}Informatics Engineering, STMIK Kaputama
Jl. Veteran No. 4A-9A, Binjai, North Sumatra, Indonesia
adlinauliarevaldi@gmail.com¹, novriyenni.sikumbang@gmail.com², lina_arliana@yahoo.com³

Abstract

Mango village is one of the areas that participates in various welfare programs for the local community, where in this village there are still people who are far from prosperous because of various factors that affect the welfare of the local community, one of which is the jobs owned by the community. Therefore, it is important for people to understand that work also greatly influences the level of welfare for their own lives, so that they are fulfilled in the economy, education and others. Therefore the author wants to create a system that can assist the government in developing community welfare programs in Mango Village by knowing the relationship between work and the level of social welfare. After carrying out the above case trials with minimum support = 25%, confidence = 100% so that the rule results that meet the support and confidence values are obtained, it can be concluded that if the assets owned are A4 (motorcycles), with T2 dependents (3-4), with M2 jobs (Private Employees) and A4 assets (Motorcycles), with P2 income (> 1,000,000 - < 2,000,000), then enter K2 (welfare stage II) with a support value of 20%, 100% confidence.

Keywords: *Apriori Algorithm, Community Welfare, Mangga Village*

1. Introduction

Mango village is one of the areas that participates in various welfare programs for the local community, where in this village there are still people who are far from prosperous because of various factors that affect the welfare of the local community, one of which is the jobs owned by the community. Therefore, it is important for people to understand that work also greatly influences the level of welfare for their own lives, so that they are fulfilled in the economy, education and others [1], [2], [3].

Therefore the author wants to create a system that can assist the government in developing community welfare programs in Mango Village by knowing the relationship between work and the level of social welfare. To carry out the analysis, the data needed is information about the results of community work income, then analyzed using the Apriori algorithm to find the relationship between the two variables. Based on the background above, the author has an interest in raising the title of thesis research with the title "**Occupational Correlation to the Level of Community Welfare Using the Apriori Algorithm**" [4], [5], [6].

Based on the background stated above, the formulation of the problem raised is: How to determine the relationship and analyze the pattern between work and the level of social welfare using the a priori algorithm? this is to process data using the a priori algorithm method into a data mining system to produce the latest information with the WEKA and PHP applications [7], [8], [9].

2. Research Methods

2.1. Data Mining

Data mining is often known as a term used to find hidden knowledge in databases. Data Mining is an analysis step of the knowledge discovery process in the database or knowledge discovery in database (KKD). According to Buulolo [10] states that Data Mining is an

activity related to data collection, the use of historical data to find knowledge, information, regularity, patterns or relationships in large data.

2.2. Apriori Algorithm

The Apriori algorithm is one of the most well-known algorithms in data mining for finding data patterns or patterns of occurrence/frequency of data. According to Buulolo [10] in his book Apriori Algorithm, the term support value is the value used to measure the occurrence of certain data compared to the total data. The formula for finding support values is as follows:

Support values for an itemset

$$Support(A) = \frac{JumlahTransaksiMengandungA}{TotalTransaksi} \times 100 \tag{1}$$

Value support for 2 (two) itemsets

$$Support(A, B) = P(A \cap B) = \frac{JumlahTransaksiMengandungA\ dan\ B}{Total\ Transaksi} \times 100 \tag{2}$$

After all high-frequency patterns are found, then we look for association rules that meet the minimum requirements for confidence by calculating the confidence of associative rules $A \rightarrow B$, where support is supporting data and confidence is confidence.

The confidence value of the $A \rightarrow B$ rule is obtained from the following formula:

$$Support(A) = \frac{\sum TransaksimengandungAdanB}{\sum Jumlahseluruhtransaksi} \times 100\% \tag{3}$$

$$Confidance(A) = \frac{\sum TransaksimengandungAdanB}{\sum TransaksimengandungA} \times 100\% \tag{4}$$

After all the high-frequency patterns are found, then the association rules that meet the minimum requirements for confidence are searched by calculating the confidence of the associative rules $A \cup B$. The confidence value of the rules $A \cup B$ is obtained by the following formula:

$$Confidence - P(B/A) = \frac{\sum J\ transaksi\ mengandung\ A\ dan\ B}{\sum transaksi} \tag{5}$$

3. Results And Discussion

3.1. Supporting Data

The following is data obtained from personnel transaction data as shown in Table 1.

Table 1: Data Desa Mangga

No.	Amount_dependent	Work	Income	Asset					Prosperity level
				Gold	Land	House	Motorcycle	Car	
1	3	Laborer	Rp. 500.000 - Rp. 1.000.000	-	-	-	√	-	Prosperous Family Stage I
2	2	Private sector employee	Rp. 1.000.000 - Rp. 2.000.000	-	-	√	√	-	Prosperous Family Phase II
3	4	Farmer	Rp. 500.000 - Rp. 1.000.000	-	√	√	√	-	Prosperous Family Phase II
4	2	Driver	Rp. 500.000 - Rp. 1.000.000	-	-	-	√	-	Prosperous Family Stage I
5	3	Self-employed	Rp. 1.000.000 - Rp. 2.000.000	-	-	√	√	-	Prosperous Family Phase II
6	1	Private sector employee	Rp. 1.000.000 - Rp. 2.000.000	√	-	-	√	-	Prosperous Family Stage I
7	2	Businessman	Rp. 1.000.000 - Rp. 2.000.000	-	-	√	√	-	Prosperous Family Phase II
8	3	Farmer	Rp. 500.000 - Rp. 1.000.000	-	√	√	√	-	Prosperous Family Phase II
9	2	Driver	Rp. 500.000 - Rp. 1.000.000	-	-	-	√	-	Prosperous Family Stage I
10	3	Self-employed	Rp. 2.000.000 - Rp. 4.000.000	-	-	√	√	-	Prosperous Family Phase II
11	4	Private sector employee	Rp. 1.000.000 - Rp. 2.000.000	√	-	√	√	-	Prosperous Family Phase II
12	2	Self-employed	Rp. 1.000.000 - Rp. 2.000.000	-	-	-	√	-	Prosperous Family Stage I
13	3	Businessman	Rp. 2.000.000 - Rp. 4.000.000	-	-	√	√	-	Prosperous Family Phase II
14	4	Farmer	Rp. 500.000 - Rp. 1.000.000	-	√	√	√	-	Prosperous Family Phase II
15	3	Private sector employee	Rp. 2.000.000 - Rp. 4.000.000	√	-	√	√	√	Prosperous Family Stage III+
16	3	Laborer	Rp. 500.000 - Rp. 1.000.000	-	-	√	√	-	Prosperous Family Phase II

17	2	Private sector employee	Rp. 1.000.000 - Rp. 2.000.000	√	-	√	√	-	Prosperous Family Phase II
18	4	Laborer	Rp. 500.000 - Rp. 1.000.000	-	-	√	√	-	Prosperous Family Phase II
19	2	Driver	Rp. 500.000 - Rp. 1.000.000	-	-	-	√	√	Prosperous Family Stage III
20	3	Self-employed	Rp. 1.000.000 - Rp. 2.000.000	√	-	-	√	-	Prosperous Family Stage I

From the research conducted to apply the correlation method or relationship to the correlation of work on the level of welfare of the residents, therefore the authors took 20 sample data in Mangga Village which can be shown in Table 1. The data is a transactional database which will be represented as Table 2, 3, 4, 5.

Table 2: Dependent Data

No.	Dependents	Code
1	1-2	T1
2	3-4	T2
3	<5	T3

Table 3: Employment data

No.	Work	Code
1	Laborer	M1
2	Private sector employee	M2
3	Farmer	M3
4	Driver	M4
5	Self-employed	M5
6	Businessman	M6

Table 1: Revenue Data

No.	Income	Code
1	<= 1.000.000	P1
2	> 1.000.000 - < 2.000.000	P2
3	>= 2.000.000	P3

Table 5: Prosperity level

No.	Prosperity level	Code
1	Welfare Stage I	K1
2	Welfare Stage II	K2
3	Welfare Stage III	K3
4	Welfare Stage III+	K4

And the data that has been formed in tabular form can be shown in Table 6.

Table 6: Data Representation

No.	Dependents			Work						Income			Asset					Prosperity level			
	T1	T2	T3	M1	M2	M3	M4	M5	M6	P1	P2	P3	A1	A2	A3	A4	A5	K1	K2	K3	K4
1	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	0	1	0	0	0	
2	1	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	1	0	0
3	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	0	1	0	0
4	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0
5	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	1	0	0
6	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1	0	0	0
7	1	0	0	0	0	0	0	0	1	0	1	0	0	0	1	1	0	0	1	0	0
8	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	0	1	0	0
9	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	0
10	0	1	0	0	0	0	0	1	0	0	0	1	0	0	1	1	0	0	1	0	0
11	0	1	0	0	1	0	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0
12	1	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0
13	0	1	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	1	0	0
14	0	1	0	0	0	1	0	0	0	1	0	0	0	1	1	1	0	0	1	0	0
15	0	1	0	0	1	0	0	0	0	0	0	1	1	0	1	1	1	0	0	0	1
16	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0
17	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	0	0	1	0	0
18	0	1	0	1	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0

19	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	1	0
20	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	1	0	0	0
Σ	8	12	0	3	5	3	3	4	2	9	8	3	5	3	13	20	2	6	12	1	1

To correlate work to the level of welfare of citizens, an a priori calculation is used. In this a priori calculation there are several processes, namely:

1. Item-set Configuration

The following is a solution based on the data that has been provided. The process of forming C1 or called 1 itemset with a minimum amount of support = 15%. With the following formula:

$$\text{Support (A)} = \frac{\Sigma \text{ transaksi mengandung A}}{\Sigma \text{ transaksi}} * 100\%$$

Table 7: Item-set

No	Count	Support
T1	8/20	40%
T2	12/20	60%
T3	0/20	0%
M1	3/20	15%
M2	5/20	25%
M3	3/20	15%
M4	3/20	15%
M5	4/20	20%
M6	2/20	10%
P1	9/20	45%
P2	8/20	40%
P3	3/20	15%
A1	5/20	25%
A2	3/20	15%
A3	13/20	65%
A4	20/20	100%
A5	2/20	10%
K1	6/20	30%
K2	12/20	60%
K3	1/20	5%
K4	1/20	5%

After obtaining 1 itemset, some data that meets the predetermined value is selected, the value itself is the number limit used to obtain the selected number, the support value is 15%, as shown in Table 8 the following:

Table 8: 1 Item-set

No	Count	Support
T1	8/20	40%
T2	12/20	60%
M2	5/20	25%
M5	4/20	20%
P1	9/20	45%
P2	8/20	40%
A1	5/20	25%
A3	13/20	65%
A4	20/20	100%
K1	6/20	30%
K2	12/20	60%

2. 2 Item-Set Combination

The process of forming C2 or called 2 itemset with a minimum amount of support = 15%. Table 9 below shows a combination of 2 itemsets with the following formula:

$$\text{Support (A, B)} = \frac{\Sigma \text{ transaksi mengandung A dan B}}{\Sigma \text{ transaksi}} * 100\%$$

Combinations of 2 itemsets that do not meet the minimum support requirements will be removed.

Table 9: 2 Item-set

Parts	Count	Support
T1&T2	0/20	0
T1&M2	3/20	15%
T1&M5	1/20	5%
T1&P1	3/20	15%
T1&P2	6/20	30%
T1&A1	2/20	10%
T1&A3	3/20	15%
T1&A4	8/20	40%
T1&K1	4/20	20%
T1&K2	3/20	15%
T2&M2	1/20	5%
T2&M5	3/20	15%
T2&P1	6/20	30%
T2&P2	3/20	15%
T2&A1	3/20	15%
T2&A3	10/20	50%
T2&A4	12/20	60%
T2&K1	2/20	10%
T2&K2	10/20	10
M2&M5	0/20	0
M2&P1	0/20	0
M2&P2	4/20	20%
M2&A1	4/20	20%
M2&A3	4/20	20%
M2&A4	5/20	25%
M2&K1	1/20	5%
M2&K2	3/20	15%
M5&P1	0/20	0
M5&P2	3/20	15%
M5&A1	0/20	0
M5&A3	2/20	10%
M5&A4	4/20	20%
M5&K1	2/20	10%
M5&K2	2/20	10%
P1&P2	0/20	0
P1&A1	0/20	0
P1&A3	5/20	25%
P1&A4	9/20	45%
P1&K1	3/20	15%

P1&K2	5/20	25%
P2&A1	4/20	20%
P2&A3	5/20	25%
P2&A4	8/20	40%
P2&K1	3/20	15%
P2&K2	5/20	25%
A1&A3	3/20	15%
A1&A4	5/20	25%
A1&K1	2/20	10%
A1&K2	2/20	10%
A3&A4	13/20	65%
A3&K1	0/20	0
A3&K2	12/20	60%
A4&K1	6/20	30%
A4&K2	12/20	60%
K1&K2	0/20	0

3. Combination of 3 itemsets

The process of forming C3 or called the minimum amount of support = 10%, the calculation results can be seen in Table 10 with the following formula:

$$\text{Support (A, B)} = \frac{\Sigma \text{transaksi mengandung A, B \& C}}{\Sigma \text{transaksi}} * 100\%$$

Table 10: 3 Item-set

Parts	Count	Support
T1, P2 & A4	5/20	25%
T1, P2 & K1	2/20	10%
T1, A4 & K1	4/20	20%
T2, P1 & A3	5/20	25%
T2, P1 & A4	6/20	30%
T2, A3 & A4	10/20	50%
M2, P2 & A1	3/20	15%
M2, P2 & A3	3/20	15%
M2, P2 & A4	4/20	20%
M2, A1 & A3	3/20	15%
M2, A3 & A4	4/20	20%
P1, A3 & A4	5/20	25%
P1, A3 & K2	5/20	25%
P1, A4 & K2	4/20	20%
P2, A1 & A3	2/20	10%
P2, A1 & A4	4/20	20%
P2, A1 & K2	2/20	10%
P2, A3 & A4	5/20	25%
P2, A3 & K2	5/20	25%
P2, A4 & K2	5/20	25%
A3, A4 & K2	12/20	60%
A4, K1 & K2	0/20	0%

4. Combination of 4 itemset

The process of forming C4 or called the minimum amount of support = 10%, the calculation results can be seen in Table 11 with the following formula:

$$\text{Support (A, B)} = \frac{\Sigma \text{ transaksi mengandung A, B, C \& D}}{\Sigma \text{ transaksi}} * 100\%$$

Table 11: 4 Itemset

Parts	Count	Support
T1, P2, A4 & K1	2/20	10%
T2, P1, A3 & A4	5/20	25%
M2, P2, A1 & A3	2/20	10%
M2, P2, A1 & A4	3/20	15%
M2, P2, A3 & A4	3/20	15%
P1, A3, A4 & K2	5/20	25%
P2, A1, A3 & A4	2/20	10%
P2, A1, A3 & K2	2/20	10%
P2, A3, A4 & K2	5/20	25%

5. Combination of 5 itemsets

The process of forming C5 or called the minimum amount of support = 10%, the calculation results can be seen in Table 12 with the following formula:

$$\text{Support (A, B)} = \frac{\Sigma \text{ transaksi mengandung A, B, C \& D}}{\Sigma \text{ transaksi}} * 100\%$$

Table 12: 5 Itemset

Parts	Count	Support
T2, M2, P2, A4, & K2	5/20	25%
M2, P2, A1, A3 & A4	2/20	10%
P1, A3, A4, K2 & P2	0/20	0%

6. Formation of Association Rules

After all the high frequency patterns are found, then the association rules are searched with the results of the frequency patterns shown in Table 13 as follows:

Table 13: Qualifying High Frequency Pattern Results

Parts	Count	Support
T2, M2, P2, A4, & K2	5/20	25%

After all high-frequency patterns are found, then we look for association rules that meet the minimum requirements for confidence by calculating the confidence or association $A \rightarrow B$, with a minimum confidence of 10%.

Table 14: Association Final Results

If antecedent then consequent	Support	Confidence	S*C
If T2, M2, P2, A4, \rightarrow K2	25%	100%	25%

7. Calculation of Lift Ratio

The lift ratio value is usually used to determine whether an association rule is valid or invalid. Expected Confidence calculation formula can be seen as follows:

$$\text{Expected Confidence} = \frac{\Sigma \text{ transaksi Mengandung Konsekuensi}}{\Sigma \text{ transaksi}}$$

$A \Rightarrow B$

(Motorcycle Assets \Rightarrow Welfare Phase I)

Number of Item B = 5 and Number of Transactions = 20

Table 15: Expected Confidence

Parts	Count	Support
If T2, M2, P2, A4, \rightarrow K2	5/20	25%

Below is the formula for calculating the Lift Ratio Value

$$Lift\ Ratio = \frac{Confidance}{Expected\ confidence}$$

A => B (Motorcycle Assets => Welfare Phase I)

Confidence value A & B = 0.25% and Expected Confidence = 25%.

Table 16: Lift Ratio

If antecedent then consequent	Support	Confidence	Expected confidence	Lift ratio
If T2, M2, P2, A4, → K2	25%	100%	25%	4%

After carrying out the above case trials with a minimum support = 25%, confidence = 100% so that the rule results that meet the support and confidence values are obtained, it is concluded that if the assets owned are A4 (motorcycles), with T2 dependents (3-4), with an M2 job (Private Employee) and has A4 assets (Motorcycles), with P2 income (> 1,000,000 - < 2,000,000), then it is included in K2 (welfare stage II) with a support value of 20%, 100% confidence.

3.2. Flowchart

The flowchart design can be described as follows:

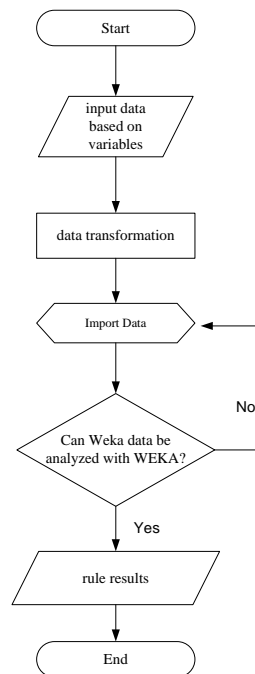


Figure 1: Flowchart

process flowchart steps:

1. In the picture above it is explained starting with start,
2. Input data according to existing variables
3. Data transformation
4. Then import research data
5. Analyzing data using WEKA data software
6. Then the association rules are generated and end with end

3.3. Use Case Diagrams

The following is a use case diagram for the data model in the form of a diagram that can explain a system process flow that will be built as shown in Figure 2.

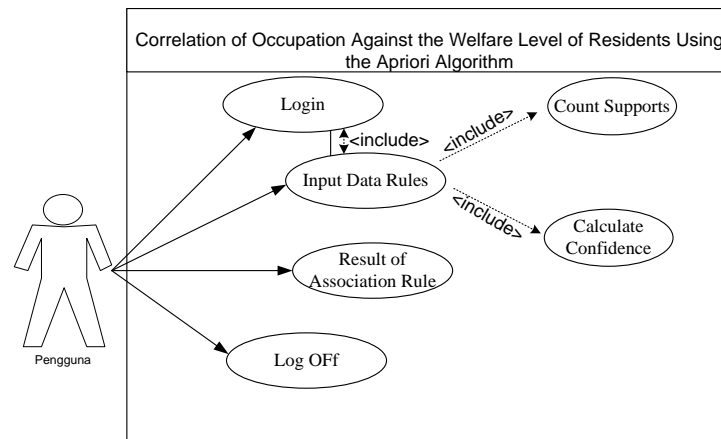


Figure 2: Use Case Diagrams

4. Conclusion

After carrying out the above case trials with a minimum support = 25%, confidence = 100% so that the rule results that meet the support and confidence values are obtained, it is concluded that if the assets owned are A4 (motorcycles), with T2 dependents (3-4), with an M2 job (Private Employee) and has A4 assets (Motorcycles), with P2 income (> 1,000,000 - < 2,000,000), then it is included in K2 (welfare stage II) with a support value of 20%, 100% confidence.

References

- [1] Annisa Ilham Fathimah, Sudin Saepudin, & Edwinanto. (2022). Penerapan Data Mining Dengan Metode Apriori Pada Penjualan Sembako (Studi Kasus: Grosir Sembako Lina). *Jurnal Rekayasa Teknologi Nusa Putra*, 8(2), 123–131. <https://doi.org/10.52005/rekayasa.v8i2.271>
- [2] Aziz Muslim, M., Prasetyo, B., Laily Harum Mawarni, E., Juli Herowati, A., Mirqotussa'adah, Hardiyanti Rukmana, S., & Nurzahputra, A. (2019). Data Mining Algoritma C4.5. In *Nucl. Phys.* (Vol. 13, Issue 1).
- [3] Badan Pusat Statistik. (2010). Modul 8 Ketenagakerjaan. Jakarta: Badan Pusat Statistik.
- [4] Budiman, E. (2016). *algoritma & Pemrograman*. CV Andi Offset, Yogyakarta.
- [5] Cara, I., Buaton, R., & Sihombing, M. (2022). Korelasi Nilai Literasi Rapot K.13 Dalam Nilai Pengetahuan Dan Keterampilan Pada Siswa Menggunakan Metode Apriory. *Jurnal Ilmiah Kaputama (JIKA)*, 6(2).
- [6] Romdon, D., & Kholil, I. (2022). Implementasi Data Mining dengan Metode Apriori Dalam Menentukan Pola Pemilihan Pemeriksaan Kimia. *TIN: Terapan Informatika Nusantara*, 2(10), 642–651. <https://doi.org/10.47065/tin.v2i10.1349>
- [7] Sabilla, M. Y., Lumbanbatu, K., & Prahmana, I. G. (2022). Implementasi Algoritma Apriori Dalam Perencanaan Persediaan Alat Kesehatan Pada Apotek. *Jurnal Teknik Informatika Kaputama (JTIK)*, 6(2), 434–447.
- [8] Sari, A. O., & Abdilah, A. (2015). *Web Programming*. Andi Offset, Yogyakarta.
- [9] Solichin. (2010). *MySQL*, Andi Offset, Yogyakarta.
- [10] Buulolo S.Kom, M.Kom, E. (2020). *Data Mining* (E. Buulolo S.Kom, M.Kom (ed.); 1st ed.). Grup Penerbit CV Budi Utama.