

## Application of Apriori Algorithm to Find Patterns of Population Mortality Data (Case Study: Disdukcapil Stabat)

Rizka Nova Fitria<sup>1\*</sup>, Novriyenni<sup>2</sup>, Ratih Puspadini<sup>3</sup>

<sup>1,2,3</sup>STMIK Kaputama

[riskanova242@gmail.com](mailto:riskanova242@gmail.com)<sup>1\*</sup>, [novriyenni.sikumbang@gmail.com](mailto:novriyenni.sikumbang@gmail.com)<sup>2</sup>, [puspadini.ratih@gmail.com](mailto:puspadini.ratih@gmail.com)<sup>3</sup>

### Abstract

The development of information technology provides great opportunities in data utilization, including in government agencies. One of the important data managed by the Population and Civil Registration Office (Disdukcapil) is population mortality data. This data not only serves as an administrative archive, but can also be analyzed to identify important patterns related to the factors causing death. This study aims to apply the Apriori algorithm in identifying association patterns from population death data based on factors such as age, gender, occupation, cause of death, and address at the Disdukcapil Stabat. The method used is data mining with the Apriori algorithm, through the stages of data processing, determining the support, confidence, and lift values until a rule is formed. The results of the study show that 173 association rules were formed, with the best rule having the highest support value of 6% and confidence of 10%. The rule states that if the age of the population is over 56 years with an address in Stabat, then the tendency of gender is male, occupation as an entrepreneur, and sudden death.

**Keywords:** Apriori algorithm, population mortality, Rapidminer

### 1. Introduction

Information technology advances in various sectors including government institutions as well as organizations and individuals, to keep up with existing developments. By utilizing information technology, data that was previously only an administrative archive can be transformed into valuable information. One important example is population death data, which is data that contains information about individuals who have died, including names, ages, genders, time of death, and the cause of death. Death data not only records the death rate, but also contains important information that can be analyzed to see patterns of causes of death, and risk factors that are developing in society. [1]

The Population and Civil Registration Office (Disdukcapil) has a strategic role in managing population data, including population death data. This data is not only important for administrative purposes, but can also be used for more in-depth analysis, such as understanding the patterns of deaths that occur in the Stabat region. This population death data includes various information, such as age, gender, address, and cause of death. However, the data is often only used as an administrative report without further utilization for in-depth analysis.

Therefore, there is a need for a system that can utilize information technology for data analysis by identifying important patterns, such as the relationship between age factors or causes of death, that can help in the formulation of public health policies. By applying a priori algorithms to this system, it is possible to identify significant patterns of association, such as patterns based on specific age groups or specific locations. The application of this algorithm can provide useful insights for the Disdukcapil and related agencies in an effort to improve the quality of health services, design prevention programs, and optimize resource allocation.

### 2. Method

Based on research conducted by "Identification of Patterns of Areas with Suicide Cases in West Java Using a Priori Algorithm". Based on the results of the experiment, this study used a minimum Support (MS) value of 2% in 2019-2021, and MS 0.2% in 2020. Meanwhile, the confidence value for 3 years is 50%. The results of the study show that there is a consistent regional pattern every year, namely experiencing suicide cases for two consecutive years (2020-2021), namely Pagaden District in Subang Regency and East Cikarang District with Jatibaru Village in Bekasi Regency.[2]

Based on research conducted by "Analysis of the Relationship Between Disease Symptoms Using a Priori Algorithm in the Health Sector". From the results of the study conducted with a Support value for both symptoms (polyuria and polydipsia) was 0.6, which indicates that 60% of all patients experienced both symptoms at the same time. The confidence that patients with polyuria also experience polydipsia is

0.75, while the confidence that patients with polydipsia also experience polyuria is 0.857. This suggests that there is a fairly strong association between the two symptoms in the context of the diagnosis of type 2 diabetes.[3]

### 2.1. Algorithm a Apriori

A priori algorithm is a basic algorithm proposed by Agrawal & Skrikant in 1994 to determine the frequent itemset for boolean association rules. A priori algorithms include a type of association rules in data mining. Rules that state associations between several attributes are often called affinity analysis or market basket analysis. Association rule analysis is a data mining technique to find the rules of a combination of items. One of the stages of association analysis that attracts the attention of many researchers to produce efficient algorithms is frequent pattern mining analysis. The importance of an association can be determined by two benchmarks, namely: support and confidence, support is the percentage of the combination of these items in the database, while confidence is the strength of the relationship between items in the association rules. [4]

Association rules capture items or events in large data that contains transaction data. With the advancement of technology, sales data can be stored in large quantities called "basket data." The association rules defined in the data basket are used for promotion, catalog design, customer segmentation and marketing targets.

The formation of association rules that meet the minimum requirements for confidence by calculating the confidence of the  $A \rightarrow B$  associative rule, where support is supporting data and confidence is confidence.

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

$$\text{Support (A)} = \frac{\sum \text{TransaksimengandungAdanB}}{\sum \text{Jumlahseluruhtransaksi}} \times 100\% \quad (1)$$

$$\text{Confidence (A)} = \frac{\sum \text{TransaksimengandungAdanB}}{\sum \text{TransaksimengandungA}} \times 100\% \quad (2)$$

Searches for the minimum qualifying combination of items from the *support* value in the database. The *support* value of an item is obtained using the following formula:

$$\text{Support (A)} = \frac{\text{Jumlah transaksi mengandung A}}{\text{Total Transaksi}} \quad (3)$$

The support value of the 2 items is obtained using the formula:

$$\text{Support (A,B)} = \frac{\sum \text{Jumlah transaksi mengandung A dan B}}{\sum \text{transaksi}} \quad (4)$$

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

$$\text{Confidence} = P(B|A) = \frac{\sum \text{J transaksi mengandung A dan B}}{\sum \text{transaksi}} \quad (5)$$

### 2.2. Data Mining

Data mining is an effort to extract useful information and patterns from a very large amount of data. The data mining process includes the stages of data collection, data extraction, data analysis, and data statistics. Data mining in other terms is guided as knowledge discovery, knowledge extraction, data or pattern analysis, information harvesting, and others. The important goal of data mining is to transform raw data into useful information to support more accurate and strategic business decision-making that is more effective and relevant.[5]

### 2.3. Factors That Cause Death

Factors that cause death are circumstances that can generally be known to every normal human being at the time when an action can cause a certain result or an action that is done intentionally or unintentionally, on the basis of the behavior of the attitude that is carried out so as to cause death. Below are the indicators of the cause of death which can be seen in the table below: [6]

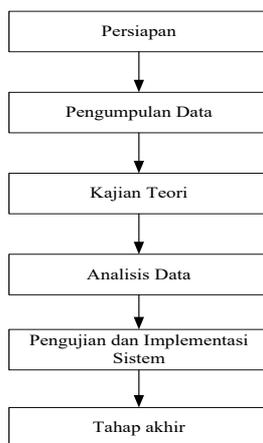
**Table 1:** Indikator penyebab kematian

Causes of Death	Indicators of Understanding
Sakit	Kematian yang disebabkan oleh penyakit akut atau umum yang tidak tergolong kronis atau menahun. Seperti infeksi saluran pernapasan akut, demam berdarah atau diare akut.
Sakit Kronis	Kematian akibat penyakit jangka panjang seperti kanker, diabetes, atau gagal ginjal.
Bunuh Diri	Kematian yang disebabkan oleh tindakan sengaja dari individu untuk mengakhiri hidupnya.

Causes of Death	Indicators of Understanding
Lanjut Usia	Kematian yang terjadi secara alami karena faktor usia lanjut tanpa penyakit tertentu.
Kematian Mendadak	Kematian tiba-tiba tanpa gejala sebelumnya, misalnya karena serangan jantung mendadak.
Covid-19	Kematian yang disebabkan oleh infeksi virus Corona yang terkonfirmasi secara medis.
Kecelakaan	Kematian yang disebabkan oleh kejadian tidak terduga seperti kecelakaan lalu lintas atau kerja.
Pembunuhan	Kematian akibat tindakan kekerasan yang disengaja oleh orang lain (kriminal/hukum).
Bencana Alam	Kematian akibat kejadian alam seperti gempa bumi, banjir, tanah longsor, tsunami, dll.
Keracunan	Kematian yang diakibatkan oleh zat berbahaya, baik sengaja maupun tidak (obat, racun).
Malnutrisi atau Kelaparan	Kematian karena kekurangan gizi parah atau tidak mendapatkan asupan makanan yang cukup.
Komplikasi Medis	Kematian yang terjadi akibat komplikasi dari tindakan medis atau kesalahan penanganan.
Penyakit Menular Lain	Kematian karena penyakit infeksi selain Covid-19, seperti TBC, HIV/AIDS, malaria, dll.
Penyakit Tidak Menular	Kematian karena penyakit non-infeksi seperti hipertensi, stroke, atau penyakit jantung.

### 3. Results and Discussion

The research method carried out is for something systematically using scientific methods and applicable sources. In the process of this research, it is shown to provide more meaningful results for the agency in determining the pattern of population mortality data in the Slabat Disdukcapil. The results of the conceptualization will be made into a research method that uses the pattern of *literature* study as intended in the following Image.



**Fig. 3:** Research Workflow

#### 3.1 Research Supporting Data

The following is the research supporting data obtained from the Disdukcapil to be used in the data processing process as shown in the table below.

**Table 3:** Data Pendukung Penelitian

No	Age	Gender	Work	Causes of Death	Address
1	55	Perempuan	Karyawan Swasta	Tua/Lanjut Usia	Pematang Jaya
2	57	Laki-laki	Karyawan Swasta	Penyakit Kronis	Stabat
3	72	Laki-laki	Petani	Bunuh Diri	Besitang
4	62	Laki-laki	Karyawan Swasta	Penyakit Kronis	Bahorok
5	78	Laki-laki	Petani	Kematian Mendadak	Sei Bingai
6	58	Laki-laki	Wiraswasta	Covid-19	Binjai
7	78	Perempuan	Wiraswasta	Tua/Lanjut Usia	Binjai
8	48	Laki-laki	Wiraswasta	Tua/Lanjut Usia	Binjai
9	58	Laki-laki	Wiraswasta	Tua/Lanjut Usia	Selesai
10	49	Laki-laki	Wiraswasta	Bunuh Diri	Binjai

### 3.2 Application of the Method

From the research conducted to apply the association rule method to analyze the pattern of population mortality data in Disdukcapil Stabat using a sample of mortality data that will be used as a support for the research, it can be seen in the Population Mortality Data table.

The following is a table of age indicators used to group respondents based on a specific age range. Each age range is assigned a code and accompanied by a description of the appropriate developmental phase which can be seen in the table below.

**Table 4: Age Data**

Code	Age	Information
U1	<12 Tahun	Masa Anak-Anak
U2	12-18 Tahun	Masa Remaja Awal
U3	18-25 Tahun	Masa Remaja Akhir
U4	26-35 Tahun	Masa Dewasa Awal
U5	36-45 Tahun	Masa Dewasa Akhir
U6	46-55 Tahun	Masa Lansia Awal
U7	>56 Tahun	Masa Lansia Akhir

The following are the indicators of work used in the data processing process, which can be seen in the table of work below.

**Table 5: Gender**

Code	Gender
JK1	Laki-laki
JK2	Perempuan

The following are the indicators of work used in the data processing process, which can be seen in the table of work below.

**Table 6: Work**

Code	Work
P1	Belum Bekerja
P2	Karyawan
P3	Petani
P4	Wiraswasta
P5	PNS
P6	MRT
P7	Pensiunan
P8	Buruh
P9	Guru
P10	Polisi
P11	Nelayan
P12	Supir
P13	TNI

The following are indicators of the causes of death used in the data processing process, which can be seen in the table of causes of death below.

**Table 7: Cause of Death**

Code	Causes of Death
K1	Sakit

Code	Causes of Death
K2	Penyakit Kronis
K3	Bunuh Diri
K4	Lanjut Usia/Tua
K5	Kematian Mendadak
K6	Covid-19
K7	Kecelakaan
K8	Bencana Alam
K9	Pembunuhan

The following is an address indicator used in the data processing process, which can be seen in the address table below.

Table 7: Address Data

Code	Address
A1	Binjai
A2	Babalan
A3	Besitang
A4	Bahorok
A5	Kabanjahe
A6	Sei Bingai
A7	Stabat
A8	Selesai
A9	Pematang Jaya
A10	Wampu
A11	Tanjung Pura
A12	Sei Ledan
A13	Kuala
A14	Salapian
A15	Gebang
A16	Batang Serangan
A17	Bukit Lawang
A18	Secanggang
A19	Pangkalan Berandan
A20	Pangkalan Susu
A21	Sirapit

And the data that has been formed in tabular form can be addressed in Tables 8 and 9. In tabular format that contains the numbers 0 if no transaction occurs and 1 when a transaction occurs, the attribute transformation is performed according to the specified simulation.

Table 8: Representation of Age, JK, Occupation Data

No	USIA							JK		PEKERJAAN					
	U1	U2	U3	U4	U5	U6	U7	JK1	JK2	P1	P2	P3	P4	P5	P6
1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0
2	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
3	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0
4	0	0	0	0	0	0	1	1	0	0	1	0	0	0	0
5	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0
6	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0
7	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0
8	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0
9	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0
10	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0

Table 9: Representation of Cause of Death and Address

No	PENYEBAB KEMATIAN	ALAMAT
----	-------------------	--------

	K1	K2	K3	K4	K5	K6	K7	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12
1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
6	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
10	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
$\Sigma$	3	2	3	7	3	1	1	5	2	1	1	1	1	2	1	1	3	1	1

From tables 8 and 9, the process of forming *support 1 itemset* will be carried out with a minimum amount of *support* = 10%.  
With the following formula:

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

The following are the results of the process of forming *support 1 itemset* with the minimum amount of *support* = 10% which can be seen in the *support table of 1 itemset* below:

**Table 10:** Representation of Cause of Death and Address

<i>ID</i>	<i>Count</i>	<i>Support</i>
U1	0/20	0
U2	0/20	0
U3	0/20	0
U4	1/20	0,05
U5	2/20	0,1
U6	5/20	0,25
U7	12/20	0,6
JK1	17/20	0,85
JK2	3/30	0,15
P1	0/20	0
P2	3/30	0,15
P3	2/20	0,1
P4	12/20	0,6
P5	1/20	0,05
P6	2/20	0,1
K1	3/30	0,15
K2	2/20	0,1
K3	3/30	0,15
K4	7/20	0,35
K5	3/30	0,15
K6	1/20	0,05
K7	1/20	0,05
A1	5/20	0,25
A2	2/20	0,1
A3	1/20	0,05
A4	1/20	0,05
A5	1/20	0,05
A6	1/20	0,05

<i>ID</i>	<i>Count</i>	<i>Support</i>
A7	2/20	0,1
A8	1/20	0,05
A9	1/20	0,05
A10	3/30	0,15
A11	1/20	0,05
A12	1/20	0,05

After obtaining 1 *itemset* selected some data that meets the predetermined value, the value itself is the limit of the number used to obtain the selected number, the *support* value is 15%, as seen in the following table:

**Table 10:** Support 1 itemset

<i>ID</i>	<i>Count</i>	<i>Support</i>
U6	5/20	25%
U7	12/20	60%
JK1	17/20	85%
JK2	3/30	15%
P2	3/30	15%
P4	12/20	60%
K1	3/30	15%
K3	3/30	15%
K4	7/20	35%
K5	3/30	15%
A1	5/20	25%
A10	3/30	15%

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

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

**Table 11:** Support 2 itemset

<i>ID</i>	<i>Count</i>	<i>Support</i>
<i>U6, JK1</i>	4/20	0,2
<i>U6, JK2</i>	1/20	0,05
<i>U6, P2</i>	1/20	0,05
<i>U6, P4</i>	4/20	0,2
<i>U6, K1</i>	1/20	0,05
<i>U6, K3</i>	1/20	0,05
<i>U6, K4</i>	3/20	0,15
<i>U6, K5</i>	0/20	0
<i>U6, A1</i>	2/20	0,1
<i>U6, A10</i>	1/20	0,05
<i>U7, JK1</i>	10/20	0,5
<i>U7, JK2</i>	2/20	0,1
<i>U7, P2</i>	2/20	0,1
<i>U7, P4</i>	5/5	0,25
<i>U7, K1</i>	1/20	0,05
<i>U7, K3</i>	1/20	0,05
<i>U7, K4</i>	4/20	0,2
<i>U7, K5</i>	3/20	0,15

<b>ID</b>	<b>Count</b>	<b>Support</b>
<i>U7, A1</i>	2/20	0,1
<i>U7, A10</i>	2/20	0,1
<i>JK1, P2</i>	2/20	0,1
<i>JK1, P4</i>	11/20	0,55
<i>JK1, K1</i>	3/20	0,15
<i>JK1, K3</i>	3/20	0,15
<i>JK1, K4</i>	5/20	0,25
<i>JK1, K5</i>	2/20	0,1
<i>JK1, A1</i>	4/20	0,2
<i>JK1, A10</i>	2/20	0,1
<i>JK2, P2</i>	1/20	0,05
<i>JK2, P4</i>	1/20	0,05
<i>JK2, K1</i>	0/20	0
<i>JK2, K3</i>	0/20	0
<i>JK2, K4</i>	2/20	0,1
<i>JK2, K5</i>	1/20	0,05
<i>JK2, A1</i>	1/20	0,05
<i>JK2, A10</i>	1/20	0,05
<i>P2, K1</i>	0/20	0
<i>P2, K3</i>	0/20	0
<i>P2, K4</i>	1/20	0,05
<i>P2, K5</i>	0/20	0
<i>P2, A1</i>	0/20	0
<i>P2, A10</i>	0/20	0
<i>P4, K1</i>	3/20	0,15
<i>P4, K3</i>	2/20	0,1
<i>P4, K4</i>	5/20	0,25
<i>P4, K5</i>	0/20	0
<i>P4, A1</i>	5/20	0,25
<i>P4, A10</i>	1/20	0,05
<i>K1, A1</i>	0/20	0
<i>K1, A10</i>	1/20	0,05
<i>K3, A1</i>	1/20	0,05
<i>K3, A10</i>	0/20	0
<i>K4, A1</i>	2/20	0,1
<i>K4, A10</i>	0/20	0
<i>K5, A1</i>	0/20	0
<i>K5, A10</i>	2/20	0,1

After obtaining 2 itemsets, some data is selected that meets the predetermined value, the value itself is the limit of the number used to obtain the selected number, the support value is 15%, as seen in the Table below.

**Table 12:** Support 2 itemset

<b>ID</b>	<b>Count</b>	<b>Support</b>
<i>U6, JK1</i>	4/20	20%
<i>U6, P4</i>	4/20	20%
<i>U6, K4</i>	3/20	15%
<i>U7, JK1</i>	10/20	50%
<i>U7, P4</i>	5/5	25%
<i>U7, K4</i>	4/20	20%
<i>U7, K5</i>	3/20	15%
<i>JK1, P4</i>	11/20	55%
<i>JK1, K1</i>	3/20	15%
<i>JK1, K3</i>	3/20	15%
<i>JK1, K4</i>	5/20	25%
<i>JK1, A1</i>	4/20	20%
<i>P4, K1</i>	3/20	15%
<i>P4, K4</i>	5/20	25%
<i>P4, A1</i>	5/20	25%

From the table above, the element above T means interconnected items, while F means no related items. The number of *itemset* frequencies must be greater than the number of *Itemset* frequencies  $\theta$ . From the table above,  $f_3$  is obtained as follows:

{U6, JK1, P4}, {U6, JK1, K4}, {U6, P4, K4}, {U7, JK1, P4}, {U7, JK1, K4}, {U7, JK1, K5}, {U7, P4, K4}, {U7, P4, K5}, {U7, K4, K5}, {JK1, P4, K1}, {JK1, P4, K3}, {JK1, P4, K4}, {JK1, P4, A1}, {JK1, K1, K3}, {JK1, K1, K4}, {JK1, K1, A1}, {JK1, K3, K4}, {JK1, K3, K4}, {JK1, K3, A1}, {JK1, K4, K4}, {JK1, K4, A1}, {P4, K1, K4}, {P4, K1, A1}, {P4, K4, A1}.

The process of forming C3 or called the minimum amount of *support* = 15%, the calculation results can be seen in table III.15 with the following formula:

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

Table 13: Support 3 itemset

ID	Count	Support
U6, JK1, P4	4/20	0,2
U6, JK1, K4	2/20	0,1
U6, P4, K4	2/20	0,1
U7, JK1, P4	4/20	0,2
U7, JK1, K4	3/20	0,15
U7, JK1, K5	2/20	0,1
U7, P4, K4	3/20	0,15
U7, P4, K5	0/20	0
U7, K4, K5	0/20	0
JK1, P4, K1	3/20	0,15
JK1, P4, K3	2/20	0,1
JK1, P4, K4	4/20	0,2
JK1, P4, A1	4/20	0,2
JK1, K1, K3	0/20	0
JK1, K1, K4	0/20	0
JK1, K1, A1	0/20	0
JK1, K3, K4	0/20	0
JK1, K3, A1	1/20	0,05
JK1, K4, A1	1/20	0,05
P4, K1, K4	0/20	0
P4, K1, A1	0/20	0
P4, K4, A1	2/20	0,1

After obtaining 3 itemsets, some data that meets the predetermined value, the value itself is the limit of the number used to obtain the selected number, the support value is 15%, as seen below:

Table 14: Support 3 itemset

ID	Count	Support
U6, JK1, P4	4/20	20%
U7, JK1, P4	4/20	20%
U7, JK1, K4	3/20	15%
U7, P4, K4	3/20	15%
JK1, P4, K1	3/20	15%
JK1, P4, K3	2/20	10%
JK1, P4, K4	4/20	20%
JK1, P4, A1	4/20	20%

From the Table above, the element above T means interrelated items, while F means no related items. The number of *itemset* frequencies must be greater than the number of *itemset* frequencies  $\theta$ . From the table above,  $f_3$  is obtained: {U7, JK1, P4, K4}, {JK1, P4, K1, K3}, {JK1, P4, K1, K4}, {JK1, P4, K1, A1}, {JK1, P4, K3, K4}, {JK1, P4, K3, A1}, {JK1, P4, K4, A1}.

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

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

**Table 15:** Support 4 itemset

ID	Count	Support
U7, JK1, P4, K4	2/20	10%
JK1, P4, K1, K3	0/20	0%
JK1, P4, K1, K4	0/20	0%
JK1, P4, K1, A1	0/20	0%
JK1, P4, K3, K4	0/20	0%
JK1, P4, K3, A1	1/20	5%
JK1, P4, K4, A1	1/20	5%

After obtaining 4 itemsets, some data that meets the specified value, the value itself is the limit of the number used to obtain the selected number, the support value is 5%, as seen in the following table:

**Table 16:** Support 4 itemset

ID	Count	Support
U7, JK1, P4, K4	2/20	10%
JK1, P4, K3, A1	1/20	5%
JK1, P4, K4, A1	1/20	5%

After all the high-frequency patterns are found, then look for the association rules with the results of the frequency patterns shown in the following table:

**Table 17:** High-frequency pattern results

ID	Count	Support
U7, JK1, P4, K4	2/20	10%

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

**Table 18:** Highest Pattern Results

ID	Count	Support
U6, JK1	4/20	20%
U6, P4	4/20	20%
U6, K4	3/20	15%
U7, JK1	10/20	50%
U7, P4	5/5	25%
U7, K4	4/20	20%
U7, K5	3/20	15%
JK1, P4	11/20	55%
JK1, K1	3/20	15%
JK1, K3	3/20	15%
JK1, K4	5/20	25%
JK1, A1	4/20	20%
P4, K1	3/20	15%
P4, K4	5/20	25%
P4, A1	5/20	25%
U6, JK1, P4	4/20	20%
U7, JK1, P4	4/20	20%
U7, JK1, K4	3/20	15%
U7, P4, K4	3/20	15%
JK1, P4, K1	3/20	15%
JK1, P4, K4	4/20	20%
JK1, P4, A1	4/20	20%

ID	Count	Support
U7, JK1, P4, K4	2/20	10%

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

After experimenting with the above case with a minimum support = 10%, confidence = 100% so that the results of the rule that meet the support and confidence values are obtained, namely:

1. If the age is >56 years old, then the work is self-employed, then the score is successful with a support value of 25%, a confidence value of 100% and a Lift ratio value of 4.00
2. If the age is 46-55 years old, the gender is male, then the job is self-employed, then the score is successful with a support value of 20%, a confidence value of 100% and a Lift ratio value of 5.00
3. If the age is >56 years old, the gender is male, then the job is self-employed, then the score is successful with a support value of 20%, a confidence value of 100% and a Lift ratio value of 5.00
4. If the age is >56 years old, the gender is male, then the cause of death is due to old age/old age, then the score is successful with a support value of 15%, a confidence value of 100% and a Lift ratio value of 6.67
5. If the age of >56 years old, the work is self-employed, then the cause of death is due to old age/old age, then the score is successful with a support value of 15%, a confidence value of 100% and a Lift ratio value of 6.67
6. If the gender is male, the job is self-employed, then the cause of death is due to illness, then the score is successful with a support value of 15%, a confidence value of 100% and a Lift ratio value of 6.67
7. If the gender is male, the job is self-employed, then the cause of death is suicide, then the score is successful with a support value of 10%, a confidence value of 100% and a Lift ratio value of 10.00
8. If the gender is male, the job is self-employed, then the cause of death is elderly/old, then the score is successful with a support value of 20%, a confidence value of 100% and a Lift ratio value of 5.00
9. If the gender is male, the job is self-employed, then the address is binjai, then the score is successful with a support value of 20%, a confidence value of 100%, and a Lift ratio value of 5.00
10. If the age is >56 years old, the gender is male and the job is self-employed, then the cause of death is due to old age, then the score is successful with a support value of 10% confidence value of 100% and a Lift ratio value of 10.00

### 3.4. Program Interface Discussion

The interface design that will be discussed in the current application is in the form of a page on Rapid Miner and a display on a web page. The interface design that will be discussed is as follows:

#### 1. Data Pages in Rapid Miner

This page is the initial view or the opening screen of the Rapid Miner software version 5.3. On this page, users can see the Rapid Miner logo as well as the license information of the software used, which is under the AGPL license version 3. In addition, this page also provides information that Rapid

Miner is provided without any warranty whatsoever and includes copyright and other sources of information that can be accessed through the official Rapid Miner website in <http://rapid-i.com>. This page usually appears when the software is in the process of loading icons and preparing the working environment before the user can start using the data analysis features available in Rapid Miner.

The image above is the initial view or homepage of the Rapid Miner software version 5.3 after the user opens the application and successfully enters the work environment. On this page, users are presented with several main options, namely creating a new process (New Process), opening a recently used process (Open Recent Process), opening a saved process (Open Process), opening the template that has been provided (Open Template), and accessing an online tutorial (Online Tutorial) to help users understand how Rapid Miner works. This page serves as the main navigation hub to easily and efficiently start various data analysis and processing activities in Rapid Miner.

The image above is the main view of the Rapid Miner application when the user starts a new process. In the middle is the main work area (Main Process) that is still empty, where users will design and organize data analysis workflows by adding various operators. On the left side there is the "Operators" panel which contains various categories of operators such as Process Control, Utility, Repository Access, Import, Export, Data Transformation, Modeling, and Evaluation. These operators can be selected and dragged to the work area to build a data analysis process. At the bottom left is the "Repositories" panel that displays the available data sources, including Samples, DBs, and Local Repository. On the right side is the "Parameters" panel that displays the settings for the operator or process that is being selected, as well as the "Context" panel that provides additional information. At the bottom, there are "Problems" and "Logs" tabs that are used to display error messages or activity logs during the process. Overall, this page is the main workspace for creating, editing, and running data analysis processes in Rapid Miner.

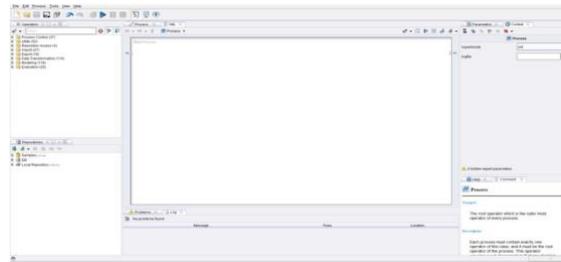


Fig. 8: Rapid Miner App View

This image is a view of the Rapid Miner software showing the data analysis process using the Association Rule Mining method. In this view, you can see a process flow diagram consisting of several interconnected operators. The process begins with the Read Excel operator which functions to read data from an Excel file, then the data is selected for its attributes through the Select Attributes operator. Furthermore, there is a Numerical to Nominal operator that converts numerical data into nominal data so that it can be further processed. After that, the Remap Binominals operator is used to remap the binomial values in the data, which usually serves to adjust the format or category values of the data. The process is then proceeded to the FP-Growth operator which is used to extract the frequent item-set pattern in the data, and ends with the Create Association Rules operator which generates the association rule based on that pattern. On the left side of the screen is the Operators panel to select the operators to use in the process, as well as the Repositories panel that shows where the data is stored and the processes that are being used. The right part of the screen displays the parameters and context of the process being built. Overall, this page shows the set of steps taken in the analysis of association rules using Rapid Miner, from data input to processing and rulemaking.

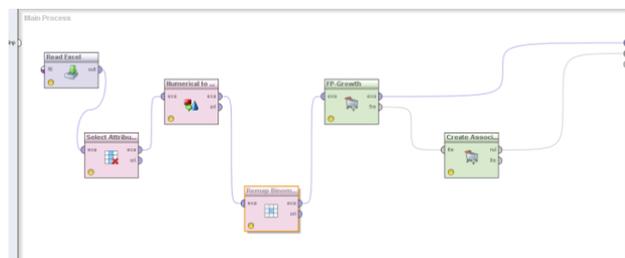


Fig. 9: Tampilan Jendela Rapid Miner

The image above is part of the Rapid Miner software interface that displays parameter settings for FP-Growth operators. On this page, users can set the custom parameters used in the frequent itemset pattern mining process. The first parameter is the option to search for the minimum number of itemsets, which is enabled with a checkmark, indicating that this feature is in use. Below it is a column to enter the min number of itemsets with the number 1000, which means that the FP-Growth algorithm will look for at least 1000 itemsets that appear frequently in the data. The second parameter is the min support set to 0.3, which indicates that only itemsets with at least 30% support in the dataset will be considered significant and used to generate association rules. This page is important because it allows users to control the sensitivity and output of the frequent pattern mining process, so that they can adjust the desired results based on the characteristics of the data and the purpose of the analysis.

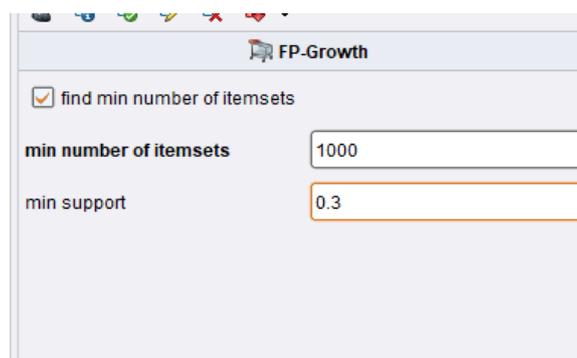


Fig. 10: FP-Growth Operator Page View

This image is part of the Rapid Miner software interface that shows parameter settings for the Create Association Rules operator. On this page, the user can choose the criteria to be used to evaluate the resulting association rules, where in this example the criteria chosen is confidence. In addition, users can also set the minimum trust value that the association rules must have in order to be accepted and displayed in the analysis results. In this figure, the minimum confidence value is set to 0.3, which means that only association rules with a minimum confidence level of 30% will be considered valid. This page is important because it allows users to control the tightness of the rules generated based on trust levels, thus helping to filter out less relevant rules and focus on stronger, more reliable patterns in the data.



## 4. Conclusions & Suggestions

### 4.1 Conclusion

Based on the research that has been carried out, it can be concluded how the application of a priori algorithm can be used to find out the relationship between age, gender, occupation, cause of death, address.

1. Using a priori algorithm, correlations between various factors such as age, gender, occupation, cause of death, and address can be implemented through the data analysis process by determining the value of support, confidence, and lift. From this process, a rule was formed that described a pattern of significant correlation in population death data.
2. The data processing application built is able to help the Disdukcapil Stabat or related agencies in finding out the pattern of relationships between mortality factors more quickly, efficiently, and accurately. This system presents an automatic linkage pattern so that it can support decision-making in efforts to prevent and handle the problem of population death.
3. The results of data processing using a priori algorithm show that of the 173 rules formed, there is a best rule with the highest support value of 6% and confidence of 10%. This proves that if the age is >56 with an address in the stabat, then the gender is male, the job is self-employed and the cause of sudden death is sudden death

### 4.2. Suggestions

Given the limitations in this study, both in terms of knowledge and data, the author provides several suggestions that are expected to be input and reference for future research, which are as follows:

1. Further research is suggested to use more and more diverse population mortality data so that the correlation patterns found with a priori algorithm become more accurate and able to describe real conditions as a whole.
2. The application system developed should continue to be improved by adding data visualization features such as graphs or interactive dashboards, so that information on population mortality patterns is easier to understand by the Disdukcapil and related agencies.
3. It is expected to use other methods such as fuzzy so that the results can be compared.

## References

- [1] F. Ridho Pratama, F. Nurahman Aziz, M. Rafli, A. Sarwana, M. Najib, and D. Nurhilman, "Pengembangan Sistem Informasi Pelayanan Akta Kelahiran Dan Kematian Pada Disdukcapil Kota Tangerang," *Jurnal MENTARI: Manajemen Pendidikan dan Teknologi Informasi*, vol. 2, no. 2, pp. 104–110, Mar. 2024, [Online]. Available: <https://journal.pandawan.id/mentari/article/view/383>
- [2] M. R. K. Fajar and S. Juanita, "Identifikasi Pola Wilayah Yang Memiliki Kasus Bunuh Diri Di Jawa Barat Menggunakan Algoritma Apriori," *Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, vol. 3, no. 2, pp. 629–638, Sep. 2024.
- [3] E. Rusmina, V. Shihombing, and P. A. Juledi, "Analisis Keterkaitan Antara Gejala Penyakit Menggunakan Algoritma Apriori dalam Bidang Kesehatan," *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 7, no. 3, pp. 337–340, Aug. 2020, doi: 10.33330/jurteks.v6i3.827.
- [4] Amna *et al.*, *Data Mining Data mining*, 1st ed., vol. 1, no. 1. Padang: PT GLOBAL EKSEKUTIF TEKNOLOGI, 2023.
- [5] T. S. Saptadi *et al.*, *Data Mining*, 1st ed. Batam: Yayasan Cendikia Mulia Mandiri, 2024. Accessed: Feb. 25, 2025. [Online]. Available: [https://www.academia.edu/124238987/Data\\_Mining\\_Konsep\\_Data\\_Mining\\_Oktober\\_2024](https://www.academia.edu/124238987/Data_Mining_Konsep_Data_Mining_Oktober_2024)
- [6] Lamintang, *Dasar Hukum Pidana di Indonesia*, 1st ed., vol. 1. Jakarta Timur: Sinar Grafika, 2024. Accessed: May 28, 2025. [Online]. Available: [https://www.google.co.id/books/edition/Dasar\\_Dasar\\_Hukum\\_Pidana\\_di\\_Indonesia/CRtEAAAQBAJ?hl=id&gbpv=0](https://www.google.co.id/books/edition/Dasar_Dasar_Hukum_Pidana_di_Indonesia/CRtEAAAQBAJ?hl=id&gbpv=0)
- [7] I. Budiman, S. Saori, R. Nurul Anwar, M. Yuga Pangestu, and Fitriani, "ANALISIS PENGENDALIAN MUTU DI BIDANG INDUSTRI MAKANAN (Studi Kasus: UMKM Mochi Kaswari Lampung Kota Sukabumi)," *Jurnal Inovasi Penelitian*, vol. 1, no. 10, pp. 2185–2190, Mar. 2021.
- [8] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. Ipnuwati, *IMPLEMENTASI DATA MINING (Clustering, Association, Prediction, Estimation, Classification)*, 1st ed., vol. 1. Indramayu Jawa Barat: CV. Adanu Abimata, 2023.
- [9] Pitrawati and A. Sanjaya, "REKAYASA PERANGKAT LUNAK PERHITUNGAN HARGA POKOK PRODUKSI METODE FULL COSTING PADA UMKM MITRA CAKE DI BANDAR LAMPUNG," *Jurnal informasi dan Komputer*, vol. 11, no. 2, pp. 154–162, Nov. 2021.