

Divorce Fact Detection Based on Internet User Behavior Using Hybrid Systems with Combination of Apriori Algorithm and K-Means Method

Sofika Enggari*, Sarjon Defit

Department of Information Systems, Faculty of Computer Science
Universitas Putra Indonesia YPTK
Padang, Sumatera Barat, Indonesia.

*sofika_enggari@upiypk.ac.id

Abstract-An ideal couple should sustain the family's ark till the end of their life without divorce. This study aims at seeking the association between divorce and internet behavior of searching negative keywords. The study observes four keywords, which are porn, sex, gay, and lesbian. We collected keyword usage data from google trend reports and obtained divorce court figures from the Religion Court of Padang. We used the apriori algorithm to reveal the association between divorce and internet behavior observing individual keyword searches and in groups. We used the K-Means algorithm in classifying negative word searches and divorce trial numbers from a group of existing data. We also investigate the combination of the apriori algorithm and the K-Means method to detect divorce facts and the behavior of internet users. The combined method has been successful in revealing the positive association between divorce facts and the behavior of internet users.

Keywords: apriori algorithm, K-Means method, hybrid system, internet user behavior

Article info: submitted March 29, 2021, revised July 19, 2021, accepted August 19, 2021.

1. Introduction

When a man loves a woman they have the intention to get married and create a family [1]. A family is the smallest community of society. A minimum family consists of a man and a woman. Usually, a family lives in a house together and settles down somewhere.

Every family wants to be happy and prosperous living in the world and the hereafter. One of the indicators of happiness is that the family integrity is built until the end of life or separating by the death. However, for one reason or other unexpected things, married couples are unable to maintain life harmony. Many families have to separate and break away. Sometimes, there are quarrels and fights in the house. This is indeed undesirable because it can be the forerunner of separation or divorce [2].

Divorce is something that is permitted but hated by Allah. Divorce is an indicator of the success of Satan's work and the best achievements of Satan in seducing humans. There are many temptations that Satan has thrown at humans, especially for a married couple who have formed a family. Satan is very happy when a family gets divorced. Satan will do everything in his power so that a legal couple who has a family will separate or divorce as soon

as possible. This was done by Satan because Allah SWT really hates divorce. For this reason, as a Muslim husband and wife it is highly recommended to be able to maintain the integrity of a family that has been fostered with all their might, not to divorce until the end of life separates. That is an example of an ideal Muslim family couple in Islam [3].

Many things and factors cause a family divorce. These factors directly or indirectly influence divorce. Among these factors are economic factors, social factors, personality mismatch, third person, lack of faith, behavior, and so on. These factors contribute greatly to the causes of divorce. If we can minimize or eliminate these causative factors, the number of divorce cases will certainly be low and very small. On the other hand, if these factors are getting bigger and often done, then the divorce case will increase or get bigger. Table 1 shows the facts of divorce that occurred in Padang City, West Sumatra Province.

One of the big causes of family divorce is behavior. This behavior can be done in real or virtual daily habits. One type of behavior that is done by a husband or wife who has a very big influence is behavior in surfing the internet. There are two types of behavior in the internet world, namely positive behavior, and negative behavior. Positive behavior that is carried out will have a positive

impact on the person who does it while negative behavior will hurt the person who does it. This negative behavior is like frequently searching for negative information. In the internet world, there is all the desired information, regardless of type, it can be positive information and it can be negative information [4] - [6].

Table 1. Number of Divorce Court Cases in Padang City, West Sumatra Province from August 2018 - July 2020

No	Month	Number of Divorce Trial
1	August 2018	210
2	September 2018	138
3	October 2018	144
4	November 2018	100
5	December 2018	71
6	January 2019	215
7	February 2019	142
8	March 2019	150
9	April 2019	111
10	May 2019	86
11	June 2019	111
12	July 2019	211
13	August 2019	148
14	September 2019	128
15	October 2019	137
16	November 2019	114
17	December 2019	16
18	January 2020	265
19	February 2020	162
20	March 2020	121
21	April 2020	19
22	May 2020	34
23	June 2020	207
24	July 2020	157
Sum		3197

Negative information that is often searched by the internet, based on the google trend report page, is a search with the keywords porn, sex, gay and lesbian. The four negative keywords are often searched for by world internet users, even in Indonesia. The four negative keywords above when searched on the Internet will display negative information as well. Table 2 shows the data on the number of negative keyword searches from the four words above based on the google trend report.

Apriori algorithm is one of the algorithms in data mining science. A algorithm functions to see the appearance of data in a large group of data. This algorithm can also be used to calculate the appearance of data pairs in a large data set. This algorithm is suitable for viewing frequent items and frequent itemset from data. In this study, the apriori algorithm is used to calculate the occurrence of the words porn, sex, gay, and lesbian individually, in pairs 2, in

pairs 3, and pairs 4 from all the data on the appearance of these data in the behavior of internet users [7] - [9].

The K-Means method is one of the algorithms in data mining science. The K-Means method is a method used to group data (clustering) from several existing data and then look at the relationship between several data attributes and other data attributes. The K-Means method is often used to make it easier to read the influence of a variable with other variables because large amounts of data are clustered properly. In this study, the K-Means method was used to classify the number of keyword searches as porn, sex, gay, lesbian, and the number of trial sessions [10] - [12].

Based on the figures in table 1 and table 2 above, it can be seen that the number of divorce hearings is quite high in Padang City, West Sumatra province and also the number of negative keyword searches is high. In this study, two methods were combined, namely the apriori algorithm and K-Means in calculating the effect of negative word searches such as porn, sex, gay, and lesbian on the number of divorce hearings in the city of Padang, West Sumatra province. Combining the two methods is called a hybrid system.

Table 2. Number of Searches for 4 Negative Words in Padang City, West Sumatra Province From August 2018 - July 2020

No	Month	Porn	Sex	Gay	Lesbian
1	August 2018	310	302	74	217
2	September 2018	177	190	100	189
3	October 2018	124	138	113	103
4	November 2018	128	143	197	180
5	December 2018	172	183	24	110
6	January 2019	123	126	98	81
7	February 2019	85	118	115	113
8	March 2019	129	137	114	151
9	April 2019	113	127	93	111
10	May 2019	108	122	71	135
11	June 2019	188	186	111	92
12	July 2019	117	110	68	74
13	August 2019	93	111	108	130
14	September 2019	111	147	101	112
15	October 2019	89	116	33	64
16	November 2019	86	101	32	72
17	December 2019	122	144	126	116
18	January 2020	90	94	117	54
19	February 2020	95	78	76	83
20	March 2020	102	109	223	89
21	April 2020	72	95	70	53
22	May 2020	112	133	187	159
23	June 2020	70	95	177	58
24	July 2020	63	100	27	37
Sum		2879	3205	2455	2583

The purpose of this research is to reveal how the apriori algorithm calculates the emergence of negative word search data for porn, sex, gay, and lesbians individually or in pairs. Then this study also aims to reveal how the K-Means algorithm can classify negative word searches and divorce trial numbers from a group of existing data. Furthermore, this study also aims to reveal how to combine the apriori algorithm with the K-Means method together (hybrid) in the detection of divorce facts and the behavior of internet users.

2. Method

a. Research Framework

In conducting this research, the authors followed the research framework as presented in Figure 1.

b. Framework Details

- a. Identification of problems
Identification of problems is the process of finding out what the main and important problems in the field must be solved in this research.
- b. Study of literature
Study of literature is the process of seeking or studying sciences and previous research related to solving previously identified problems.

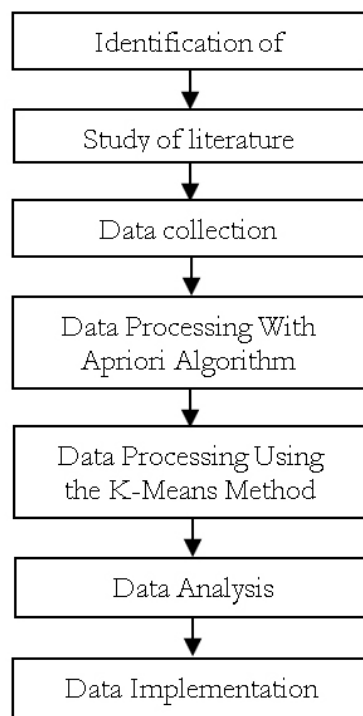


Figure 1. Research Framework

c. Data Collection

Data collection is the process of collecting field data that will be used in problem-solving. In this study, the data were sourced from the religious courts of Padang City and the google trend page. Data for divorce court figures were taken from the religious court of Padang City from August

2018 to July 2020 and for internet keyword search data, it was obtained from the google trend page for four words, namely the words PORNO (Porn), Sex, HOMO (Gay), and Lesbian from August 2018 to with the month of July 2020. The following data attributes were collected. The data collected attributes can be seen in Table 3.

Table 3. Data Collected Attributes

No	Data Attributes	Data Types
1	Month	Text
2	PORNO (porn)	Text
3	Sex	Integer
4	HOMO (gay)	Integer
5	Lesbian	Integer
6	Divorce Trial Number	Integer

d. Data Processing With Apriori Algorithm

The data that has been collected in the field is processed using the apriori algorithm. The following are the steps of the apriori algorithm, namely:

- 1) Determine the minimum value of support and confidence value using the formul 1, 2 and 3 [13].

$$\text{Support}(A) = \frac{\text{Number of Transactions Containing } (A)}{\text{Total Transactions}} \quad (1)$$

$$\text{Support}(A \cup B) = \frac{\text{Number of Transactions Containing } (A \cup B)}{\text{Total Transactions}} \quad (2)$$

$$\text{Confidence}(B/A) = \frac{\text{Number of Transactions Containing } (A \cup B)}{\text{Total Transactions } A} \quad (3)$$

Where:

A = Transaction A

B = Transaction B

- 2) Iteration 1: count items from support (transactions that contain all items) by scanning the database for 1-itemset, after 1-itemset is obtained, from 1-itemset is above minimum support if it meets minimum support, 1-itemset it will be a high-frequency pattern [14].
- 3) Iteration 2: to get 2-itemset, 3-itemset, 4-itemset a combination of the previous k-itemset must be done, then scan the database again to count items that contain the support. The itemset that meets the minimum support will be selected as the candidate's high frequent pattern [15].
- 4) Set the k-itemset value of the support that has met the minimum support for the k-itemset.
- 5) So that the item set meets the minimum support and does not meet.
- 6) Carry out the process for the next iteration until no more k-itemset meets minimum support.

e. Data Processing Using the K-Means Method

The data that has been collected in the field is processed using the K-Means method. The following are the steps for the K-Means method, namely:

- 1) Determine how many clusters will be created which are called values [16]
- 2) Determine the mean (centroid) random value (random) for each predetermined cluster.
- 3) Determine the nearest cluster center on each data record with the centroid value using the formula 4 [17]:

$$d_{\text{Euclidean}}(x, y) = \sqrt{\sum (x_i - y_i)^2} \quad (4)$$

Where: $d_{\text{Euclidean}}(x, y)$ = the distance value for each record with the centroid value, $x = x_1, x_2, x_3, \text{etc}$, $y = y_1, y_2, y_3, \text{etc}$.

- 4) Determine the closest cluster for each data record by comparing the closest distance value that has been obtained previously and then updating the cluster center value (centroid) using the formula 5 [18]:

$$\text{Cluster Center} = \frac{\sum a_i}{n} \quad (5)$$

Where: Cluster Center = The cluster center value, a_i = Value on each cluster, n = Number of clusters

- 5) Repeating steps 3 to 5 until there is no data transfer from one cluster to another [19] especially in the biomedical domain, where information overload is a major problem. In this paper, we propose a novel biomedical text summarization system that combines two popular data mining techniques: clustering and frequent itemset mining. Biomedical paper is expressed as a set of biomedical concepts using the UMLS metathesaurus. The K-means algorithm is used to cluster similar sentences. Then, the Apriori algorithm is applied to discover the frequent itemsets among the clustered sentences. Finally, the salient sentences from each cluster are selected to build the summary using the discovered frequent itemsets. For the evaluation step, we selected randomly 100 biomedical papers from the BioMed Central database full-text, and we evaluated the performances of our system by comparing the resulting summaries with the abstracts of these papers using the ROUGE metrics in term of recall, precision, and F-measure. We also compared the obtained summaries with those achieved by five well-known summarizers: TextRank, TextTeaser, SweSum, ItemSet Based Summarizer, Microsoft AutoSummarize, and two baselines: summarization using only the frequent itemsets mining (FRQ-CL).

f. Data Analysis

The data that has been processed and generated using the apriori algorithm and the K-Means method are then

analyzed and conclusions are drawn from the results of the analysis. Data processing for analysis used the Rapid Miner software version 9.7.002 [20].

g. Data Implementation

The data that has been analyzed and concluded will then be implemented in the field which will later be useful and very helpful for an advisory institution for marriage and other family matters as well as internet managers [21] [22].

3. Result

a. Results of Data Processing Using the Apriori Algorithm

The data collected in the field according to those in Table 1 and 2 are converted into value of 1 and 0, by calculating the average value of each column and comparing each record value with the average value of each column. Table 3 shows the results of converting data in Table 2. The data conversion result from table 2 can be seen in Table 4.

Table 4. Data Conversion Results from Table 2

No	Month	Porn	Sex	Gay	Lesbian
1	August 2018	1	1	0	1
2	September 2018	1	1	0	1
3	October 2018	1	1	1	0
4	November 2018	1	1	1	1
5	December 2018	1	1	0	1
6	January 2019	0	0	0	0
7	February 2019	1	0	1	1
8	March 2019	0	0	1	1
9	April 2019	0	0	0	1
10	May 2019	1	1	0	1
11	June 2019	0	0	1	0
12	July 2019	0	0	0	0
13	August 2019	0	1	1	1
14	September 2019	0	0	0	1
15	October 2019	0	0	0	0
16	November 2019	1	1	1	0
17	December 2019	0	0	1	1
18	January 2020	0	0	0	0
19	February 2020	0	0	1	0
20	March 2020	0	0	0	0
21	April 2020	0	0	0	1
22	May 2020	0	0	1	1
23	June 2020	0	0	1	0
24	July 2020	0	0	0	0

The value of support is calculated using Formula 1 and 2 with the following classification:

- Individual support, namely support A (Porn), support B (Sex), support C (Gay), and support D (Lesbian),
- Support pair 2, namely support A U B (Porn-Sex), A U C (Porn-Gay), support A U D (Porn-Lesbian), support B U C (Sex - Gay), support B U D (Sex-Lesbian), support C U D (Gay-Lesbian)
- Support in pairs 3, namely support A U B U C (Porn-Sex-Gay), support A U B U D (Porn-Sex-Lesbian), support A U C U D (Porn-Gay-Lesbian)
- Support pair 4, namely support A U B U C U D (Porn-Sex-Gay-Lesbian)

Table 5, 6, 7, and 8 show the value of support for 1, 2, 3, and 4 itemset respectively.

Table 5. Value of Support 1 Itemset

Aug 2018 – July 2020	Porn	Sex	Gay	Lesbian
1 Itemset	7	7	9	8
Number Transaction	24	24	24	24
Support Value	0.292	0.292	0.375	0.333

Table 6. Value of Support 2 Itemset

Aug 2018 – July 2020	Porn - Sex	Porn - Gay	Porn - Lesbian	Sex - Gay	Sex - Lesbian	Gay - Lesbian
2 Item Sets	7	4	5	4	5	4
Number Transaction	24	24	24	24	24	24
Support Value	0.292	0.167	0.208	0.167	0.208	0.167

Table 7. Value of Support 3 Itemset

Aug 2018 – July 2020	Porn – Sex – Gay	Porn – Sex – Lesbian	Porn – Gay – Lesbian	Sex – Gay – Lesbian
3 Item Sets	4	5	4	3
Number Transaction	24	24	24	24
Support Value	0.167	0.208	0.167	0.125

Table 8. Value of Support 4 Itemset

Aug 2018 – July 2020	Porn – Sex – Gay - Lesbian
4 Item Sets	3
Number Transaction	24
Support Value	0.125

Table 9. Itemset that meets and do not meet the Minimum Support value

No	Item Set	August 2018 – July 2020	Support	Qualify /Not Qualify
1	1 Item Sets	Porn	0.292	Qualify
2		Sex	0.292	Qualify
3		Gay	0.375	Qualify
4		Lesbian	0.333	Qualify
5	2 Item Sets	Porn – Sex	0.292	Qualify
6		Porn – Gay	0.167	Not Qualify
7		Porn – Lesbian	0.208	Not Qualify
8		Sex – Gay	0.167	Not Qualify
9		Sex – Lesbian	0.208	Not Qualify
10		Gay – Lesbian	0.167	Not Qualify
11	3 Item Sets	Porn – Sex – Gay	0.167	Not Qualify
12		Porn – Sex – Lesbian	0.208	Not Qualify
13		Porn – Gay – Lesbian	0.167	Not Qualify
14		Sex – Gay – Lesbian	0.125	Not Qualify
15	4 Item Sets	Porn – Sex – Gay – Lesbian	0.125	Not Qualify

Next, calculate the minimum support value. In this study, the minimum support value is 6 or $\frac{6}{24} = \frac{1}{4} = 0.25\%$.

By obtaining the minimum support value in Table 8, the next step is to compare the support value for each data line with the minimum support value so that it can be seen which itemset meets and which does not meet. The results of this comparison can be seen in Table 9.

Next, perform data processing using the Rapid Miner application version 9.7.002 and so that the 8 best rules are obtained as presented in Figure 2.

In addition to the 8 best rules in figure 2, the Rapid Miner application also produces an associated rule or Association Rule as in figure 3.

To produce rules as in Figures 2 and 3, the block design built on the Rapid Miner application is as shown in Figure 4.

```

W-Apriori
Apriori
=====
Minimum support: 0.1 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:

Size of set of large itemsets L(1): 4
Size of set of large itemsets L(2): 6
Size of set of large itemsets L(3): 4
Size of set of large itemsets L(4): 1

Best rules found:

1. SEX=true 7 ==> PORN=true 7   conf: (1)
2. PORN=true 7 ==> SEX=true 7   conf: (1)
3. SEX=true LESBIAN=true 5 ==> PORN=true 5
   conf: (1)
4. PORN=true LESBIAN=true 5 ==> SEX=true 5
   conf: (1)
5. SEX=true GAY=true 4 ==> PORN=true 4   conf: (1)
6. PORN=true GAY=true 4 ==> SEX=true 4   conf: (1)
7. SEX=true GAY=true LESBIAN=true 3 ==> PORN=true
   3   conf: (1)
8. PORN=true GAY=true LESBIAN=true 3 ==> SEX=true
   3   conf: (1)

```

Figure 2. 8 Best Rules Generated

```

AssociationRules
Association Rules
[LESBIAN] --> [GAY] (confidence: 0.500)
[SEX] --> [GAY] (confidence: 0.571)
[PORN] --> [GAY] (confidence: 0.571)
[SEX] --> [GAY, PORN] (confidence: 0.571)
[PORN] --> [GAY, SEX] (confidence: 0.571)
[SEX, PORN] --> [GAY] (confidence: 0.571)
[LESBIAN, SEX] --> [GAY] (confidence: 0.600)
[LESBIAN, PORN] --> [GAY] (confidence: 0.600)
[LESBIAN, SEX] --> [GAY, PORN] (confidence:
0.600)
[LESBIAN, PORN] --> [GAY, SEX] (confidence:
0.600)
[LESBIAN, SEX, PORN] --> [GAY] (confidence:
0.600)
[LESBIAN] --> [SEX] (confidence: 0.625)
[LESBIAN] --> [PORN] (confidence: 0.625)
[LESBIAN] --> [SEX, PORN] (confidence: 0.625)
[SEX] --> [LESBIAN] (confidence: 0.714)
[PORN] --> [LESBIAN] (confidence: 0.714)
[SEX] --> [LESBIAN, PORN] (confidence: 0.714)
[PORN] --> [LESBIAN, SEX] (confidence: 0.714)
[SEX, PORN] --> [LESBIAN] (confidence: 0.714)
[GAY, LESBIAN] --> [SEX] (confidence: 0.750)
[GAY, SEX] --> [LESBIAN] (confidence: 0.750)
[GAY, LESBIAN] --> [PORN] (confidence: 0.750)
[GAY, PORN] --> [LESBIAN] (confidence: 0.750)
[GAY, LESBIAN] --> [SEX, PORN] (confidence:
0.750)
[GAY, SEX] --> [LESBIAN, PORN] (confidence:
0.750)
[GAY, PORN] --> [LESBIAN, SEX] (confidence:
0.750)
[GAY, SEX, PORN] --> [LESBIAN] (confidence:
0.750)
[SEX] --> [PORN] (confidence: 1.000)
[PORN] --> [SEX] (confidence: 1.000)
[GAY, SEX] --> [PORN] (confidence: 1.000)
[GAY, PORN] --> [SEX] (confidence: 1.000)
[LESBIAN, SEX] --> [PORN] (confidence: 1.000)
[LESBIAN, PORN] --> [SEX] (confidence: 1.000)
[GAY, LESBIAN, SEX] --> [PORN] (confidence:
1.000)
[GAY, LESBIAN, PORN] --> [SEX] (confidence: 1.000)

```

Figure 3. Association Rules

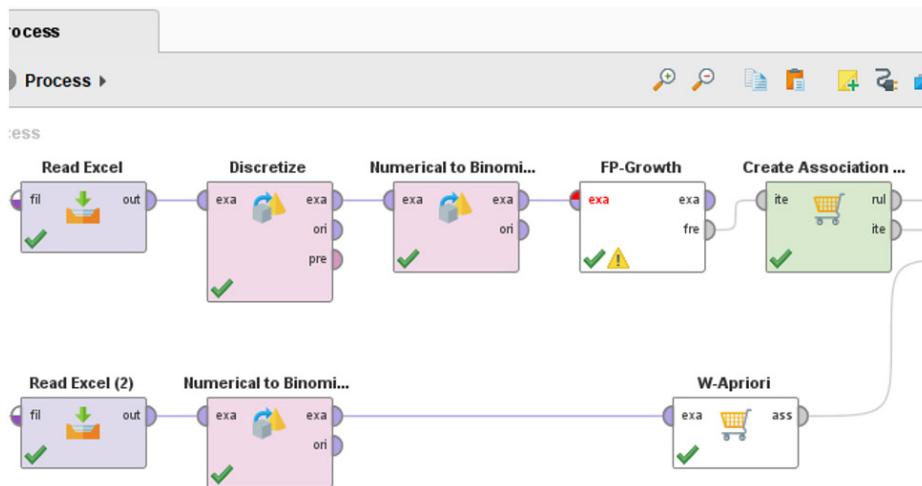


Figure 4. Apriori Algorithm Rapid Miner Block Design

b. Results of Data Processing Using the K-Means Method

The data collected in tables 1 and 2 are then grouped (clustered) using the K-Means method.. In this study, 3 clusters were determined for each data attribute. After determining the number of clusters, then determine the middle value (centroid) of each cluster for each attribute. The final scores using the Rapid Miner application can be seen in table 10.

After obtaining the final middle value as in table 10, the results of grouping each row (a record) of data for each attribute are obtained as in table 11.

The results of grouping data as in table 11 are converted into quality values to analyze and calssify them into low, medium, and high values. Table of 12 shows results from the change in numerical values to quality values.

Table 10. Middle Value (centroid)

Atribute	Cluster 0	Cluster 1	Cluster 2
Porn	17.530	27.333	71.0
Sex	73.250	23.503	34.29
Gay	28.214	43.444	7.0
Lesbian	41.429	78.0	28.375
Divorce Trial Numbers	50.833	224.5	133.071

Table 11. Results of Data Grouping

No	Month	Porn	Sex	Gay	Lesbian	Divorce
1	August 2018	cluster_2	cluster_0	cluster_1	cluster_1	cluster_1
2	September 2018	cluster_1	cluster_2	cluster_1	cluster_0	cluster_2
3	October 2018	cluster_1	cluster_2	cluster_0	cluster_0	cluster_2
4	November 2018	cluster_1	cluster_2	cluster_1	cluster_0	cluster_2
5	December 2018	cluster_1	cluster_2	cluster_0	cluster_0	cluster_0
6	January 2019	cluster_1	cluster_1	cluster_1	cluster_2	cluster_1
7	February 2019	cluster_0	cluster_1	cluster_1	cluster_2	cluster_2
8	March 2019	cluster_0	cluster_1	cluster_1	cluster_0	cluster_2
9	April 2019	cluster_0	cluster_1	cluster_0	cluster_0	cluster_2
10	May 2019	cluster_0	cluster_1	cluster_0	cluster_2	cluster_0
11	June 2019	cluster_1	cluster_2	cluster_0	cluster_2	cluster_2
12	July 2019	cluster_0	cluster_1	cluster_0	cluster_2	cluster_1
13	August 2018	cluster_0	cluster_1	cluster_0	cluster_2	cluster_2
14	September 2018	cluster_0	cluster_1	cluster_1	cluster_2	cluster_2
15	October 2018	cluster_0	cluster_1	cluster_0	cluster_2	cluster_2
16	November 2018	cluster_0	cluster_1	cluster_1	cluster_2	cluster_2
17	December 2018	cluster_0	cluster_1	cluster_0	cluster_2	cluster_0
18	January 2019	cluster_0	cluster_1	cluster_1	cluster_2	cluster_1
19	February 2019	cluster_0	cluster_1	cluster_0	cluster_2	cluster_2
20	March 2019	cluster_0	cluster_1	cluster_0	cluster_2	cluster_2

No	Month	Porn	Sex	Gay	Lesbian	Divorce
21	April 2019	cluster_0	cluster_1	cluster_0	cluster_2	cluster_0
22	May 2019	cluster_0	cluster_1	cluster_0	cluster_0	cluster_0
23	June 2019	cluster_0	cluster_1	cluster_0	cluster_2	cluster_1
24	July 2019	cluster_0	cluster_1	cluster_2	cluster_2	cluster_2

Table 12. Results of Changes in Data Value to Numbers

No	Month	Porn	Sex	Gay	Lesbian	Divorce
1	August 2018	Hight	Hight	Hight	Hight	Hight
2	September 2018	Moderate	Moderate	Hight	Moderate	Moderate
3	October 2018	Moderate	Moderate	Moderate	Moderate	Moderate
4	November 2018	Moderate	Moderate	Hight	Moderate	Moderate
5	December 2018	Moderate	Moderate	Moderate	Moderate	Low
6	January 2019	Moderate	Low	Hight	Low	Hight
7	February 2019	Low	Low	Hight	Low	Moderate
8	March 2019	Low	Low	Hight	Moderate	Moderate
9	April 2019	Low	Low	Moderate	Moderate	Moderate
10	May 2019	Low	Low	Moderate	Low	Low
11	June 2019	Moderate	Moderate	Moderate	Low	Moderate
12	July 2019	Low	Low	Moderate	Low	Hight
13	August 2018	Low	Low	Moderate	Low	Moderate
14	September 2018	Low	Low	Hight	Low	Moderate
15	October 2018	Low	Low	Moderate	Low	Moderate
16	November 2018	Low	Low	Hight	Low	Moderate
17	December 2018	Low	Low	Moderate	Low	Low
18	January 2019	Low	Low	Hight	Low	Hight
19	February 2019	Low	Low	Moderate	Low	Moderate
20	March 2019	Low	Low	Moderate	Low	Moderate
21	April 2019	Low	Low	Moderate	Low	Low
22	May 2019	Low	Low	Moderate	Moderate	Low
23	June 2019	Low	Low	Moderate	Low	Hight
24	July 2019	Low	Low	Low	Low	Moderate

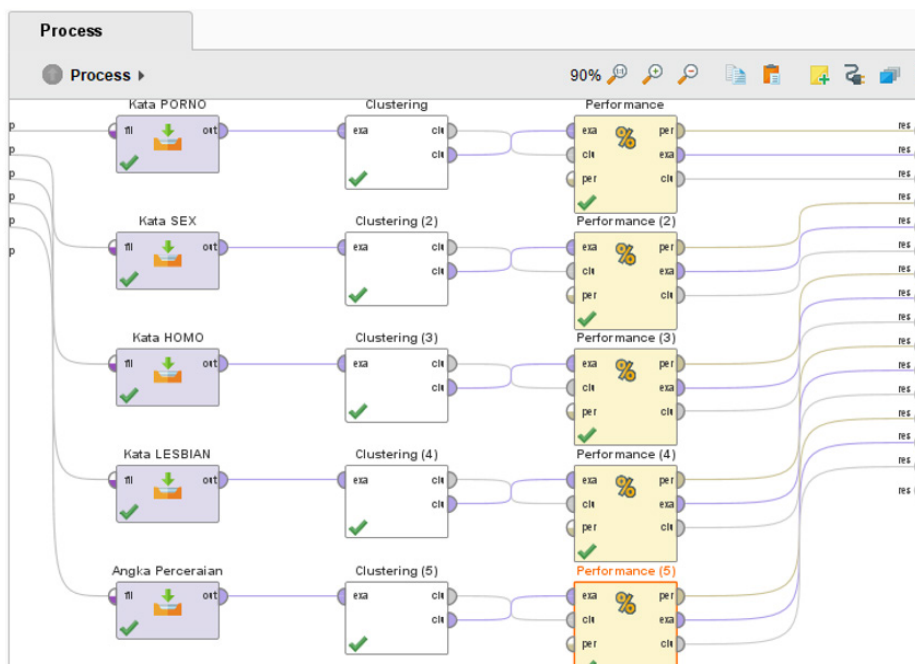


Figure 5. Rapid Miner Block Design of K Means Method

To produce grouping data as in tables 11 and 12 using the Rapid Miner application, a block design is needed, as presented in Figure 5.

4. Discussion

a. Discussion of Apriori Algorithm Results

Based on the results of data testing using the Apriori Algorithm method, it can be analyzed that the 8 best rules are found among the 35 existing rules, namely:

1. When someone searches for the word Sex, he/she will simultaneously search for the word Porn with a confidence value of 1
2. When someone searches for the word porn, he/she will simultaneously search for the word sex with a confidence value of 1
3. When someone searches for the word Sex, he/she will simultaneously search for the word Lesbian and Porn with a confidence value of 1
4. When someone searches for the word Porn, he/she will simultaneously search for the word Lesbian and Sex with a confidence value of 1
5. When someone searches for the word Sex, he/she will simultaneously search for the word Gay and Porn with a confidence value of 1
6. When someone searches for the word Porn, he/she will simultaneously search for the word Gay and Sex with a confidence value of 1
7. When someone searches for the word Sex, he/she will simultaneously search for the word Gay, Lesbian, and Porn with a confidence value of 1
8. When someone searches for the word Porn, he/she will simultaneously search for the word Gay, Lesbian, and Sex with a confidence value of 1

b. Discussion of the Results of the K-Means Method

Based on the results of data testing with the K-Means method, the word search clustering can be analyzed against the divorce trial figures as follows:

1. The number of divorce hearings is in the High cluster if the search number for the word GAY is also in the High cluster this happened in August 2018, January 2019, and January 2020.
2. The number of divorce hearings is in the Medium cluster if the number of word searches for PORN is in the Medium cluster, the SEX word search number is in the Medium cluster, the GAY word search number is in the Medium cluster, and the LESBIAN word search number is in the Medium cluster. This happened in September 2018, October 2018, and November 2018.
3. The number of divorce hearings is in the Low cluster if the number of word searches for PORN is in the Medium / Low cluster, the number of SEX word searches is in the Medium cluster, the number of word searches for GAY is in the Medium / Low cluster, and the number of word searches for LESBIAN is in the Medium / Low cluster. This

happened in December 2018, May 2019, December 2019, April 2020, and May 2020.

5. Conclusion

This study has successfully combined the apriori algorithm with the K-Means method in detecting divorce facts and the behavior of internet users. The results show that from the 35 rules generated by the apriori algorithm method, the best eight rules show up, meaning that the eight rules have the highest value to describe the true association. The number of divorce hearings is in the cluster High if the search number for the word GAY (gay) is also in the cluster High. The number of divorce hearings is in the cluster Medium if the number of searches for PORN (porn), SEX, GAY, and LESBIAN is all in the cluster Medium. The number of divorce hearings is in the cluster Low if the number of searches for PORN, GAY, and LESBIAN is in the cluster Medium / Low, and the number of searches for SEX is in the cluster Medium.

Reference

- [1] A. Wiratri, "Rethinking the meaning of family in Indonesian society (Revisiting the concept of family in Indonesian society)," *J. Kependud. Indones.*, vol. 13, no. 1, pp. 15–26, 2018, [Online]. Available: <http://ejurnal.kependudukan.lipi.go.id/index.php/jki/article/view/305/pdf>.
- [2] A. Khumas, J. E. Prawitasari, and S. Retnowati, "Model Explaining the Intention of Divorce for Muslim Women in South Sulawesi," *J. Psikol.*, vol. 42, no. 3, p. 189, 2015, doi: 10.22146/jpsi.9908.
- [3] N. Najwah, "Criteria for Choosing a Spouse (Hadith Hermeneutics Study)," *J. Stud. Ilmu-ilmu Al-Qur'an dan Hadis*, vol. 17, no. 1, p. 95, 2018, doi: 10.14421/qh.2016.1701-05.
- [4] N. S. Mudawamah, "Internet User Behavior: A Case Study of the Student of the Department of Library and Information Science, Uin Maulana Malik Ibrahim," *J. Kaji. Perpust. dan Inf.*, vol. Vol. 4 No., pp. 107–113, 2020.
- [5] A. Susanto, "Factors Affecting the Behaviour of Internet Use of Villager Pasar Vi, Kualanamu, Deli Serdang, North Sumatra," *J. Penelit. Pos Inform.*, vol. 5, no. 1, pp. 65–86, 2015, doi: 10.17933/jppi.2015.050100.
- [6] P. M. Yusup, N. Komariah, N. Prahmatmaja, and S. CMS, "Utilization of the Internet for Livelihoods among Rural Youth," *Baca J. Dokumentasi Dan Inf.*, vol. 40, no. 2, p. 217, 2019, doi: 10.14203/j.baca.v40i2.491.
- [7] K. Henrique de Jesus Prado, L. Santos Souza, I. Dias de Jesus Junior, and M. Colaço Júnior, "Applied Intelligent Data Analysis to Government Data Related to Criminal

- Incident: A Systematic Review,” *J. Appl. Secur. Res.*, vol. 15, no. 3, pp. 297–331, 2020, doi: 10.1080/19361610.2020.1716511.
- [8] A. -, F. Marisa, and D. Purnomo, “Application of the Apriori Algorithm to Sales Data in BM Warehouse Stores,” *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 1, no. 1, pp. 1–5, 2016, doi: 10.31328/jointecs.v1i1.408.
- [9] V. N. Budiyasari, P. Studi, T. Informatika, F. Teknik, U. Nusantara, and P. Kediri, “Implementation of Data Mining in Eyeglass Sales Using the Apriori Algorithm,” *Indones. J. Comput. Inf. Technol.*, vol. 2, no. 2, pp. 31–39, 2017.
- [10] A. Bastian, H. Sujadi, and G. Febrianto, “Application of the K-Means Clustering Analysis Algorithm in Human Infectious Diseases (Case Study of Majalengka Regency),” no. 1, pp. 26–32.
- [11] J. Jaroji, D. Danuri, and F. P. Putra, “K-Means To Determine Prospective Bidik Misi Scholarship Recipients in Polbeng,” *INOVTEK Polbeng - Seri Inform.*, vol. 1, no. 1, p. 87, 2016, doi: 10.35314/isi.v1i1.129.
- [12] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, “Application of the K-Means Algorithm for Drug Data Clustering,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [13] I. P. Astuti, “AApriori algorithm to find the relationship between school majors and student graduation rates,” *J. Tek. Inform.*, vol. 12, no. 1, pp. 69–78, 2019, doi: 10.15408/jti.v12i1.10525.
- [14] A. Sanjani; Fahmi, Hasanul;Sindar, “Implementation of Data Mining Sales of Clothing Products Using the Apriori Algorithm,” *Indones. J. Appl. Informatics*, vol. 4, no. 4, 2019.
- [15] D. Listriani, A. H. Setyaningrum, and F. Eka, “Application of the Association Method Using the Apriori Algorithm in the Consumer Shopping Pattern Analysis Application (Case Study of the Gramedia Bintaro Bookstore),” *J. Tek. Inform.*, vol. 9, no. 2, pp. 120–127, 2018, doi: 10.15408/jti.v9i2.5602.
- [16] F. Nur, M. Zarlis, and B. B. Nasution, “Application of the K-Means Algorithm to New Vocational High School Students for Department Clustering,” *InfoTekJar (Jurnal Nas. Inform. dan Teknol. Jaringan)*, vol. 1, no. 2, pp. 100–105, 2017, doi: 10.30743/infotekjar.v1i2.70.
- [17] W. Dhuhita, “Clustering Using the K-Mean Method to Determine the Nutritional Status of Toddlers,” *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160–174, 2015.
- [18] Y. Ma, X. Zhang, X. Di, T. Ren, H. Yang, and B. Cai, “Analysis and Identification of Students with Financial Difficulties: A Behavioural Feature Perspective,” *Discret. Dyn. Nat. Soc.*, vol. 2020, 2020, doi: 10.1155/2020/2071025.
- [19] O. Rouane, H. Belhadeif, and M. Bouakkaz, “Combine clustering and frequent itemsets mining to enhance biomedical text summarization,” *Expert Syst. Appl.*, vol. 135, pp. 362–373, 2019, doi: 10.1016/j.eswa.2019.06.002.
- [20] D. C. G. Putri, J. S. Leu, and P. Seda, “Design of an unsupervised machine learning-based movie recommender system,” *Symmetry (Basel)*, vol. 12, no. 2, pp. 1–27, 2020, doi: 10.3390/sym12020185.
- [21] Hendri, H., Enggari, S., Mardison, Putra, M. R., & Rani, L. N. (2019). Automatic System to Fish Feeder and Water Turbidity Detector Using Arduino Mega. *Journal of Physics: Conference Series*, 1339(1).
- [22] Hendri, H., Awal, H., & Mardison. (2021). Solar-Cell Implementation for Supporting Tourist Facilities and Tourism Promotion Media. *Journal of Physics: Conference Series*, 1783(1), 012058.