ISSN 2614-0047

Application of rule association with algorithm apriori of disaster in Indonesia residential fires

Muhammad Muhajir [1,*], Gusmayyeni [2], Redita Anggita Sari [3], Tusriana Rahmatika [4]

Department of Statistics, Faculty of Mathematics and Natural Science, Islamic University of Indonesia
[1] muhammad.muhajir.stat89@gmail.com *; [2] gusmayyeni29@gmail.com; [3] reditaanggita118@gmail.com; [4] tusrianarahmatika@gmail.com
* corresponding author

ARTICLE INFO

Article history
Received September 21, 2017
Revised October 17, 2017
Accepted November 4, 2017

Keywords
Data mining
Association rules
Apriori
Fire settlements

ABSTRACT

Data mining is a technique of decision making by means of extracting information based on historical data of existing data in a large database. One of technique in data mining, is association rule algorithm where the method is searching for a set of items that frequently appear together. This study will use data association rule mining method for data processing Fire Disaster settlements in Indonesia because we want to know what information is often occur together in the event of fire disaster settlement. From the analysis associative relationship, event's pattern that occur from residential fires in Indonesia which the data is from the beginning of January 2015 to June 2015, support the highest value that the event of catastrophic fires in settlements in the afternoon resulted in broken homes with a value of 0.8148148 support and confident value of 1.0292398.

1. Introduction

Flame of fire both small and large in place, situation and time undesired detrimental and generally difficult to control. Fire is also included in one of the categories conditions or emergency situations within the house, company, etc [1]. Fire disaster is a disaster that is mostly caused by human negligence. Although in some cases can also occur as a spark of fire that comes from lightning strikes [2].

Fire incident intensification of settlement lately also triggered by the seasons. While the number of objects that are not saved, due to many things. Fire engines such as delays in the place of the crime scene, the limited fire engine away water sources and the difficulty of access to the scene. Various constraints and limitations of the above has been the record from year to year, but almost minimal with progress in the form of a concrete solution. Some other things that also be important for the handling of the fire accident is recognition, understanding and mitigation capabilities are still low, even nil in the community.

Judging from fires in Indonesia event data, event data residential fire disasters in the last five years, there were 979 fire cases recorded from August 24 2011 until June 30, 2015. In 2011 there were 11 cases of fire. Presentation of cases of fires in 2012 began to rise to 53 cases of fire, for the year 2013 has increased very rapidly that there were 398 cases of fire. For the year 2014 increased again to 471 fire cases. Whereas in 2015, from the beginning of January until June 30, there were 43 cases of fire National Disaster Management Agency (BNPB). This indicates that fires potency continued to increase from year to year. This is certainly not a proud achievement for Indonesia. In the case of fire, the cause of the fire source such as an electric short circuit. Gas cylinder explosion, discarded cigarettes or sparks, etc.

One of the efforts to prevent or reduce the presence of residential fires is that by studying the pattern of linkages between events with the characteristics of the occurrence of events such as the time of the incident, location of the incident, victims, and losses due to the incident. So the factors that may affect the fire can be seen. One method that can be used is by using the Association rule Apriori algorithm.

Association rule (rule associative) is one of the main techniques in data mining and is the form most commonly used in finding a pattern or patterns from a data set [3]. Association rule (rule associative) trying to discover certain rules that associate the data with other data. To search for association rule from a data set, first of all we should seek first the so-called "frequent itemset" (a collection of items that often appear together) [4].

2. Material and Methodology

In this study used data settlement fires in Indonesia in early January to June 2015. This data was obtained from the National Disaster Management Agency (BNPB). In this case there are 43 residential fire event data to be analyzed. Data following the landslide in Indonesia:

Table 1. Genesis residential fires in Indonesia 1 January-June 2015

No	Disaster	Date	Time	Longitude	Latitude	Location	Victim	Damage
1	Fire Settlement	30/06/2015	12.00 AM	106.796	-6.14545	West Jakarta	63 family 290 people evacuated	35 material losses
2	Fire Settlement	06/05/2015	03.55 AM	106.907	-6.32262	East Jakarta	none	1 unit building collapsed
3	Fire Settlement	05/05/2015	13.55 PM	106.816	-6.18562	Central Jakarta	none	1 unit of houses were damaged
- - -								
41	Fire Settlement	02/01/2015	10.40 AM	106.814	-6.1506	West Jakarta	None	1 unit of houses were damaged
42	Fire Settlement	01/01/2015	00.52 AM	106.929	-6.11166	North Jakarta	None	1 unit of houses were damaged
43	Fire Settlement	01/01/2015	01.05 AM	106.828	-6.14186	South Jakarta	None	1 unit of houses were damaged

Variables used in this paper can be done for coding is as follows;

1. Time: Variable time with regard to the settlement of immortality catastrophic events. The division of time variable is divided into three, namely;
   a. Morning: 00:00 to 11:59 o'clock pm
   b. Day: 12:00 to 15:59 o'clock pm
   c. Afternoon: at 4:00 p.m. to 17:59 pm
   d. Night: at 6:00 p.m. to 23:59 pm
2. Location: The variable is the location of the scene of the fire disaster may facilitate then action of the analysis.

3. Victim: Victim variables are objects that are vulnerable to catastrophic fires. Victims carried classification into three (nil / no casualties, injuries either mild, moderate, or severe, and death)
4. Damage: variabe loss is the impact of disaster events fire. Losses can be divided into; a. Building: houses, shops / stalls, stalls, shops, offices, companies.
   a. Vehicle: Motorcycle and Car
   b. In the investigation (still follow-up)
   c. Land
5. Description: a chronology of events which occurred because the fire could be short circuit currents listring, in the investigation, the explosion (gas stove), and fire (derived from cigarette butts, combustion, wax).

3. Association Rule

Association rule mining is one of data mining techniques to discover the rules of the association between a combination item with other items that often appear together with a certain amount or frequency of a set of data. I = {i1, i2, ..., im} is a set of items. For example, items such as butter, bread, milk is purchased at a supermarket; and Ai = v is an item, where v is the value of the attribute domain Ai, in relation R (A1, ..., An). A collection of items referring to the term itemset, X called itemset if a subset of I.

D = {ti, ti + 1, ..., tn} is the set of transactions, called database transactions, where each transaction has a tid t and t-itemset t = (tid, t-itemset). For example, shopping cart checkout consumer who has examined referred to as one transaction; tuple (v1, .., vn) of relation R (A1, ..., An) as a transaction. Transaction t containing itemset X for all items, which i∈ X, and i is t-itemset [5].

Analysis of the association is also known as one of data mining techniques that became the basis of a variety of other data mining techniques. In particular one of the stages of the analysis of the association called high frequency pattern analysis attracted the attention of many researchers to produce an efficient algorithm.

Association rule is an implication of the form $X \rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. X (or Y) can consist of a single item or the whole collection of items. X can be regarded as antecedent and Y as the consequent. Each association rule has two main measure of quality, namely the support and confidence.

Support the rule X → Y is support (X∩Y) = P (X∩Y) is defined as [6]:

$$\text{Supp}(X \rightarrow Y) = \text{support}(X \cap Y) = P(X \cap Y) = \frac{n(X \cap Y)}{n(D)}$$

with:

P (X∩Y) = probability of occurrence of X and Y simultaneously

n (X∩Y) = the number of occurrences of X and Y simultaneously on transactions

n (D) = total number of transactions in the database D

Confidence rule $X \rightarrow Y$ is defined as [7] :

$$\text{confidence}(X \rightarrow Y) = P(Y|X) = \frac{P(X \cap Y)}{P(X)}$$

With:

P (Y X) =c onditional probability Y occur given X has occurred

P (X) = probability of occurrence itemset X

Besides the two measures, one of the better ways to determine the strength of an association rule is to look at the value of the lift. lift is defined as [6]:

$$\text{Lift}(X \rightarrow Y) = \text{Interes}(X \rightarrow Y) = \frac{P(X \cap Y)}{P(X)\,P(Y)}$$

The association rule mining can be viewed as a two step process [8], namely:

1. Find all frequent itemsets: Each of these itemsets will occur at least as frequently as a predetermined minimum support count.
2. Generate strong association rules from the frequent itemsets: The rules must satisfy minimum support and confidence. These rules are called strong rules.

3.1. Apriori Algorithm

Apriori is a seminal algorithm proposed by R. Agarwal and R. Srikant in 1994 [4] for mining frequent itemsets for Boolean association rules. The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent itemset properties. The following lines state the steps in generating frequent itemset in Apriori algorithm [9]. The stages of apriori algorithm is described in Fig. 1.

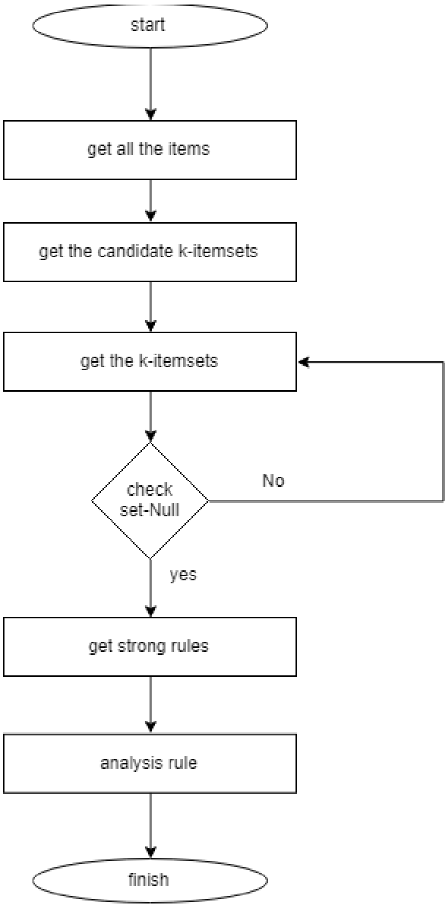

Fig. 1. Flowchart of apriori algorithm

The first step in association rule mining is to find frequent itemsets. K-itemset is defined as an itemset as k items, Lk as the set of frequent k-itemsets, and Ck as the set of candidate k-itemset. Pseude-code following a priori algorithm used to generate all frequent itemsets and pruning frequent itemset unattractive in a transaction database [10].

1. k = 1.
2. Lk = $\{ i|i \in I \wedge \sigma(\{i\}) \geq N \times \text{minnsup} \}$. {Get all the frequent 1-itemset }
3. Repeat
4. k = k + 1
5. Ck = apriori-gen (Lk-1). {Produced candidates itemsets }

6. for each transaction t∈Ddo
7. Ct = subset(Ckt). {Determine all the candidate itemsetsCk on each t }
8. for each candidate itemset c
9. p(c) = p (c)+1 {add in support count}
10. end for
11. end for
12. Lk = {c|c ∈ Ck ∧σ(c) ≥ N × minnsup}.{ Remove the candidates is less thanminsup}
13. Until Lk = Ø
14. Result = U Lk

Immediately generated a strong association rules (strong) from the frequent itemsets (which strong association rules meet the minimum support and minimum confidence). This can be done using the equation of confidence.

4. Results

Residential fire disaster events of 1 January to 30 June 2015 there were 43 cases of fire. Of the 43 cases of fires will be analyzed using algorithms arules priori to obtain a pattern of rules. In this analysis using the software R which determines the value of minimum support, minimum confident, and the main line. The value to be shown is the value that is equal to or greater than a predetermined minimum support. First, the data summary will be known residential fire disaster. Summary data is a summary regarding some information about residential fire data to be analyzed. Fig 2 shows an output of fire disaster data summary settlement.

```
> summary(data)
      event              attribute
 event15:   5    building      :41
 event17:   5    shorted       :31
 event31:   5    morning       :21
 event1 :   4    night         :12
 event10:   4    east jakarta: 7
 event11:   4    day           : 6
 (Other):144     (Other)       :53
```

Fig. 2. Output fire disaster data summary settlement

From the above data shows that the output of residential fires 15,17,31 to have the attributes of 5 pieces, 1,10,11 events have attributes incidence of 4 pieces, and other residential fires have attributes with a total of less than 4 attributes as much as 144 pieces. Other information that can be known is the attribute of the building appear as much as 41 times, shorted appeared 31 times, morning appeared 21 times, night mumcul 12 times, day appears 6 times, and east jakarta appears 7 times of the total incidence of landslides observed as many as 43 events.

Next will be the grouping attributes in an accident with a split order data. Software R will categorize the attributes into the same incident. Fig. 3 shows the results of data output in the event of fire split the first settlemen.

```
> splitdata
$event1
[1] morning      west jakarta building     shorted
32 Levels: affected afternoon building cilacap cisolok day ... west jakarta
```

Fig. 3. Output split residential fire disaster data

The output of note that the attributes that go event1 residential fires that morning, west jakarta, building, shorted. To display the attributes of accompanying events and events that are not accompanied by the initiation of a value of zero (0) and one (1) will be determined by the rules of the matrix. The output produced is as follows.

Vol. 1, No. 2, December 2017, pp. 52-59

```
> as(aturan,"matrix")
        affected afternoon building central jakarta cilacap cisolok   day
event1     FALSE     FALSE     TRUE            FALSE   FALSE   FALSE FALSE
event10    FALSE     FALSE     TRUE            FALSE   FALSE   FALSE FALSE
event11    FALSE     FALSE     TRUE            FALSE   FALSE   FALSE FALSE

        night north jakarta rembang shorted sidoarjo situbondo solok
event1  FALSE         FALSE   FALSE    TRUE    FALSE     FALSE FALSE
event10 FALSE         FALSE   FALSE   FALSE    FALSE     FALSE FALSE
event11 FALSE         FALSE   FALSE    TRUE    FALSE     FALSE FALSE
```

Fig. 4. Output rule matrix residential fire disasters

Fig. 4 shows the attributes that accompany the event. FALSE logic shows there is no attribute in the incidence and value of TRUE indicates there are attributes in the incident. Attributes are sorted alphabetically from a to z. Event1 describes buildings there are attributes and attribute building, and shorted with logic TRUE. Whereas, to attribute affected, afternoon, central jakarta, cilacap, cisolok, day, night, north jakarta, rembang, shorted, sidoarjo, situbondo, and solok with a logic FALSE.

Tests were first carried out is by determining the minimum support = 0.1; Minimum confident = 0.1; and and a main line = 2 with the following output.

```
> inspect(aturan.ap)
   lhs                           rhs             support   confidence lift
1  {north jakarta}            => {shorted}       0.1162791 1.0000000  1.3870968
2  {shorted}                  => {north jakarta} 0.1162791 0.1612903  1.3870968
3  {north jakarta}            => {building}      0.1162791 1.0000000  1.0487805
4  {building}                 => {north jakarta} 0.1162791 0.1219512  1.0487805
5  {south jakarta}            => {building}      0.1162791 1.0000000  1.0487805
6  {building}                 => {south jakarta} 0.1162791 0.1219512  1.0487805
7  {investigation}            => {building}      0.1162791 1.0000000  1.0487805
8  {building}                 => {investigation} 0.1162791 0.1219512  1.0487805
9  {day}                      => {building}      0.1395349 1.0000000  1.0487805
10 {building}                 => {day}           0.1395349 0.1463415  1.0487805
11 {east jakarta}             => {shorted}       0.1627907 1.0000000  1.3870968
12 {shorted}                  => {east jakarta}  0.1627907 0.2258065  1.3870968
13 {east jakarta}             => {building}      0.1627907 1.0000000  1.0487805
14 {building}                 => {east jakarta}  0.1627907 0.1707317  1.0487805
15 {night}                    => {shorted}       0.2325581 0.8333333  1.1559140
16 {shorted}                  => {night}         0.2325581 0.3225806  1.1559140
17 {night}                    => {building}      0.2325581 0.8333333  0.8739837
18 {building}                 => {night}         0.2325581 0.2439024  0.8739837
19 {morning}                  => {shorted}       0.3255814 0.6666667  0.9247312
20 {shorted}                  => {morning}       0.3255814 0.4516129  0.9247312
21 {morning}                  => {building}      0.4883721 1.0000000  1.0487805
22 {building}                 => {morning}       0.4883721 0.5121951  1.0487805
23 {shorted}                  => {building}      0.6976744 0.9677419  1.0149489
24 {building}                 => {shorted}       0.6976744 0.7317073  1.0149489
25 {north jakarta,shorted}    => {building}      0.1162791 1.0000000  1.0487805
26 {building,north jakarta}   => {shorted}       0.1162791 1.0000000  1.3870968
27 {building,shorted}         => {north jakarta} 0.1162791 0.1666667  1.4333333
28 {east jakarta,shorted}     => {building}      0.1627907 1.0000000  1.0487805
29 {building,east jakarta}    => {shorted}       0.1627907 1.0000000  1.3870968
30 {building,shorted}         => {east jakarta}  0.1627907 0.2333333  1.4333333
31 {night,shorted}            => {building}      0.2093023 0.9000000  0.9439024
32 {building,night}           => {shorted}       0.2093023 0.9000000  1.2483871
33 {building,shorted}         => {night}         0.2093023 0.3000000  1.0750000
34 {morning,shorted}          => {building}      0.3255814 1.0000000  1.0487805
35 {building,morning}         => {shorted}       0.3255814 0.6666667  0.9247312
36 {building,shorted}         => {morning}       0.3255814 0.4666667  0.9555556
```

Fig. 5. Continued output combinations of two itemsets

From the output of Fig. 5, for a pattern of a combination of two itemsets which shows the relationship of two attributes, eg in case A then there occurs B. Establishment of two itemsets frequency pattern, formed from the type of items-items that meet the minimum support that is by combining all of the items into two patterns combination. In the syntax has been determined that the minimum support = 0.1, then the displayed minimum support value equal to or more than 0.1, the value of which is less than 0.1 are automatically eliminated by the system. The following description of the calculation of point 1, and 2 that can result as above.

$$\text{Support (south jakarta} \cap \text{building)} = \frac{contains\ a\ number\ of\ events\ south\ jakarta\ and\ building}{total\ events}$$

$$= \frac{5}{43} = 0.1162791$$

Support (building ∩ south jakarta) $= \frac{contains\ a\ number\ of\ events\ south\ jakarta\ and\ building}{total\ events}$

$= \frac{5}{43} = 0.1162791$

Confidence P(south jakarta|building) $= \frac{total\ events\ building\ and\ south\ jakarta}{total\ events\ south\ jakarta}$

$= \frac{5}{5} = 1$

Confidence P(building|south jakarta) $= \frac{total\ events\ building\ and\ south\ jakarta}{total\ events\ building}$

$= \frac{5}{41} = 0{,}1219512$

From the above calculation in mind that although support equal value but the value is not always the same confident. This occurs because the value of the support shows the probability appears attributes A and B together from all events while the value of certain or confident showed strong relationships between items within a priori.

From output Fig. 5, the importance of the final result associative relationship patterns with the highest support value if there is a fire settlement with the occurrence of a shorted, in the morning, it will cause damage to building with a value of 0,3255814 support and confident value of 1. If there is a settlement due to fire condition of the building, in the morning, then cause a shorted with the value of 0,3255814 support and confident value for 0,667. The next case of residential fires due to the condition of the building, with the caused by shorted then a fire broke out there early morning to support the value of 0,3255814 and confident value of 0,4827. The following table Association Final Rule.

Table 2. Table Final Association Rule

No	Rules	Supp	Conf	Support X Confidence
1	If fires settlement "shorted" and "morning", then "building"	32%	100%	32%
2	If fires settlements "building", and "morning", then " shorted "	32%	67%	21%
3	If fires settlements "building", and " shorted", then in the "morning"	32%	48%	15%

5. Conclusion

Based discussion can be concluded, using the association rule with a priori algorithm that can be applied to residential fire disasters in some rules such as; iresidential fires often occur with the support value at 0,32 and confident value of 1 due to a short circuit or a short circuit that occurred on the morning of the types of losses obtained in the form of the building. Also, in case of fire settlement with the type of loss of the building, in the morning it is caused by a shorted with a support value of 0,32 and a confident value of 0,67. If there is a fire settlement with the type of building damages caused by the short circuit occurred in the morning at 0,32 day with support value and confident value at 0,48.

References

[1] N. Wagner and V. Agrawal, "An agent-based simulation system for concert venue crowd evacuation modeling in the presence of a fire disaster," *Expert Syst. Appl.*, vol. 41, no. 6, pp. 2807–2815, 2014.

[2] H. B. Cheng, Z. Y. He, H. T. Hu, X. Q. Mu, and B. Wang, "Risk assessment and early warning of lightning disaster for traction power supply system of high-speed railway," *J. China Railw. Soc.*, vol. 35, no. 5, pp. 21–26, 2013.

[3] F. Fadlina, "Data mining untuk analisa tingkat kejahatan jalan dengan algoritma association rule metode apriori (Studi kasus di Polsekta Medan Sunggal)," *Maj. Ilm. Inf. dan Teknol. Ilm.*, vol. 3, no. 1, pp. 144–154, 2014.

[4] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules," in *Proc. 20th int. conf. very large data bases, VLDB*, 1994, pp. 487–499.

[5] C. Zhang and S. Zhang, *Association Rule Mining : Models and Algorithms*. Sydney: Springer, 2002.

[6] S. Brin, R. Motwani, and C. Silverstein, "Beyond market baskets: Generalizing association rules to correlations," *Acm Sigmod Rec.*, vol. 26, no. 2, pp. 265–276, 1997.

[7] Y. Zhao, *R and data mining: Examples and case studies*. Elsevier, 2015.

[8] V. Kumar and A. Chadha, "Mining association rules in student's assessment data," *Int. J. Comput. Sci. Issues*, vol. 9, no. 5, p. 211, 2012.

[9] M. A. Anwar and N. Ahmed, "Knowledge mining in supervised and unsupervised assessment data of students' performance," in *2011 2nd International Conference on Networking and Information Technology IPCSIT*, 2011.

[10] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to data mining*, First edit. Boston, MA: Addison-Wesley Longman Publishing, 2005.