Statistical Analysis to Select Evacuation Route

Z Musyarof, DY Sutarto, DR Atika and RB Fajriya Hakim

Department of Statistics, Faculty of Mathematics and Natural Science, Islamic University of Indonesia, Jalan Kaliurang KM 14.5, Sleman, Indonesia

E-mail: zakymusyarof@yahoo.com, hakimf@fmipa.uii.ac.id

Abstract. Each country should be responsible for the safety of people, especially responsible for the safety of people living in disaster-prone areas. One of those services is provides evacuation route for them. But all this time, the selection of evacuation route is seem does not well organized, it could be seen that when a disaster happen, there will be many accumulation of people on the steps of evacuation route. That condition is dangerous to people because hampers evacuation process. By some methods in Statistical analysis, author tries to give a suggestion how to prepare evacuation route which is organized and based on people habit. Those methods are association rules, sequential pattern mining, hierarchical cluster analysis and fuzzy logic.

1. Introduction

We have anxiety panic attacks because we are creatures of habit [10]. We can say that habit is like a routine work. Habit makes us behave in a similar that habit when we were panic. One of the panic moment that may happens is when a disaster happen, because disaster is threaten lives.

Humans often experience anxiety attacks because human beings are accustomed to doing routine things. The habit causes us to behave similarly when we experience a panic. One moment of panic that may occur is when disaster strikes. Based on common practice, then in the face of disaster, people will behave the same as the disaster threatens their safety.

These things happened when Kelud Mountain, Kediri, erupted. Traffic jam occur in main evacuation route, Wates, Kediri, Province of East Java. Officers had difficulty to untie the traffic jam because of the large number of vehicles, doesn't comparable to existing roads. The traffic jam made people scared. They screamed to call out for help from officers, while hail is happened. Sand and volcanic ash is also fell. On the sky, bright and red storm is happened. Thunder is also heard by all the people [3].

All this time, the selection of evacuation route doesn't well organized. People is forced as soon as possible to goes to main road, and unfortunately main road is limited especially in developing countries. With that condition, many accumulation of people on the steps of evacuation route when a disaster happen. That will cause occurrence of traffic jam and hamper evacuation process. And of course will endangers people lives.

There are many ways to describe emergency management and the importance purpose is saving lives and property in disasters. Science and technology were applied in planning and managing the extreme events that can injure or kill large numbers of people, do extensive damage to property and disrupt community life [4]. Statistics as a science had been applied in disaster management [9]. This

method will use a statistical method to make simulation method in managing the evacuation road for mitigation when Merapi Volcano erupt. One of many Statistics methods is association analysis, which is useful for discovering interesting relationships which is hidden in large data sets. This relationships can be displayed in the form of association rules or sets of frequent items (frequent road passed through by local residents). For illustrative purposes, our discussion of association rules will be represented in the term of market transaction.

Let $I = \{i_1, i_2, ..., i_d\}$ be the set of all items in a market transaction data and $T = \{t_1, t_2, ..., t_N\}$ be the set of all transactions recorded in database. Each transaction t_i contains a subset of items which is chosen from I. In association analysis method, a set of zero or more items is termed an itemset. The transaction width is defined as the number of items present in a transaction. A transaction t_j is said to contain an itemset X if X is a subset of t_j. An important property of an itemset is its support count, which refers to the number of transactions that contain a particular itemset. Mathematically, the support count, s(X), for an itemset X can be stated as follows:

$$s(X) = |\{t_i | X \subseteq t_i, t_i \in T\}|,$$

where the symbol |.| denote the number of elements in a set. An association rule is an implication expression of the form $X \rightarrow Y$, where X and Y are disjoint itemsets. The strength of an association rule can be measured in terms of its support and confidence. Support determines how often a rule is applicable to a given data set, while confidence determines how frequently items in Y appear in transactions that contain X. The formal definition of these metrics are:

$$\text{Support, } S(X \rightarrow Y) = s(X \cup Y)/N$$
$$\text{Confidence, } C(X \rightarrow Y) = s(X \cup Y)/s(X)$$

Support is an important measure because a rule that has very low support may occur simply by chance. A low support rule is also likely to be uninteresting from a business perspective because customers seldom buy together. Support is often used to eliminate uninteresting rules. Support also has a desirable property that can be exploited for the efficient discovery of association rules. Confidence, on the other hand, measures the reliability of the inference made by a rule. For a given rule $X \rightarrow Y$, the higher the confidence, the more likely it is for Y to be present in transactions that contain X. Confidence also provides an estimate of the conditional probability of Y given X. In the selection of an evacuation route, this method will be applied to see which road in that route frequently traveled together.

In this case, support determines how often a rule is applicable to the road taken by the locals while to save themselves. Confidence determines how frequently road Y traveled after road X taken during evacuation process. A rule that has very low support may occur simply by chance or we could say that the attempt to save themselves, they are just guessing which road can be passed safely without having been familiar with the road conditions as best as possible which evacuation route to be taken.

Association rules says that when someone do A, and he or she do B too will get a support and confidence value. The bigger support and confidence, then more often the occurrence of $A \rightarrow B$ happen from all. In selecting for evacuation route, this method will use to see association rules of activities and driving social habit. This method will sees is it true that if someone drive on A road will drive on B road too or will drive on other roads too. The rule is different in each village.

This method is used to select the evacuation route, therefore, should pay attention to the patterns formed road sequence. Evacuation route to be sequenced from a starting location to the point closest to the epicenter of the safest distance. Method pattern formed associations combined with the sequential pattern.

Roads conditions in each other is different, there are good roads and there are damaged roads

too. Each road will be grouped using hierarchical cluster analysis. Grouping will produce some graded group, very prone to very safe. Variables that will be used is road width, road condition and ease of access road. These variables doesn't have an indicator to be evaluated, so these variables evaluated using fuzzy logic.

The selection of evacuation road is done by activities and driving social habit by considering roads conditions in each village. Besides those, population in each village is also considered. The selection of evacuation road was directed away from disaster-prone areas of last disaster happen.

2. Related Research

Information from IKONOS images can be to create a map of evacuation routes and the impact area of cold lava floods. The preparation of evacuation route consider to eight parameters. Those parameters are disaster-prone area, slope, length and width of roads, material of roads, the existence of bridges and roads direction [2].

Association rules method can also used in education database. This method is used to seek for unique rule of Teacher Certification in data of Nomor Unik Pendidik dan Tenaga Kependidikan (NUPTK), codes of teachers in Indonesia. They found there are 184 rules in the data [1].

Fuzzy logic has been widely applied in real life, one of all is using fuzzy logic to find best supplier for a clothing store. Researcher use Fuzzy Principal Component Analysis and Analytic Hierarchy Process (AHP). Those two method are methods of decision-making that use qualitative and quantitative measurement. Qualitative measurement define issue and quantitative measurement provide an assessment. This methods use primary data from questionnaires and observation [7].

A specific application of "automated geo-spatial data acquisition and mapping" offers the transformation of verbally given geo-information into a situation map. For a successful transformation it is essential to develop a domain specific knowledge base. The presented Disaster Management Data Model is based on the Command and Control Information Exchange Data Model. Although the C2IEDM is a sophisticated standard, the DM points out important considerations that have to be taken into account for disaster management ontologies in general [8]. Another application is used in health. Epidemics, chronic diseases which are the major social disasters follow strategic-virulent disasters that affect the ecosystem of a spatial zone probabilistic study is made on the health demographic data. Using the Collocation rule, the effected area of dengue is found and results are obtained. Using the Participation index the symptoms that influence spreading the dengue were found and results are obtained. i.e., spatial knowledge for the health campaign. Identifying dengue effected area using spatial maps as spatial knowledge i.e. to identify spreading of dengue was proposed and proved [6].

Discovering association rules in sequences is an important data mining problem that is useful in many scientific and commercial domains. Predicting events and identifying sequential rules that are inherent in the data help domain experts to learn from past data and make informed decisions for the future. Several different approaches have been investigated for sequential data mining. Algorithms for discovering associations in sequential data, and episodal associations, use all frequent episodes. The entire set of association rules is produced and significance criterion such as J-measure for rule ranking are used to determine the valuable rules. An approach that uses temporal constraints on transactional sequences was presented in. We present a new approach that uses Minimal Occurrences With Constraints And Time Lags (MOWCATL), to find relationships between sequences in the multiple data sets. In addition to the traditional frequency and support constraints in sequential data mining, MOWCATL uses separate antecedent and consequent inclusion constraints, along with separate antecedent and consequent maximum window widths, to specify the antecedent and consequent patterns that are separated by a maximum time lag. The MINEPI algorithm was the first approach to find minimal occurrences of episodes [5].

3. Research Methodology

In the process of selecting evacuation route, three methods of Statistics will used be used. First method is association rules, one of many methods of data mining. This method is used to see the rules of

people when driving and activity. Selecting evacuation route process will be based on those rules that will be obtained. Those rules can called as local people habit. Those rules were later obtained will be different for each village, so also will decrease accumulation of people on the evacuation route sections risk. Each rule have support and confidence in percentage. Support is got from:

$$Supp(X) = \frac{X\ incident\ (transaction)}{All\ incidents\ (transactions)}$$

And confidence is got from:

$$Conf(X \rightarrow \mathrm{Y}) = \frac{Supp(X\ U\ Y)}{Supp(X)}$$

With sequential mining, sequence of lines considered. The route that obtained is not distorted from the nearest lane to the lane farthest point of disaster and back to the nearest road again. The route begins to form on the road closest to the disaster epicenter.

The next method is two methods that are combined, those two methods are hierarchical clustering analysis and fuzzy logic. Fuzzy logic is used to assess the assessment variables. Those variables are road condition, road width and ease of access road. Those three variables doesn't have indicator to rate the assessment variables, so the assessment is conducted using simulation of fuzzy logic. Fuzzy set will be divided to seven stratified groups, those groups are:

1. very unsafe,
2. unsafe,
3. bit unsafe,
4. normal,
5. bit safe,
6. safe, and
7. very safe.

Fuzzy set for the three variables is different. Road condition variable is divided into:

1. very bad,
2. bad,
3. bit bad,
4. normal,
5. bit good,
6. good, and
7. very good.

Road width variable is divided into:

1. very narrow,
2. narrow,
3. bit narrow,
4. normal,
5. bit wide,
6. wide, and
7. very wide.

Ease of access road variable is divided into:

1. very hard,
2. hard,
3. bit hard,
4. normal,
5. bit easy,
6. easy, and
7. very easy

Disaster affected area from the last disaster is also used as a consideration to select evacuation route. The first selecting evacuation route process is decide some driving local people habit. The rule (habit) with the largest support becomes the main recommendation to be the main evacuation route. Smaller support rules used as routes that local people is used to reach the main evacuation route. Those all rules directed away from disaster affected area from the last disaster. All rules routes are grouped into seven determined groups. Evacuation route selection consider to which group they are grouped. Population of villages is also considered to select evacuation route.

For example, one of some main evacuation route of Merapi Volcano Evacuation Route, Sleman, Yogyakarta, is used. All the evacuation route is as shown in **Figure 1**. Merapi Volcano evacuation route is used only to make it easier, as authors live near Merapi Volcano.

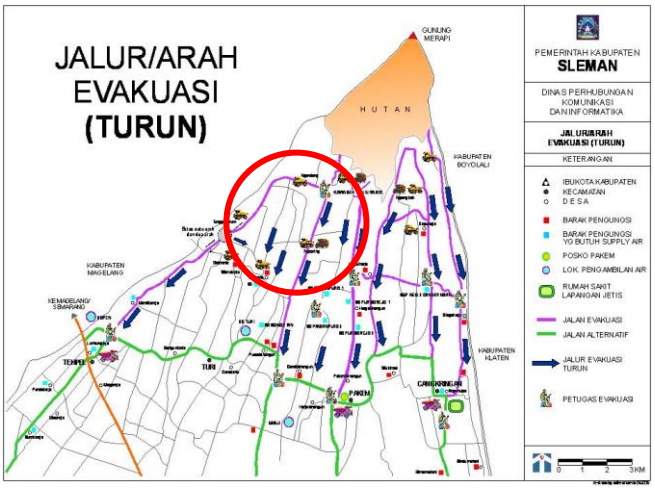

Figure 1. Merapi Volcano evacuation route. **Source :** Sleman District Government (slemankab.go.id)

Evacuation route that used as a simulation is route circled in red, the routes is as shown in **Figure 2**.

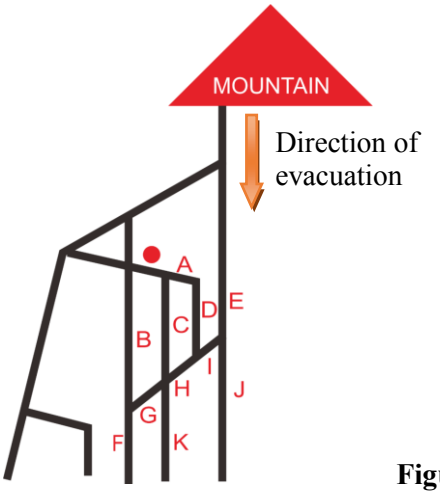

Figure 2. One of Merapi Volcano evacuation routes.

● : starting point of Merapi volcano evacuation route simulation

For example, the road condition of B is shown **Figure 3**, road condition of C is shown **Figure 4** and road condition of D is shown **Figure 5**.

Figure 3. Road B condition.

Figure 4. Road C condition.

Figure 5. Road D condition.

And the affected area of the last Merapi Volcano eruption in 2010 is as shown in **Figure 6**.

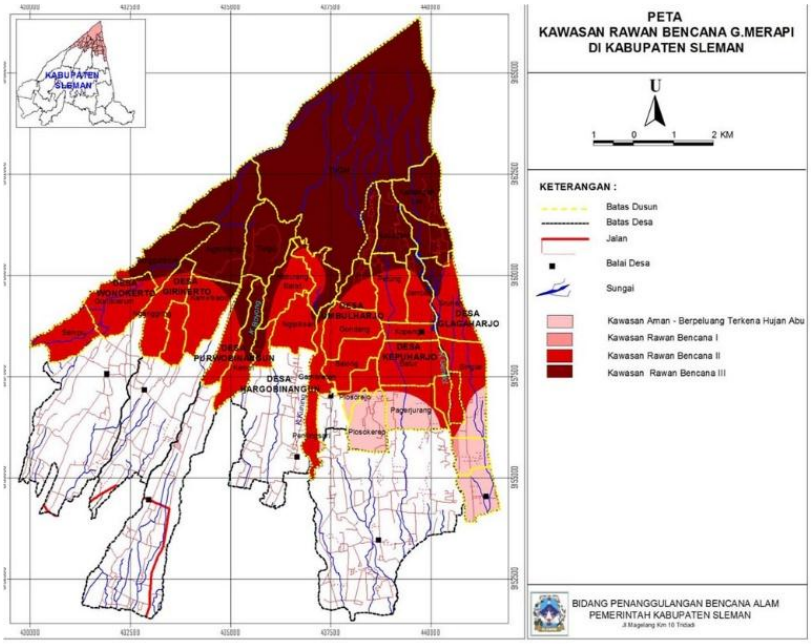

Figure 6. Affected area of Merapi Volcano eruption (based on 2010 eruption).
Source : Sleman District Government (slemankab.go.id)

4. Simulation Study

This study will use a simulation with local people living in the disaster-prone areas. A thousand people were then placed in one area (in Fig. 1 colored red dots). Then circumstances made as if it is similar to when Merapi Volcano will reach the highest point of warning. After being given a sign to immediately evacuate, then the thousand people hasten to save themselves down the road that they consider the best way to save themselves. After being in a safe place that is determined, then every person who follows the simulation, telling which line he passes to reach a safe point. The result is then recorded as follows.

Table 1. Example of the route that taken by local residents
(This data is not real, used only to show how the data, these data is not exist.)

Travel (person)	Route
1	C, H, I, J
2	C, K
3	C, G, F
4	A, D, I, J
5	B, G, K
...	...
997	B, F
998	A, D, H, K
999	C, H, I, J
1,000	C, K

Then we get the following measurements of rules for two routes,
$A \rightarrow I$: 0.17,
$C \rightarrow I$: 0.1,
$C \rightarrow F$: 0.32,

...
$B \rightarrow F$: 0.24,

Those support values is not real, those four support value is just a made-up and there are no support value for another route.
From those rules, seems that rule of "$C \rightarrow F$" get the highest value of support. So, rule "$C \rightarrow F$" is the first recommendation to become main evacuation route. However, to reach route F from C, people have to go through route G. Thus the existing rules instead of $C \rightarrow F$ but became $C \rightarrow G \rightarrow F$. Therefore, this rules recommends that the main evacuation route is $C \rightarrow G \rightarrow F$.

The second biggest support is $B \rightarrow F$ rule route, so that rule become second recommendation to become main evacuation route or ordinary evacuation route to reach main evacuation route for a while. The third is $A \rightarrow I$ rule route and become third recommendation, and so on until the smallest support. Road condition of those all rule route are also considered. It should be noted that this method using sequential mining, so the results that obtained is started from the closest from volcano to farthest.

From direct observation of the existing roads, it appears that the road B (**Figure 3**) has poor road conditions and can harm people when they pass through in large numbers and panic. The condition of road C (**Figure 4**) rather bad, while the condition of the road D (**Figure 5**) look good, but after the barrel located in the middle of the road, the left side of the road was severely damaged. Those three roads have same width. By considering the real roads conditions, then the best choice is road C. Direct observation was also conducted to assess the other roads.

From observation, seems that route B (**Figure 3**) is damaged and endanger local people if they pass through that route. Route C (**Figure 4**) is bit damaged and route D (**Figure 5**) is in very good condition but after the drum, there is landslide on the left side of the road. Those three route is also have same width.

Authors saw directly roads then assessed in the opinion of authors (in fuzzy set). In simulation, roads condition for each route is below:

Table 2. Roads condition for each route.

Route	Road Condition	Width	Ease of Access
A (1)	Bad	Normal	Very easy
B (2)	Bad	Normal	Easy
C (3)	Bit bad	Normal	Easy
D (4)	Very bad	Normal	Easy
E (5)	Good	Wide	Hard
F (6)	Bit bad	Bit wide	Bit hard
G (7)	Normal	Bit narrow	Bit hard
H (8)	Bad	Bit narrow	Bit hard
I (9)	Normal	Bit narrow	Easy
J (10)	Very good	Very wide	Hard
K (11)	Good	Wide	Hard

Using hierarchical clustering (single linkage, manhattan), got dendrogram below:

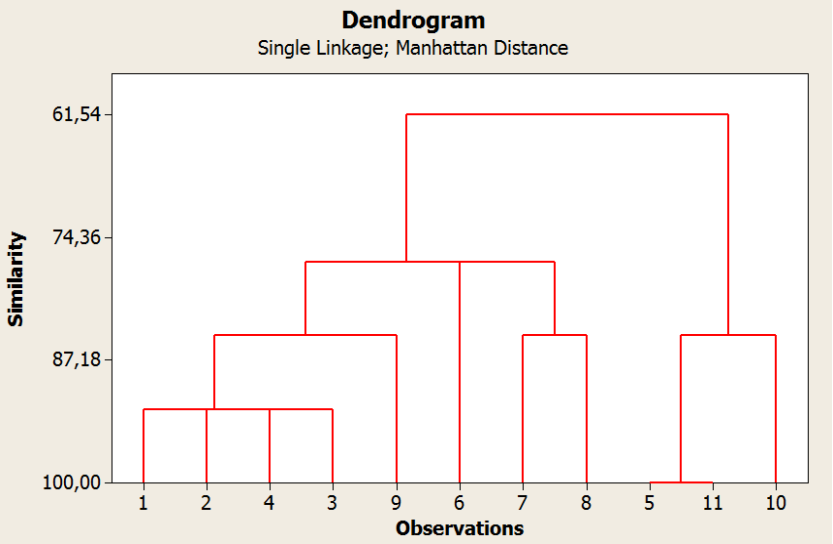

Figure 6. Dendrogram of roads condition

From dendrogram, choosed object is grouped into 4 groups, got groups:

Table 3. Groups.

Group	Ease of Access
1	1, 2, 4, 3, 9
2	6
3	7, 8
4	5, 11, 10

From the result of association rules, got rules C → G → F become the first recommendation to become main evacuation route. From the hierarchical clustering, road C is grouped into Cluster 1, G is grouped into Cluster 3 and F is grouped into Cluster 2. There's no road from road C, G and F that is grouped into worst cluster, so route C → G → F still can become the first recommendation to become main evacuation route.

The second evacuation route and so on can be determined in the same manner. These results can then be provided to the competent authorities in the evacuation problem. Thus dividing routes could consider road capacity and the number of actual residents in order to avoid excess capacity at one point causing congestion and excessive panic.

5. Result and Discussion

Results and conclusions are as follows:

1. in simulation, from the result of association rules, got rules C → G → F become the first recommendation to become main evacuation route. From the hierarchical clustering, road C is grouped into Cluster 1, G is grouped into Cluster 3 and F is grouped into Cluster 2. There's no road from road C, G and F that is grouped into worst cluster, so route C → G → F still can become the first recommendation to become main evacuation route. The second evacuation route and so on can be determined in the same manner.
2. selection of evacuation route with this method can decrease accumulation of people on the steps of the evacuation route risk,

3. if this method is used in real life, the hardest part is to record activity and driving people data. Same person must recorded from first place, then their travel road, until their destination. And because data mining analysis need big data to obtain relevant result, and
4. for the development of this method, small area estimation method can be used to estimate the disaster-prone area and many simulation possibility of evacuation process, so probability of people safe on every part of evacuation route can be calculated.

6. References

Results and conclusions are as follows:

[1] Amiruddin P et. al. 2010 *Penerapan Association Rules Mining Pada Data Nomor Unik Pendidik dan Tenaga Kependidikan Untuk Menemukan Pola Sertifikasi Guru* (*Association Rules Mining Aplication To Find Teacher Sertification Rules In Nomor Unik Pendidik dan Tenaga Kependidikan Data*) 15-16

[2] Ardana D M S and Purwanto T H. 2012 *Penentuan Jalur Evakuasi dan Dampak Banjir Lahar Dingin Gunung Merapi* (*Select An Evacuation Route and The Impact Of Cold Lava Flood From Merapi Mountain*) 149

[3] BeritaJatim.com April 24th 2014 *Warga histeris, jalur evakuasi macet total*, http://m.beritajatim.com/peristiwa/198365/warga_histeris,_jalur_evakuasi_mmace_total.html#.U1jF5_mSxmw

[4] Drabek 1991 *Introduction to emergency management* ch 1

[5] Harms S K. et al *Discovering Sequential Association Rules with Constraints and Time Lags in Multiple Sequences* 1-2

[6] Kameswara R et. Al 2013 *Spatial mining system for disaster management* 6

[7] Muhajir M 2011 *Pendekatan Metode Fuzzy Principal Component Analysis dan Analytic Hierarchy Process dalam Pemilihan Supplier Terbaik Produk Bahan Baku Baju*" (*Fuzzy Principal Component Analysis and Analytic Hierarchy Process Methods Approach to Select Best Raw Material Dress Products Supplier*) 17

[8] Lucas et. al 2007 *Information Mining for Disaster Management* 6

[9] Savoiu G *The Holistic Concepts of Disaster Management and Social Cohesion* Statistics and Method, Scientific Bulletin – Economic Sciences Vol. No. 10/Issue 1

[10] The Secret of Mindpower and NLP April 22th 2014 *Anxiety Panic Attacks* http://www.the-secret-of-mindpowerandnlp .com/an.xiety-panic-attacks.html