Spatio-Temporal Associative Mining for Earthquake Data Distribution in Indonesia

Renovita Edelani, Ali Ridho Barakbah, Tri Harsono

Politeknik Elektronika Negeri Surabaya
Kampus PENS, Jl. Raya ITS, Sukolilo Surabaya
E-mail: redelani1410@gmail.com, {ridho, trison}@pens.ac.id

Abstract

Indonesia has the highest seismic activity in the world because it traversed by three meeting plates (Indo-Australian, Eurasian and Pacific Plate) and located in the Ring of Fire area. The shaking events of an earthquake are very strong, spread in all directions and had a seismic connection between each contiguos event based on location or time occur. The impact of an earthquake can destroy property, triggers another disaster and take so many victims. There is a seismic association between provinces in Indonesia. In this paper, we present a new earthquake Spatio-temporal mapping system based on the association confidence value from the result of associative mining process on earthquake data distribution in Indonesia. The system proposed three main functions which are (1) Data Acquisition which taken from four data provider, then preprocess and combine it become one, (2) Associative Mining process to get the rule of association earthquake between provinces in Indonesia, and (3) Earthquake Association Spatio-Temporal Model from the highest confidence value and Visualization. We use data from several earthquake data providers from 1900 until 2018. To perform our proposed Spatio-temporal earthquake association mapping system, we divided the data to become a 5-year discrete partition. After that, we mining the rule and get the highest confidence value from each period. This confidence value is used for modeling and visualization of our Spatio-temporal mapping system. As a result of this study, we manage to generate earthquake association risk mapping from 13 provinces. The low-level risk provinces are North Sumatra, Bengkulu, NTT, Gorontalo and West Papua. The moderate risk provinces are Aceh, Papua, North Sulawesi, and Maluku. The high-level risk provinces are East Java, Bali, NTB, and North Maluku.

Keywords: Earthquake, Association Rule, Spatio-Temporal Visualization, Risk-Mapping

1. INTRODUCTION

An earthquake is the shaking events of the surface of the earth due to the sudden release of energy which is characterized by broken rock layers in the earth's crust [1]. An earthquake is one of the most destructive natural disasters. Many provinces in Indonesia are in the active plate subduction zone. If a large earthquake occurs, it is almost certain that a large number of fatalities, material and traumatic losses will occur in the community. The consequences of the earthquake have been tremendous because they cover a very wide area, penetrating the territorial boundaries of the state, even between continents. The impact of this disaster inevitably must be minimized. Therefore, Indonesia needs careful planning in terms of rapid development and disaster management. In 2018 there were 11,577 earthquake events in various magnitudes and depth occurred in Indonesia [2]. The country of Indonesia belongs to one of the countries whose territory is on the Ring of Fire or the Ring of Fire so that it becomes the area that has the highest seismic activity on earth. Indonesia is also an area prone to earthquakes because it is traversed by three meeting plates. The three tectonic plates are the Indo-Australian tectonic plate, the Eurasian Plate and the Pacific Plate and the small plate of the Philippines.

Movement of a plate can cause other plate activity. So, if an earthquake occurs in an area caused by the movement of the plates on the earth, it is very likely that an earthquake will occur in other areas that are on the path of the meeting plate. One proof is the statement from the Head of the Center for Meteorology, Climatology, and Geophysics (BMKG) Center, Prof. Dwikorita Karnawat, the earthquake that occurred in Banten province on January 23, 2018, had a considerable risk because the plate movement in the area also covered the Sulawesi region, Papua and Ambon. "So the risk is quite extensive," he said. The potential for Pacific Ocean Plate and Eurasian Plate collisions to occur. The Pacific Ocean plate is east of Sulawesi to the west towards the Eurasian Plate. It's also potential, "Dwikorita said on Tuesday (1/23/2018) [3]. This seismic connection can be caused by the arrangement of the plates that cause earthquakes in Indonesia and other factors. With the analysis of the earthquake association, we can visualize it into an earthquake disaster mapping.

The fundamental problem raised in this study is the need for earthquake risk mapping and relationships of earthquake events in several regions in Indonesia. Analysis of earthquake data has been carried out a lot but rarely discusses seismic associations between regions in a country.

2. RELATED WORKS

Earthquake analysis is a hot topic now. Lots of researchers are involved in this field of research. Because the earthquake is one of the natural disasters that occur at any time and has a very harmful and destructive effect on human life. Besides, with easy access to earthquake data throughout the world, many researchers try to discovery new finding from the earthquake data. Of the

various research in these topics, there are three topics related to this research, namely: earthquake connection analysis with other factors, earthquake data analysis, and earthquake risk mapping.

On the topic of earthquake connection analysis with other factors, there are Aqdas Ikram and Usman Qamar [4] used the Association Rule method to obtain seismic rules that will later be applied to the expert system to predict earthquakes, Bin Hu et al. [5] analyzed the relationship between the impact of earthquakes on earthquake victims and their living environment, Wang Xi et al. [6] examined the relationship of credible associations and the probability of crises from 9 influential earthquakes in the world to predict the possibility of the next major earthquake, Yuan Zhou and Lianxiong Gao [7] used apriori for analysis spatial association pattern during point and line feature, and description algorithm process, Jin A. Lee et al. [8] experimented to found the unknown characteristics of earthquakes using association rule mining methods global earthquake data occurred, Zhong Li et al. [9] had been done the research by using the association rule method between the abnormal variations of earthquake precursory information and the occurrences of earthquakes, and also the correlations among precursory data subjects.

Related to earthquake analysis, there are Amin Endah Suliswati et al. [10] which calculated earthquake density values using the automatic clustering method, Ali Ridho Barakbah et al. [11] which analyzed cluster-oriented seismicity predictions of earthquakes in Indonesia and semantic interpretation of predicted results projected into each region, Mohammad Nur Shodiq et al. [12] performed earthquake predictions using the Automatic Clustering and Neural Networks methods, Ken-ichi Fukui et al. [13] found the damage patterns in Fuel Cell and Earthquake Occurrence Patterns by Co-Occurring Cluster Mining, Evaldas Luksys et al. [14] proposed tools for enabling scientists analyze and interpret large-scale datasets about earthquakes in a way which will complement current analytical tools and thinking thus, complement current effort on understanding the event itself.

Dealing with earthquake risk mapping, there is Arna Fariza et al. [15] which maps the seismic risk of East Java province were used the Analytical Hierarchy Process method, Man Li et al. [16] were mapping earthquake risk throughout the world based on the earthquake vulnerability table of each region, Mapping the earthquake risk in Indonesia in 2015 by Badan Nasional Penanggulangan Bencana (BNPB- National Disaster Management Agency Indonesia) [17] based on the hazard index, vulnerability index and capacity index, Daniela Faur et al. [18] created a methodology used to quantitatively evaluate the impact of an earthquake on a region, considering multi temporal high resolution optical images, Junxiang Zhang [19] reformed earthquake map using a fuzzy approach by integration of fuzzy set theory, risk analysis and geographical information system (GIS) technology, F. Benjamin Zhan et al. [20] has made a geospatial visual analytical system for mapping medium-term earthquake probabilities at a global scale, Xiang Ding et al. [21] developed a system named The Catastrophe Earthquake Risk Estimation System (CERES

for Windows) to catastrophe earthquake risk assessment applied especially in Asia based on GIS, Qing-Quan Tan et al. [22] built information collection is proposed and implemented, which is based on GIS technology.

From all research above, our research dealing with earthquake risk mapping from the result of earthquake data analysis using Associative Mining method.

3. ORIGINALITY

In this paper, we present a new earthquake Spatio-temporal mapping system based on the highest association confidence value from the result of associative mining process on earthquake data distribution in Indonesia. The earthquake data was taken from 1900 to 2018 from several sources of earthquake data providers namely the Meteorology, Climatology and Geophysics Agency (BMKG), United State Geological Survey (USGS), European-Mediterranean Seismological Center (EMSC), and International Seismological Center (ISC). In this system, we using an experimental scenario by dividing earthquake data from 1900-2018 into 5 years trial period. The system has three main functions, there are:

The first function, data retrieval and preprocessing is the process of retrieving the data from several providers, merge it to make the most complete earthquake data, after that preprocessing the data like cleaning, integration, and reduction of the data. In this paper, we used a more complete dataset than other studies by combining earthquake data from 4 data providers where other researchers just using one provider such as USGS or BMKG. Data cleaning is the process of erasing data that is incomplete or has no value on the depth and magnitude attributes. Data integration is needed because each data provider records earthquake events differently. This process homogenizes the format of all earthquake data attributes. Data from all providers are combined so that there is a redundancy of earthquake data from the same earthquake source. The process of data reduction involves calculating the proximity between data by using Dynamic Time Warping so that it can be known which data has the same earthquake point source (has a small distance value) and is selected into one data only.

The second function is the Association rule mining process, which aims to find earthquake associations rules for each province in Indonesia that produce antecedent provinces and consequent provinces with their confidence values. With this method, we discovered the earthquake relation between provinces in Indonesia based on the occurrence day of events. When there are several earthquake events in 7 days, Association Rule would find which provinces that have a cause-effect relationship.

The third function is modeling and visualization of earthquake association between each province in Indonesia based on the highest confidence value. Modeling of earthquake risk based on Association Rule results is the new method we developed in this research. The level of earthquake risk in each consequent province is taken from the confidence

value of earthquake cause-effect relation with the antecedent province. In this study, earthquake risk is categorized into 3 i.e low, moderate and high. This categorization is based on Richter scale graphic representation.

4. SYSTEM DESIGN

To achieve the objectives of this research, the system design was developed to explain or break down all the work done by the researcher. The system design proposed in this study in Figure 1. The system design explains the workflow of the research, there are (1) Data Acquisition and Preprocessing, (2) Associative Mining, and (3) Earthquake Association Spatio-Temporal Model and Visualization. Each process is explained in section 4.1-4.3.
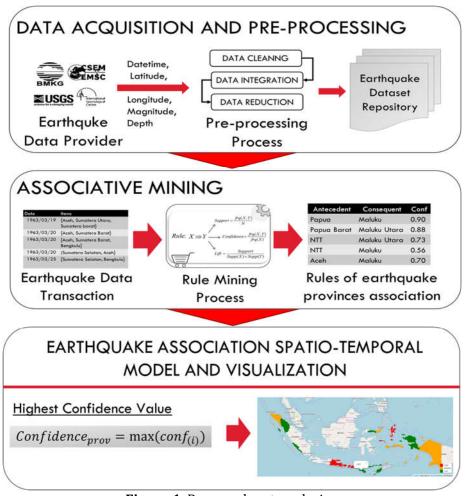

Figure 1. Proposed system design

4.1 Data Acquisition and Preprocessing

At this stage, data will be taken from several earthquake data providers and will be processed in such a way in several stages, namely data cleaning, integration, and reduction as well as determining the provincial location of each data so that the data is ready for use at the next stage. Earthquake dataset

is a collection of earthquake data that has attributes such as DateTime, latitude, longitude, magnitude, and depth of the earthquake. This data is taken from ANSS (Advanced National Seismic System) or USGS (United States Geological Survey), BMKG (Meteorology, Climatology and Geophysics Agency), European-Mediterranean Seismological Center (EMSC), and International Seismological Center (ISC). The earthquake data will be taken starting from 1900 - 2018, between latitude -11 - 6 and longitude 94.0 - 142.0. This research retrieves and manually downloads data from each earthquake provider website.

The dataset obtained from the data provider is not all complete and there are missing values such as the magnitude and depth of the earthquake. Data cleaning was deleting the data that had a missing value. The data provided by each earthquake data provider is very diverse and has its own characteristics. From the attributes provided from these data, there are several that are needed by this study. Data Integration process dealing with changes the date and time format for each dataset to dd / MM / yyyy where between the date and time are separated or not one column. Dataset had the date, time, latitude, longitude, depth and magnitude attributes. After the data integration process, data from all data providers are combined into one so that there is redundancy of earthquake data with the same source. One earthquake source can be recorded differently by several providers because the recording sensor of each provider has a different location and method of calculating the earthquake source value. To determine the data has the same source, the DateTime, latitude, longitude and magnitude attributes are used to calculate the distance between data. These four attributes are taken because they have close distance values with other data that have the same earthquake source. The depth attribute is not used because the value of this attribute differs too much from one data to another. The DateTime attribute is changed in the form of a timestamp first. After that, the distance between the data will be calculated using the Dynamic Programming method (the calculation method used by DTW). Dynamic time warping (DTW) is a technique for finding optimal alignment between two data sequences (time-dependent) given within certain limits. The purpose of DTW is to find the best mapping with minimum distance using Dynamic Programming (DP) [23]. Stages to reduce data are:

1. Data is sorted by date and time.
2. Calculate the distance between $data_{(n)}$ (timestamp, latitude, longitude, and magnitude atribut) with $data_{(n-1)}$ using Dynamic Time Warping method [23].
   a. If the distance ≤ 10 then $data_{(n)}$ have the same cluster with $data_{(n-1)}$.
   b. If the distance > 10 the $data_{(n)}$ have different cluster with $data_{(n-1)}$.

These data after being labeled as a group, one data will be taken from each group using the modus calculation method. Where data that has the closest distance to the midpoint or centroid is considered as data that

represents the source of the earthquake. Centroid value is obtained from the average value of latitude, longitude and magnitude of the data that is in one group.

After passing through several preprocessing stages, the dataset will be stored in the earthquake database or repository. The province of each data is taken from the latitude and longitude values. The process is carried out by calculating the closest distance from the latitude and longitude points of the earthquake data to the midpoint of each province, after taking three provinces closest to the point of the earthquake data, calculate the distance of the earthquake data with the edge points of the three closest provinces, the province of that earthquake data is the closest edge provinces.

4.2 Associative Mining

Association Rule is the main techniques or procedures in Market Basket Analysis to find facts about the relationship between items in a data set and display in the form of Association Rule [24]. Association Rule will find certain patterns that associate one data with another. To find the Association Rule from a data set, the first stage that must be done is to look for frequent itemset first. After all the frequent itemset patterns are found, then look for associative rules or linkages rules that meet the predetermined requirements [25].

Association Rule requires a variable size that decided by the user to determine how much output the user wants. Support and confidence are a measure of the trust and usefulness of a pattern that has been found [24]. Whereas Lift ratio is a measure to determine the strength of the formed association rules [26].

The Apriori algorithm is an algorithm that performs frequent itemset searches using the Association Rule technique [27]. This algorithm was introduced by Agrawal and Srikant in 1994. Apriori was designed to operate a database containing transaction data. The Apriori algorithm uses the knowledge of the frequency of previously known attributes to process further information. The Apriori algorithm determines which candidates might appear by paying attention to minimum support and minimum confidence. Support is the visitor's value or the percentage of a combination of items in the database. Support formulas are as follows:

$$\text{Supp(X)} = \left(\frac{\text{number of transaction containing X}}{\text{Total transaction}} \right) \quad (1)$$

While confidence is the value of certainty, the strong relationship between items in an apriori. Confidence can be searched after the pattern of frequency of appearance of an item is found. The formula for calculating confidence is as follows:

$$\text{confidence}(X \rightarrow Y) = P(Y|X) = \frac{\text{support}(X \cup Y)}{\text{support}(X)} = \frac{\text{support\_count}(X \cup Y)}{\text{support\_count}(X)} \quad (2)$$

Lift is a measure that shows how important or useful a rule [26]. Lift is measured based on their respective antecedent and consequent support. The greater the value of the lift, the stronger the association or connection between itemset on the rule. The lift value can be obtained by the following formula:

$$\text{Lift}(X \rightarrow Y) = \frac{\text{support}(X \cup Y)}{\text{support}(X) \times \text{support}(Y)} \quad (3)$$

Or

$$\text{Lift}(X \rightarrow Y) = \frac{\text{confidence}(X \rightarrow Y)}{\text{support}(Y)} \quad (4)$$

The process of finding earthquake association rules from provinces in Indonesia uses the Association Rule method, there are 3 stages: The first stage that must be done is the formation of transaction data because this method requires data in the form of data transactions. Earthquake transaction data is formed based on the determined n-day earthquake event. For example, transaction data for 7 earthquake days means earthquakes that occur during 7 days will be used as one transaction data. Similarly, if the data transaction for one day, then the earthquake that occurred on that day into one transaction data.

The second stage, Association Rule process for transaction data to get the rule from the specified transaction data. The stage is to find support values from each province, after determining the first combination of 2 provinces, rediscovering the second combination support value for 3 provinces, do that until there are no more combinations for each province after that calculate the desired provincial association confidence. The minimum support value inputted on the parameter is a frequent itemset determinant of earthquakes in each province. The confidence value obtained is used as a percentage of the probability of an earthquake occurring in other provinces and later this confidence value is used as a basis for obtaining an earthquake risk value.

The next stage is to measure the strong bond of the rules that have been obtained used measurements using the Lift ratio. The Lift will give the strength value of the association from the earthquake of each province in Indonesia. The lift value greater than 1 is considered to be the valid rule or shows the inter-relation earthquake between the provinces.

4.3 Earthquake Association Spatio-Temporal Model and Visualization

In this research Spatio Temporal Associative Mining Visualization based on earthquake association between provinces in Indonesia, provinces that get the risk from other provinces will be visualized. This paper will divide the years of an experiment to become a 5-year discrete partition, such as 1900-1904, 1905-1909, until 2015-2018. The value of confidence is considered as the level of earthquake risk in a province. The confidence value is the strong amount of the relationship between earthquake events each provinces. The

risk level is taken based on the greatest confidence value of all the rules found. This research proposes modeling by considering that risk is something that has the greatest value of all values that have been found.

The Province that get the earthquake risk is the consequent or LHS and the province that gives the risk of an earthquake is the antecedent or RHS. The equation below is used to calculate earthquake association rule between antecedent and consequent province:

$$\text{Total Transaction earthquake events} = \frac{\text{total days of each period}}{n} \qquad (5)$$

$$\text{Support}(Y) = \frac{\text{number of transaction containing Y}}{\text{Total transaction earthquake events}} \qquad (6)$$

$$\text{Support }(X, Y) = \frac{\text{number of transaction containing X with Y}}{\text{Total Transaction earthquake events}} \qquad (7)$$

$$\text{Conf}(Y \rightarrow X) = \frac{\text{Support }(X, Y)}{\text{Support}(Y)} = \text{Conf}_{(X \text{ when become consequent})} \qquad (8)$$

where,
n = 7 (days earthquake event will be use as one transaction)
Y = Antecedent province
X = Consequent province

The formula below is used to calculate the risk value in the province of each period based on the value of the largest confidence of the association rule:
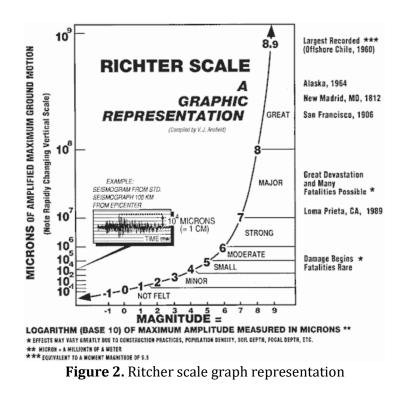
$$\text{Risk}_{(X \text{ in each period})} = \max\left(\text{conf}_{(X \text{ when became consequent})}\right) \qquad (9)$$

And the next formula is to get the risk value of province from the highest or largest value of all period:

$$\text{Risk}_{(X \text{ in all period})} = \max(\text{Risk}_{(X \text{ in each period})}) \qquad (10)$$

In this stage, we categorize the risk of earthquakes of the consequent province based on highest confidence modelings. Categorizing earthquake risk into 3 levels is based on Richter scale graphic representation in Figure 2 [29]. The Richter scale is used to assess the magnitude of an earthquake based on the amount of energy released.

In the Figure 2, the damage caused by magnitude is divided into not felt with magnitudes below 2 SR, minor with magnitudes between 2-4 SR, small with magnitudes between 4-5 SR, moderate with magnitudes between 5-6 SR, strong with magnitudes between 6-7 SR, major with magnitude between 7-8 SR, and great with magnitude above 8 SR.

Figure 2. Ritcher scale graph representation

This study simplifies 7 categories into 3 main risk levels, namely low, moderate and high. We classified the categories of NOT FELT, MINOR and SMALL as listed in Figure 2 as Low-Level of Risk. The MODERATE and STRONG categories are grouped into Moderate Risk. The MAJOR and GREAT categories are classified as High Risk. In this risk modeling, the color displayed on the map is based on the risk level, red means the risk value is ≥ 0.7 with a high level of risk, yellow is between 0.5 and 0.7 with a moderate-risk level and green < 0.5 with a low-risk level.

Table 1. Earthquake risk category based on highest convidence value

Scale	Level of Risk	Confidence/Risk Value
I	Low	x < 0.5
II	Moderate	0.5 ≤ x < 0.7
III	High	x ≥ 0.7

5. EXPERIMENT AND ANALYSIS

In this paper, the stages of implementing the proposed system design were part of the experimental process. Also, the explanation and analysis of the experimental results of Spatial-Temporal Associative Mining visualization based on the highest value of confidence.

5.1 Data Acquisition and Preprocessing

Earthquake dataset from each provider is taken between latitude -11 - 6 and longitude 94 - 142. Researchers obtained several sources data from the Indonesian Meteorology, Climatology and Geophysics Agency (BMKG), United States Geological Survey (USGS), which collected data from the Advanced

National Seismic System (ANSS), European-Mediterranean Seismological Center (EMSC), and International Seismological Center (ISC). Each provider has different characteristics, providers can record time, latitude, longitude, depth and magnitude from one source earthquake in a different way.

Table 2. List of earthquake data providers

Providers	Year Available	Amount of Data
BMKG	October 2008-2018	58,561
USGS/ANSS	1931-2018	66,857
EMSC	October 2004 – 2018	14,823
ISC	1900-2018	186,658
Total		326,899

Therefore, to combine data from each of these providers, we must have our own approach or method. In this study, the process of preprocessing data have several stages.

The first stage, collecting and retrieving data from each earthquake data provider where the location parameters are latitude -11-6 and longitude 94-142. In this process, researchers do it manually and take quite a long time because the form of output from each provider is different. The second stage, data cleaning. At this stage, if the data is incomplete or it does not have a value on the depth and magnitude attribute then it will be discarded.

The third stage is the process of data integration. This process homogenizes the format of each attribute in the dataset. Data taken has attributes date, time, latitude, longitude, depth, and magnitude. The fourth stage, the process of reducing data that has been combined from all sources. Dynamic Time Warping is used to calculate the proximity of data, where data that has the closest distance is considered to have the same earthquake source. In the data group, one data is taken that represents the source of the earthquake. The last stage was adding the province attribute to each data. The entire preprocessing stage reduced almost half of the amount of data from 326,899 to 188,803.

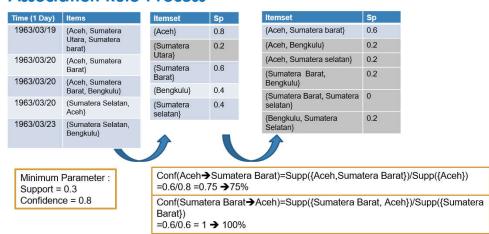**Table 3**. The total data after preprocessing process from 1900-2018

Raw Data	Data Cleaning	Data Reduction
326,899	308,363	188,803

5.2 Associative Mining Process

Earthquake data is converted into earthquake transaction data, this transaction data consists of time (days) as transactions and provinces as an itemset. To get the earthquake association pattern for each province, parameters such as earthquake data parameters are taken, minimum support, minimum confidence, minimum lift and choice of time (transactions) of earthquake events. The Association Rule parameters set for this calculation were minimum support 0.1, minimum confidence 0.1, minimum lift 1 and earthquake event transactions that occur for 7 days. If the transaction is in n-

number of days then the provincial relationship to the earthquake occurred in those n-days.

The algorithm to calculate the frequent itemset is Apriori if subset S appears k times, another subset S' which contains S also appears k times or less. So, if S does not pass with the specified support limit, so does S '. There is no need to count S' again. Subset S here is a subset of provinces that exist in each date transaction. After obtaining subsets that meet the minimum support value, we will calculate the value of confidence between the rules made from these subsets. A rule that is considered valid or strong is where the minimum lift value is equal to one.

Figure 3. Process of Association Rule mining on earthquake transaction data

From Table 4 it is explained that if an earthquake occurs in the antecedent province, the probability value of earthquake will occur in the consequent province is the confidence value. For example, Papua is a consequent province where it gets the risk of an earthquake if the Maluku province had an earthquake before, the risk value is 0.7.

Table 4. Example of provinces earthquake association rule result

Antecedent	Consequent	Confidence
Maluku	Papua	0.70
North Maluku	North Sulawesi	0.56
North Maluku	NTT	0.73
Maluku	NTT	0.88
Maluku	North Maluku,NTT	0.90

5.3 Spatio Temporal Associative Mining Model and Visualization

Calculation of earthquake risk values for each province is taken based on the results of earthquake association rules for each province. These rules provide information on the antecedent, consequent provinces and the value of the province's relationship confidence.

In this research, to visualize the result of association rule from all period and made it become the risk value of the province, we purposed Spatio Temporal visualization. The experiments based on the highest confidence value risk to develop the risk value of the earthquake. The risk level of each province was taken based on the value of the biggest confidence from the antecedent province that gave the consequent province the risk. The experimental results obtained from 1900-1904 to 1960-1964 were not produce an association rule, and besides that period got several rules.

Figure 4. Spatio-Temporal Associative visualization with highest confidence value

Some of the province that has risk from the impact of association earthquake with another province, namely:

a. Aceh

   From the experiments conducted, Aceh has a moderate risk level and it has a seismic relationship with the provinces of North Sumatra, North Maluku, and Maluku. the risk of this province is from relations with North Sumatra emerged in the 2005-2009 trial period with a confidence value of 0.5 with a moderate risk level. The weakest earthquake bond is from Maluku with a confidence value of 0.23.

b. North Sumatera

   North Sumatra Province has a low earthquake risk and it has an earthquake connection with Aceh, Maluku, and North Sulawesi. the highest level of confidence in each province is 0.43 with Aceh in the 2005-2009 trial period, so North Sumatra has a low potential earthquake risk. The lowest level of confidence is Maluku with a confidence value of 0.25.

c. Bengkulu

   Bengkulu has low earthquake risk level. Bengkulu has seismic relations with Maluku that appear 3 times in 2000-2004, 2010-2014, and 2015-2018. North Maluku that appears 3 times too in 1990-1994, 2005-2009, and 2015-2018. North Sulawesi provinces that appear 2 times in 1980-1984 and 2005-2009. Bengkulu province has a low earthquake risk level with the largest confidence value of 0.36 with Maluku where the

rule found in 2000-2004, 2010-2014, and 2015-2018 period. The lowest bond of the earthquake with North Maluku and North Sulawesi with 0.3 confidence value. Association with North Maluku appears in 1990-1994, 2005-2009, and 2015-2018 period. Also, relation with North Sulawesi is found in 1980-1984 and 2005-2009 period.

d. East Java

East Java has a high level of earthquake risk and relation with the provinces of Bali, NTB, Maluku, Papua, and North Sulawesi. East Java has a high level of earthquake risk because the largest confidence value found is 0.8 with Maluku and Bali. If there is more than one antecedent province, it means East Java get the impact of an earthquake when this Maluku and Bali province had occurred the earthquake first in 7 days. However, the appearance of the rule or rule was only one time and dominated in 2000-2004. The weakest confidence value is from Papua with 0.28 and occur in 1995-1999.

e. Bali

Similar to the province of East Java, the province of Bali also has a high risk of an earthquake with the highest confidence value of 0.7 from NTB and NTT. If there is more than one antecedent province, it means Bali get the impact of the earthquake when NTB and NTT province had occurred the earthquake first in the range 7 days. However, the rule that emerged was only one time during the experiment, 2000-2004. This province has a seismic relationship with East Java, NTB, NTT, and Maluku. The weakest bond of earthquake impact with 0.5 value from NTT province.

f. Nusa Tenggara Barat

Related to the previous two provinces, NTB province also has a high level of earthquake risk because the largest confidence value is 0.79, which is greater than 0.7. This greatest value of confidence came from Bali and NTT. If there is more than one predecessor province, that means NTB has an earthquake impact when the Bali and NTT provinces have experienced earthquakes in the range of 7 days. The appearance of the rule in this province was only in 2000-2004. This province has seismic relations with Bali, East Java, NTT, North Sulawesi and Maluku. The lowest confidence value is 0.54 with Maluku and East Java province.

g. Nusa Tenggara Timur

NTT Province also has many association rules with low risk. Although it has seven association rules with the other province, the largest value of confidence is only 0.42 so that the province is considered to have a low level of risk. This value came from North Maluku and Maluku in the 1990-1994 trial period. When there is more than one antecedent province, it means North Maluku and Maluku must occur an earthquake first after that it will give the impact to NTT. NTT has an earthquake relationship with Bali appear in 2000-2004, Bali and NTB appear in 2000-2004, Maluku emerge 5 times in the 1970-1974, 1980-

1984, 1985-1989, 1990-1994, and 2005-2009. North Maluku emerge 5 times too in the 1975-1979, 1980-1984, 1985-1989, 1990-1994, and 2000-2004. NTB appear in 2000-2004 period. Where North Sulawesi give the lowest confidence 0.27 in 2005-2009 period.

h. Gorontalo

Gorontalo Province only emerged the rule from 1990-1994 where it was associated with North Maluku province with a small confidence level of only 0.19.

i. West Papua

West Papua Province also has only one rule that emerged in 1990-1994 which was related to Maluku with a confidence level of 0.16. A very low level of confidence makes this province has a low level of earthquake risk.

j. Papua

Papua has a moderate risk level. This province also has quite a lot of earthquake relations with other provinces. The level of confidence is varied, but the more frequent associations with a province appear, the greater the confidence value does not reach 0.5. Papua has a seismic relationship with the provinces of East Java, Maluku, North Maluku, and North Sulawesi. The biggest value of confidence that arises for this province is the relationship with the East Java province with a value of 0.62 which appeared once in 1995-1999 so that it has a moderate risk level. North Sulawesi give the weakest bond of an earthquake with confidence value 0.33 between one of these 1970-1974 and 1990-1994 period. Where Maluku appear 7 times in 1970-1974, 1980-1984, 1990-1994, 1995-1999, 2005-2009, 2010-2014, and 2015-2018 with confidence value 0.46. North Maluku appear 4 times in 1965-1969, 1970-1974, 1980-1984, and 1990-1994 period with confidence value 0.37.

k. Maluku

Maluku has a moderate risk level. Maluku Province also includes a list of provinces that have a lot of rule appearances. The province has a seismic risk relationship with Aceh, Bengkulu, East Java, Bali, North Maluku, NTT, Papua, West Papua, and North Sumatra. The biggest confidence value of this province is 0.68, which is related to the earthquake in West Papua province. The most significant relationship occurred 7 times with North Maluku province with confidence 0.61 and Papua with confidence 0.67. Relations with the NTT province appeared 5 times with the largest confidence value of 0.66. Relations with Bengkulu province were found 3 times with the highest confidence value of 0.53. Maluku Province has a moderate level of earthquake risk with the largest confidence value of 0.68 from West Papua and the lowest confidence value from East Java with confidence value 0.39.

l. North Maluku

North Maluku Province has a high seismic risk level, also has many relations with the provinces of Aceh, Bengkulu, Gorontalo, Maluku, NTT,

Papua and North Sulawesi. This province has a seismic relationship with Bengkulu with a rule appearance of 3 times and the largest confidence value of 0.56. The relationship between earthquake and Maluku emerged 7 times with the highest confidence value of 0.54. Seismic relations with NTT appeared 5 times with the highest confidence value of 0.56. Papua gave the province earthquake risk 4 times with the highest confidence value of 0.63. With Aceh in 2005-2009 where confidence value 0.44. Finally, the North Sulawesi rule was found 5 times with the greatest confidence value of 0.54. North Maluku Province has a high seismic risk level where the value of the largest antecedent confidence is 0.72 from Gorontalo. This province has a relationship with more than one antecedent province such as Maluku and NTT than appear in 1990-1994 with the value 0.56, along with Maluku and Papua emerge in 1985-1989 and 1990-1994 with value 0.61.

m. North Sulawesi

North Sulawesi Province has a moderate level of earthquake risk. This province has seismic relations with Bengkulu, East Java, NTB, NTT, North Maluku, Papua, and North Sumatra. Relationships that often arise 5 times, namely with North Maluku province where the value of the biggest confidence is 0.47. After that followed by Bengkulu with 2 times occurrence and the biggest confidence value of 0.5. Although this province has links with many provinces, the value of the greatest confidence is only 0.5. so, the risk level of this province is a moderate level of risk.

5.4 Earthquake Risk in Indonesia

Mapping the earthquake risk based on earthquake associations between each province can be done if the earthquake frequency was increasing or usually the province has a high frequency of earthquake events. In this paper, researchers found 13 provinces in Indonesia had earthquake association risk. Examples of provinces where seismic associations have been found because seismic activities of the province were increasing in certain periods are East Java, Bali, NTB, Aceh, North Sumatra, West Papua, and Gorontalo. While the provinces of Maluku, North Maluku, NTT, North Sulawesi and Papua had a high frequency of earthquake events because they located on active plates. For Bengkulu province, the seismic activity depends on some active period and in the recent period had a high frequency of seismicity.

Of all the related works mentioned in this paper, only Badan Nasional Penanggulangan Bencana (BNPB- National Disaster Management Agency) and Minister for Public Works and Human Settlements Indonesia [17][28] has formed earthquake risk mapping in all regions of Indonesia based on the physical condition of the earthquake event. Fig. 4 was Indonesia earthquake map 2017 and it will be use for next 500 years. Earthquake risk in this map was not based on provinces but areas that include some provinces. From all the results of seismic risk mapping conducted by BNPB, it can be seen that the

west coast of Sumatra, south coast of Java, areas around NTT, Maluku, North Maluku, Papua, West Papua, North Sulawesi have high hazard values with more orange colors on the map.

Sumatra had a complex tectonic structure and is considered to be one of the most active earthquake zones on earth, which has 55 active faults sources. Java had 37 active fault sources and is dominated by earthquakes related to subduction of faults on land and calmer seismicity compared to Sumatra. Sulawesi Island was composed of an active geological structure that moves at different speeds and has 48 earthquake sources. The Maluku sea area was a unique example of the closure of an ocean basin as a result of the collision of two oceanic arcs, the Halmahera and Sangihe arcs, the results of the collisions of these two arcs make the Maluku and North Maluku regions have many active faults. The island of Papua had complex geological conditions as a result of interactions between the Australian and Carolina tectonic plates. Maluku and Papua locations had 79 sources of earthquake. An additional area group is Nusa Tenggara - the Banda Sea which is known to have 75 earthquake sources.
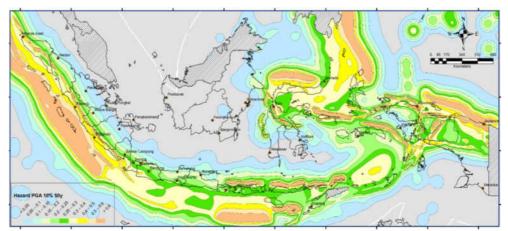

Figure 5. Indonesia Earthquake Hazard Map 2017 (for 500 yrs)

The results of earthquake risk mapping based on seismic associations between provinces in Indonesia had some similarities and differences risk with the mapping of earthquake risk based on the physical characteristics of the earthquake.

Provinces in the Sumatra region that were found to have risks were Aceh (moderate), North Sumatra (low) and Bengkulu (low) but on Indonesia Earthquake Hazard Map, had a high-level earthquake risk in the sea area. Based on this paper, East Java, Bali, and NTB had a high risk because they had strong earthquake connections with each other. NTT province had low level of risk. Meanwhile, according to the Indonesia earthquake hazard map, the Java island region had a moderate risk but Nusa Tenggara - Banda Sea region included had a moderate to a high level of earthquake risk in the earthquake hazard map. In the eastern part of Java, there are three faults (Pasuruan, Probolinggo, and Baluran Faults), but the activity of this fault was found at least 6 times in the past 400 years.

In the Maluku islands, Maluku province had a moderate risk but it is different from an earthquake hazard map that had a high-level risk. North Maluku had the same high-level risk as to the Indonesia earthquake hazard map. North Sulawesi Province had a moderate risk where the earthquake hazard map has a higher risk than the results of the study. Gorontalo province had low risk where in the earthquake hazard map had high risk. Papua Province had a moderate risk and West Papua had low-risk where the earthquake hazard map in this region had a low to high level of risk.

6. CONCLUSION

Several provinces in Indonesia were found to have seismic relations with one another. This relationship or association can be used as a basis for calculating seismic risk from a province. This paper presents a new approach to modeling and visualizes the risk of an earthquake from each province in Indonesia based on the association rule. From the experiment using the highest confidence value risk modeling, it was concluded the experiments from 1900-2018 with the divided of 5 years trial period was found 13 provinces from the total of 34 provinces in Indonesia that had earthquake association risk. The total of low-risk provinces is 5 or 38.5% based on the total provinces had association risk, the moderate-risk is 4 or 30.8%, and the high-risk is also 4 or 30.8%.

Low-level risk provinces are North Sumatra, Bengkulu, NTT, Gorontalo and West Papua. The highest value of confidence in the province is below 0.5. The moderate level risk that has the confidence value between $0.5 \le conf < 0.7$: (1) Aceh, Papua, and North Sulawesi, (2) Maluku is included in this category, where the emergence of rule and seismic relations are many with other provinces. The time span of the occurrence of relations with other provinces also reached 7 times. The high-level risk provinces are : (1) East Java, Bali, and NTB with the largest confidence value of ≥ 0.7, but these three provinces only have associations just between these three with the emergence of the rule in 2000-2004. (2)North Maluku, this province has the highest confidence value of 0.72 and had a seismic relationship with some provinces and the rules was found in several years.

This risk modeling has a weakness, where this modeling only focuses on the greatest value of confidence from the relationship between earthquake provinces but does not take into how many provinces that provide risk and appearance of rules in the province. So that why this modeling is unfair for risk modeling, for example, the provinces of East Java, Bali, and NTB are considered to have a high level of earthquake risk which has the highest confidence value ≥ 0.7, but only has one rule appearance in each of its provincial relations trial period. Maluku Province which has a moderate risk level with the largest confidence value of 0.68 but has many seismic relationships with other provinces where the rule of appearance can be up to 7 times.

The comparison with the earthquake hazard map made by BNPB turned out that all the provinces discovered in this paper had an earthquake risk but

have some similarities and different levels of risk. The provinces that had the same level of risk as earthquake hazard maps are Bali, NTB, North Maluku, Papua, and West Papua. The provinces that have an earthquake risk level lower than earthquake hazard map are Aceh, North Sumatera, Bengkulu, NTT, Maluku, North Sulawesi, and Gorontalo. East Java is the one province that has a higher level of risk than earthquake hazard map.

From this risk-mapping, we know how the relationship earthquake between province in Indonesia has important information about earthquake characteristic of each province. We can find out which provinces always have an earthquake impact on other provinces and conduct preventive efforts to reduce these impacts in the consequent province.

REFERENCES

[1]   Sunarjo, Muhammad Taufik Gunawan dan Sugeng Pribadi, **Gempabumi Edisi Populer**, *Badan Meteorologi Klimatologi dan Geofisika (Jakarta),* 2012.

[2]   Ryana Aryadita Umasugi, "**Selama 2018, Gempa di Indonesia Meningkat 4.648 Kali Dibanding 2017**," 29 December 2018. [Online].   Available: https://megapolitan.kompas.com/read/2018/12/29/10303711/sel ama-2018-gempa-di-indonesia-meningkat-4648-kali-dibanding-2017. [Accessed 15 July 2019].

[3]   Dewi Divianta, "**Gempa Banten Akan Pengaruhi Lempeng Lainnya, Seberapa Bahaya?**," Liputan 6, 24 Januari 2018. [Online]. Available: https://www.liputan6.com/regional/read/3236388/gempa-banten-akan-pengaruhi-lempeng-lainnya-seberapa-bahaya.   [Accessed   10 April 2019].

[4]   Aqdas Ikram, Usman Qamar, **Developping an expert system based on association rules and predicate logic for earthquake prediction**, *Knowledge-Based Systems,* Vol. 75, No. C, pp. 87-103, 2015.

[5]   Bin Hu, Ya-Fei Peng, Yuanyuan An, Rui-Yi Li, Guo-Ming Lu, **Association analysis on casualty and residential environment data of earthquake**, *International Conference on Wavelet Active Media Technology and Information Processing (ICWAMTIP)*, Chengdu, pp. 308-313, 2012.

[6]   Wang Xi, Bo Xiao, Lin Zhi-qing, Chen Hao, **An empirical research of critical incident of earthquake disaster based on credible Association mining**, *International Conference on Management Science & Engineering*, Rome, pp. 1345-1352, 2011.

[7]   Yuan Zhou, Lianxiong Gao, **An Apriori Based Algorithm Associated Point Line Pattern Applied in Seismic Spatial Data**, *International Conference on Artificial Intelligence: Technologies and Applications*, Bangkok, pp. 187-190, 2016.

[8]   Jin A Lee, JongGyu Han, Kwang Hoon Chi, **Mining quantitative association rule of earthquake data**, *International Conference on Hybrid Information Technology*, Daejeon, pp. 349-352, 2009.

[9]   Zhong Li, Jianqin An, Huichao Yin, Yutang Tian, Weihao Yu, **Study on Association Rules Between Earthquake Event and Earthquake Precursory Information Anomalies**, *International Congress on Image and Signal Processing, BioMedical Engineering and Informatics,* Beijing, pp. 1-6, 2018.

[10]   Amin Endah Suliswati, Ali Ridho Barakbah, Tri Harsono and Yuliana Setyowati, **Earthquake density measurement using Automatic Clustering**, *Knowledge Creation dan Intelligent Computing (KCIC)*, Malang, pp. 102-110, 2014.

[11]   Ali Ridho Barakbah, Tri Harsono and Amang Sudarsono, **Automatic Cluster-oriented Seismicity Prediction Analysis of Earthquake Data Distribution in Indonesia**, *International Journal on Advanced Science Information Technology,* Vol. 9, No. 2, pp. 587-593, 2019.

[12]   Muhammad Nur Shoddiq, Dedy Hidayat Kusuma, Mirza Ghulam Rifqi, Ali Ridho Barakbah, Tri Harsono, **Neural Network for Earthquake Prediction Based on Automatic Clustering in Indonesia**, *International Journal of Informatics Visualization,* Vol. 2, No. 1, pp. 38-43, 2018.

[13]   Ken-ichi Fukui, Daiki Inaba, Masayuki Numao, **Discovery of Damage Patterns in Fuel Cell and Earthquake Occurrence Patterns by Co-Occurring Cluster Mining**, *Twenty-Eighth AAAI Conference on Artificial Intelligence,* Québec, pp. 19-26, 2014.

[14]   Evaldas Luksys, Eleana Asimakopoulou, Nik Bessis, **Development of tools for data analysis of earthquakes**, *International Conference on Intelligent Networking and Collaborative Systems*, Salerno, pp. 406-410, 2014.

[15]   Arna Fariza, Nanda Pratyaksa Abhimata, Jauari Akhmad Nur Hasim, **Earthquake Disaster Risk Map in East Java, Indonesia, using Analytical Hierarchy Process - Natural Break Classification**, *Knowledge Creation and Intelligent Computing (KCIC)*, Manado, pp. 141-147, 2016.

[16]   Man Li, Zhenhua Zou, Guodong Xu, and Peijun Shi, **Mapping Earthquake Risk of the World**, *Journal of World Atlas of Natural Disaster Risk, IHDP/Future Earth-Integrated Risk Governance Project Series*, pp. 25-39, 2015.

[17]   Lilik Kurniawan, Bernardus Wisnu Widjaja and Willem Rampangilei, **Risiko Bencana Indonesia (RBI)**, *Badan Nasional Penanggulangan Bencana (Jakarta)*, Ed.1, pp. 38, 53-61, 2016.

[18]   Daniela Faur, Mihai Datcu, **A Rapid Mapping Approach to Quantify Damages Caused by the 2003 Bam Earthquake Using High Resolution Multitemporal Optical Images**, *International Workshop*

on the Analysis of Multitemporal Remote Sensing Images, Annecy, pp. 1-4, 2015.

[19] Junxiang Zhang**, A fuzzy approach for reforming earthquake map**, *International Conference of the North American Fuzzy Information Processing Society*, Chicago, pp. 383-388, 2003.

[20] F. Benjamin Zhan, Zhongliang Cai, Yiqing Zhu, Jiangcun Zhou, **A geospatial analytical system for mapping global medium-term earthquake probabilities**, *International Conference on Systems and Informatics,* Yantai, pp. 2276-2279. 2012.

[21] Xiang Ding, Xiaoqing Wang, Long Wang, Youhua Zheng, **The development of catastrophe earthquake risk estimation system based on GIS**, *IEEE International Geoscience and Remote Sensing Symposium,* Vancouver, pp. 4352-4354, 2011.

[22] Qing-Quan Tan, Qun Liu, Hua-Chun Luo, Bo Liu, Jian Liu, **Design and implementation of building information collection system for earthquake disaster scenario construction based on GIS**, *International Conference on Computer Science and Network Technology*, Dalian, pp. 51-54, 2017.

[23] Meinard Muller, **Information Retrieval for Music and Motion**, *Springer(Germany),* Vol. XVI, pp. 69-74, 2007.

[24] Indra Budi, Stéphane Bressan, Nasrullah, **Co-Reference Resolution for Indonesian Languange Using Association rules**, *The Eighth International Conference on Information Integration and Web-based Applications Services*, Yogyakarta, pp. 117-126, 2006.

[25] Jiawei Han, Micheline Kamber and Jian Pei, **Data Mining: Concepts and Techniques**, *Morgan Kaufmann (Waltham)*, Ed.3, pp. 243-256, 2000.

[26] Roberto J. Bayardo Jr., Rakesh Agrawal, **Mining the Most Interesting Rule**, *KDD '99 Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, San Diego, pp. 145-154, 1999.

[27] Rakesh Agrawal, Ramakrishnan Srikant, **Fast Algorithms for Mining Association Rule,** *IBM Almaden Research Center*, California, pp. 1-32, 1994.

[28] Tim Pusat Studi Gempa Nasional, **Peta Bahaya dan Sumber Gempa Indonesia Tahun 2017**, *Kementrian Pekerjaan Umum dan Perumahan Rakyat (Bandung)*, 2017.

[29] John P. Rafferty, "**Richter scale,**" 27 September 2017. [Online]. Available: https://www.britannica.com/science/Richter-scale. [Accessed 21 Juni 2019].