OPEN ACCESS
Scholars Bulletin
Abbreviated Key Title: Sch Bull
ISSN 2412-9771 (Print) |ISSN 2412-897X (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: http://scholarsbulletin.com/

Subject Category: Computer Science

Clustering of Earthquake Data Using Kohonen Self Organizing Maps (SOMs) Algorithm

Herry Derajad Wijaya*, Saruni Dwiasnati

Faculty of Computer Science, Mercu Buana University, Jl. Meruya Selatan No.1, RT.4/RW.1, Meruya Sel., Kec. Kembangan, Kota Jakarta Barat, Daerah Khusus Ibukota Jakarta 11650, Indonesia

DOI:10.21276/sb.2019.5.7.11 | Received**:** 17.07.2019 | **Accepted:** 25.07.2019 | **Published:** 30.07.2019

***Corresponding author:** Herry Derajad Wijaya

Abstract

An earthquake is the result of a sudden release of stored energy n the Earth's crust that creates seismic waves. In term of the earthquake, this study aimed to cluster which areas were the most affected by earthquake occured in Java Province in 2017. The algorithm used in this study was Self Organizing Maps algorithm (SOMs) with Cohonen as a type of Artificial Neural Networks (ANN) that is trained using unsupervised learning in decision making. In addition, the clustering results through its algoritm are functioned as a base of determining the eartquake pattern criteria and which areas often occurred in order to be able to mitigate earthquake that causes fatal impacts.

Keywords: Earthquake, Clustering, Self Organizing Map, Artificial Neural Networks.

INTRODUCTION

Indonesia is one of the countries in the Asian continent that has the largest islands throughout 3200 miles (5,120 km2) which consists of 13,667 major and minor islands. Astronomically, Indonesia archipelago located in the Latitude 07o LU - 12o LS and Longitude 95o BT - 141o BT. It may potentially cause an earthquake disaster. The number of earthquake disaster in Indonesia is quite a lot compared to other countries after Japan. An earthquake is the result of a sudden release of stored energy n the Earth's crust that creates seismic waves. Unfortunately, earthquakes cannot be definitely detected by existing technologies, but those can be accordingly measured by a seismometer, commonly known as a seismograph using richter standard scale. If it values below or approximately 3, it will be hard to recognize. But if the magnitude has reached more than 7 on the Richter scale, it will potentially cause critical damage.

The causes of eartquakes are tectonics, volcanic activity, meteors, avalanches or nuclear explosion. The earthquake caused by mountain activity called the volcanic earthquake while the Earthquake that occurs due to plate movement is called the tectonic earthquake. Eartquake in Indonesia is relatively occured every day with magnitude of richter scale 1-3 which known as minor earthquake. So, it needs to be identified which areas are affected by the impact of the disaster and how depth those earthquakes need to be classified. Indeed, each category of earthquake has a magnitude scale which can be visualized by a *cluster.*

Clustering method is a method used to analyze non-directive/unsupervised data as one of data mining patterns to classify data with similar characteristics to similar region and data with different characteristics to different region too. If all these regions can be combined, the entry data entered into an appropriate group. Clustering method has two approaches; partition-based clustering and hierarchical clustering. Partition-based clustering classifies data by separating the data to be analyzed into available clusters. Meanwhile, hierarchical clustering classifies data by creating a hierarchy (dendogram) where the data that has similarities will be placed into a close hierarchy. Besides, there is also a clustering approach with the automatic mapping (Self-Organizing Map/SOM) either with or without Kohonen-based. Clustering with partition approach consists of several algorithms namely K-Means and Mixture Modeling (Mixture Modeling). Meanwhile, clustering with the Hierarchy Approach consists of agglomerative hierarchical clustering and divisive hierarchical clustering. Agglomerative performs a clustering process from N clusters into one cluster unit, in which N is the amount of data while divisive processes on the reverse from one cluster to N cluster.

Cluster analysis used in this research paper is Kohonen Self Organizing Maps (SOMs). SOM is an efficient algorithm in visualizing data by reducing its dimensions from n-dimensional input to a lower dimension while maintaining its original topology relationship. In addition, SOM is a non-parametric approach that does not require an assumption of population distribution. In this study, it will be clustered all earthquake disasters occurred in Java province in 2017 then identified which areas are vulnerable to earthquake. Based on these data, the patterns and characteristics of the earthquake that occurred during one year were obtained. The data will be imported from Excel to the Rapid Miner application. The purpose of grouping earthquake data is to mitigate earthquakes, so that, it does not have an impact after a disaster or at least reducing the risk of victims. The indicators used are four, namely the strength of the earthquake, the depth of the epicenter, the duration of the earthquake and the magnitude. In short, this study was to find out which areas were the most affected by the earthquake disaster in Java Province in 2017. Therefore, the society would be more cautious in the following year.

LITERATURE STUDY

The term of data mining was known since 1990 as one of the stages in Knowledge Discovery in Databases (KDD). Knowledge discovery is a process that has two stages, namely KDD (Knowledge Discovery in Databases) and CRISP-DM. The process consists of data cleaning, data integration, data selection, data transformation, data mining, pattern evaluation and knowledge presentation [1]. The model which is produced from the data transformation process will be used by data analysis to publish new knowledge by using statistical analysis, machine learning and information visualization.

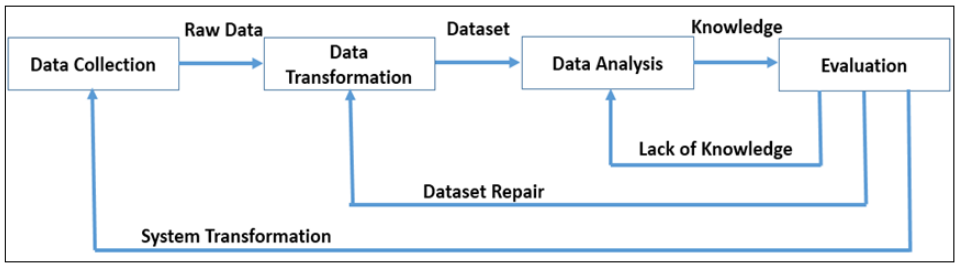

Fig-1: Information Flow in Data Mining

Clustering is a process to place a set of data records into a group (cluster), so that, in one cluster consists of a data record with the same characteristics and different clusters [2]. Cluster Validity, is a testing evaluation process that provides list of performance criteria based on cluster centroids.The stages of the clustering process is divided into 5 stages [2]:

1. Pattern representation
2. Measuring defined differences
3. Clustering
4. Data abstraction
5. Output Assessment

Self Organizing Map (SOM) with Kohonen or called topology-preserving map is one of clustering method that is most often used in any field and is one of the reliable visualization tools to project complex relationships from a large amount of data into less data. The SOM architecture consists of two layers, namely the input layer and the output layer. Each neuron in the input layer is connected to each neuron in the output layer. Then, each neuron in the output layer represents a class (cluster) of the input given. In line with that, this study aimed how to classify the data that will be identified where the vulnerable place to earthquake in Java region. The following are the stages of the Self Organizing Map algorithm with Kohonen [3, 4]:

1. Initialize randomly W_{ji} with ventor input code $x_1, x_2, x_3, \ldots x_n$.
2. Initialize neuron output with amount y1, y2, y3, ... yn.
3. Determine the weight of the output neurons with the value between the initial understanding rate and the decrease factor.
4. Determine the index J to D (j) minimum.
5. Repeat the 6 to 9 steps until there is no weight value update or has reached the condition of the Min error.
6. Random selection of one data derived from the input vector as data training.
7. Find the closest distance from each output neuron to the input neuron using the euclidian distance formula.

$$D(j) = \sum (W_{ji} - X_i)^2 \quad \ldots\ldots\ldots\ldots (1)$$

8. Of all the weights (D_j), found the closest distance with another distance, the result of the weight (D_j) is called winning neurons.
9. Each modified w_{ji} use the formula with the equation as follows:

$$W_{ji}^{baru} = W_{ji}^{lama} + \propto (X_i - W_{ji}^{lama}) \quad \ldots\ldots\ldots\ldots\ldots (2)$$

10. Data that has been modified then updates the bias weight (error).
11. Save values that have convergent results. The architecture of SOM can be seen in the following figure:

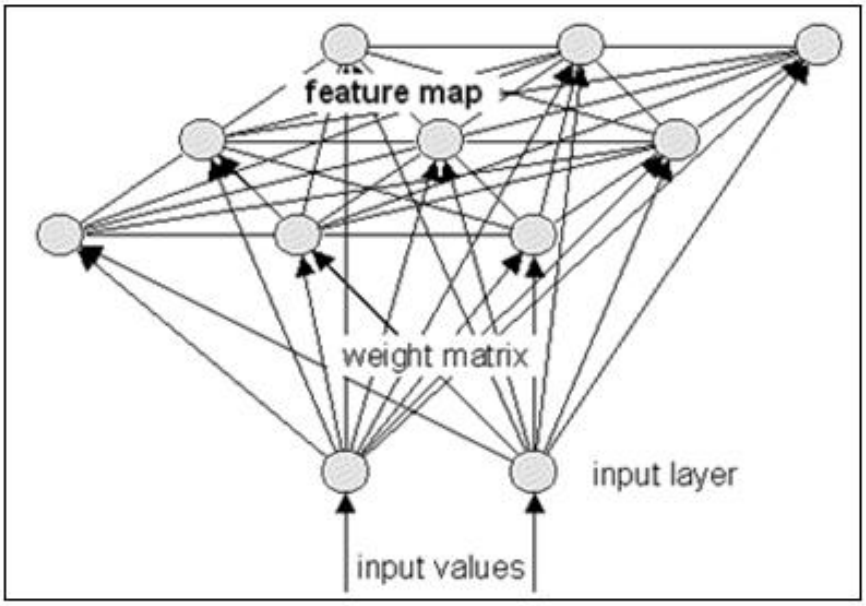

Fig-2: SOM Architecture

Rapid Miner is a software with a GUI display (Graphical User Interface) found by Dr. Markus Hofmann from the Institute of Technology Blanchardstown and Ralf Klinkenberg. Rapid Miner is an open-source software that functions to analyze big data into data mining, text mining or analyzing various cases to predict a decision. It provides a GUI (Graphic User Interface) to create an analytical pipeline pattern. The data is then compiled using Excel in order to easily read by Rapid Miner to run the analysis automatically and provide the required results. Rapid Miner has several advantages as following:

1. The syntax uses Java programming language, so it enables to run existing various operating systems.
2. The process of compiled data is made as a model to produce trees operators.
3. Application of internal XML is to ensure the standard format of required data changes.
4. Multi-layer design is able to bring up the appropriate data display in data processing.
5. Able to process large data with data mining algorithms such as decision tree, self-organization map and support vector machine.
6. The graphic forms generated in this application are histogram diagrams, tree charts and 3D Scatter plots.
7. Many variations of plugins such as text plugins to perform text analysis.

Mean Square Error (MSE) is a formula to evaluate cluster results, to measure the error rate by calculating the number of squares of the distance of the input vector against winning neurons which divided by the number of weights specified. The formula as follows [5]:

$$MSE = \sum_{t=1} (\frac{y - y^t}{n})^2 \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

MSE functions to measure how many error occurs in initializing of weights randomly during the training process, where the instability of the data position will affect the level of convergence. The smaller the value of the MSE, the more convergent the level.

METHODOLOGY

This study processed the secondary data derived from the official website of the Meteorology, Climatology and Geophysics Agency (BMKG) *http://bmkg.go.id*. The methodology is how to implement the Kohonen Self Organizing Map algorithm (SOMs) to find out earthquake areas in Java Province in 2017. There are four indicators namely the strength of the earthquake, the depth of the epicenter, the duration of the earthquake and the magnitude. The stages of the methodology in this study are explained in the scheme below:

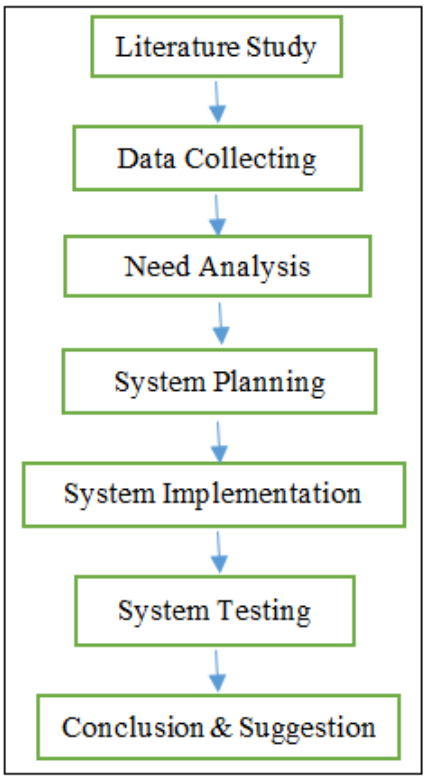

Fig-3: Stages of Research Methodology

FINDINGS

Based on the results of data analysis using the algorithm of Kohonen Self Organizing Maps (SOM), it was found that Earthquake data can be used to provide information on which the most areas that easily affected by earthquake disaster as the information in the following year.

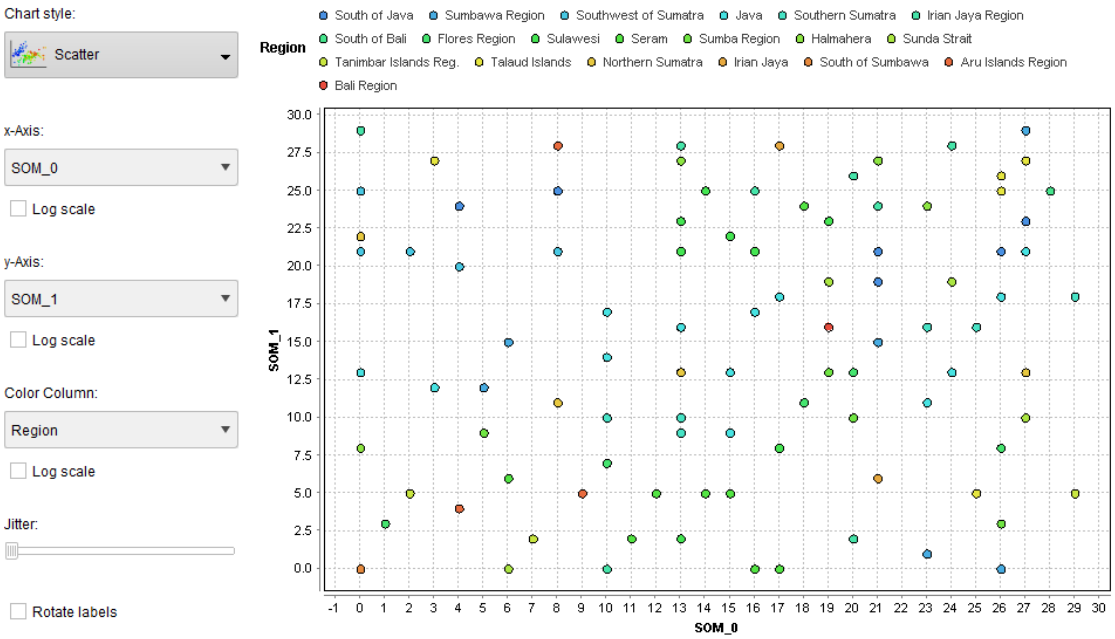

Fig-4: Model of Kohonen Self Organizing Maps (SOMs)

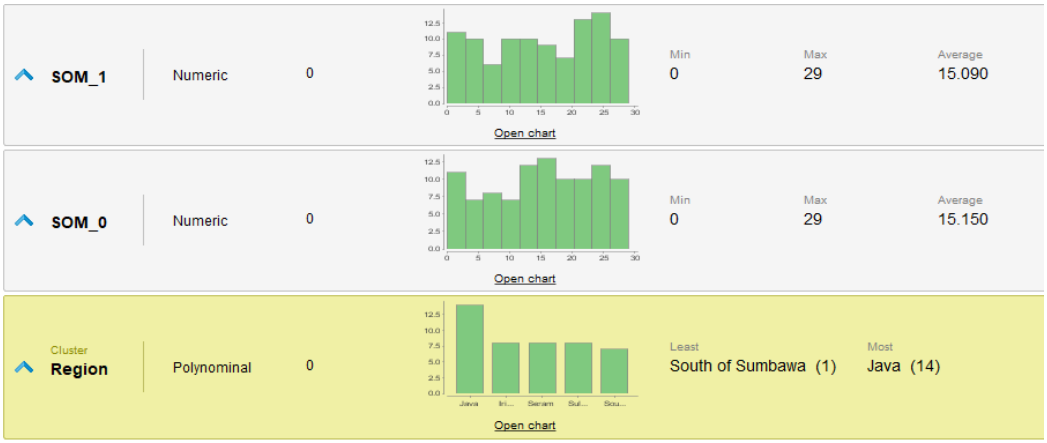

Fig-5: Graph of Kohonen Self Organizing Maps (SOMs)

CONCLUSION

The algorithm of Kohonen Self Organizing Maps (SOMs) can be functioned as the use of Earthquake data in finding which areas are the most affected by its disaster. Besides, this algorithm is able to perform the other disaster data such as floods, volcanic eruptions, tornadoes, etc.

REFERENCES

1. Firdaus, D. (2017). Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer. *Format*, *6*(2), 91-97.
2. Pandian, P. (2013). Multi-objective programming approach for fuzzy linear programming problems. *Applied Mathematical Sciences*, *7*(37), 1811-1817.
3. Terkait, P. P. (2011). Sistem Clustering Kecerdasan Majemuk Mahasiswa Menggunakan Algoritma Self Organizing Maps (SOM). STMIK Duta Bangsa. Surakarta.
4. Sote, A. M., & Pande, S. R. (2015). Web page clustering using self-organizing map. *International Journal of Computer Science and Mobile Computing*, *4*(1), 78-84.
5. Cahyani, A. D., Khotimah, B. K., & Rizkillah, R. T. (2014). Perbandingan Metode SOM (Self Organizing Map) Dengan Pembobotan Berbasis RBF (Radial Basis Function). *Jurnal Teknologi Technoscientia*, *7*(1), 85-92.