Journal Of Data Science
Volume 3, Number 02, 2025, DOI 10.58471/Science.v3i02
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

Ground Acceleration Clustering Using Self-Organizing Map Method

Siska Simamora[1], Amran Manalu[2], Paska Marto Hasugian[3]
[1,2]Fakultas Teknologi dan Bisnis, Universitas Putra Abadi Langkat, [3]Fakultas Ilmu Komputer, Universitas Katolik Santo Thomas

Article Info
Keywords:
Peak Ground Acceleration (PGA),
Ground Acceleration Clustering,
Self-Organizing Map (SOM)

ABSTRACT
Peak Ground Acceleration (PGA) is an important parameter in seismic studies because it is directly related to the level of shaking felt on the earth's surface. Analysis of ground acceleration data is needed to identify patterns, group regions based on their seismic characteristics, and support earthquake disaster mitigation efforts. This study uses the Self-Organizing Map (SOM) method, which is an unsupervised learning approach based on artificial neural networks that can map high-dimensional data into a two-dimensional map representation without losing its topological structure. The ground acceleration dataset used in this study consists of key seismic parameters such as depth, magnitude, source distance, and PGA values. The SOM learning process is carried out iteratively to produce a cluster map that groups earthquake data into several groups with different ground acceleration characteristics. The results show that the SOM method is able to identify ground acceleration distribution patterns more clearly than conventional methods, by producing clusters that represent variations in PGA from low to high. These findings can provide important contributions to earthquake risk mapping, regional spatial planning, and the formulation of more accurate disaster mitigation strategies.

Corresponding Author:
Siska Simamora
Fakultas Teknologi dan Bisnis, Universitas Putra Abadi Langkat

INTRODUCTION

Peak Ground Acceleration (PGA) is a crucial parameter in seismic studies because it directly relates to the level of shaking felt on the Earth's surface. Analysis of ground acceleration data is essential for identifying patterns, categorizing regions based on their seismic characteristics, and supporting earthquake disaster mitigation efforts. This study employed the Self-Organizing Map (SOM) method, an unsupervised learning approach based on artificial neural networks capable of mapping high-dimensional data into a two-dimensional map representation without losing its topological structure. The ground acceleration dataset used in this study comprises key seismic parameters such as depth, magnitude, source distance, and PGA values. The SOM learning process is carried out iteratively to generate cluster maps that group earthquake data into several groups with different ground acceleration characteristics. The results show that the SOM method is able to identify ground acceleration distribution patterns more clearly than conventional methods, by producing clusters that represent variations in PGA from low to high. These findings can provide important contributions to earthquake risk mapping, regional spatial planning, and the formulation of more accurate disaster mitigation strategies.

Earthquakes are a natural disaster with the potential for significant damage to infrastructure, the environment, and human safety. The resulting tremors often result in material losses and loss of life, making understanding earthquake characteristics crucial. One of the key parameters in seismological studies is Peak Ground Acceleration (PGA). PGA represents the intensity of vibrations occurring on the ground surface due to seismic activity, and its value significantly determines the extent of the impact felt by buildings and humans. In the context of disaster mitigation, PGA serves as a key indicator in determining earthquake-resistant building design standards, determining disaster-prone zones, and developing safe spatial planning policies. The higher the PGA value, the greater the potential for structural damage. Therefore, an in-depth analysis of the distribution of ground acceleration values in various regions is essential to understand the distribution of earthquake shaking and the potential risks it may pose. This study also plays a crucial role in assisting governments, researchers, and civil engineering practitioners in formulating more effective, data-driven mitigation strategies, thereby minimizing the impact of earthquakes.

Earthquake data analysis methods are generally conducted using statistical approaches or deterministic modeling using mathematical formulas based on seismological theory. However, ground acceleration data is highly complex and variable. This complexity arises because PGA values are influenced by many factors, such as earthquake magnitude, source depth, distance from the epicenter, and local geological conditions that vary across regions. This variation results in data with a non-uniform distribution, making it difficult to analyze using traditional approaches alone. In such circumstances, more adaptive analysis methods are needed to accurately identify data distribution patterns. One approach widely used in modern research is clustering, a method of grouping data based on the degree of similarity in characteristics between samples. With clustering, large and complex data can be simplified into several groups with similar characteristics, thus facilitating interpretation and further analysis. This approach is particularly relevant in the context of ground acceleration studies, as it allows for mapping of areas with similar shaking levels and provides an initial overview of zones with high seismic risk that require more attention.

This study uses the Self-Organizing Map (SOM) method, an unsupervised neural network algorithm. SOM is designed to map high-dimensional data into a simpler two-dimensional map representation while maintaining its topological structure. This means that data with similar characteristics will be grouped close together on the resulting map, while dissimilar data will be placed in more distant areas. The main advantage of SOM over traditional clustering methods is its ability to reveal non-linear and complex patterns in the data, making it highly suitable for analyzing ground acceleration phenomena influenced by many variables. Thus, SOM functions not only as a clustering method but also as a visualization tool that provides an intuitive picture of the relationships between data. Through two-dimensional mapping, the distribution pattern of ground acceleration can be seen more clearly, facilitating the identification of areas with similar potential for shaking, and helping to distinguish groups of earthquakes with extreme characteristics from those with general characteristics.

The main objective of this research is to cluster ground acceleration data using the SOM method by utilizing the main seismic parameters, namely magnitude, depth, source distance, and PGA. Through this process, it is hoped that a clustering map will be obtained that provides a comprehensive picture of ground acceleration variations in various regions. The results of this research are expected to provide an important contribution to the development of seismic risk studies in Indonesia, especially in the formulation of data-based disaster mitigation policies. In addition, the results of clustering with SOM can also be the basis for regional spatial planning that considers the level of seismic vulnerability, so that infrastructure development can be carried out more safely. Thus, this research not only provides theoretical contributions in the fields of computer science and seismology, but also practical contributions for policymakers and development planners. Ultimately, SOM-based ground acceleration mapping is expected to improve understanding of the spatial distribution of potential earthquake shaking, as well as become a strategic step in reducing the impact of losses caused by earthquake disasters.

RESEARCH METHODS

This study uses an unsupervised learning approach with the Self-Organizing Map (SOM) algorithm to cluster ground acceleration data. SOM was chosen because of its ability to project high-dimensional data onto a two-dimensional map while maintaining the topological structure between the data. Thus, data with similar characteristics are grouped close together, while dissimilar data are separated further apart.

Research Data

The dataset used in this study consists of 1,457 earthquake data obtained from official seismic records. Each data set includes key parameters affecting ground acceleration, namely:

Magnitude (Mag): the strength of an earthquake.

Depth: the depth of the earthquake epicenter from the earth's surface.

Source Distance (Dmin): the closest distance between the earthquake epicenter and the recording station.

Peak Ground Acceleration (PGA): the maximum recorded ground acceleration.

All data first goes through a preprocessing stage, which includes:

Data cleaning to remove duplicate or anomalous data.

Data normalization uses the min-max scaling method so that all variables are in the range [0,1], so that each variable has a balanced contribution in the SOM learning process.

SOM Process

The SOM implementation stage is carried out with the following steps:

1. Initialize neuron weights on a two-dimensional map consisting of a grid of a certain size. The initial weights are chosen randomly.
2. Calculation of the Euclidean distance between the input vectors (magnitude, depth, source distance, and PGA) with the weight of each neuron in the map.
3. Identify the Best Matching Unit (BMU), which is the neuron with the closest distance to the input vector.
4. The weight updates on the BMU and its neighboring neurons using the neighbor function and a learning rate that decreases with iteration.
5. Iterations are repeated until the weight values converge, so that the data distribution pattern on the SOM map is stable.

Journal Of Data Science
Volume 3, Number 02, 2025, DOI 10.58471/Science.v3i02
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

RESULTS AND DISCUSSION

For Clustering testing using the SOM Algorithm with the principle of testing all data against predetermined weights and the process is carried out in the form of repetition to produce new weights which will be used as a reference in the next iteration in determining the next data group, this will continue to be done until all datasets occupy the predetermined cluster positions.

The required parameters for testing using SOM are alpha and Epsilon. Alpha determines how much change is applied to the network weights at each iteration. The value will continuously decrease as the number of iterations increases, to ensure that the learning process becomes more stable and converges towards a better solution. Epsilon is the smallest error tolerance value used as a stopping condition in SOM training. When the error change between two consecutive iterations becomes smaller than a predetermined value, the training will be stopped, because it is considered to have reached convergence. Description of the results of Clustering Formation with SOM with the need for 1457 input data used as a dataset, 4 clusters will be formed with initial weights selected randomly, a learning rate of 0.5 and a max epoch of 100. The weights used as Neuron Output are

[[8.80180000e+04 4.60000000e+00 3.69500000e+03 3.13346902e+00]
[1.00000000e+01 4.00000000e+00 2.45500000e+03 2.58557934e+02]
[9.23200000e+03 4.70000000e+00 1.36300000e+03 6.35824149e+01]
[6.29790000e+04 5.10000000e+00 1.98300000e+03 6.26110402e+00]
[1.22111000e+05 4.40000000e+00 7.57000000e-01 1.84273980e+00]]

Continued with Clustering Formation using the SOM principle where the iteration process by calculating the proximity between data (x) to the specified weight data by making changes to the weight and will stop when it meets the minimum learning rate requirements specified. The following description is the clustering formation process for testing each data expressed in each test in the form of iterations.

Table 1 Formation of Clustering Iteration 1

Training Data (X1)	1.01870000, 5.6000000, 2.36600000, 8.58272392
Distance Value	10173.44525, 13447.44958, 10181.75633, 135272.1485 10209.19041
Winning Neuron	1
Old Weight	5111, 5.45, 2188.5, 364.15228005
New Weight	5111, 5.45, 2188.5, 364.15228005

Table 2 Formation of Clustering Iteration 2

Training Data (X2)	1.13260000, 4.60000000, 2.68300000,4.52133175
Distance Value	6242.794,12294.8748, 11327.06934, 134132.21124, 11377.0456
Winning Neuron	1
Old Weight	7.90775000, 5.06750000, 2.41102500, 2.20629747
New Weight	7.90775000, 5.06750000, 2.41102500, 2.20629747

Table 3 Clustering Formation Iteration 3

Training Data (X3)	29.51, 4.5, 2634, 302.7949117
Distance Value	7881.8230, 23590.28277, 323.7715, 145429.07753
Winning Neuron	3
Old Weight	17.90155, 4.619, 2449.55, 357.15631933
New Weight	17.90155, 4.619, 2449.55, 357.15631933

Table 4. Formation of 95 Iteration Clustering

Training Data (X95)	1.00000000, 4.00000000, 1.90000000, 7.16208409
Distance Value	72057.45733, 86913.7837,71334.55430, 221016.939
Winning Neuron	5
Old Weight	58.27950172, 4.59418384, 1310.91680, 798.477489
New weight	58.27950172, 4.59418384, 1310.91680, 798.477489

Table 5 Clustering Formation Iteration 100

Training Data (100)	1.285590, 4.300000, 2.422000, 1.637399
Distance Value	119788.6347, 79338.0068, 128543.38884, 80540.4351
Winning Neuron	1
Old Weight	4.92234805, 4.68399073, 1.96500037, 9.01196655
New Weight	4.92234805, 4.68399073, 1.96500037, 9.01196655

The test results are the same for alldataset then a cluster is formed with SOM with the following table:

Table 6 Clustering Results Using the SOM Algorithm

No	Depth	Mag	Dmin	PGA	Cluster SOM
1	10187	5.6	2366	85.8272392	1
2	11326	4.6	2683	45.21331755	1
3	29.51	4.5	2634	302.7949117	3
4	9221	4.8	3743	61.4175169	1
5	54874	4.9	1945	6.79338988	2
6	204591	4.2	1202	0.843758365	4
7	233212	5	5.46	1.058993772	4
8	63029	4.7	4546	5,106631635	2
9	46177	4.6	0.971	7.348097588	2
10	134961	4.8	1857	1.972064952	4
...	...	...	...	...	...
1539	117.6	5.2	1873	668.8851894	3
1540	31319	4	1288	9.074672877	2
1541	21087	5.1	2427	26.29991737	1

Journal Of Data Science
Volume 3, Number 02, 2025, DOI 10.58471/Science.v3i02
ESSN 3025-2792 (Online)
https://ejournal.seaninstitute.or.id/index.php/visualization

No	Depth	Mag	Dmin	PGA	Cluster SOM
1542	184158	4.2	0.763	0.969479984	4
1543	42226	4.1	1294	6.435463009	2
1544	145614	4.3	1914	1.389334783	4
1545	35	4.4	2743	273.1443466	3
1546	146223	4.3	1749	1.381730211	4
1547	10065	4.2	3634	41.35170335	1

Visualization for Mapping Data Groups as a whole is described in Figure 1 below.This :

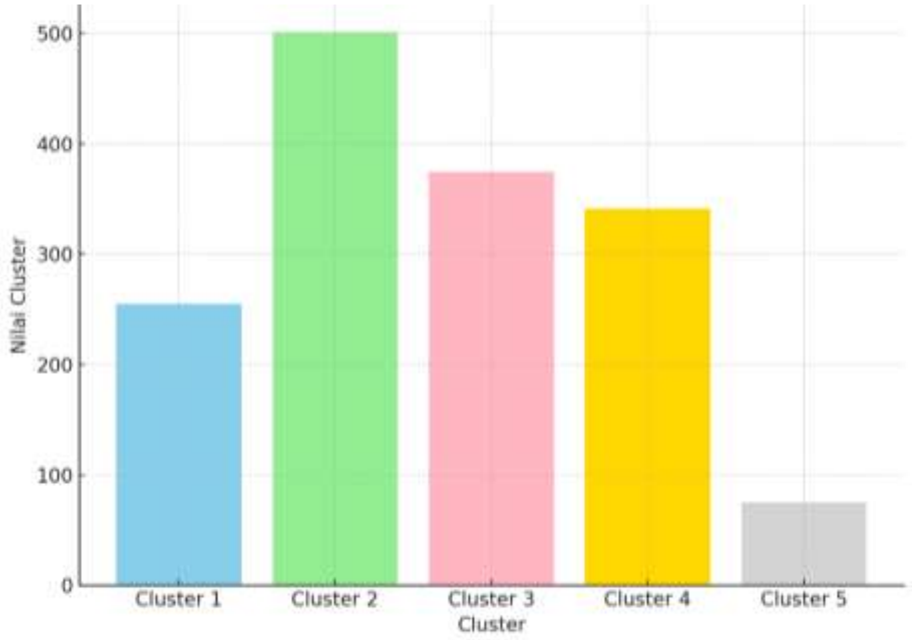

Figure 1. Clustering Visualization With SOM

PictureFigure 1 is a visualization of the results of data clustering using the Self-Organizing Map (SOM) Algorithm, which divides the data into five clusters based on feature similarities between the data. Each cluster represents a set of data that share the same characteristics. To facilitate interpretation, each cluster is given a different color to strengthen the visualization. Each cluster is occupied by datasets that have been grouped based on feature similarities. Cluster 1 contains 255 data, Cluster 2 is the largest with 501 data, Cluster 3 contains 375 data, Cluster 4 is occupied by 341 data, and Cluster 5 is the smallest with 75 data. The variation in the number of data in each cluster reflects the differences in patterns between groups, with Cluster 2 dominating and Cluster 5 being the smallest group.

Discussion

The clustering testing process using the Self-Organizing Map (SOM) algorithm is carried out using the principle of unsupervised learning, where each data item is tested against predetermined neuron weights. At each iteration, the input data is compared with the network weights to calculate the closeness using Euclidean distance. The neuron with the smallest distance is designated as the Best Matching Unit (BMU) or the winning neuron, and then the winning neuron's weights are updated based on the learning parameters. This process occurs repeatedly, where the new weights formed in one iteration are used as a reference in the next iteration. Iterations continue until a stopping condition is met, namely when the weight change is very small or when the maximum number of epochs is reached.

Two important parameters in this testing are alpha and epsilon. The alpha parameter represents the learning rate, which is the magnitude of the change applied to the network weights each time an update is made. The alpha value gradually decreases with increasing iterations to ensure stable and convergent learning. Meanwhile, epsilon is used as the minimum error tolerance value, which determines when to stop the training process. If the error change between successive iterations is smaller than epsilon, then the training is considered to have reached the point of convergence. In this study, the initial learning rate was set at 0.5 with a maximum number of iterations of 100.

The results of clustering with SOM using a dataset of 1,457 data points and randomly selected initial weights indicate a significant weight adaptation process. In the initial iteration, the winning neuron is determined based on its closest distance to the input data, as seen in the example of the test results for data X1, X2, and X3. The selected neuron then undergoes weight adjustment to increasingly align with the data distribution. A similar process occurs in the 95th to 100th iterations, where weight changes become smaller and show signs of convergence. This aligns with the SOM principle, where a gradually decreasing learning rate leads the model toward a stable state.

After the entire dataset was iterated, the data were grouped into four main clusters. The distribution of the number of data in each cluster showed significant variation. Cluster 1 contained 255 data points, Cluster 2, the largest group, contained 501 data points, Cluster 3 contained 375 data points, Cluster 4 contained 341 data points, and Cluster 5, the smallest cluster, contained 75 data points. The dominance of Cluster 2 indicates that most of the data points have similar characteristics and are concentrated in this group, while the presence of Cluster 5 indicates the presence of data points with unique characteristics, but the number is relatively small.

Overall, the clustering results using SOM successfully mapped variations in ground acceleration based on depth, magnitude, source distance, and PGA. This model was able to group homogeneous data into one cluster, while simultaneously separating data with different characteristics into other clusters. This shows that SOM is effective in identifying complex PGA distribution patterns, because this method not only produces representative clusters but also maintains topological connectivity between data. Thus, these clustering results provide important information for seismic risk analysis, where large clusters represent general earthquake characteristics, while small clusters can be considered as extreme data groups that require special attention in disaster mitigation planning.

CONCLUSION

The results of ground acceleration clustering tests using the Self-Organizing Map (SOM) algorithm demonstrate that this method is able to work effectively in grouping complex earthquake data. The iterative learning process with a weight update mechanism produces a convergent model, where the gradually decreasing learning rate makes the system more stable with increasing iterations, while the use of the epsilon value as an error tolerance limit ensures that training stops when the error changes are no longer significant. From a total of 1,457 earthquake data analyzed, five clusters were formed with different number distributions, where the largest cluster contains more than five hundred data points and the smallest cluster contains only about seventy-five data points. The difference in the number of members in each cluster illustrates the variation in ground acceleration patterns, with large clusters representing common earthquake characteristics and small clusters marking groups of data with extreme or rare characteristics. Overall, SOM has proven capable of mapping the distribution pattern of ground acceleration clearly based on the parameters of depth, magnitude, source distance, and PGA. The results of this clustering not only provide an overview better understanding of PGA variations, but can also be used as a basis for seismic risk mapping, spatial planning, and earthquake disaster mitigation strategies in vulnerable areas.

REFERENCE

Maharani, Y. N., Rizkianto, Y., & Ikhsan, I. (2023). ANALISIS KERENTANAN TERHADAP BAHAYA GEMPABUMI DAN TSUNAMI DENGAN METODE SELF ORGANIZING MAP DI KABUPATEN BANTUL. *Jurnal Mineral, Energi, dan Lingkungan*, *7*(2), 23-32.

Dermawan, I., Salma, A., Kurniawati, Y., & Mukhti, T. O. (2023). Implementation of the Self Organizing Maps (SOM) Method for Grouping Provinces in Indonesia Based on the Earthquake Disaster Impact. *UNP Journal of Statistics and Data Science*, *1*(4), 337-343.

Sulianta, F. (2023). *Basic Data Mining from A to Z*. Feri Sulianta.

Firmansyah, H., Saputra, S. A., & Falah, M. (2025). Pengaruh dimensionality reduction menggunakan Self-Organizing Maps terhadap klasifikasi Naive Bayes pada data kampanye pemasaran. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, *2*(1b), 1853-1862.

Machdi, I., MNSPM, A. R., Wijayanto, A. W., Putra, A. P., & Pramana, S. (2023). Implementasi Big Data Dan Kecerdasan Artifisial Untuk Statistik Ofisial. *ARTIFICIAL*, 47.

Pratama, A. K., Ashaury, H., & Umbara, F. R. (2023). Klasifikasi Data Gempa Bumi Di Pulau Jawa Menggunakan Algoritma Extreme Gradient Boosting. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(4), 2923-2929.

Dermawan, I., Salma, A., Kurniawati, Y., & Mukhti, T. O. (2023). Implementation of the Self Organizing Maps (SOM) Method for Grouping Provinces in Indonesia Based on the Earthquake Disaster Impact. *UNP Journal of Statistics and Data Science*, *1*(4), 337-343.

Meyer, S. G., Reading, A. M., & Bassom, A. P. (2022). The use of weighted self-organizing maps to interrogate large seismic data sets. *Geophysical Journal International*, *231*(3), 2156-2172.

Danielsen, A. S., Johansen, T. A., & Garrett, J. L. (2021). Self-organizing maps for clustering hyperspectral images on-board a cubesat. *Remote Sensing*, *13*(20), 4174.