www.ijarp.com ISSN 2456-9992 Page: 01-10

International Journal Advanced Research Publications

Research Article

Volume: 02
Issue: 06

APPLICATION OF SELF-ORGANIZING MAPS FOR CLUSTERING REGENCIES AND MUNICIPALITIES IN CENTRAL JAVA BASED ON FACTORS CONTRIBUTING TO DIVORCE

Lisa Rotin Milawati, Di Asih I Maruddani*, Suparti

Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro.

Article Received: 2 May 2026, Article Revised: 22 May 2026, Published on: 12 June 2026

***Corresponding Author: Di Asih I Maruddani**

Department of Statistics, Faculty of Science and Mathematics, Universitas Diponegoro.

Doi: https://doi-doi.org/101555/ijarp.5222

ABSTRACT

The increasing number of divorce cases in Central Java during 2017–2019, along with the various social problems arising from divorce, highlights the need for effective and region-specific divorce prevention strategies. Since each district and city has distinct characteristics related to the factors contributing to divorce, clustering analysis can provide valuable insights for designing targeted interventions. This study applies the Self-Organizing Maps (SOM) algorithm, an unsupervised neural network approach, to cluster districts and cities in Central Java based on divorce-related factors. The analysis uses data from 35 regencies and municipalities in Central Java Province in 2019 and incorporates 13 variables representing the causes of divorce. Several clustering scenarios consisting of 3, 4, 5, and 6 clusters were evaluated. The SOM model was trained using a learning rate of 0.50, Euclidean distance, a maximum of 1,000 iterations, an initial neighborhood radius of 1, a bubble neighborhood function, and linearly decreasing learning rates and neighborhood radii. Initial weights were randomly selected from the input vectors. Clustering performance was assessed using the silhouette coefficient. The results indicate that the three-cluster solution provides the best partition, achieving the highest silhouette coefficient of 0.39547. The resulting clusters consist of 13 regencies/municipalities in Cluster 1, 10 in Cluster 2, and 12 in Cluster 3. These findings provide a basis for developing more targeted and region-specific policies to address divorce-related issues in Central Java.

KEYWORDS: Self-Organizing Maps, clustering, divorce factors, silhouette coefficient, Central Java.

INTRODUCTION

Divorce is the legal dissolution of a marriage due to certain causes through a court decision initiated by one or both parties involved in the marriage (Simanjuntak, 2015). Divorce does not merely signify the termination of the marital relationship between husband and wife; it is also accompanied by various positive and negative consequences. However, divorce tends to generate more adverse effects, particularly when the marriage has produced children, who often become the primary victims of their parents' separation.

Data from the Central Java Provincial Statistics Agency (BPS Central Java, 2020) indicate a continuous increase in divorce cases in Central Java Province from 2017 to 2019. In 2017, the number of divorce cases reached 69,857, increasing to 75,557 cases in 2018 and further rising to 82,758 cases in 2019. The existence of serious social problems resulting from divorce, combined with the increasing trend of divorce cases, highlights the need for effective divorce prevention strategies. Such strategies are expected to be more effective when tailored to the specific conditions of each region, as every district and municipality possesses unique characteristics. Therefore, regional clustering analysis is needed to identify areas with similar characteristics so that appropriate interventions can be developed according to their specific conditions.

Clustering is a process of grouping objects into several clusters in such a way that objects within the same cluster are relatively homogeneous, whereas objects belonging to different clusters are relatively heterogeneous. Previous studies have examined divorce-related clustering analysis. For example, Hidayati and Aridinanti (2012) clustered regencies and municipalities in East Java based on divorce-causing factors in 2010 using single linkage, complete linkage, average linkage, and Ward’s methods. However, these clustering approaches require certain assumptions, including that the sample adequately represents the population and that multicollinearity is absent among the variables. Consequently, their effectiveness may be limited when these assumptions are violated.

According to Halim and Widodo (2017), Self-Organizing Maps (SOM) is a high-dimensional data analysis method that does not require strict statistical assumptions and is capable of producing visual representations of data objects. Brabazon et al. (2015) defined SOM as an artificial neural network employing unsupervised learning, enabling the network to adapt and self-organize in response to input signals. The fundamental principle of SOM is that the network receives input signals and automatically maps their representations into a set of responses while preserving the topological relationships among the inputs. As a result, similar input patterns are mapped to neighboring locations on the output map.

The objective of this study is to apply the Self-Organizing Maps method to cluster regencies and municipalities in Central Java Province based on factors contributing to divorce and to provide a spatial visualization of the resulting clusters.

MATERIAL AND METHODS

Study Area and Data

This study used secondary data on divorce cases across 35 regencies and municipalities in Central Java Province, Indonesia, for the year 2019. The data were obtained from the Central Java Statistics Agency (BPS Central Java) through the publication *Central Java in Figures 2020*. The unit of analysis consisted of all regencies and municipalities in Central Java Province.

Thirteen variables representing the proportion of divorce cases attributed to different causes were used in the analysis. These variables included: continuous disputes and quarrels (X_1), economic problems (X_2), abandonment by one spouse (X_3), imprisonment (X_4), domestic violence (X_5), forced marriage (X_6), apostasy (X_7), drug addiction (X_8), gambling (X_9), adultery (X_{10}), physical disability (X_{11}), alcoholism (X_{12}), and polygamy (X_{13}).

Self-Organizing Maps (SOM)

The clustering analysis was performed using the Self-Organizing Maps (SOM) algorithm, an unsupervised artificial neural network introduced by Kohonen. SOM projects high-dimensional data onto a lower-dimensional map while preserving the topological relationships among observations.

The SOM training procedure consisted of the following stages:

Initialization

Weight vectors for each neuron were initialized by randomly selecting values from the input vectors. The following SOM parameters were specified before training:

(1) Initial learning rate (α_0): 0.50

(2) Maximum number of iterations (T): 1,000

(3) Initial neighborhood radius (σ_0): 1

(4) Distance measure: Euclidean distance

(5) Neighborhood function: Bubble neighborhood function

(6) Learning-rate decay: Linear

(7) Neighborhood-radius decay: Linear

Competition Stage

For each input vector (x), the Euclidean distance between the input vector and every neuron weight vector (w_j) was calculated as

$$d_j = \|\mathbf{x} - \mathbf{w}_j\|$$

The neuron with the smallest distance was selected as the Best Matching Unit (BMU) or winning neuron.

Adaptation Stage

The winning neuron and its neighboring neurons updated their weight vectors according to

$$\mathbf{w}_j(\mathbf{t}+\mathbf{1}) = \mathbf{w}_j(\mathbf{t}) + \alpha_t h_{cj}(t)[\mathbf{x}(\mathbf{t}) - \mathbf{w}_j(\mathbf{t})]$$

where $\mathbf{w}_j(\mathbf{t}+\mathbf{1})$ denotes the weight vector of neuron (j) at iteration (t), α_t is the learning rate, and$h_{cj}(t)$ is the neighborhood function.

The bubble neighborhood function was defined as

$$h_{cj}(t) = \begin{cases} 1 & \text{untuk neuron ke} - j \in N_c(t) \\ 0 & \text{lainnya} \end{cases}$$

where $N_c(t)$represents the set of neurons located within the neighborhood radius of the winning neuron.

Parameter Updating

During training, the learning rate and neighborhood radius were decreased linearly according to

$$\alpha_t = \alpha_0 \left(1 - \frac{t}{T}\right)$$

$$\sigma_t = \sigma_0 \left(1 - \frac{t}{T}\right)$$

The training process continued until the maximum number of iterations was reached.

Cluster Validation

To determine the optimal number of clusters, four clustering scenarios were evaluated, consisting of 3, 4, 5, and 6 clusters. Cluster quality was assessed using the Silhouette Coefficient (SC).

The silhouette value for observation (i) was calculated as

$$s(i) = \frac{b_i - a_i}{\max(b_i, a_i)}$$

where (a_i) is the average distance between observation (i) and all other observations within the same cluster, and (b_i) is the minimum average distance between observation (i) and observations in other clusters.

The overall Silhouette Coefficient was computed as

$$SC = \frac{1}{N}\sum_{1}^{N} S(i)$$

where (N) denotes the total number of observations. A higher SC value indicates better cluster separation and cohesion.

Cluster Profiling and Spatial Visualization

After determining the optimal clustering solution, cluster profiling was conducted to characterize the dominant divorce factors within each cluster. Subsequently, the clustering results were integrated with administrative boundary data of Central Java Province and visualized spatially using QGIS version 3.10.

The resulting thematic map was developed as a web-based geographic information system (Web GIS) to facilitate the interpretation and dissemination of regional clustering patterns.

Software

All clustering analyses and cluster validation procedures were conducted using RStudio (R version 3.3.1) with the *kohonen* and *cluster* packages. Spatial visualization and Web GIS development were performed using QGIS version 3.10.

RESULTS AND DISCUSSION

Clustering Using Self-Organizing Maps

A total of 35 regencies and municipalities in Central Java Province were clustered into 3, 4, 5, and 6 groups based on 13 divorce-related factors. Accordingly, the number of neurons in the output layer was set to 3, 4, 5, and 6, respectively. The clustering process employed an initial learning rate of 0.50, a maximum of 1,000 iterations, a rectangular map topology, an initial neighborhood radius of 1, Euclidean distance as the distance measure, and a bubble neighborhood function. Both the learning rate and neighborhood radius were decreased linearly during the training process. Initial weight vectors were generated by randomly selecting values from the input vectors.

Weight Initialization

Since the clustering scenarios consisted of 3, 4, 5, and 6 clusters based on 35 observations and 13 variables, the initial weight matrices had dimensions of 3×13, 4×13, 5×13, and 6×13, respectively. Weight vectors were initialized using the kohonen package in R, which randomly selects values from the input data by default.

SOM Training

In the kohonen package, the number of neurons is represented by the grid dimensions specified through the xdim and ydim parameters. The dimensions of the grid are not required to be equal. In this study, grid sizes of 3×1, 2×2, 5×1, and 3×2 were used for the 3-, 4-, 5-, and 6-cluster solutions, respectively. During the training process, each input vector was mapped onto the neurons in the output layer. Since the neurons represent clusters, the objective of training was to obtain optimal weight vectors that serve as cluster centroids, indicated by the completion of the maximum number of iterations.

Clustering Results

The quality of each clustering solution was evaluated using the Silhouette Coefficient (SC). The results are presented in Table 1

Table 1. Silhouette Coefficient for Different Numbers of Clusters.

Number of Clusters	Silhouette Coefficient
3	0.39547
4	0.34472
5	0.29880
6	0.32875

Based on the Silhouette Coefficient values, the best clustering solution was obtained with three clusters, yielding the highest SC value of 0.39547.

Table 2. Membership of the Three-Cluster Solution.

Cluster	Number of Members	Cluster Members
1	13	Purbalingga, Banjarnegara, Purworejo, Wonosobo, Boyolali, Klaten, Sukoharjo, Semarang Regency, Kendal, Batang, Pekalongan Regency, Surakarta Municipality, and Salatiga Municipality
2	10	Kebumen, Magelang Regency, Wonogiri, Rembang, Pati, Kudus, Temanggung, Magelang Municipality, Semarang Municipality, and Pekalongan Municipality
3	12	Cilacap, Banyumas, Karanganyar, Sragen, Grobogan, Blora, Jepara, Demak, Pemalang, Tegal Regency, Brebes, and Tegal Municipality

Cluster Profiling

Cluster profiling was conducted using the optimal three-cluster solution to identify the characteristics of each cluster. The profile of each cluster was determined by comparing the mean values of the divorce-related variables across clusters. A score of 1 was assigned to the cluster with the lowest mean value for a variable, a score of 2 to the cluster with the intermediate mean value, and a score of 3 to the cluster with the highest mean value.

Table 3. Comparison of Cluster Scores.

Variable	Mean Value of Cluster 1	Mean Value of Cluster 2	Mean Value of Cluster 3	Score C1	Score C2	Score C3
X1	0.45273	0.66462	0.36273	2	3	1
X2	0.24017	0.09543	0.43566	2	1	3
X3	0.24694	0.19079	0.15721	3	2	1
X4	0.03611	0.03102	0.02324	3	2	1
X5	0.00857	0.00489	0.00737	3	1	2
X6	0.00458	0.00122	0.00330	3	1	2
X7	0.00212	0.00463	0.00117	2	3	1
X8	0.00282	0.00176	0.00117	3	2	1
X9	0.00193	0.00136	0.00149	3	1	2
X10	0.00121	0.00073	0.00290	2	1	3
X11	0.00154	0.00082	0.00144	3	1	2
X12	0.00084	0.00181	0.00170	1	3	2
X13	0.00044	0.00090	0.00062	1	3	2
Total Score	31	24	23			

Cluster 1 obtained the highest total score (31), indicating that it consists of regencies and municipalities with relatively high levels of divorce-related problems. Cluster 2 obtained the second-highest score (24), representing regions with moderate divorce levels, while Cluster 3 had the lowest score (23), indicating relatively low divorce levels.

Cluster 1 Profile

Cluster 1 consists of 13 regencies and municipalities characterized by the highest proportions of divorces caused by abandonment by one spouse (X3), imprisonment (X4), domestic violence (X5), forced marriage (X6), drug addiction (X8), gambling (X9), and physical disability (X11). These factors were more prominent in Cluster 1 than in the other clusters.

Cluster 2 Profile

Cluster 2 consists of 10 regencies and municipalities characterized by the highest proportions of divorces caused by continuous disputes and quarrels (X1), apostasy (X7), alcoholism (X12), and polygamy (X13).

Cluster 3 Profile

Cluster 3 consists of 12 regencies and municipalities characterized by the highest proportions of divorces caused by economic factors (X2) and adultery (X10).

The characteristics of the clusters can also be observed through the SOM code plot, as illustrated in Figure 1. Each segment of the fan-shaped diagram represents a variable, while the segment length indicates the relative contribution of that variable. The code plot shows that variables X3, X4, X5, X6, X8, X9, and X11 are more dominant in Cluster 1 than in Clusters 2 and 3. Variables X1, X7, X12, and X13 are more dominant in Cluster 2, whereas variables X2 and X10 are more dominant in Cluster 3.

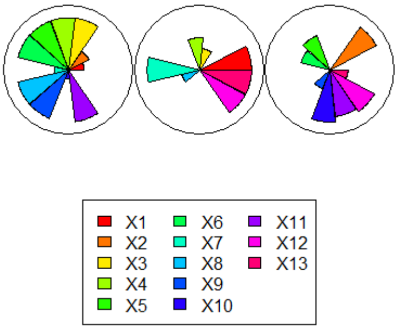

Figure 1. SOM Code Plot.

Visualization of Clustering Results Using QGIS

The clustering results were spatially visualized using QGIS version 3.10. The visualization process required a spatial map in shapefile (.shp) format, a basemap, the *qgis2web* plugin, the *QGIS Cloud Plugin*, and a web browser.

The resulting Web GIS provides a spatial representation of the clustering results across regencies and municipalities in Central Java Province. As shown in Figure 2, regions belonging to Cluster 1 are represented by the darkest color, Cluster 2 by an intermediate color, and Cluster 3 by the lightest color.

The Web GIS visualization enables users to explore the spatial distribution of divorce-related characteristics across Central Java and supports policymakers in identifying region-specific patterns that may require different intervention strategies.

Figure 2. Web GIS Interface.

CONCLUSION

The Self-Organizing Maps (SOM) method can be applied to cluster regencies and municipalities in Central Java Province based on factors contributing to divorce. Among the clustering scenarios consisting of 3, 4, 5, and 6 clusters, the best clustering result was obtained with three clusters, yielding the highest silhouette coefficient (0.39547). Cluster 1 consisted of 13 regencies/municipalities characterized by relatively high divorce rates, Cluster 2 comprised 10 regencies/municipalities with moderate divorce rates, and Cluster 3 included 12 regencies/municipalities with relatively low divorce rates.

To effectively reduce divorce rates, preventive measures should be tailored to the dominant divorce factors within each cluster. In Cluster 1, greater attention should be directed toward issues related to abandonment by one spouse, imprisonment, domestic violence, forced marriage, drug addiction, gambling, and physical disability. In Cluster 2, interventions should focus on continuous disputes and quarrels, apostasy, alcoholism, and polygamy. Meanwhile, in Cluster 3, economic problems and adultery were identified as the predominant factors contributing to divorce.

Furthermore, QGIS successfully provided a spatial visualization of the clustering results in the form of a Web GIS map, enabling users to access and explore the information anytime and anywhere through an internet connection. This visualization can serve as a useful tool for policymakers in designing more targeted and region-specific divorce prevention strategies.

REFERENCES

1. Badan Pusat Statistik Provinsi Jawa Tengah. 2020. *Provinsi Jawa Tengah dalam Angka 2020*. Semarang: BPS Jawa Tengah.
2. Brabazon, A., O'Neill, M., dan McGarraghy, S. 2015. *Natural Computing Algorithms*. Berlin: Springer.
3. Budhi, G. S., Liliana, dan Haryanto, S. 2008. Cluster Analysis untuk Memprediksi Talenta Pemain Basket Menggunakan Jaringan Saraf Tiruan Self-Organizing Maps (SOM). *Jurnal Informatika*, 9(1): 23-32.
4. Budiyanto, E. 2016. *Sistem Informasi Geografis dengan Quantum GIS*. Yogyakarta: Penerbit Andi.
5. Fausett, L. 1994. *Fundamental of Neural Network : Architecture, Algorithm and. Application*. New Jersey: Prentice-Hall.
6. Halim, N. N. dan Widodo, E. 2017. Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self-Organizing Maps. *Prosiding SI MaNIs*, 1(1): 188-194.
7. Haykin, S. 1999. *Neural Networks: A Comprehensive Foundation.* England: Pearson Education.
8. Hennig, C. 2016. *Clustering Strategy and Method Selection*. Dalam Hennig dkk. (Penyunting), *Handbook of Cluster Analysis*. Boca Raton: CRC Press.
9. Hidayati dan Aridinanti. 2012. *Pengelompokan Kabupaten/Kota di Jawa Timur berdasarkan Faktor-Faktor Penyebab Perceraian Tahun 2010.*http://digilib.its.ac.id/ITS-Undergraduate-3100012045894-/17888 (28 Maret 2020).
10. Kalter, N. dan Rembar, J. 1981. The Significance of A Child's Age at The Time of Parental Divorce. *American Journal of Orthopsychiatry*, 51(1): 85-100.
11. Kassambra, A. 2017. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning.* New York: STHDA.
12. Kohonen, T. 2001. *Self-Organizing Maps*. Third Edition. Berlin: Springer.
13. Kusumadewi, S. 2003. *Artificial Intelligence: Teknik dan Aplikasinya.* Yogyakarta: Graha Ilmu.
14. Pemerintah Indonesia. 1974. *Undang–Undang No. 1 Tahun 1974 Tentang Perkawinan*. Lembaran Negara RI No. 1 Tahun 1974. Jakarta: Sekretariat Negara.
15. Samarasinghe, S. 2007. *Neural Network for Applied Sciences and Engineering*. New York: Auerbach Publications.
16. Santosa, A. D., Priyandari, Y., dan Suletra, I. W. 2013. Perancangan Sistem Pendukung Keputusan Pemetaan Profil Kecamatan di Kabupaten Sragen Berdasarkan Indeks Pembangunan Manusia Menggunakan Algoritma Self-Organizing Maps. *Prosiding Seminar Nasional IENACO*, Universitas Muhammadiyah Surakarta: 28 Maret 2013.
17. Simanjuntak, P. N. H. 2015. *Hukum Perdata Indonesia*. Jakarta: KENCANA.