Jurnal Matematika dan Statistika serta Aplikasinya, Vol. 13, No. 2, 64-72, 2025
DOI:https://10.24252/msa.v13i2.57741

Grouping Of Regencies/Cities In West Sumatra Province Based On Economic Development Indicators Using The Self-Organizing Maps (SOM) Algorithm

Wiwil Dzil Izzatil[1], Chairina Wirdiastuti[2]*, Syafriandi[3]
[1] *Program Studi Statistika, FMIPA, Universitas Negeri Padang*
[2] *Program Studi Statistika, FMIPA, Universitas Negeri Padang*
[3] *Program Studi Statistika, FMIPA, Universitas Negeri Padang*
***Corresponding author**: chairinawirdi@unp.ac.id

***Submission date**: 10 June 2025, **Revision**: 11 July 2025, **Accepted**: 19 October 2025

ABSTRACT
Economic development is an important aspect in improving the standard of living of the community. To measure economic development progress in a region, relevant indicators are needed, one of which is Gross Regional Domestic Product (GRDP) per capita. In West Sumatra Province, there are disparities in GRDP per capita between regions. Therefore, clustering analysis is necessary to assist local governments in determining development priorities, formulating more targeted development policies, and allocating resources efficiently. This study aims to cluster West Sumatra regions using the Self-Organizing Maps algorithm based on economic development indicators. The optimal number of clusters was determined using three validation approaches: the Connectivity index, Dunn's index, and the Silhouette index. The selected cluster configuration shows the lowest Connectivity value, a Dunn index close to 1, and the highest Silhouette value. The analysis results identified three clusters: Cluster 1 consists of 6 districts/cities with moderate economic development, Cluster 2 includes 7 districts/cities showing high economic development, and Cluster 3 consists of 6 other districts/cities categorized as having low economic development.

KEYWORDS
Economic Development, Clustering, Self-Organizing Maps.

1. INTRODUCTION

Development encompasses various efforts undertaken to improve the well-being of society in all aspects, both physical, such as infrastructure, and non-physical, such as education, health, and the economy. In this context, economic development plays a crucial role in enhancing the quality of life for communities. Economic development can be understood as a sustainable process aimed at fostering long-term growth in per capita income, with the expectation of improving overall social well-being [1].

To assess the progress of economic development in a region, appropriate and relevant indicators are needed. These indicators play an important role in evaluating and comparing the level of development and welfare of communities in various regions, as well as understanding development patterns in each region [2]. Gross Regional Domestic Product (GRDP) is one of the main indicators used to measure the level of economic development, as it reflects the total value of goods and services produced in a region within a certain period [3]. The Gross Regional Domestic Product in various regions of West Sumatra still shows significant disparities. Some regions have advanced economic development with continuously improving infrastructure. Meanwhile, other regions still face obstacles in economic development. This can be seen in **Figure 1**. Based on **Figure 1**, it can be seen that in 2023 Bukittinggi City recorded the highest per capita GRDP, which amounted to Rp. 89.74 million, followed by

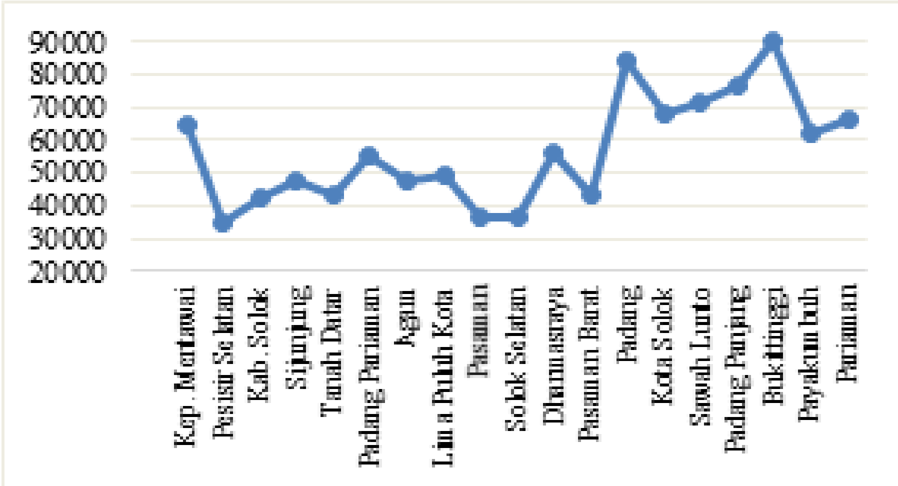

Figure 1. GRDP per capita of West Sumatra Province in 2023

Padang City at Rp. 84.53 million. Meanwhile, Pesisir Selatan Regency has the lowest GRDP per capita of Rp. 34.31 million [4]. This shows the inequality of economic development between regions in West Sumatra Province. This imbalance is caused by differences in economic activity and geographical conditions in each region. Therefore, an analysis is needed to group the regions in West Sumatra Province based on economic development indicators.

The clustering of regions based on these indicators is expected to assist the local government in determining data-based development priorities and identifying patterns that emerge in these regional groups. One of the relevant methods for clustering is Self-Organizing Maps (SOM). According to [5] Self-Organizing Maps is an unsupervised learning algorithm that maps high-dimensional data into a two-dimensional space so that group patterns can be visualized easily.

This algorithm has been used in various previous studies for regional clustering. Research [6] used the SOM method to group provinces in Indonesia based on educational aspects. The results of the analysis obtained 2 clusters that show the characteristics of the educational aspects of each province. Furthermore, research [7] compared the SOM algorithm with K-Means with the results obtained a smaller DBI value in the SOM algorithm, meaning that the SOM algorithm is better than K-Means.

Although previous studies have demonstrated the effectiveness of self-organizing maps (SOM) in various regional groupings, research on district or city groupings based on economic development indicators remains limited. Most studies analyzing economic development in Indonesia use conventional clustering methods, such as K-means and K-harmonic means. For instance, a study conducted by [8] used the K-harmonic means method to cluster districts/cities in Indonesia based on the human development index. Meanwhile, a study in East Java used the Similarity Weight and Filter Method (SWFM) to group regions based on economic development indicators [9]. In addition, research on economic development clustering in Indonesia is still focused on specific regions such as East Java, Central Java, and overall national analysis.

Based on the description above, a study was conducted to cluster the West Sumatra region using the SOM algorithm based on economic development indicators. This research is expected to be useful for local governments in formulating development policies that are more targeted, and can allocate resources efficiently. The objective of this study is to group regencies/cities in West Sumatra based on economic development indicators using SOM, in order to inform data-driven regional development strategies.

2. LITERATURE REVIEW

2.1 Cluster Analysis

Cluster analysis is the process of grouping data based on certain similarities [10]. The main purpose of clustering is that data (objects) with similar patterns will be put into the same cluster, while objects with different patterns will be put into different clusters, so that objects in one cluster have a high level of similarity, while objects in different clusters have a high level of difference [11].

There are two main types of cluster analysis, Hierarchical and Non-hierarchical. The hierarchical method is a gradual grouping of objects based on similarity without knowing the number of clusters from the beginning. Meanwhile, the non-hierarchical method is a clustering method by determining the number of clusters in advance and the clustering results depend on the selected center [12].

2.2 Determination of the Number of Clusters

The number of clusters is determined using the internal validation method. This method is a numerical measurement approach used for data clustering without involving external factors [13]]. There are three types of internal validation, namely the Connectivity, Dunn, and Silhouette indices [14]. The Connectivity index has a range of values between 0 and ∞. Optimization of this index in determining the best number of clusters is achieved when the resulting value is lower than the value of the cluster formed. The equation for finding the connectivity index is as follows.

$$Conn\ (c) = \sum_{i=1}^{N}\sum_{j=1}^{L} X_{i,nni(j)} \tag{1}$$

Next, the Dunn index is a comparison between the closest distance between observations in different clusters and the farthest distance in one cluster. The Dunn index can be formulated in the following equation.

$$Dunn = \frac{d_{min}}{d_{max}} \tag{2}$$

Then for the Silhouette index value, where a value close to 1 indicates that the cluster formed is arguably good. If the value is close to -1, the resulting cluster is considered less than optimal. The Silhouette index can be written as the following equation.

$$eq83S_{(i)} = \frac{b_{(i)} - a_{(i)}}{max\left(a_{(i)}, b_{(i)}\right)} \tag{3}$$

2.3 Artificial Neural Network

Artificial neural networks are a type of machine learning system that can be used to model complex processes and make predictions. Artificial Neural Network (ANN) is an information processing approach that mimics the way biological nervous systems, such as the brain, process information. Like humans, ANN learn through examples or patterns. ANN can also recognize relationships between the inputs and outputs of a process without requiring a physical representation of the system being studied [15].

In ANN, there are two types of learning methods: supervised and unsupervised. In supervised learning, a pattern or input vector is provided sequentially, along with the corresponding output target vector. During training, the weights are updated using a specific learning algorithm. In contrast, unsupervised learning does not use training data with a specific target. Instead, the network independently creates groups and determines the characteristics of each group so that certain units in the network can be associated with new input patterns [16].

2.4 Self-Organizing Maps

One of the algorithms used in artificial neural networks for clustering is the self-organizing map (SOM), which Teuvo Kohonen first introduced in 1982 [17]. This algorithm forms clusters based on data characteristics or features, and does not require supervision during the training process (target output) [15].

The algorithm comprises two main layers: the input layer and the output layer. Each neuron in the input layer is connected to all neurons in the output layer via specific weights. The input layer neurons act as input data in the clustering process, while the output layer neurons represent the cluster results. During the process, the unit with the weight closest to the input pattern (usually calculated based on the minimum Euclidean distance) is selected as the winner and its weight, along with those of its neighbouring neurons, is updated based on their distance from the winning neuron. The SOM output shows the similarity of characteristics between members in the same cluster [18].

2.5 Data Normalization

Since the data has values with different ranges, a normalization process is needed. Data normalization involves changing the scale of attribute values to be smaller, ensuring that each value has balanced weight [19]. One applicable method is min-max normalization, a data adjustment technique that retains information without losing content and has a scale ranging from 0 to 1 [20]. The formula for min-max normalization is:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{4}$$

3. METHODOLOGY

This study uses secondary data from the 2023 publication of the West Sumatra Province Central Bureau of Statistics. The unit of analysis was the 19 districts/cities in West Sumatra Province. This study uses variables that form economic development indicators, consisting of ten variables, as shown in the **Table 1**. The Self-Organizing Maps (SOM) algorithm was chosen as

Table 1. Research Variables

Variable	Description	Unit
X_1	Gross Regional Domestic Product per Capita	Thousands
X_2	Human Development Index (HDI)	Percent
X_3	School Participation Rate (SPR) for ages 7–12	Percent
X_4	School Participation Rate (SPR) for ages 13–15	Percent
X_5	School Participation Rate (SPR) for ages 16–18	Percent
X_6	Open Unemployment Rate (OUR)	Percent
X_7	Labor Force Participation Rate (LFPR)	Percent
X_8	Percentage of Poor Population	Percent
X_9	PLN Electricity Supply	Percent
X_{10}	Safe Drinking Water Supply	Percent

the clustering method for this study because it can map multidimensional data into an easy-to-understand, two-dimensional space. Additionally, the SOM algorithm preserves proximity relationships (topology) between data points, ensuring that areas with similar economic characteristics are placed close together on the map. This reflects the similarity of their geographical and economic conditions. Previous research has demonstrated that SOM outperforms traditional clustering methods, such as K-means clustering [7].

3.1 Analysis Procedure

The steps of the Self-Organizing Maps algorithm are as follows.

1. Initialization
   a. Initial weights (w_{ij}) by randomly selecting small numbers
   b. Learning rate (α)
   c. Maximum Epoch
   d. Radius value (R), the value of R used is $R = 0$, which means that only the winning input vector performs the learning process.
2. Perform steps 3-8 when the stop condition is not met
3. Perform steps 4-6 for each input vector (x_i)

4. Calculate the distance between the input vector x_i and the weight vector at each input w_{ij} using the following equation

$$D_j = \sum_{i=1}^{n} (w_{ij} - x_i)^2 \tag{5}$$

5. After calculating the distance for each input, determine the winning input by identifying the input vector with the minimum distance value from the weight vector

6. Update the weight value w_{ij} for the winning input using the following equation:

$$w_{ij}\ (new) = \ w_{ij}\ (old) + \ \alpha\ (x_i - w_{ij}\ (old)) \tag{6}$$

7. Update the learning rate.

$$\alpha\ (new) = 0,5\ \times\ \alpha\ (old) \tag{7}$$

8. Update the radius value R

9. Check the stop condition. The iteration process can stop if the new weights differ very little from the old weights or if the obtained values have converged.

4. RESULT & DISCUSSION

4.1 Data Normalization

In this study, the initial stage of cluster analysis using the SOM method was data normalization. Normalization was performed because there were significant differences in data scales. The results of data normalization can be seen in **Table 2**.

Table 2. Normalization Data

Regency/City	$\mathbf{X_1}$	$\mathbf{X_2}$	···	$\mathbf{X_{10}}$
Kepulauan Mentawai	0.55	0.00	···	0.00
Pesisir Selatan	0.00	0.36	···	0.58
Solok	0.15	0.34	···	0.44
Sijunjung	0.24	0.36	···	0.04
⋮	⋮	⋮	⋮	⋮
Pariaman	0.58	0.77	···	0.83

4.2 Determining the Number of Clusters

The number of clusters was determined using three approaches, namely the Connectivity index, Dunn, and Silhouette. The selected clusters were those with the lowest Connectivity value, a Dunn value close to 1, and the highest Silhouette value. The internal validation results are shown in **Table 3**. Based on **Table 3**, the internal validation results show the smallest Connectivity value of 10.9155 in cluster 3, a Dunn value close to 1 of 0.5406 in cluster 7, and the highest Silhouette value of 0.3204 in cluster 3. This indicates that the optimal number of clusters is 3.

Table 3. Internal Validation Test Results

Cluster Size	**3**	**4**	**5**	**6**	**7**
Connectivity	10.9	21.6	23.1	27.1	29.5
Dunn	0.34	0.29	0.34	0.49	0.54
Silhouette	0.32	0.17	0.20	0.18	0.24

4.3 Analysis Self-Organizing Maps

SOM networks are used to group input patterns into several clusters. SOM networks require a training process that aims to minimize the average distance between each object and the nearest unit. **Figure 2** shows the number of training processes from

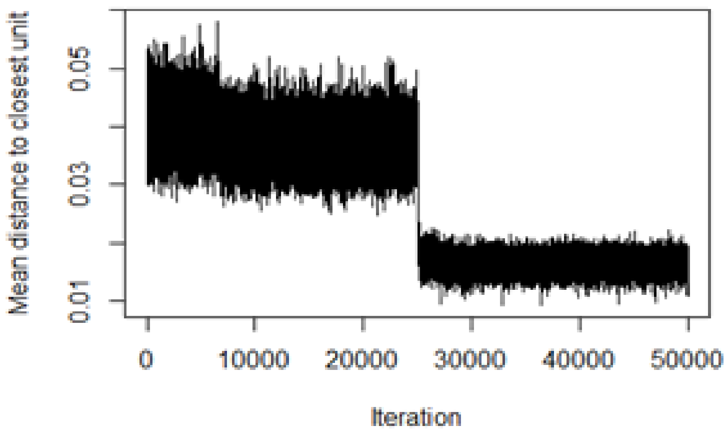

Figure 2. Training Progress Chart

the SOM method, where the horizontal axis shows the number of iterations performed during training, and the vertical axis shows the average distance to the nearest unit. In this study, the iteration process will stop after reaching 50,000 iterations. It can be seen that at the 25,000th iteration and above, the process begins to show convergence. As the number of iterations increases, the average distance between the input data and the nearest cluster unit (mean of distance cluster unit) decreases, leading to an improvement in clustering quality. The training is considered to have reached convergence when the average distance between cluster units is below 0.02.

The results of regional grouping in West Sumatra Province based on economic development indicators are shown in **Table 4**. It shows the number of members in each cluster. Next, the clustering results were visualized using a SOM map. The results

Table 4. Number and Members of Clusters

Cluster	Number of Members	Cluster Members
1	6	Kab. Pesisir Selatan, Kab. Tanah Datar, Kab. Padang Pariaman, Kab. Agam, Kab. Pasaman, Kab. Pasaman Barat
2	7	Kota Padang, Kota Solok, Kota Sawahlunto, Kota Padang Panjang, Kota Bukittinggi, Kota Payakumbuh, Kota Pariaman
3	6	Kab. Kepulauan Mentawai, Kab. Solok, Kab. Sijunjung, Kab. Lima Puluh Kota, Kab. Solok Selatan, Kab. Dharmasraya

of this visualization are in the form of a fan diagram, as shown in **Figure 3**.

This study uses a hexagonal topology with a $3x4$ grid. The resulting fan diagram illustrates the characteristics of each cluster, thereby showing the distribution of variables on the map. Each circle represents a unit in the SOM map, with color segments indicating the contribution of variables ($X_1 - X_{10}$) to each cluster. The variables dominating this cluster can be seen from the color and length of a particular color segment within the circle; the longer the color segment, the greater the influence of the variable on that unit. Based on **Figure 3**, it can be seen that three clusters are formed, represented by different colors: cluster 1 (red), cluster 2 (yellow), and cluster 3 (green). For a clearer picture of the characteristics of each cluster, refer to **Table 5**.

Table 5 illustrates the characteristics of each cluster. The cluster with the highest average score on an indicator shows the most superior characteristics compared to other clusters on that indicator. Cluster 1 is a transitional region with moderate conditions across most indicators. However, it has low GRDP and labor participation. This area has the potential to be prioritized for accelerated development, as it already has a solid foundation. Cluster 2 is an advanced region with the highest GRDP and

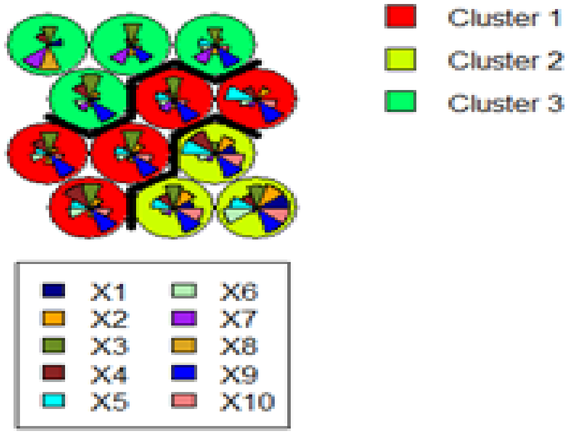

Figure 3. SOM Map

Table 5. Cluster Characteristics

Variable	Cluster 1	Cluster 2	Cluster 3
X_1	43,309.67	73,944.86	49,530.33
X_2	73.11	80.15	71.36
X_3	99.63	99.55	99.81
X_4	97.05	97.73	96.42
X_5	85.28	91.34	73.88
X_6	5.47	5.79	3.96
X_7	68.04	69.49	74.73
X_8	6.36	4.06	7.59
X_9	99.53	99.91	94.57
X_{10}	85.23	96.78	73.55

Human Development Index (HDI), the best infrastructure, and the lowest poverty rates. However, it still faces unemployment issues. Cluster 3 exhibits conflicting conditions: high labor participation, low HDI, and high poverty, indicating that most of the population works in low-skilled sectors.

The clustering results point to targeted policy directions: Cluster 1 requires economic acceleration programs through infrastructure development and skill training. Cluster 2 needs to focus on innovation and sharing benefits with surrounding areas. Cluster 3 necessitates a fundamental transformation through improved education and a shift in the economic structure from traditional to modern sectors..

5. CONCLUSION

Using the Self-Organizing Maps (SOM) algorithm, grouping regions in West Sumatra Province based on economic development indicators and found three clusters. Cluster 1 consists of six districts/cities with moderate economic development characteristics. These areas are moderately developed but need improvement, particularly in terms of further education and poverty reduction. Cluster 2 consists of seven districts/municipalities with high economic development characteristics. In other words, this cluster is the most developed region with the best economic indicators. Cluster 3 consists of six other districts/municipalities characterized by low economic development. These regions are less developed and lag behind others. Therefore, the government must improve economic development and the quality of life of its people to reduce inequality between regions in West Sumatra Province.

REFERENCES

[1] M. Hasan and M. Azis, *Pembangunan Ekonomi & Pemberdayaan Masyarakat*, kedua ed. Makassar: CV. Nur Lina dan Pustaka Taman Ilmu, 2018, [Online]. [Online]. Available: http://eprints.unm.ac.id/10706/1/Bukupembangunanekonomicontohfix.pdf

[2] L. Arsyad, "Ekonomi pembangunan dan pembangunan ekonomi," in *Ekonomi Pembangunan Lanjutan*. Jakarta: Universitas Terbuka, 2014, ch. 1, pp. 1–37, [Online]. [Online]. Available: http://repository.ut.ac.id/3975/

[3] A. Novelia and A. Kinanti, "Pengelompokkan kabupaten/kota berdasarkan indikator produk domestik regional bruto (pdrb) menggunakan metode self organizing map," *Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 13, no. 3, pp. 411–418, 2024.

[4] BPS Sumatera Barat, "Sumatera barat dalam angka 2023," Berita Resmi Statistik, Badan Pusat Statistik, 2023.

[5] T. Kohonen, "The self-organizing map," *Neurocomputing*, vol. 21, no. 1–3, pp. 1–6, 1998.

[6] M. F. Faqih and I. F. Mahdy, "Penerapan self organizing maps dalam pengelompokkan provinsi di indonesia berdasarkan aspek pendidikan," *Jurnal Riset Statistika*, pp. 93–102, 2024.

[7] L. R. Iyohu, I. Djakaria, and L. O. Nashar, "Perbandingan metode k-means clustering dengan self-organizing maps (som) untuk pengelompokan provinsi di indonesia berdasarkan data potensi desa," *Jurnal Statistika dan Aplikasinya*, vol. 7, no. 2, pp. 195–206, 2023.

[8] A. D. Sofia and A. Kudus, "Pengelompokan kabupaten/kota berdasarkan indikator indeks pembangunan manusia 2022 menggunakan k-harmonic means clustering," *Jurnal Riset Statistika*, pp. 163–172, 2023.

[9] R. A. Lutfi and I. Zain, "Pengelompokan kabupaten/kota berdasarkan indikator pembangunan ekonomi dan potensi daerah provinsi jawa timur menggunakan similarity weight and filter method (swfm)," *Jurnal Inferensi*, vol. 1, no. 2, pp. 49–56, 2018.

[10] M. W. Talakua, Z. A. Leleury, and A. W. Taluta, "Analisis cluster dengan menggunakan metode k-means untuk pengelompokkan kabupaten/kota di provinsi maluku berdasarkan indikator indeks pembangunan manusia tahun 2014," *BAREKENG Jurnal Ilmu Matematika dan Terapannya*, vol. 11, no. 2, pp. 119–128, 2017.

[11] A. C. Benabdellah, A. Benghabrit, and I. Bouhaddou, "A survey of clustering algorithms for an industrial context," *Procedia Computer Science*, vol. 148, pp. 291–302, 2019.

[12] M. Azmi, A. A. Putra, D. Vionanda, and A. Salma, "Comparison of the performance of the k-means and k-medoids algorithms in grouping regencies/cities in sumatera based on poverty indicators," *UNP Journal of Statistics and Data Science*, vol. 1, no. 2, pp. 59–66, 2023.

[13] A. R. Ivansyah, F. Fitri, Y. Kurniawati, and T. O. Mukhti, "Implementation self organizing maps method in cluster analysis based on achievement sustainable development goal/sdg's west sumatera province," *UNP Journal of Statistics and Data Science*, vol. 1, no. 5, pp. 480–487, 2023.

[14] N. N. Halim and E. Widodo, "Clustering dampak gempa bumi di indonesia menggunakan kohonen self organizing maps," in *Prosiding Seminar Nasional Integrasi Matematika dan Nilai Islam (SI MaNIS)*, vol. 1, no. 1, 2017, pp. 188–194, [Online]. [Online]. Available: http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62

[15] A. R. Gunawan and Z. Rais, "Applied the self-organizing maps (som) method for clustering educational equity in south sulawesi," *ARRUS Journal of Mathematics and Applied Science*, vol. 4, no. 1, pp. 6–19, 2024, [Online]. [Online]. Available: https://doi.org/10.35877/mathscience2607

[16] E. Setyowati and S. Mariani, "Penerapan jaringan syaraf tiruan dengan metode learning vector quantization (lvq) untuk klasifikasi penyakit infeksi saluran pernapasan akut (ispa)," in *Prosiding Prisma Seminar Nasional Matematika*, vol. 4, 2021, pp. 514–523, [Online]. [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/article/view/44356

[17] T. Kohonen, *Self-Organizing Maps*, 1995, vol. 7, no. 2.

[18] A. B. T. Dananjaya and M. I. Irawan, "Analisis sentimen pada komentar terhadap kebijakan perjalanan domestik yang dikelompokkan menggunakan metode self-organizing maps," *Jurnal Sains dan Seni ITS*, vol. 12, no. 1, 2023.

[19] G. A. B. Suryanegara, Adiwijaya, and M. D. Purbolaksono, "Peningkatan hasil klasifikasi pada algoritma random forest untuk deteksi," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 1, no. 10, pp. 114–122, 2021.

[20] N. Sepriyanti, R. S. Nahampun, M. H. Zikri, I. Ambarani, and A. Rahmadeyan, "Implementation of k-means clustering to group poverty levels in riau province," in *Seminar Nasional Penelitian dan Pengabdian Masyarakat*, 2022, pp. 59–65, [Online]. [Online]. Available: https://journal.irpi.or.id/index.php/sentimas