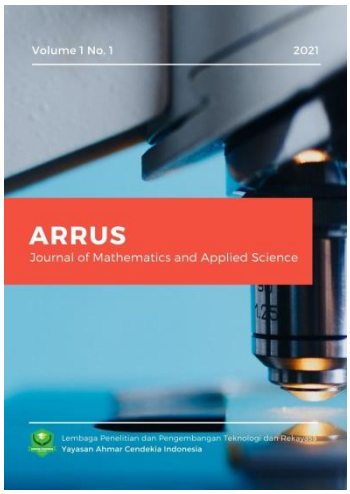
https://doi.org/10.35877/mathscience4782

OPEN ACCESS

ISSN : 2776-7922 (Print) / 2807-3037 (Online)



*Corresponding author: Dinda Warfani Hasbullah, Statistics Study Program, Faculty of Mathematics and Natural Sciences, Makassar State University, Indonesia.

E-mail: dindawarfani24@gmail.com

RESEARCH / REVIEW ARTICLE [Calibri 11pt]

A Hybrid Neural Network Approach Using SOM and LVQ for Mapping Crime Clusters in Indonesia

Zulkifli Rais, Sitti Masyitah Meliyana, Dinda Warfani Hasbullah*

Statistics Study Program, Faculty of Mathematics and Natural Sciences, Makassar State University, Indonesia.

Abstract: Crime ratehigh crime rates in Indonesia are one of the important issues that need to be addressed with data-based strategies. This study aims to group provinces in Indonesia based on crime patterns using Self-Organizing Map (SOM) and classify the results using Learning Vector Quantization (LVQ). The results of the clustering analysis using SOM show that the optimal number of clusters is two, as supported by validation using Connectivity, Dunn Index, and Silhouette Score. Cluster 1 consists of 31 provinces with lower crime rates, while Cluster 2 includes 3 provinces with higher crime rates. To improve understanding of the clustering results, classification was carried out using the LVQ method, which produced an accuracy of 91.43%.

Keywords: Cluster Analysis, SOM, Internal Validation, Classification, LVQ, Crime.

1. INTRODUCTION

Multivariate statistical analysis is used to process data with more than two variables and is divided into dependency and interdependence analysis.(Riswan & Khairudin, 2019). Dependency analysis explains the influence of variables on other variables, such as multiple regression and discriminant analysis. Meanwhile, interdependence analysis reveals the relationship between variables or groups them, such as factor and cluster analysis.(Kunto & Hasana, 2006).

Cluster analysis is one of the analyses in multivariate statistics that is used in grouping objects that have similarities so that the members in one cluster are homogeneous and between the clusters themselves are heterogeneous.(Azrahwati et al., 2022). There are various clustering analysis methods, ranging from simple methods to complex methods that use artificial intelligence such as Artificial Neural Network (ANN).(Rahayu & Hakim, 2019).

Artificial Neural Network (ANN) is a computing system whose architecture and operation are inspired by biological neurons in the brain. ANN can be described as a mathematical and computational model used for clustering and classification of data.(Cape, 2015). One of the ANN methods for clustering is Self Organizing Map, while to classify data, one of the ANN methods that can be used is Learning Vector Quantization.

Self Organizing Map (SOM) is a neural network method that does not require special supervision (unsupervised)(Rahayu & Hakim, 2019). SOM is used to group (cluster) data based on the characteristics of data features.(Aziz & Mustakim, 2022). This method can be used for large and small data, as well as to display clustering results in lower dimensions.(Annas et al., 2022).

https://doi.org/10.35877/mathscience4782

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Learning Vector Quantization is a classification method on each output unit by presenting a class. LVQ has several advantages, such as lower error values compared to other methods and the resulting model can be updated gradually.(Melisa et al., 2022).

The crime rate in an area can endanger the safety and welfare of the community. Crime, as a violation of civil or criminal law, is a serious problem that disrupts public order.(Sembiring & Faza, 2021). Understanding crime patterns and characteristics is important for effective prevention policies. Every year, the increase in crime cases in Indonesia produces data on various crimes, such as murder, assault, rape, theft, fraud, and narcotics.(Mayona et al., 2022).

Several previous studies that have applied the Self Organizing Map (SOM) and Learning Vector Quantization (LVQ) methods include: Zulfahmi et al. (2023) mapped the level of crime vulnerability using SOM and found two clusters. Novianto & Andayani (2023) used the SOM method to group provinces in Indonesia based on crime indicators and obtained five clusters with a DBI value of 0.3793. Tomasouw et al. (2021) applied the LVQ method to detect drug abuse, with the best classification results achieving an accuracy of 86.7%.

2. Literature Review

Hypotheses: Entrepreneurial competencies comprise those skills that are deeply rooted in a person's background as well as skills that can be acquired and learned at work (Man & Lau, 2005).

2.1. Multivariate Analysis

Multivariate analysis is a series of statistical methods used to process several measurements related to an individual or object simultaneously.(Simamora, 2005). This method is useful in complex situations with an unlimited number of variables and varying levels of measurement. With multivariate analysis, all variables can be examined simultaneously to assess their respective effects on the target variable.(Sihombing, 2022). Multivariate analysis is divided into two main methods, namely dependency and interdependence methods.

The dependency method is used when the research has variables that can be categorized as dependent variables.(Sihombing, 2022). Statistical analysis included in this method includes regression analysis, variance analysis, canonical correlation analysis, discriminant analysis, and logit analysis. Meanwhile, the interdependence method is used when the measured variables cannot be clearly distinguished between dependent and independent variables, so the focus is on the reciprocal relationship between variables. One of the analysis techniques in this method is factor analysis(Himayati et al., 2020).

2.2. Clustering

Clusteringis the process of grouping data into a set or cluster, where each cluster contains data with the same characteristics and is different from other clusters.(Munawar, 2015). Clustering aims for understanding, by forming groups that capture the natural structure of the data as an initial stage of analysis, or for use, by finding prototype clusters that contain the data.(Kusumah et al., 2017). Clustering results must have high homogeneity within clusters (within-clusters) and high heterogeneity between clusters (between-clusters).(Halim & Widodo, 2017).

2.3. Artificial Neural Network(ANN)

Artificial Neural Network (ANN) is an artificial representation of the human brain that simulates the human learning process.(Kusmaryanto, 2014). ANN is designed to generalize the mathematical model of human understanding, with neurons as the information processing elements. Signals flow through connections between neurons, where each connection has a weight that affects the signal sent. In addition, each neuron applies an activation function to the resulting signal.(Alkhairi et al., 2019).

https://doi.org/10.35877/mathscience4782

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

2.4. *Self Organizing Map(SOM)*

Self Organizing Map (SOM) is an unsupervised machine learning algorithm introduced by Teuvo Kohonen in 1982.(Kohonen, 1982). The SOM network consists of two layers, namely the input layer and the output layer, where each node in the input layer is connected to all nodes in the output layer, but the output nodes are not connected to each other.(Winalda, 2017)..

The stages of the SOM algorithm according to Hermadi, Sitanggang & Edward (2006) and Halim & Widodo (2017) include random weight initialization, competition to determine the winner node, cooperation in the environment of the winning node, and weight adaptation with a learning rate value that decreases each epoch. Testing is stopped when the weight change is very small, indicating that convergence has been achieved.

2.5. *Cluster Validation*

Cluster validation is a procedure to evaluate the results of cluster analysis quantitatively and objectively. There are three approaches to cluster validation.(Hermadi et al., 2006): external, which evaluates based on the expected structure; internal, which evaluates based on the characteristics of the data itself; and relative, which compares the clustering structure with alternatives obtained from parameter adjustments. Researchers use internal validation because crime data is quantitative.

Some internal validation methods include:

(1). **Connectivity Index:** Measures the number of best clusters with smaller values indicating better clustering.

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

with,

$nn_{i(j)}$= nearest neighbor observation i to j and L

$nn_{i(j)}$as a parameter that determines the number of neighbors that contribute to the Connectivity measurement.

(2). **Silhouette Index:** Measures confidence in clustering, with values close to 1 indicating good clustering and values close to -1 indicating poor clustering.

$$s(i) = \frac{b(i) - a(i)}{(a(i), b(i))}$$

with,

$a(i)$= average distance between i and all other observations in the same cluster.

$b(i)$= average distance between I and the observation in the nearest cluster

(3). **Dunn Index:** Measures the ratio between the smallest distance between clusters and the largest distance within a cluster. A higher ratio indicates better cluster quality.

$$C = \frac{d_{min}}{d_{max}}$$

with,

d_{min}= the smallest distance between observations in different clusters

d_{max}= the largest distance in each data cluster

2.6. *Learning Vector Quantization(LVQ)*

Learning Vector Quantization(LVQ) is a training method used for supervised learning in a competitive, single-layer artificial neural network architecture.(Mardiana & Kalsum, 2021). LVQ is a classification method where each output unit represents a class. The classes generated from this competitive layer are determined only by the distance between the input vectors. If two input vectors have almost the same distance, the competitive layer will group them into the same class.(Melisa et al., 2022).

https://doi.org/10.35877/mathscience4782

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

2.7. *Classification Accuracy*

One of the methods used to evaluate classification accuracy is the Apparent Error Rate (APER). APER is a method for evaluating classification accuracy by measuring the proportion of samples that are incorrectly classified.(Johnson & Wichern, 2007). The first step before calculating APER is to divide the data into two parts: a training set to form a classification and a testing set to test the obtained classification.(Haykin, 1999). Classification errors can be seen in the Confusion Matrix table.(Suhendra et al., 2020).

Table 1. Confusion Matrix

Observation Results	Estimate	
	y1	y2
y1	n_{11}	n_{12}
y2	n_{21}	n_{22}

So the APER value can be calculated using the formula:

$$APER = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} = \frac{total\ obyek\ yang\ salah\ diklasifikasikan}{total\ sampel}$$

Accuracy $= 1 - APER$

2.8. *Crime*

Crime comes from the word "Crimen" which means crime or an act that violates the law. Criminal acts include behavior that is economically and psychologically detrimental, and is contrary to the law and social norms.(Putra et al., 2021). Light, Keller, and Calhoun (1994) identified four types of crime: 1) Violent Crime, 2) Victimless Crime, 3) Organized Crime, and 4) White Collar Crime. According to the Criminal Code (KUHP), some forms of crime listed are: 1) Murder (Article 338 of the Criminal Code), 2) Theft (Article 262 of the Criminal Code), 3) Indecent Acts (Article 289 of the Criminal Code), 4) Assault (Article 351 of the Criminal Code), 5) Kidnapping (Article 328 of the Criminal Code), and 6) Corruption (Article 209 of the Criminal Code).

3. Research Method and Materials

3.1. *Types of research*

This type of research is applied research with a quantitative approach, which involves taking and collecting relevant data, then analyzing it using the SOM and LVQ algorithms.

3.2. *Data source*

The type of data used in this study is secondary data obtained and collected from previously existing data sources. The data comes from the 2023 Criminal Statistics publication taken from the website of the Central Statistics Agency (www.bps.go.id)

3.3. *Operational Definition of Variables*

The variables used in this study consist of 7 (seven) crimes, namely, Murder, Assault, Rape, Kidnapping, Theft, Narcotics and Corruption.

3.4. *Research Procedures*

Data analysis techniques are carried out to process data and information into the research process. In this study, the following data analysis techniques were used:

(1). Standardize the data.

https://doi.org/10.35877/mathscience4782

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

(2). Conduct descriptive statistical analysis of the variables used. Descriptive statistical analysis to provide an overview of the data used. This includes the amount of data from each variable, as well as the minimum and maximum values of each variable.

(3). Performing clustering with the SOM method. The steps include:

(a). Initialization is in the form of weights obtained randomly for each node. After the weights are given, the network is given input.$(w_{ij})(w_{ij})(x_i)$.

(b). For each output node j, compute the value of the distance function. Determine the winning node j that minimizes all output nodes.$D(w_j, x_n)D(w_j, x_n)$

$$D(w_j, x_n) = \sqrt{\sum_{i=1}^{n}(w_{ij} - x_{ni})^2}$$

(c). Identify all nodesoutputj in the neighborhood of the winning node Jdefined byenvironment of size R. for these nodes the weights are updated.

$$w_{ij,baru} = w_{ij,lama} + \eta(x_i - w_{ij,lama})$$

(d). In the process of obtaining new weights, a learning rate value is required (i.e. . The learning rate value at each epoch will decrease to .$\eta)0 \leq \eta \leq 1\eta(i+1) = 0{,}5\eta$

(e). The condition for stopping the test is done by calculating the difference between the (new) and (old) weights. If the value only changes slightly, it means that the test has reached convergence so it can be stopped.$w_{ij}w_{ij}w_{ij}$

(4). Determine the best cluster by using internal validation, such as connectivity, dunn index, and silhouette value.

(5). Classifying clustering results using the LVQ method. The steps include:

(a). Initialize the initial weights, parameters, and enter the input and target data.

(b). Forming the optimal ANN architecture with the LVQ method using training data.

(c). Reclassify the testing data based on the best architecture of the LVQ method formed from the training data.

(6). Calculating the classification accuracy of the LVQ method on training data and testing data using the APER formula.

(7). Interpret the analysis results.

4. Results and Discussion

For Results, provide sufficient detail to allow the results to be meaningful and informative. For Discussion, this should explore the significance of the results of the work, not repeat them. A combined Results and Discussion section is often appropriate. Avoid extensive citations and discussion of published literature.

4.1. Descriptive Analysis

Descriptive analysis is used to provide an overview of crime data that occurred in Indonesia in 2022. The data used in this study covers various types of crimes recorded throughout the year, including murder, assault, rape, kidnapping, theft, narcotics, and corruption.

Table 2. Descriptive Statistics of Crime Data

Variables	Min	1st Qu	Median	Mean	3rd Qu	Max
Murder	1	9.25	17.50	25.12	34.50	102
Persecution	97	313.50	600	977.50	1117.80	5142
Rape	9	20	31.50	42.88	58	135
Kidnapping	1	8.25	21.50	43.29	49.75	233
Theft	90	334	638	1094	1133	6630
Narcotics	28	273.50	553	923.40	1311.80	5006
Corruption	0	4.25	7	7.68	10.75	31

https://doi.org/10.35877/mathscience4782

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

4.2. Training Progress

4.2.1. Initialization

At this stage, the initialization process is carried out in the form of random weights () withw_{ij} $i = 7$is the number of variables and is the number of clusters.$j = 2$

Table 3. Random Weight

Wij	j	
i	1	2
1	4	33
2	22	6
3	19	13
4	24	26
5	29	1
6	3	9
7	18	28

4.2.2. Sampling

Samplingdone by taking data (provinces) with input vectors which will be processed by calculating the distance vector dj.

Table 4. Input Vector

Province	X1	X2	X3	X4	X5	X6	X7
Aceh	-0.049	0.042	2,908	0.145	0.342	0.257	-0.116

Description: X1 (Murder); X2 (Assault); X3 (Rape); X4 (Kidnapping); X5 (Theft); X6 (Darcotics) and X7 (Corruption)

4.2.3. Matching

Next, the distance vector dj is calculated, which is obtained by adding the difference between the weight vector () in Table 4.1 and the input vector () in Table 4.$w_{ij}x_i$

$$d_j = \sqrt{\sum_{i=1}^{n}(w_{ij} - x_i)^2}$$

$$d_1 = \sqrt{\begin{matrix}(4 + 0{,}049)^2 + (22 - 0{,}042)^2 + (19 - 2{,}908)^2 + (24 - 0{,}145)^2 \\ +(29 - 0{,}342)^2 + (3 - 0{,}257)^2 + (18 + 0{,}116)^2\end{matrix}}$$

$$= \sqrt{2483{,}556}$$

$$= 49{,}835 \text{(random weight 1)}$$

$$d_2 = \sqrt{\begin{matrix}(33 + 0{,}049)^2 + (6 - 0{,}042)^2 + (13 - 2{,}908)^2 + (26 - 0{,}145)^2 \\ +(1 - 0{,}342)^2 + (9 - 0{,}257)^2 + (28 + 0{,}116)^2\end{matrix}}$$

$$= \sqrt{2765{,}446}$$

$$= 52{,}588 \text{(random weight 2)}$$

4.2.4. Matching

Weight updates are performed on the distance vector$d_j$with minimum value. In the previous stage, the distance was obtained as a random weight with a minimum value, which is 49.835. The process of obtaining new weights requires a learning rate value, which is . In this study, a learning rate of 0.05 was used. The learning rate value at each iteration will decrease to .$d_1(\eta) 0 \leq \eta \leq 1 \eta(i + 1) = 0{,}5\,\eta$

$$w_{ij,new} = w_{ij,old} + \eta(x_i - w_{ij,old})$$

$$w_{1,1} = 4 + 0{,}05(-0{,}049 - 4) = 3{,}748$$

$$w_{2,1} = 22 + 0{,}05(0{,}042 - 22) = 20{,}902$$

$$w_{3,1} = 19 + 0{,}05(2{,}908 - 19) = 18{,}195$$

$$w_{4,1} = 24 + 0{,}05(0{,}145 - 24) = 22{,}807$$

$$w_{5,1} = 29 + 0{,}05(0{,}342 - 29) = 27{,}567$$

$$w_{6,1} = 3 + 0{,}05(0{,}257 - 3) = 2{,}863$$

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

$$w_{7,1} = 18 + 0{,}05(-0{,}116 - 18) = 17{,}094$$

So that the new weight is obtained as follows.

Table 5. New Weight

Wij	j	
i	1	2
1	3,748	33
2	20,902	6
3	18,195	13
4	22,807	26
5	27,567	1
6	2,863	9
7	17,094	28

4.3. Clustering with SOM Method

In this study, the SOM method is used to group provinces in Indonesia based on the patterns of crime that occur. The clustering results are visualized in the form of a fan diagram, which illustrates how provinces in Indonesia are distributed into several groups based on the characteristics of the crimes observed. To obtain a more comprehensive mapping, grouping was carried out with various numbers of clusters, namely 2, 3, 4, and 5 clusters. Each configuration produces a different grouping pattern, which shows variations in crime rates in each region.

PictureThe following shows the clustering results in the form of a fan diagram, which illustrates the distribution of provinces based on the number of clusters used:

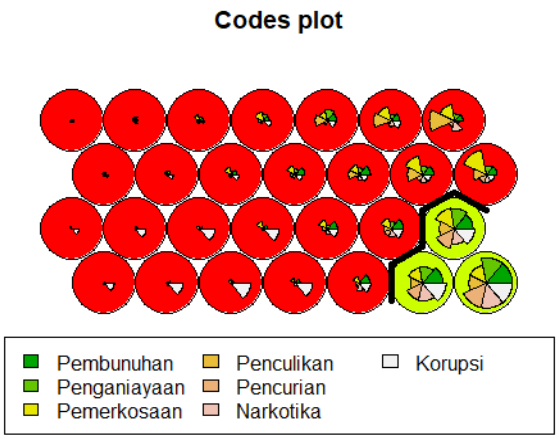


Figure 1. Fan Diagram For 2 Clusters

Based on the clustering results with 2 clusters, the provinces in Indonesia are divided into two main groups. Cluster 1 consists of 31 provinces, while Cluster 2 includes 3 provinces with higher crime characteristics than the other groups. The following table summarizes the results of the grouping of provinces based on the number of clusters used:

Table 6. Result of Grouping 2 Clusters

Cluster	Number of Members	Cluster Members
1	31	Aceh, West Sumatra, Riau, Jambi, South Sumatra, Bengkulu, Lampung, Bangka Belitung Islands, Riau Islands, DKI Jakarta, West Java, Central Java, DI Yogyakarta, Banten, Bali, East Nusa Tenggara, West Nusa Tenggara, West Kalimantan, Central Kalimantan, South Kalimantan, East Kalimantan, North Kalimantan, North Sulawesi, Central Sulawesi, Southeast Sulawesi, Gorontalo, West Sulawesi, Maluku, North Maluku, West Papua and Papua
2	3	North Sumatra, East Java and South Sulawesi

https://doi.org/10.35877/mathscience4782

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

4.4. *Optimum Cluster Validation*

Table 7. Cluster Validation Test Results

Number of Clusters	*Connectivity*	*Dunn*	*Silhouette*
2	**10,2282**	**0.2718**	**0.5714**
3	14,4290	0.2528	0.4058
4	24,7194	0.1332	0.2571
5	29,9044	0.1019	0.2570

Based on Table 7, cluster validation using various methods shows different results. The connectivity method shows that the best crime grouping occurs when using 2 clusters, with the lowest value of 10.2282. The Dunn index method also shows that the best grouping is 2 clusters, marked by the highest value of 0.2718. Meanwhile, the silhouette method shows the best results with 2 clusters, with the highest value of 0.5714. Based on these results, this study decided to use 2 clusters in grouping provinces based on crime rates.

4.5. *Optimum Cluster Profiling*

After the clustering process using SOM, cluster profiling was carried out to understand the characteristics of each group based on the average number of criminal cases. The results of the average calculation for each cluster are shown in Table 8.

Table 8. Average Results of Each Cluster

Cluster	X1	X2	X3	X4	X5	X6	X7
1	20.1	689	38.2	36.2	693	666	6.77
2	77.3	3955	91.7	117	5239	3587	17

Description: X1 (Murder); X2 (Assault); X3 (Rape); X4 (Kidnapping); X5 (Theft); X6 (Darcotics) and X7 (Corruption)

From the table above, Cluster 1 consists of 31 provinces with lower crime rates, while Cluster 2, which includes 3 provinces (North Sumatra, East Java, and South Sulawesi), has a much higher crime rate. Cluster 2 has a significantly higher number of assault cases (3,955 cases), theft (5,239 cases), and narcotics (3,587 cases) compared to Cluster 1.

This profiling shows that Cluster 2 requires more attention in security policies and law enforcement, while Cluster 1 still requires supervision to prevent an increase in crime. By understanding the characteristics of each cluster, crime prevention strategies can be formulated more effectively according to the conditions in each region.

4.6. *Classification using the LVQ method*

In this study, LVQ was trained to classify data into two classes. The dataset used consisted of 34 samples with 7 features as predictors. The model was trained using a cross validation approach with 3 parameter variations as shown in Table

Table 9. Parameter Accuracy Value

Size	k	Accuracy
4	3	0.9143
4	5	0.9143
4	7	0.9143
6	3	0.9143
6	5	0.9143
6	7	0.9143
8	3	0.9143
8	5	0.9143
8	7	0.9143

The training results show that all parameter combinations produce the same accuracy value in the cross validation process, namely 91.43%.

Table 10. Confusion Matrix

Confusion Matrix		Actual Data	
		*Cluster*1	*Cluster*2
Prediction Data	*Cluster*1	31	3
	*Cluster*2	0	0

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Based on the table, it shows that there are 31 clusters "1" predicted as cluster "1" and 3 clusters "2" predicted as cluster "1". In addition, no data was successfully predicted as cluster "2", either from cluster "2" itself or from cluster "1". So the accuracy calculation is obtained as follows:

$$Akurasi = \frac{31+0}{31+3+0+0} \times 100\%$$
$$Akurasi = \frac{31}{34} \times 100\%$$
$$Akurasi = 91{,}18\%$$

5. Conclusion

For Conclusions, the main conclusions of the study may be presented in a short Conclusions section, which may stand alone.

5.1. Conclusion

From the results of the research and discussion in the previous chapter, the following points can be concluded.

(1). Based on the results of cluster validation with the connectivity method, it states that the best grouping of criminal acts occurs when using 2 clusters, with the lowest value of 10.2282. The Dunn index method also shows that the best grouping is 2 clusters, marked by the highest value of 0.2718. And the silhouette method also shows the best results with 2 clusters, with the highest value of 0.5714. Therefore, this study uses 2 clusters to group provinces based on criminal acts.

(2). LVQ was applied to classify the clustering results obtained through the SOM algorithm. By using the cross-validation method and trying various combinations of k parameter values, this model managed to achieve an overall accuracy rate of 91.18%. Although this model showed excellent performance in recognizing the majority class (Cluster 1), which was seen from the sensitivity reaching 100%, it was unable to detect the minority class (Cluster 2), which was reflected in the specificity reaching 0%.

5.2. Suggestion

In this study, of course, there are still shortcomings, so suggestions for further research can consider the use of other algorithms that are more robust to class imbalance considering the limitations of LVQ in handling imbalanced data.

References

Alkhairi, P., Damanik, IS, & Windarto, AP (2019). Application of Artificial Neural Networks to Measure the Correlation of Lecturer Workload to the Increase in the Number of Publications. Proceedings of the National Seminar on Information Science Research (SENARIS), 1(September), 581. https://doi.org/10.30645/senaris.v1i0.65

Annas, S., Uca, Irwan, Safei, RH, & Rais, Z. (2022). Using k-Means and Self Organizing Maps in Clustering Air Pollution Distribution in Makassar City, Indonesia. Jambura Journal of Mathematics, 4(1), 167–176. https://doi.org/10.34312/jjom.v4i1.11883

Aziz, S., & Mustakim, M. (2022). Implementation of Self Organizing Map Algorithm to Identify Family Welfare Level Grouping Patterns in Siak Regency. Indonesian Journal of Informatic Research and Software Engineering (IJIRSE), 2(2), 82–92. https://doi.org/10.57152/ijirse.v2i2.431

Azrahwati, Nusrang, M., Aidid, MK, & Rais, Z. (2022). K-Means Cluster Analysis for Grouping Districts in South Sulawesi Province Based on Village Potential. ARRUS Journal of Mathematics and Applied Science, 2(2), 73–82. https://doi.org/10.35877/mathscience739

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Halim, NN, & Widodo, E. (2017). Clustering of earthquake impacts in Indonesia using Kohonen self-organizing maps. Proceedings of SI MaNIS (National Seminar on Integration of Mathematics and Islamic Values), 1(1), 188–194. http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/62

Haykin, S. (1999). Neural Networks: A Comprehensive Foundation. Upper Saddle River, NJ : Prentice Hall.

Hermadi, I., Sitanggang, IS, & Edward. (2006). Clustering using Self Organizing Maps (case study: PPMB IPB Data). Wisnujati.

Himayati, Switrayni, NW, Komalasari, D., & Fitriyani, N. (2020). Orthogonal Rotation Analysis on Factor Analysis Technique Using Procrustes Method. Eigen Mathematics Journal, 03(01), 45–55. https://doi.org/10.29303/emj.v3i1.66

Johnson, R. A., & Wichern, D. W. (2007). Applied Multivariate Statistical Analysis. Prentice Hall, Inc.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. Biological Cybernetics, 43, 59–69.

Kunto, YS, & Hasana, SN (2006). CHAID ANALYSIS AS A STATISTICAL TOOL FOR MARKET SEGMENTATION (Case Study on Al-Hidayah Sharia Cooperative). Journal of Marketing Management, 1(2), 88–97. http://puslit2.petra.ac.id/ejournal/index.php/mar/article/view/16538

Kusmaryanto, S. (2014). Backpropagation Artificial Neural Network for Face Recognition Feature Extraction Method. 8(2), 193–198.

Kusumah, RD, Warsito, B., & Mukid, MA (2017). Comparison of K-Means and Self Organizing Map Methods (Case Study: Grouping of Districts/Cities in Central Java Based on the 2015 Human Development Index Indicator). Gaussian Journal, 6(3), 429–437. http://ejournal-s1.undip.ac.id/index.php/gaussian

Light, D., Keller, S. I., & Calhoun, C. J. (1994). Sociology. New York : McGraw-Hill.

Mardiana, Y., & Kalsum, TU (2021). Implementation of Artificial Neural Networks with Learning Vector Quantization (LVQ) Algorithm for Detecting Fingerprint Characteristics. Journal of Computer, Information and Technology (JKOMITEK), 1(2), 444–450. https://doi.org/10.53697/jkomitek.v1i2.432

Mayona, Y., Buaton, R., & Simanjutak, M. (2022). Data Mining Clustering of Crime Rates Using the K-Means Algorithm Method (Case Study: Binjai District Attorney's Office). Kaputama Informatics Journal (JIK), 6(3), 2548–9739.

Melisa, Pardede, AMH, & Sihombing, M. (2022). Implementation of Learning Vector Quantization (LVQ) in Identifying Original Palm Sugar with Mixed Palm Sugar. Sci-Tech Journal, 1(1), 39.

Munawar, G. (2015). Implementation of Self Organizing Map (SOM) Algorithm for Student Clustering in Project Course (Case Study: JTK POLBAN). Proceedings of the 6th Industrial Research Workshop and National Seminar, 66–78.

Novianto, MD, & Andayani, S. (2023). Cluster Analysis using SOM Method for Grouping of Provinces in Indonesia based on National Crime Indicators. 9, 39–53. https://www.numbeo.com/

Putra, AD, Martha, GS, Fikram, M., & Yuhan, RJ (2021). Factors Affecting the Crime Rate in Indonesia in 2018. Indonesian Journal of Applied Statistics, 3(2), 123. https://doi.org/10.13057/ijas.v3i2.41917

Rahayu, B., & Hakim, RBF (2019). Landslides in Indonesia Using Kohonen Self Organizing Maps (Som). 1–17.

Riswan, & Khairudin. (2019). Multivariate Statistics. AURA.

Sembiring, A.P., & Faza, S. (2021). Predicting Crime Rate Based on News Articles Using Support Vector Machine (SVM) Method. 9th Applied Business and Engineering Conference, 10–19.

Sihombing, SO (2022). Introduction to Multivariate Analysis Methods. NEM Publisher.

Simamora, B. (2005). Multivariate analysis of marketing. Gramedia Pustaka Utama.

ISSN : 2776-7922 (Print) / 2807-3037 (Online)

Suhendra, MA, Ispriyanti, D., & Sudarno. (2020). Accuracy of Classification of Prosperous Family Card Granting in Semarang City Using Binary Logistic Regression Method and Chaid Method. Gaussian Journal, 9(1), 64–74. https://doi.org/10.14710/j.gauss.v9i1.27524

Tanjung, DH (2015). Artificial Neural Network with Backpropagation to Predict Asthma Disease. Creative Information Technology Journal, 2(1), 28. https://doi.org/10.24076/citec.2014v2i1.35

Tomasouw, BP, Aulele, SN, & Rijoly, ME (2021). Application of Learning Vector Quantization (LVQ) Method to Detect Drug Abuse. Contemporary Mathematics and Applications (ConMathA), 3(1), 36. https://doi.org/10.20473/conmatha.v3i1.26940

Winalda, D. (2017). Testing and Implementation of Self-Organizing Map in Tourism Case Study. 4(2), 3235–3242.

Zulfahmi, RN, Daul, MK, Al Ayyubi, M., Julianta, IW, Pradnyana, & Dwi Bekti, R. (2023). Mapping of Crime Rate Vulnerability Using the Self Organizing Map Method. INSOLOGI: Journal of Science and Technology, 2(5), 872–881. https://doi.org/10.55123/insologi.v2i5.2566