Journal of Natural Sciences and Mathematics Research
J. Nat. Scien. & Math. Res. Vol. 10 No. 2 (2024) 167-177
ISSN: 2460-4453 (Online); 2614-6487 (Printed)
Available online at http://journal.walisongo.ac.id/index.php/jnsmr

Enhancing the productivity of irrigated rice fields in West Nusa Tenggara through utilizing Multilayer Perceptron (MLP) and Self-Organising Maps (SOM)

Azzahra Fajriani Chaerunnisa[1*], and Mujiati Dwi Kartikasari[1]

[1] Department of Statistics, Universitas Islam Indonesia, Indonesia

* Corresponding author's e-mail: azzahra.chaerunnisa@students.uii.ac.id

ABSTRACT

As Indonesia's population grows, ensuring a stable food supply becomes increasingly important. Recent changes in weather patterns have significantly impacted food production, particularly rice farming. In West Nusa Tenggara (NTB), a key area for rice production, maintaining consistent output is crucial. However, varying responses to unpredictable weather have led to significant differences in productivity across NTB's regencies and cities. This study aims to enhance the productivity of irrigated rice fields in NTB by predicting productivity levels for 2023 to 2024 using the best multilayer perceptron (MLP) model. We will compare 5 MLP model architectures to identify the optimal model for the prediction process. We will use the prediction results to cluster areas regionally through the self-organizing map (SOM) algorithm. We used the Davies-Bouldin Index (DBI) to determine the optimal number of clusters. This research compared DBI values for cluster counts of 2, 3, 4, and 5, determining the optimal cluster number by the smallest DBI value. The lowest DBI is 0.391 observed for 3 clusters. From this clustering, Cluster 1 consists of 7 regencies/cities with the lowest productivity level, Cluster 2 contains 1 regency with a moderate productivity level, and Cluster 3 includes 2 regencies/cities with the highest productivity level. The study concludes that the 7 regencies/cities in Cluster 1, identified as having low productivity require greater focus from local governments to optimize land area and paddy yields to enhance productivity in those areas.

Keywords:
Multilayer Perceptron; Self-Organizing Map; Davies-Bouldin Index.

Introduction

The conversion of almost every green open space into rice fields or paddies makes Indonesia one of Southeast Asia's rice granaries. Although Indonesia is known for its stable food security, there are still many threats to food production, especially rice. These threats include the high cost of rice production per hectare, frequent pest attacks, and unpredictable extreme weather (Fahrezi et al., 2023). Food security is one of the government's development and poverty reduction objectives. Rice is critical to global food security because it is the staple food for more than half of the world's population. Almost all Indonesians consume rice as their primary food—the demand for rice increases yearly with Indonesia's growing population. Despite being the third-largest rice producer in the world, Indonesia still imports rice to meet the needs of its ever-increasing population (Hutajulu et al., 2022).

The high domestic demand for rice and the low productivity of this essential food commodity have led the government to continue importing it. According to CNBC Indonesia, in 2023, Indonesia issued permits to import 35 tons of rice from other major rice-producing countries (Yanwardhana, 2023). Contrary to the national policy, the Regional Government of West Nusa Tenggara (NTB) stated there is no option to accept imported rice because the current rice food security position in NTB is surplus and can still serve as a national rice granary.

However, differences in population density and climate across NTB's regencies and cities result in unequal production of high-quality rice. Additionally, there is a warning from the Minister of Agriculture regarding the El Niño phenomenon, which will cause droughts lasting until 2025. El Niño is a condition in which the dry season feels hotter and drier than usual, potentially affecting the quality of harvested rice due to reduced water content (Imansyah, 2023). Drought has affected nine regencies/cities in NTB, with eight at the alert level and one at drought emergency response status.

The NTB regional government needs to plan appropriate programs early to minimize the worst-case scenarios regarding the availability of staple foods. Therefore, the author plans to forecast the productivity levels of rice paddies in each regency/city in NTB, providing the NTB Regional Government with a projection for the implementation of appropriate anticipatory programs. However, the geographical differences across NTB regencies and cities necessitate different approaches to addressing these issues. Forecasting is necessary to predict future rice productivity levels. Given that rice is a staple food for Indonesians, particularly in NTB, its fluctuations are significant. The unpredictable weather in NTB and the recent increase in land reclamation activities transforming rice fields into residential areas influence this fluctuation in productivity levels. This study will discuss the Multilayer Perceptron (MLP) method, one of many methods available for forecasting analysis.

Based on its characteristics, the MLP method has advantages in determining better weight values than other methods (Hermanianto et al., 2017). The MLP method can solve linear and nonlinear pattern problems without prior knowledge, and its learning algorithm is simple to implement (Marwala & Lagazio, 2018). Additionally, there is literature on the MLP method comparing its performance in forecasting and classifying data across 30 different studies, concluding that the MLP method's performance in learning data patterns is excellent, with an average modeling accuracy of 91.98% (Pardede et al., 2022).

The region of NTB consists of two large islands, Lombok and Sumbawa, and several small islands around Lombok and Sumbawa. The distance between the provincial capital of West Nusa Tenggara, located in the city of Mataram on Lombok Island, and the neighboring island is quite far, resulting in unequal regional development across various sectors of life in West Nusa Tenggara, including the agricultural sector. Sorting each regency or city according to their monthly productivity figures for 2023–2024 is one method to deal with this. This study profiles regency/city groups using Self-Organizing Maps (SOM), one of the Artificial Neural Networks (ANN) methods. This study employs SOM clustering as an unsupervised learning analysis, meaning it operates without supervision throughout the learning process. SOM can represent data in a two-dimensional topological map, facilitating the analysis of data relationships and patterns (Kartikasari, 2021).

Methods

In this study, the flowchart illustrated in Figure 1 uses forecasting and clustering methods. Both methods are Artificial Neural Network methods, namely MLP, used to forecast with various combinations of hidden layer nodes that will be compared to each other for modeling accuracy. The other one is Self-Organizing Maps, which are used to cluster the forecasted data, with the Davies-Bouldin Index as the evaluation criteria to determine the optimal number of clusters.

This study utilized secondary data on the productivity levels of irrigated rice fields from the raw data archives of the BPS of NTB Province. The Integrated Statistics Center of the BPS NTB issued an Official Data Usage Permit (Surat Perjanjian Penggunaan Data/SPPD) to acquire this dataset, which is not widely available. Data were collected in February 2023..

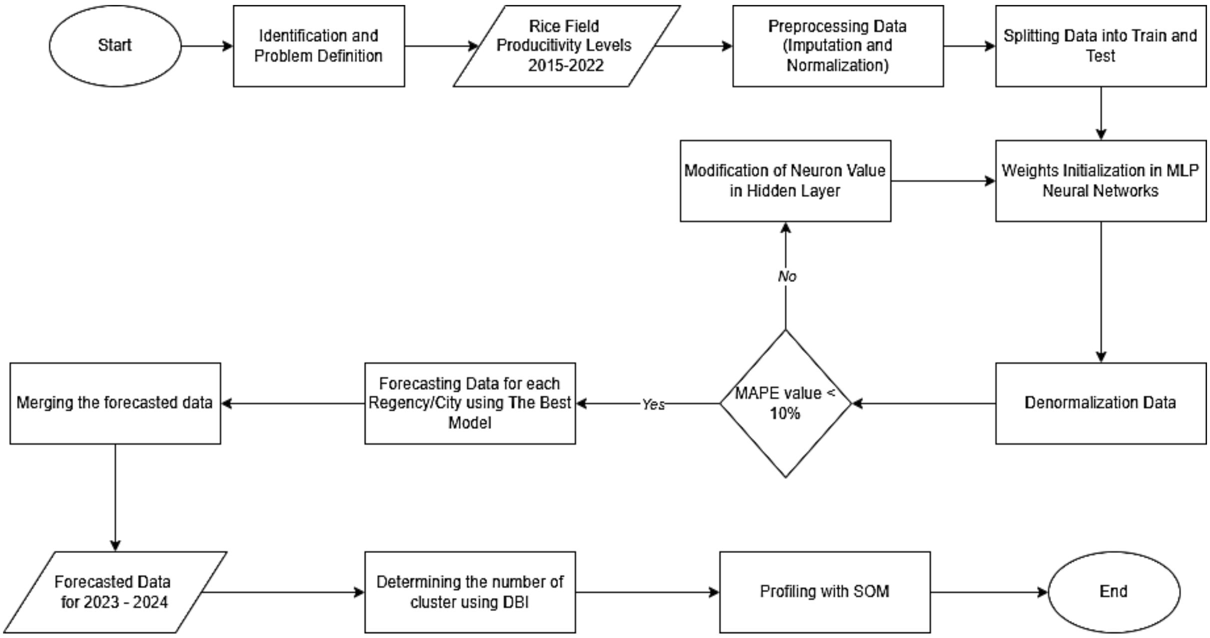

Figure 1. Research Flowchart

Multilayer Perceptron

The most commonly used Neural-Network for data prediction is Feed-Forward MLP. This method consists of one or more hidden layers located between the input and output layers. The networks with multiple layers can solve more complex problems than the single layer networks, owing to their sophisticated learning capabilities (Ardilla, 2016). Feed-forward is a learning algorithm used by Artificial Neural Network (ANN) methods, including the MLP. Typically, MLP has three or more layers: the input layer, hidden layer(s), and output layer, as presented in Figure 2.

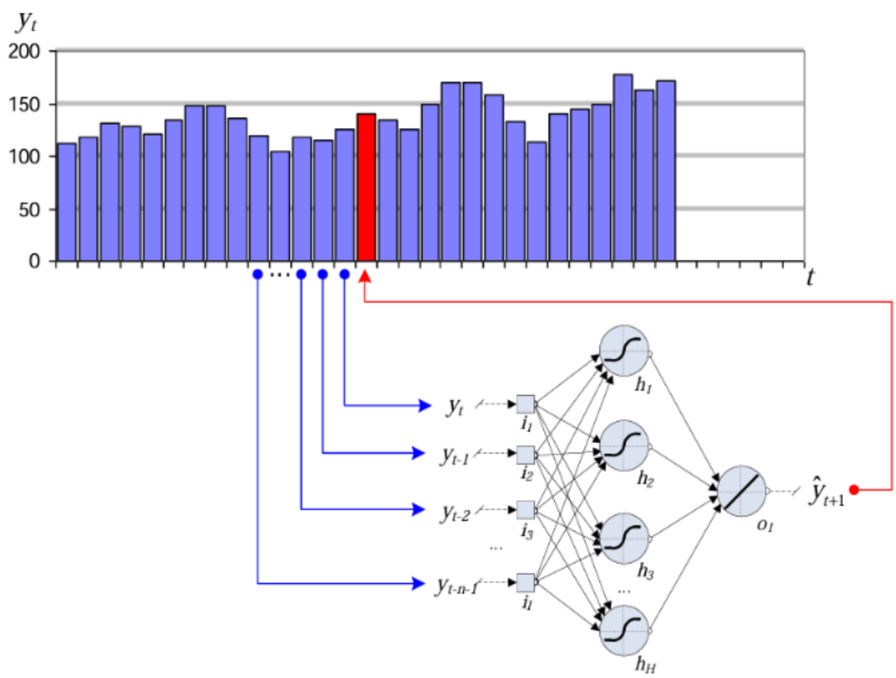

Figure 2. MLP's Architecture

The process begins with each neuron of the input layer receiving the input value, then the input units are passed to the hidden layer(s). In the hidden layer, the value from the input layer will be calculated using the activation function in each neuron of the hidden layer, and after the calculation, the results are passed to the next layer until a predicted value is obtained. The calculation of the predicted value is defined as

$$\hat{y}_{t+1} = \beta_0 + \sum_{h=1}^{H} \beta_h g\left(\gamma_{0i} + \sum_{i=1}^{I} \gamma_{hi} y_i\right) \tag{1}$$

where $\hat{y}_{t+1}$ is the forecasted value with γ_{hi} is weights on the input layer and β_h is weight on hidden layer. The 0 denotes the bias feature of each layer. The $g(\cdot)$ is a transfer function.

The activation function commonly used in MLP methods is the logistic-sigmoid function.

$$\sigma(x) = \frac{1}{1+e^{-x}} \tag{2}$$

The MLP approach incorporates many non-linear functions, which are more intricate than a single perceptron. The wholly connected arrangement, in which every neuron is connected to every other neuron in the layer below, is the most frequently used in MLP (Gouravlohar, 2024).

Mean Absolute Percentage Error (MAPE)

This research's error measurement uses mean absolute percentage error (MAPE). MAPE is a method that calculates the error of predicted value compared to the actual data in percentage form. The smaller MAPE value indicates a more accurate prediction value.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left(\left|\frac{Y_t - \hat{Y}_t}{Y_t} \times 100\right|\right) \tag{3}$$

The model is evaluated using objective criteria based on MAPE value. In this research, we used the criteria of Lewis (1982); there are four levels of accuracy of MAPE including highly accurate forecasting (< 10%), good forecasting (10% - 20%), reasonable forecasting (20% - 50%), and inaccurate forecasting (>50%) respectively.

Davies-Bouldin Index (DBI)

The Davies-Bouldin Index is usually used to evaluate time-series clustering to calculate how tightly a cluster is formed. DBI is the ratio of the intra-cluster distance to the inter-cluster distance (Guthikonda, 2005). The intra-cluster must be lower while the inter-cluster is higher to achieve an optimal value for good clustering results (Ashari et al., 2022):

Step 1 : Calculating the intra-cluster distance as the standar deviation between each data on a cluster with the center of cluster

$$S_i = \left\{\frac{1}{T_i}\sum_{j=1}^{T_i}\left|X_j - A_i\right|^2\right\}^{\frac{1}{2}} \tag{4}$$

where S_i denotes intra-cluster distance, the $X_j = [X_1, X_2, \dots, X_{T_i}]$ is vector value cluster which the T_i is amount of data in each cluster, and A_i is the center of each cluster.

Step 2 : Calculate the inter-cluster distance between each Cluster pair.

$$M_{ij} = \left|A_i - A_j\right|^2 \tag{5}$$

Step 3 : For each pair of cluster, the dispersion is summed and then divided by inter-cluster distance.

$$R_{ij} = \frac{S_i + S_j}{M_{ij}} \tag{6}$$

Step 4 : Determine the maximum ratio between pair of cluster

$$R_k = \max_{i \neq j} R_{ij} \tag{7}$$

Step 5 : Calculating the DBI value as an average of maximum similarity clusters measure

$$DBI = \frac{1}{K} \sum_{k=1}^{K} R_k \tag{8}$$

Assume there is a K pair of clusters, where R_k denotes the maximum ratio between pairs of cluster.

Self-Organizing Map

Kohonen created the Self-Organizing Map algorithm (SOM), a particular kind of artificial neural network used for clustering. It comprises a grid of nodes, each representing a location in the multi-dimensional input space. During clustering, a new point is categorized with the node that is its closest neighbor. The grid is educated so that nodes near one another look more similar than nodes connected by a lengthy path. Because of this, the grid has topological information, and one training point might affect several nodes (Van Gassen et al., 2015).

Results and Discussions

The rice field productivity data are measured in quintals per hectare (Ku/Ha) and recorded monthly from 2015 to 2022. The diagram shows that productivity levels show significant fluctuations in each regency/city. However, at the 36 first periods, the plot shows the same data for all the regencies/cities as presented in Figure 3.

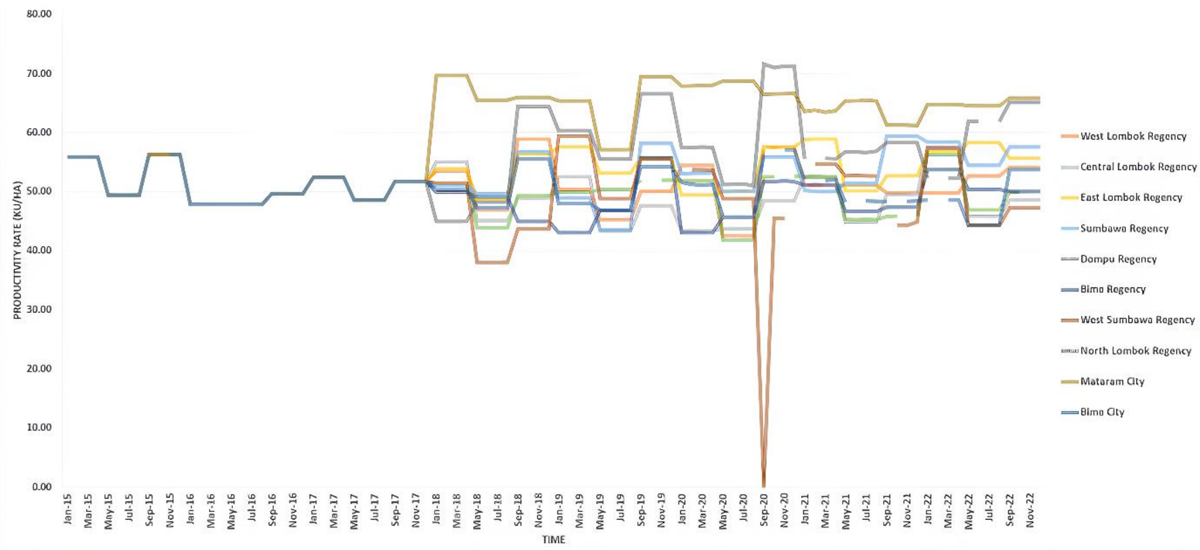

Figure 3. Productivity Levels of Rice Field in NTB

In Figure 3, there are several empty data points. This indicates the presence of missing values in several regencies/cities. Therefore, the data needs to go through the preprocessing step first. The first step is the data cleaning with imputation the mean value through the missing data point. If the data no longer has missing values, the next step is normalizing the dataset into the range [0,1] using the min-max scaling method.

After the dataset has normalized, divide it into train-test sets with an 80:20 ratio. The train set contains 77 rows of data and the test set contains 19 rows. The data forecasting process using MLP. We set 5 models with different combinations of hidden layers. One of those combinations is the 12-5-3-1 architecture of MLP, which means the input layer has 12 neurons representing the frequency of data. The first hidden layer has 5 neurons, the second has 3 neurons, and 1 is in the output layer. The MLP architecture is obtained from these combinations, as in Figure 4.

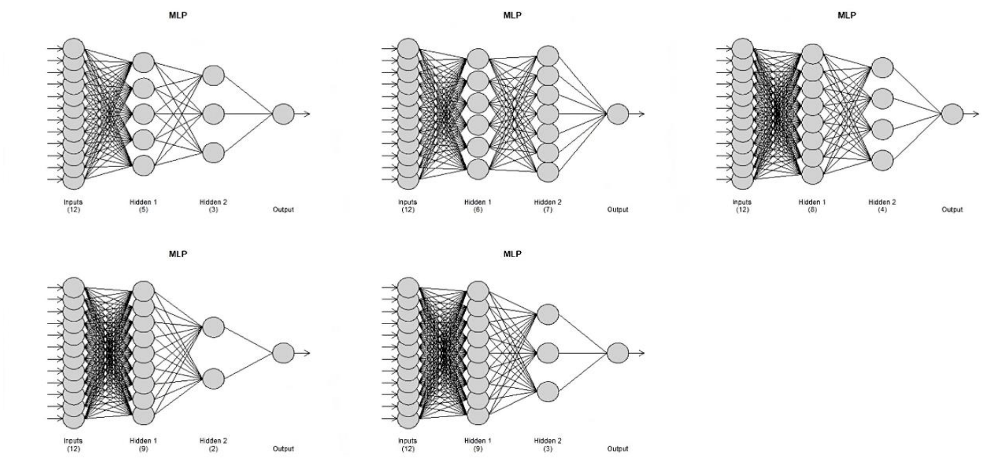

Figure 4. Architecture of MLP Models

After building the five models, they are compared by looking at the smallest MAPE value. The MAPE value, calculated using training data, measures how well a model trains a network on a data pattern. The MAPE values for training data in each regency/city are presented in Table 1.

Table 1 MAPE Value of Train Data

Regency/City	Hidden Layers				
	(5-3)	(6-7)	(8-4)	(9-2)	(9-3)
West Lombok Regency	0.18775	0.24163	0.16225	0.09903	0.11198
Central Lombok Regency	0.12804	0.10599	0.06691	0.06620	0.08652
East Lombok Regency	0.08306	0.06030	0.04158	0.03975	0.03999
Sumbawa Regency	0.40457	0.37982	0.26699	0.27449	0.24172
Dompu Regency	0.20191	0.17309	0.15173	0.09430	0.11307
Bima Regency	0.11994	0.13722	0.07181	0.08074	0.08923
West Sumbawa Regency	50.44055	42.84351	33.30865	41.05781	41.70621
North Lombok Regency	0.29726	0.23612	0.17892	0.20031	0.19641
Mataram City	0.61481	0.22034	0.15134	0.24025	0.17702
Bima City	0.20407	0.23234	0.11059	0.10101	0.10089

However, the best model to use in the following forecasting process can be determined by comparing MAPE values with the testing data. This MAPE testing comparison is carried out because the primary purpose of dividing data into training and testing divisions is to see the validation of a model using a model with testing data so that the best model will be determined based on the smallest MAPE value on the testing data as presented in Table 2.

Based on Table 2, the best model for forecasting the productivity rate is the model that has the smallest MAPE values than the other. In more detail, the optimal model for West Lombok is the MLP (12-9-3-1), whereas for Central Lombok and Bima City, the best model is the MLP (12-8-4-1). The MLP (12-9-2-1) is the preferred model for East Lombok and Mataram City, while the best model for the remaining five regencies is the MLP (12-6-7-1). Paddy productivity values were forecast after determining the best model for each regency/city. This forecasting was carried out to predict the value of the following 24 periods, namely January 2023 to December 2024. The forecasting results are presented in Figure 5.

Table 2 MAPE values of Test Data

Regency/City	Hidden Layers				
	(5-3)	**(6-7)**	**(8-4)**	**(9-2)**	**(9-3)**
West Lombok Regency	0.00264	0.00196	0.00234	0.00216	0.00134
Central Lombok Regency	0.05149	0.03456	0.02728	0.03114	0.05362
East Lombok Regency	0.00514	0.00534	0.00914	0.00461	0.00842
Sumbawa Regency	0.02198	0.01699	0.02262	0.02294	0.01844
Dompu Regency	0.06418	0.03303	0.03696	0.08954	0.04936
Bima Regency	0.00462	0.00391	0.00417	0.00558	0.00452
West Sumbawa Regency	0.00977	0.00427	0.00697	0.00641	0.00568
North Lombok Regency	0.0272	0.01975	0.02657	0.02736	0.02835
Mataram City	0.00301	0.00327	0.00325	0.00254	0.00247
Bima City	0.04658	0.02694	0.01499	0.03554	0.02052

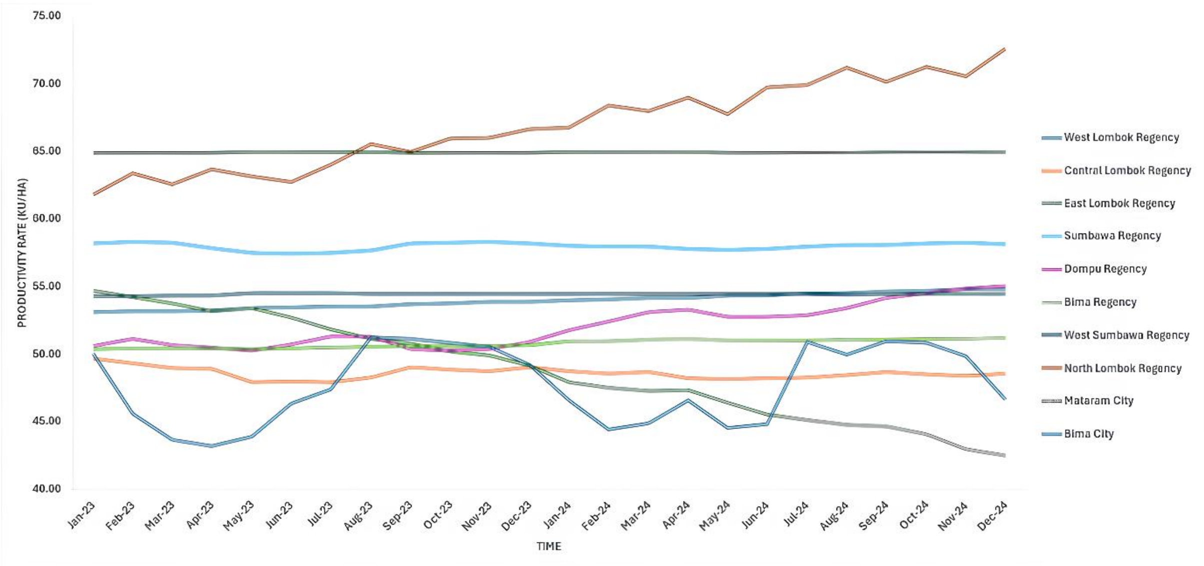

Figure 5. Forecasted Data for 2023 to 2024

Figure 5, An overview of the forecasted productivity levels of paddy fields in each region of NTB Province, using the optimal model, reveals various graph fluctuations. In West Lombok, Dompu, Bima, and North Lombok regencies, the forecasted results show an upward trend from the start to the end of the prediction period. In contrast, East Lombok Regency displays a downward trend in paddy productivity forecasts from January 2023 to December 2024. Meanwhile, the forecasted values in Bima City show highly significant changes in each period. The forecast results for the remaining regencies/cities are stable, with the graphs fluctuating within a specific range without significant changes from one period to the next.

After obtaining the forecast data for January 2023 to December 2024, further analysis is conducted to group each regency into clusters using the Self-Organizing Map method. As shown in Figure 5, the horizontal axes, P1, P2, etc., present the period of forecast. We have P1 for January 2023, P2 for February 2023, and P24 for December 2024. Before clustering, the optimal number of clusters is determined using the Davies-Bouldin Index (DBI). The DBI values for each number of clusters are presented in Table 3.

Table 3 Davies-Bouldin Index

n-Cluster	DBI
2	0,57042
3	0,39124
4	0,57140
5	0,43306

Table 3 presents the DBI values for 2 to 5 clusters. The smallest DBI value determines the optimal number of clusters. Table 3 indicates that the smallest DBI value is for three clusters. Therefore, the optimal number of clusters is three. Using this information, the analysis proceeds by grouping the regencies into three clusters using the SOM method.

After determining the optimal number of clusters, the next step is to analyze the regional grouping using Self-Organizing Maps (SOM). The SOM is trained through several iterations. The SOM iteration process will stop when the membership within a cluster converges or no longer changes. In this case, the SOM training was conducted using up to 1000 iterations, as presented in Figure 6.

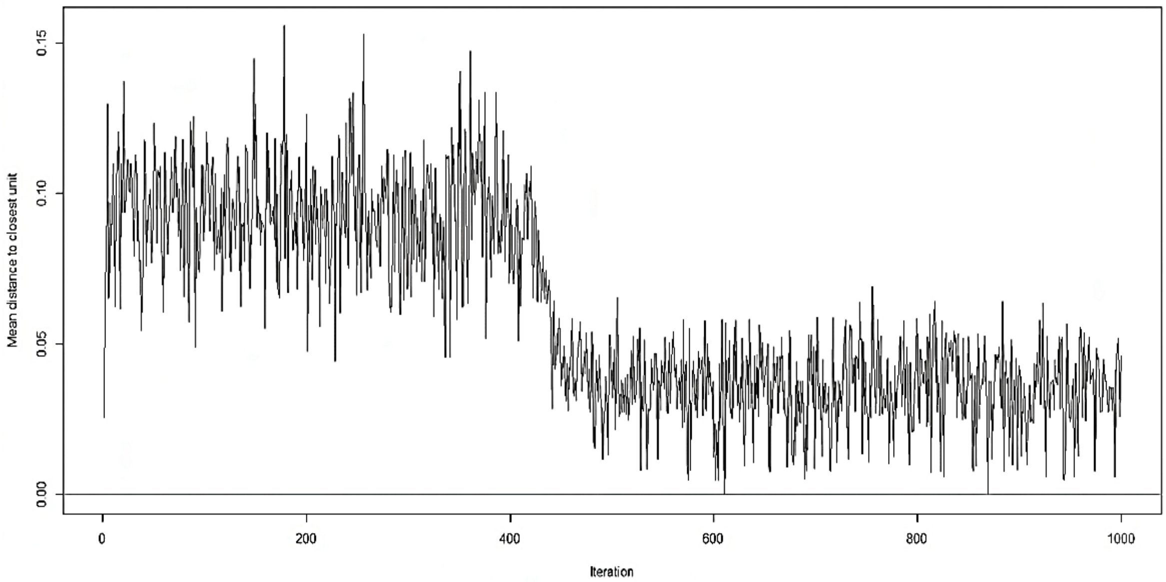

Figure 6. SOM's Training Progress

This training progress is conducted to determine the amount of time required for a cluster to reach optimality. The more iterations performed during training, the smaller the intra-cluster distance becomes, resulting in better clustering outcomes. From Figure 6, it can be observed that after approximately 500 iterations, the training begins to stabilize with a mean distance of the closest unit below 0.05. After 600 iterations, the average intra-cluster distance reaches 0.00.

In Figure 7, a line diagram represents the time frames of the 24-period forecast data for each regency/city. The diagram uses a hexagonal display with a 3×3 grid size. The line diagram shows the values of the corresponding variables, namely the rice paddy productivity values and the time variable as the X variable. In the diagram, three different colors represent the clustering results for each regency/city, with each color indicating different characteristics.

Table 4 shows that Cluster 1 consists of 7 regencies/cities, Cluster 2 consists of 1 regency, and Cluster 3 consists of 2 regencies or cities. Next, the average for each period is calculated from these clusters to determine the predicted productivity levels for P1 to P24, representing the forecasting period, January 2023 through December 2024—the total of each cluster's average as a profiling cluster presented in Table 5.

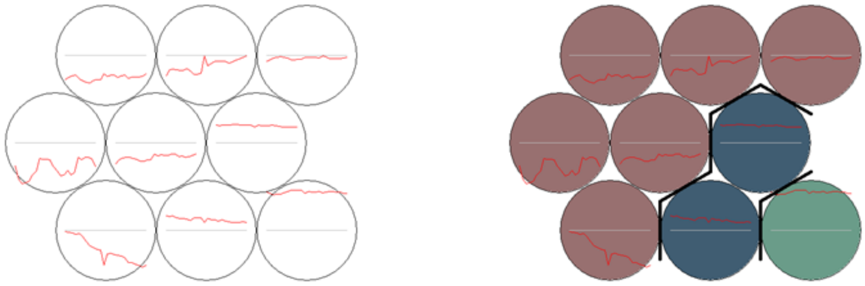

Figure 7. Self-Organizing Maps for Time Series

Table 4 Members of each cluster

Cluster	Amount Of Members	Members
1	7	West Lombok Regency, Central Lombok Regency, East Lombok Regency, Dompu Regency, Bima Regency, West Sumbawa Regency, and Bima City
2	1	Sumbawa Regency
3	2	North Lombok Regency and Mataram City

Table 5 Profiling Cluster

Variable	Average (Ku/Ha)		
	Cluster 1	Cluster 2	Cluster 3
P1	51.83	58.17	63.36
P2	51.17	58.27	64.13
P3	50.72	58.23	63.75
P4	50.53	57.84	64.29
P5	50.54	57.46	64.04
P6	50.88	57.44	63.84
P7	50.99	57.49	64.48
P8	51.50	57.65	65.25
P9	51.43	58.20	64.93
P10	51.29	58.25	65.44
P11	51.21	58.29	65.46
P12	51.03	58.19	65.79
P13	50.62	58.03	65.85
P14	50.34	57.97	66.65
P15	50.53	57.94	66.46
P16	50.74	57.76	66.93
P17	50.24	57.74	66.32
P18	50.16	57.76	67.33
P19	51.02	57.94	67.41
P20	50.94	58.04	68.06
P21	51.23	58.08	67.54
P22	51.16	58.21	68.07
P23	50.92	58.22	67.73
P24	50.45	58.15	68.76
Total	1221.47	1391.32	1581.88

Table 5 shows that cluster 1 has seven regencies/cities with the smallest average cluster value. Cluster 2 consists of one regency with a medium cluster average; in cluster 3, there are 2 regencies/cities with the highest average value. Thus, based on the results of this grouping, it is known that there are seven regencies/cities that are predicted to have a low level of irrigated rice field productivity until the end of 2024. This can be a concern for the local government to be able to increase efforts to increase irrigated rice field production in every green open land so that the level of irrigated rice field productivity can increase in future periods. Given that the annual increase in population necessitates the periodic maintenance of rice quality and quantity, the local government must act.

Conclusion

According to this research, the predicted paddy field productivity levels in each region of NTB Province exhibit various fluctuations in their graphs. Based on the predicted levels of irrigated rice field productivity data for 2023–2024, the grouping analysis shows that three clusters are the best number for putting NTB regencies and cities together. The smallest DBI value demonstrates this. The SOM algorithm resulted in regional clustering with the following distribution: cluster 1 includes 7 regencies/cities, cluster 2 includes 1 regency, and cluster 3 includes 2 regencies/cities. Based on the profiling of the three clusters formed, it is identified that Cluster 3 has the highest productivity in paddy fields, followed by Cluster 2 with moderate productivity levels, and lastly, Cluster 1, which has relatively low productivity.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

Ardilla, Y. (2016). *Metode Hibrida ARIMA dan Multilayer Perceptron untuk Peramalan Jangka Pendek Konsumsi Listrik di Jawa Timur*. Institut Teknologi Sepuluh Nopember.

Ashari, I. F., Banjarnahor, R., Farida, D. R., Aisyah, S. P., Dewi, A. P., & Humaya, N. (2022). Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies. *Journal of Applied Informatics and Computing*, *6*(1), 07–15. https://doi.org/10.30871/jaic.v6i1.3485

Fahrezi, D. S. B., Hadianto, A., & Raswatie, F. D. (2023). Dampak Perubahan Iklim terhadap Produksi Padi di Indonesia. *UT - Resources and Environmental Economic*.

Gouravlohar. (2024). *Feedforward neural network*.

Guthikonda, S. M. (2005). *Kohonen Self-Organizing Maps*. Wittenberg Universit.

Hermanianto, A. B., Buono, A., & Akhmad, F. (2017). *Optimasi Multi-Layer Perceptron Pada Model Prediksi Karakteristik Musim Hujan Dan Kemarau Di Kabupaten Pacitan*.

Hutajulu, C. P., Sugiarto, D., & Sholihah, B. (2022). Peramalan Pasokan Beras Dengan Menggunakan Metode Holt-Winter dan Multilayer Perceptron. *Journal of Intelligent Informatics*, 1–6.

Imansyah, N. (2023). *Pemda tak ada opsi beras impor masuk NTB*. ANTARA Kantor Berita Indonesia.

Kartikasari, M. D. (2021). Self-Organizing Map Menggunakan Davies-Bouldin Index dalam Pengelompokan Wilayah Indonesia Berdasarkan Konsumsi Pangan. *Jambura Journal of Mathematics*, *3*(2), 187–196. https://doi.org/10.34312/jjom.v3i2.10942

Lewis, C. D. (1982). *Industrial and business forecasting methods*. Butterworth Scientific.

Marwala, T., & Lagazio, M. (2018). Multi-layer Perceptron and Radial Basis Function for Modeling Interstate Conflict. In *Militarized Conflict Modeling Using Computational Intelligence* (pp. 43–64). Springer. https://doi.org/https://doi.org/10.1007/978-0-85729-790-7_3

Pardede, D., Hayadi, B. H., & Iskandar. (2022). Kajian Literatur Multi Layer Perceptron Seberapa Baik Performa Algoritma Ini. *Journal of Ict Aplications and System*, *1*(1), 23–35. https://doi.org/10.56313/jictas.v1i1.127

Van Gassen, S., Callebaut, B., Van Helden, M. J., Lambrecht, B. N., Demeester, P., Dhaene, T., & Saeys, Y. (2015). FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A*, *87*(7), 636–645. https://doi.org/10.1002/cyto.a.22625

Yanwardhana, E. (2023). *RI Tambah Impor Beras 1,5 Juta Ton, Ini Alasan Pemerintah*. CNBC

Indonesia.