International Journal of Innovative Computing, Information and Control
Volume 18, Number 2, April 2022

ICIC International ©2022 ISSN 1349-4198
pp. 561–576

SPATIAL DATA MODELING USING MADM FOR CLASSIFICATION OF FOOD SELF-SUFFICIENCY REGIONS

Anik Vega Vitianingsih[1,*], Robert Marco[2], Anastasia Lidya Maukar[3]
Erri Wahyu Puspitarini[4] and Seftin Fitri Ana Wati[5]

[1]Department of Informatics
Universitas Dr. Soetomo
Jl. Semolowaru 84, Surabaya 60118, Indonesia
*Corresponding author: vega@unitomo.ac.id

[2]Department of Informatics
Universitas Amikom Yogyakarta
Jl. Ring Road Utara, Condong Catur, Sleman, Yogyakarta 55281, Indonesia
robertmarco@amikom.ac.id

[3]Department of Industrial Engineering
President University
Jl. Ki Hajar Dewantara, Kota Jababeka, Cikarang Baru, Bekasi 17550, Indonesia
almaukar@president.ac.id

[4]Department of Informatics Engineering
STMIK Yadika Bangil
Jl. Bader No. 9, Kalirejo Bangil, Pasuruan 67153, Indonesia
www.erri@stmik-yadika.ac.id

[5]Department of Information Systems
UPN "Veteran" Jawa Timur
Jl. Rungkut Madya No. 1, Surabaya 60294, Indonesia
seftin.fitri.si@upnjatim.ac.id

Received July 2021; revised January 2022

Abstract. *A population increase without equivalent rice production can lead to a decrease in food security. Efforts are required to identify agricultural land for its self-sufficient rice field areas. It is presented in this research how spatial data modeling can be used to categorize and predict food self-sufficiency zones utilizing Multi-Attribute Decision-Making (MADM) technology on Geographical Information System (GIS) technology. The classification of food self-sufficient areas uses the Weighted Product (WP) method applying multi-attribute parameters of agricultural production, total food demand, and the area of the agricultural sub-districts. The Naïve Bayes method predicts food self-sufficiency based on several parameters: seed type, fertilizer, season, and terrain type. The results of the method test show superiority in classifying food self-sufficient areas by having an average coefficient value in the kappa index test of 0.78. The trial results conclude that it was determined that this method has a high degree of agreement strength when used for spatial data analysis of the food self-sufficient areas classification utilizing the MADM methodology.*
Keywords: GIS, Spatial data modeling, Food self-sufficiency, MADM, WP, Naïve Bayes

DOI: 10.24507/ijicic.18.02.561

1. **Introduction.** Rice is a staple food in many countries throughout the world, and it is one of the most widely consumed grains in the world. Because of this, mapping rice fields in a timely and efficient manner is critical to ensuring agricultural sustainability and food security in the future. The agricultural land mapping remains challenging in fragmented landscapes, such as rice-growing areas, because the information on rice farming areas is still dominated by small-scale agriculture compared to large-scale agriculture. Thus, land use is one of the functions in accelerating the production of agricultural products aimed at meeting food needs and improving people's welfare [1]. Based on the Food and Agriculture Organization (FAO) survey, it is estimated that the growth rate of agricultural production declines to 1.5% between 2015 and 2030, further to 0.9% between 2030 and 2050. Thus, it is necessary to apply a spatial pattern to producing information on the distribution/mapping of rice fields, which is very much needed as a strategic policy of food security [2].

Spatial data analysis is essential for monitoring and controlling agricultural land mapping. In recent decades, there has been an increase in research interest in presenting MADM-based models for assessing spatial data in domains such as healthcare [3,4], agriculture [5,6], and population [7]. It was developed based on climatic, soil, and topographical conditions to determine the rank of various suitability factors and weights as a map of the suitability of production and rice fields [8]. In order to determine the appropriateness of rice farming land based on spatial climate maps, researchers employed Extracting Criteria Maps for Agro-climatic Zoning and weighted overlay as a spatial analysis technique, which was also applied in determining the suitability of other crops [9].

In geocoding and mapping GIS, spatial data modeling is the act of analyzing spatial data in order to design a decision-making system that is utilized for stakeholder policy development and implementation [10,11]. At present, the rapid development of the GIS through the integration process and precise analysis can be performed using different methods. The model approach uses MADM to determine the factors and their weights for mapping the suitability of rice farming land [12], such as analytical hierarchical process [13-15], and simple additive weighting [16]. Meanwhile, modeling and analyzing spatial patterns through a machine learning-based Artificial Intelligence (AI) algorithm used for mapping the suitability of rice farming land, includes Naïve Bayes and radial basis function networks [12], decision tree [17], Bayesian [18], support vector machine and random forest [19].

The suitability analysis of land mapping and the preparation of land use maps using GIS is the most practical application in land resource planning and management [20]. GIS technology has been widely used in evaluating the suitability of agricultural land mapping because it leads to the rapid creation of static maps and map estimates by combining several information data to produce a layer suitability map [20-23]. Based on previous research, GIS technology uses spatial analysis to identify agricultural land suitability with spatial, temporal, and spatial-temporal methods. The development of sustainable rice was analyzed by integrating the logistic regression and multi-criteria land evaluation, such as characteristics of local land-use conversions [24]. An agricultural spatial data-driven Bayesian autoregressive framework was utilized to create a predictive smoothing model for the Self-Sufficiency Index (SSI) as a subset of clusters, which was then used to test the model's predictions [18,24]. However, the approach and parameters offered in this study, namely, the multi-criteria parameter approach, were not used in earlier studies to explore the need for supporting factors in the analysis process. AI using mathematical modeling is suitable to produce a mapping distribution of agricultural land areas with multi-class classification and experts to determine criteria, weighting, and ranking attributes.

According to the most relevant literature and theory of the methodologies utilized in this study [25], the categorization of agricultural land mapping areas based on food self-sufficiency status is the most appropriate classification. Several literature studies have attempted to improve results in scientifically mapping an area. Also, previous researchers have suggested developing mathematical models, GIS MADM methods, and AI. Thus, in the theoretical background section, we will discuss research on MADM, artificial intelligence, Geographic Information Systems (GIS), and combinations of these technologies. A variety of multi-criteria decision-making methods, including the Analytic Network Process (ANP), Simple Additive Weighting (SAW), and Vlse Kriterijumska Optimizacija I Kompromisno Resenje-Analytical Hierarchical Process (VIKOR-AHP), were used in a GIS environment to investigate an ecological model framework with the goal of selecting a suitable location for agricultural land use [16]. Another study combined Geographic Information System (GIS) technology with Multi-Criteria Decision-Making (MCDM) and the Analytic Hierarchy Process (AHP) to determine the suitability of agricultural land for crop development in a different part of the world [15]. According to this study [17], which makes use of the MCDM spatial method and the AHP-based GIS, the value of each criterion layer is calculated by multiplying the parameters for each factor obtained from the pair comparison matrix by adding weights, and the appropriate evaluation of several criterion factors affecting agricultural land is performed.

Application of the AHP approach is used to rank various appropriateness factors in order to make comparisons. The weights obtained as a result of the analysis are utilized to create a suitability map layer on the ArcGIS 10.1 platform, using the weighted sum overlay tool. Furthermore, a map is made that describes the suitability of rice production based on specific regions [8]. [19] proposed machine learning using Support Vector Machine (SVM) and Random Forest (RF) classification techniques to map the spatial distribution of rice fields in order to map the spatial distribution of rice. [18] presented a predictive smoothing model to determine the Self-Sufficiency Index (SSI) based on the Bayesian autoregressive framework by utilizing available agricultural data in each region. The researchers devised a fuzzy multi-criteria decision-making technique that was combined with Geographic Information Systems (GIS) to assess optimal rice-growing regions in the Amol District of Iran. In accordance with the FAO framework and expert opinion [20], it included soil qualities, meteorological conditions, terrain, and accessibility. In accordance with the findings of the literature study, there are still a limited number of studies that combine different methodologies for mapping agricultural land.

There are various difficulties in mapping land suitable for rice growing based on food self-sufficiency status, which is a difficult task. One issue is spatial information about the surrounding population, which is reflected in the demand for rice as a food security strategy to agricultural productivity as a result of increased agricultural productivity. Then, geographic information about the surrounding environment, the network structure, the qualities of the surrounding environment in relation to climatic conditions, and pest attacks are required, and the network structure is required. Several studies have stated that population density is the most significant criterion for food security [2,26]. Another study stated that essential factors in agricultural yield models are climate, soil properties, and water availability [27]. There is an analysis related to land suitability that must be applied in the final decision to meeting the needs and reflecting local conditions well [2,6], which is used to produce information on spatial mapping and the areas of rice fields as a strategic form of food security [2]. Previous studies have not used the proposed multi-parameter criteria for modeling spatial data with WP and Naïve Bayes methods. The authors proposed a spatial data modelling using MADM to define the mapping of agricultural areas based on the scope of food self-sufficiency category to address the challenges of mapping rice farming areas to determine food self-sufficiency status. This proposed approach is still very limited so far.

Multi-Attribute Decision-Making (MADM) approaches are commonly used to find the best solution, choose a single option, or rate options from most to least appropriate [28]. As one of the MADM methods, the Weighted Product (WP) method aims to evaluate and compare to the rest through the multiplication of ratios related to every criterion and select the most applicable alternatives [29]. This method is more straightforward and more efficient [28]. The WP method is considered suitable for both single and multi-dimensional problems/having high subjectivity [30], and produces a short calculation time [31]. In addition, the WP method has a moderate agreement strength category, which can be applied for modelling spatial data using GIS for regional classification [4]. While the use of Naïve Bayes classification in determining the class based on the hypothesis, there is no dependence between attributes in maximizing the posterior probability [32,33]. This method can quickly build simple structures without learning procedures and has a shorter computation time, resulting in higher efficiency [34]. Naïve Bayes is one of the algorithms that have advantages and outperforms many sophisticated classifications, especially when the attributes are not strongly correlated [33,35,36]. Meanwhile, limited studies combine Naïve Bayes classification with weighting features [37-39].

The results of this study could be part of an effort to observe, monitor, and control food self-sufficiency as a strategic policy of food security in developing torpical countries. The mapping results can help stakeholders, or the food security agency classify and predict self-sufficient food areas. AI is used as a framework in spatial data modeling, using GIS technology to visualize the classification of food self-sufficient areas. From implementation and testing results, it can be concluded that web-GIS applications of mapping food self-sufficiency in Mojokerto district can provide information on the productivity of rice products, determine the regional potential for self-sufficiency, and predict areas of potential self-sufficiency. The analysis results using the WP and Naïve Bayes methods based on the parameters of land area, productivity, population, irrigation system, rainfall, and agricultural equipment in the Mojokerto district show that the prediction of self-sufficiency is good. Kohen Kappa index is 0.78, and the analysis results determine the number of areas with abundant agricultural products and high self-sufficiency.

2. **Method.** When GIS and MADM approaches are used for decision-making, a powerful tool is created that may be used to handle a variety of challenges, including the selection of a feasible location [40]. Identifying the most desirable from a small number of choices based on a predefined quality [41] is a useful approach for comparative analysis. Using MADM, decision-making systems using spatial data can be equipped to do spatial data analysis [42,43]. MADM is capable of integrating and managing geographic data as well as attribute data. Agricultural land mapping classification based on food self-sufficiency status is the major data used in the spatial data modelling discussed in this section. Figure 1 shows a flowchart depicting the stages of the spatial data modelling process for classification.

Step 1: This process is necessary to determine the necessity for spatial datasets and attribute data in the spatial shapefile dataset (**.shp*). This paper uses two types of datasets, namely spatial datasets including district maps in each sub-district and quantitative attribute datasets. The base map spatial datasets of the Mojokerto Regency consist of 18 sub-districts with information coverage at the village level. The quantitative attribute dataset for food self-sufficiency spatial data modeling (Table 1) contains attributes, such as population (households/sub-district), land in hectare (Ha), productivity in quintal (= 100 kg) per hectare (Qt/Ha), Plant Pest Organisms (Pest), and Rainfall (Month). The

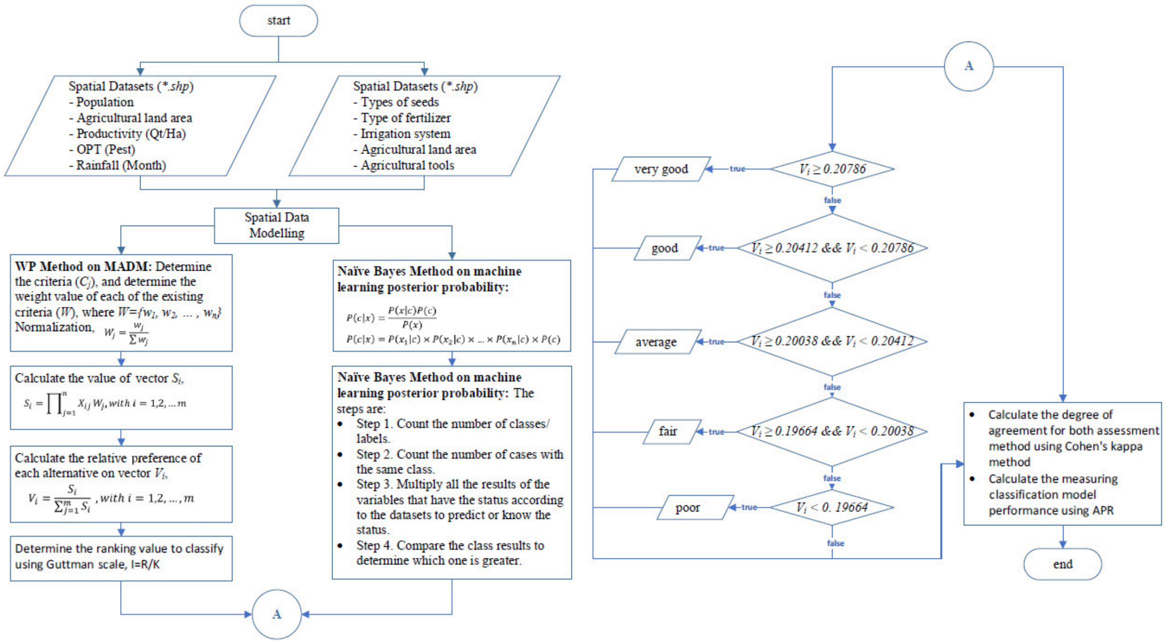


Figure 1. Flowchart of spatial data modeling for food self-sufficiency classification

quantitative attribute dataset for spatial data modelling predicting food self-sufficiency (see Table 2) contains attributes, such as types of seeds, type of fertilizer, irrigation system, agricultural land area, and agricultural tools.

Step 2: The spatial data modelling to determine food self-sufficiency areas using the WP method on the MADM model is explained in Section 2.1. The WP method is part of the MADM model in decision making which will process the criterion value of each parameter to get the V_i preference value. The spatial data modelling for predicting food self-sufficiency using the Naïve Bayes method on machine learning is explained in Section 2.2.

Step 3: Compute the ranking value to determine the classification of food self-sufficiency areas using the Guttman scale as described in Section 2.4. The classification value comprises the level of food self-sufficiency in each region, with circumstances ranging from very good, good, average, fair, and poor.

Step 4: Calculate the degree of agreement for both assessment methods using Cohen's kappa method based on the process of Section 2.5. Then, calculate the measuring of classification model performance using APR based on Section 2.6.

2.1. **Multiple-Attribute Decision-Making (MADM).**

MADM in the field of spatial analysis is part of a Multi-Criteria Decision-Making system (MCDM) and Multi-Objective Decision-Making (MODM) [44]. MADM is used for discrete retrieval, where alternative decision support systems are predetermined [45]. The Weighted Product (WP) approach is a prominent weighting method that is used as part of a decision-making system that employs MADM multi-parameter criteria to make decisions [30]. In addition, WP method has a limited set of decision alternatives that provide explanations for several decision criteria. WP method's primary process is multiplication, which is used to connect attribute ratings in situations where each attribute must be ranked with attribute weights in order to be considered. This process has similarities to the normalization process [46,47]. The weight is computed based on the level of importance. The more important, the higher the weight value, value of 1 is "very unimportant" and 5 is "very important".

The WP method approach is to assign a score to each alternative multiplied by the weighted value for each parameter attribute, with the following steps.

Step 1: Determine the criteria (C_j) of rice farming land that has the suitability status of a food self-sufficient area based on expert judgment. In MADM, using expert weight rationality directly influences the accuracy of the decision results [48].

Step 2: Determine the weight value of each existing criteria (W) or relative importance of each criterion (C_j) given by experts. The process in Equation (1) normalizes the criterion weight (W), $\sum w_j = 1$, with $W(w_1, w_2, \ldots, w_n)$ as the weighted importance value of each criterion.

$$W = \{w_1, w_2, \ldots, w_n\} \tag{1}$$

Step 3: Simplify the weight criteria according to Equation (2). Normalize or increase the weights to produce a value of $w_j = 1$ where $j = 1, 2, \ldots, n$ criteria and $\sum w_j$ is the sum of weights.

$$W_j = \frac{w_j}{\sum w_j} \tag{2}$$

Step 4: Calculate the value of vector S_i as an alternative preference based on Equation (3).

$$S_i = \prod_{j=1}^{n} X_{ij} W_j, \text{ with } i = 1, 2, \ldots, m \tag{3}$$

where S_i is the result of decisions normalization on the i-th alternative (preference criteria), and X_{ij} is an alternative rating per attribute (value of the criteria). The weight attribute is represented by W_j, and the number of criteria is represented by n. The W_j variable is the rank of positive value for the profit attribute and negative value for the cost attribute in the profit and cost attributes, respectively.

Step 5: Calculate the vector V_i value, using Equation (4), as the relative preference of each alternative on vector V by dividing each number of vector values S with the total value of vector S.

$$V_i = \frac{S_i}{\sum_{j=1}^{m} S_i}, \text{ with } i = 1, 2, \ldots, m \tag{4}$$

2.2. **Naïve Bayes.** The Naïve Bayes technique is a straightforward probability classification approach that calculates the likelihood of a new observation being classified into a predetermined category based on previous observations [34,47]. On the basis of this assumption, the classification can be estimated by computing the conditional probability density function and the posterior probability density function [49], to determine the posterior probability using Equation (5) and Equation (6) [50].

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \tag{5}$$

$$P(c|x) = P(x_1|c) \times P(x_2|c) \times \cdots \times P(x_n|c) \times P(c) \tag{6}$$

where $P(c|x)$ is defined as the posterior probability of class (c, target) given predictor (x, attribute), $P(c)$ is defined as the probability of the preceding class, and $P(x)$ is defined as the prior probability of the predictor. The $P(x|c)$ variable denotes the possibility, which is the class probability given the predictor in the case of the possibility.

2.3. **Spatial dataset.** This section explains the weighting process for various attributes using the WP method as shown in Table 1. In order to establish the level of importance/influence on the classification of each spatial dataset, a weighted value will be assigned to each one. The level of importance used for weighting in each attribute [51] is as follows: the value of X_i is 95 for category "Very good"; value 85 for category "Good";

TABLE 1. Weighting parameters of self-sufficiency attributes using WP method

Attribute	Parameter	Category	Weight value
Population (X_1)	< 500	Very good	95
	500-1000	Good	85
	> 1000	Average	75
Agricultural land area (X_2)	> 250	Very good	95
	250-200	Good	85
	200-150	Average	75
	150-100	Fair	65
	100-0	Poor	55
Productivity (Qt/Ha) (X_3)	> 90	Very good	95
	$\leq 90 - > 70$	Good	85
	$\leq 70 - > 50$	Average	75
	$\leq 50 - > 30$	Fair	65
	< 30	Poor	55
OPT (Pest) (X_4)	0-8%	Very good	95
	8%-15%	Good	85
	15%-25%	Average	75
	25%-45%	Fair	65
	$> 45\%$	Poor	55
Rainfall (Month) (X_5)	≥ 150 mm	Very good	95
	< 150 mm – ≥ 100 mm	Good	85
	< 100 mm – ≥ 50 mm	Average	75
	< 50 mm	Fair	65

value 75 for category "Average"; value 65 for category "Fair"; value 55 for category "Poor".

By analyzing the data presented in Table 2, the Naïve Bayes approach is used to calculate the weights assigned to each feature of self-sufficiency prediction.

2.4. **The Guttman scale.** When evaluating a classification, the Guttman scale can be used [52] to determine its importance. In order to draw conclusions from qualitative data [53], this scale is used as a basis for measurement [52]. It also helps to reduce uncertainty from an intervention outcome value in the projected categorization value [54]. The sort of dataset that employs scores/weights in the analysis process will produce a value based on the uncertainty factor of the class of variables described, which may be assessed using the Guttman scale [55] based on Equation (7).

$$I = \frac{R}{K} \tag{7}$$

The I variable is the result of the interval value derived from the R variable, which denotes the range of values in the data set. Very good, good, average, fair, and poor are among the potential classifications that will be generated; the K variable is the number of such classifications. As shown in this paper, the value of the R variable can be calculated by looking at the range of values between the maximum value of V_i and the V_i lowest value.

2.5. **Method consistency examination.** For determining the consistency of the two methods used in this experiment, the Cohen's kappa approach was employed. It specifies

Table 2. Weighting parameters of self-sufficiency prediction attributes using Naïve Bayes

<table>
<tr><th>Attribute</th><th>Parameter</th><th>Category</th></tr>
<tr><td rowspan="3">Types of seeds</td><td>Hybrid</td><td>Very good</td></tr>
<tr><td>Superior</td><td>Good</td></tr>
<tr><td>Local</td><td>Average</td></tr>
<tr><td rowspan="3">Type of fertilizer</td><td>Organic and Inorganic (Mix)</td><td>Very good</td></tr>
<tr><td>Inorganic</td><td>Good</td></tr>
<tr><td>Organic</td><td>Average</td></tr>
<tr><td rowspan="3">Irrigation system</td><td>Technical Irrigation Rice Fields</td><td>Very good</td></tr>
<tr><td>Semi-Technical Irrigation Rice Fields</td><td>Good</td></tr>
<tr><td>Rainfed Rice Fields</td><td>Average</td></tr>
<tr><td rowspan="5">Agricultural land area</td><td>> 250</td><td>Very good</td></tr>
<tr><td>250-200</td><td>Good</td></tr>
<tr><td>200-150</td><td>Average</td></tr>
<tr><td>150-100</td><td>Fair</td></tr>
<tr><td>100-0</td><td>Poor</td></tr>
<tr><td rowspan="3">Agricultural tools
TR2: Tractor
RT/TRAY: Rice Transplanter with tray</td><td>TR2 +RT/TRAY (Mix)</td><td>Very good</td></tr>
<tr><td>TR2</td><td>Good</td></tr>
<tr><td>RT/TRAY</td><td>Average</td></tr>
</table>

that this measurement should be utilized for qualitative data based on Equation (8) [56].

$$K = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \tag{8}$$

The measuring coefficient between the WP and Naïve Bayes methods is denoted by the K variable. The percentage of the number of consistent measurements used for comparisons between methods is denoted by the $\Pr(a)$ variable. The percentage change is denoted by the $\Pr(e)$ variable. The method, based on the range of coefficient values, gives results "poor" agreement strength if the value of the variable $K < 0.21$, "fair" for value between 0.21 and 0.40, "moderate" for value between 0.41 and 0.60, "good" for value between 0.61 and 0.80, "very good" for value between 0.81 and 1.00.

2.6. **Confusion matrix measuring model.** The confusion matrix consists of two positive and two negative classes comparing the actual and classification data [57,58] as seen in Table 3. This study uses the confusion matrix which is illustrated in Table 3 to measure the model used by obtaining the values of accuracy, precision and recall. Precision and recall are commonly defined as the ratio of correctly classified events (usually referred to as true positives in classification) to important occurrences (precision), or actual events (recall) [60].

Table 3. Confusion matrix

Actual data	Predicted classification	
	Positive (+)	Negative (−)
Positive (+)	True positives (TP)	False negatives (FN)
Negative (−)	False positives (FP)	True negatives (TN)

3. **Results and Discussion.** Table 4 represents the findings of the Guttman scale examination using Equation (9) as the result of the classification scale value using the WP method.

$$\begin{cases} \text{very good}, & \text{if } V_i \geq 0.20786 \\ \text{good}, & \text{if } V_i \geq 0.20412 \text{ and } V_i < 0.20786 \\ \text{average}, & \text{if } V_i \geq 0.20038 \text{ and } V_i < 0.20412 \\ \text{fair}, & \text{if } V_i \geq 0.19664 \text{ and } V_i < 0.20038 \\ \text{poor}, & \text{if } V_i < 0.19664 \end{cases} \tag{9}$$

Table 4. The findings of the Guttman scale examination

WP method
$R = V_{i_{Max}} - V_{i_{Min}} = 0.21160 - 0.1929 = 0.0187$ $K = 5$ and $I = \frac{0.0187}{5} = 0.00374$
Assessment very good criteria: Highest score $- I = 0.21160 - 0.00374 = 0.20786$ *Assessment good criteria: Very good criteria* $- I = 0.20786 - 0.00374 = 0.20412$ *Assessment average criteria: Good criteria* $- I = 0.20412 - 0.00374 = 0.20038$ *Assessment fair criteria: Average criteria* $- I = 0.20038 - 0.00374 = 0.19664$

Table 5 shows implementation datasets of self-sufficiency attributes assessment from Jolotundo villages.

Table 5. Weighted product implementation datasets

Village	Attributes (X)				
	Population (X_1)	Land area (X_2)	Productivity (X_3)	OPT (X_4)	Rainfall (X_5)
Jolotundo	75	95	75	96	75

Step 1: The WP method requires weights and attributes to determine food self-sufficiency.

Step 2: The decision-maker assigned the preference weights for each attribute (X_i) as in Table 6.

Table 6. Weights of each self-sufficiency attribute preferences

Weight	Attribute (X_i)					
	Population (X_1)	Land area (X_2)	Productivity (X_3)	OPT (X_4)	Rainfall (X_5)	$\sum w_i$
w	95	75	65	80	95	395

Step 3: The normalization is performed using Equation (2), and the result can be seen in Table 7.

Table 7. Result of normalization of self-sufficiency attributes

Weight	$\left(\frac{X_1}{\sum W}\right)$	$\left(\frac{X_2}{\sum W}\right)$	$\left(\frac{X_3}{\sum W}\right)$	$\left(\frac{X_4}{\sum W}\right)$	$\left(\frac{X_5}{\sum W}\right)$	$\sum w_i$
w	0.24	0.19	0.24	0.16	0.16	1.00

Step 4: Calculate S vector using Equation (3), with the elaboration of Equation (10). The result of vector calculations of each village for self-sufficiency attributes are shown in Table 8.

$$S_i = \left(X_1{}^{\wedge attribute\ weight\_x_1}\right) * \left(X_2{}^{\wedge attribute\ weight\_x_2}\right) * \left(X_3{}^{\wedge attribute\ weight\_x_3}\right) \\ * \left(X_4{}^{\wedge attribute\ weight\_x_4}\right) * \left(X_5{}^{\wedge attribute\ weight\_x_5}\right) \tag{10}$$

Step 5: Determine the preference (V_i) using Equation (4) and the results are shown in Table 8.

Table 8. Preference calculation results

Vector S_i	Vector V_i
81.03	0.21160

Preference (V_i) is used to determine the distribution of the mapping classification of food self-sufficient areas. Figure 2(a) shows the analysis map of the irrigation system of each village, with green color for technical irrigation conditions, yellow for semi-technical conditions, and blue for rain-fed conditions. Figure 2(b) shows population data for each village. The yellow color indicates population less than 1,000 households, the green for more than 1,000 but less than 2,000 households, the red for more than 2,000 but less than 3,000 households, the gray for more than 3,000 households, and the turquoise green for not populated.

Data of the rice planting area is shown in Figure 2(c). The yellow color represents the planting area less than 200 Hectares (Ha), the green for more than 200 Ha, and the gray for no rice planting area. Figure 2(d) shows each village's rice harvest productivity data. The yellow color represents the yield of less than 50 Qt/Ha, the green for more than 100 Qt/Ha, and the gray for no rice planting land area. Figure 2(e) displays the deployment of agricultural tools in every village with red color for the area with RT/TRAY tools, the yellow for TR2 only, the green for a combination of RT/TRAY and TR2, and the gray for the area with no subsidy due to no agricultural land for rice. Figure 2(f) shows the results of the self-sufficiency classification analysis using the WP method with the blue for very good self-sufficiency, the green for good, the yellow for quite good, the orange for poor, and the red for very poor.

Using the Naïve Bayes method, the potential self-sufficiency area uses five parameters: types of seeds, type of fertilizer, irrigation system, agricultural land area, and agricultural tools. Using Equations (5) and (6), the status of each village can be determined as in Table 7. For a numerical example, the calculation is performed for Village 11 with the following steps.

Step 1: Compute the probability of the appearance of "Yes" status and the appearance of "No" status.

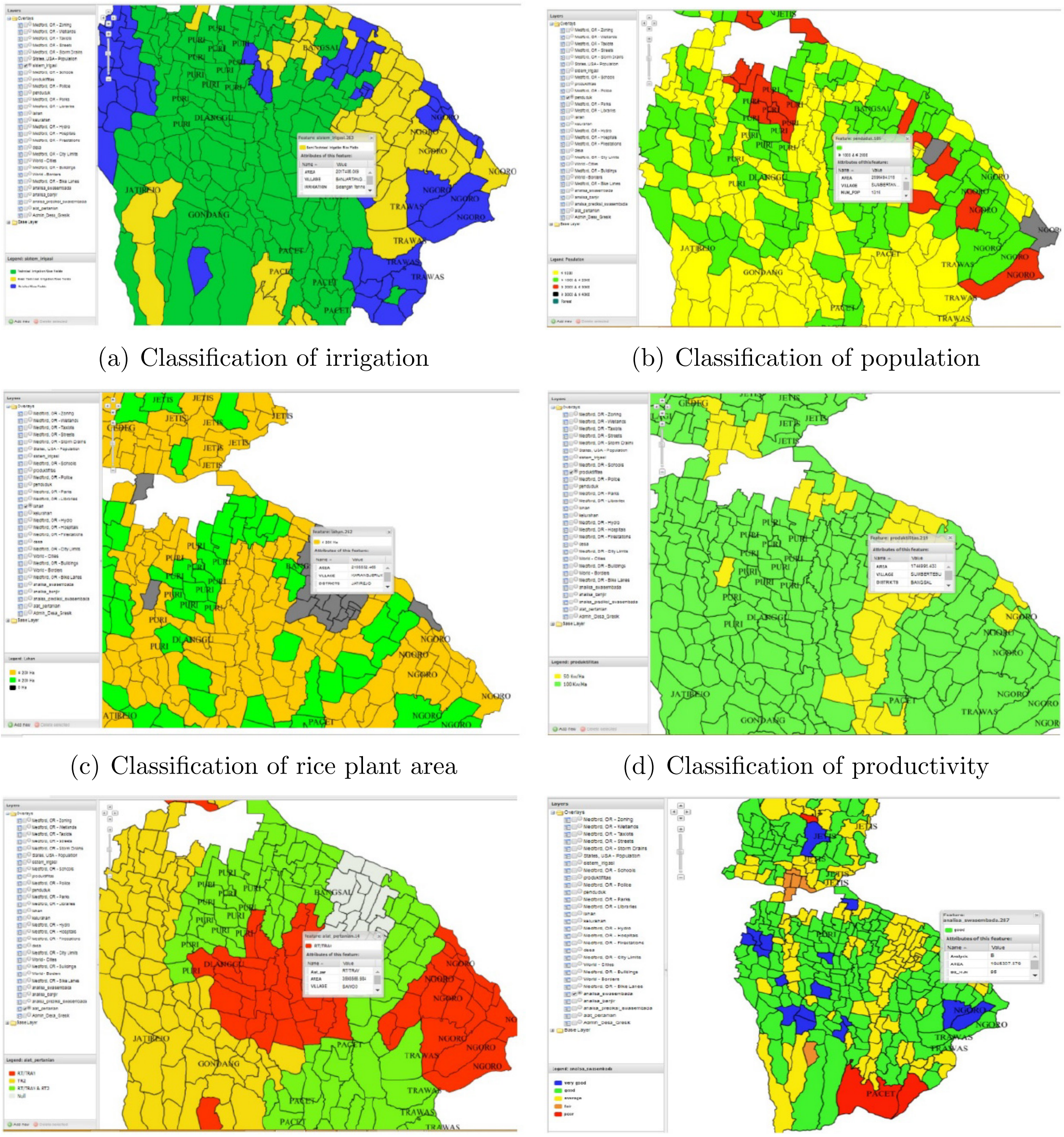
Step 2: Compute the probability of the appearance of "Yes" status when X variable is established ($P(Yes|X)$) and the probability of the appearance of "No" status when X is established ($P(No|X)$). Here, x_1 = superior; x_2 = mix; x_3 = technical irrigation; x_4 = 300-400; x_5 = TR2; x_6 = '?'.

Step 3: Compute the $P(Yes|X)$ and $P(No|X)$ using Equation (11) and Equation (12).

$$P(Yes|X) = 0.081;\ P(No|X) = 0$$

$$P(Yes|X) = P(x_1|Yes) \times P(x_2|Yes) \times P(x_3|Yes) \times P(x_4|Yes) \\ \times P(x_5|Yes) \times P(Yes) \tag{11}$$

$$P(No|X) = P(x_1|No) \times P(x_2|No) \times P(x_3|No) \times P(x_4|No) \times P(x_5|No) \times P(No) \tag{12}$$



(a) Classification of irrigation (b) Classification of population

(c) Classification of rice plant area (d) Classification of productivity

(e) Classification of agricultural tool (f) Self-sufficiency analysis results

FIGURE 2. (color online) Mapping classification results using the WP method

Step 4: Compare the $P(Yes|X)$ and $P(No|X)$. Since the $P(Yes|X)$ is greater than the $P(No|X)$, the status of Village 11 is "Yes".

Figure 3 shows blue color for areas predicted to be self-sufficient food, and yellow for being able to be self-sufficient.

Using the WP method on Equation (4) resulting Table 8 and Naïve Bayes method on Equation (5) resulting in process Equations (11) and (12), evaluate the classification performance on the analysis result. Measuring algorithm performance in classification metrics usually revolves around using precision and recall evaluation frameworks [59]. To evaluate categorical classifiers for areas of food self-sufficiency uses precision, recall, and performance metric accuracy. Precision is intended to assess the accuracy of the classification findings, whereas recall is intended to measure the completeness of the classification results. The accuracy of the categorization process, on the other hand, is the most commonly used confusion metric [61].

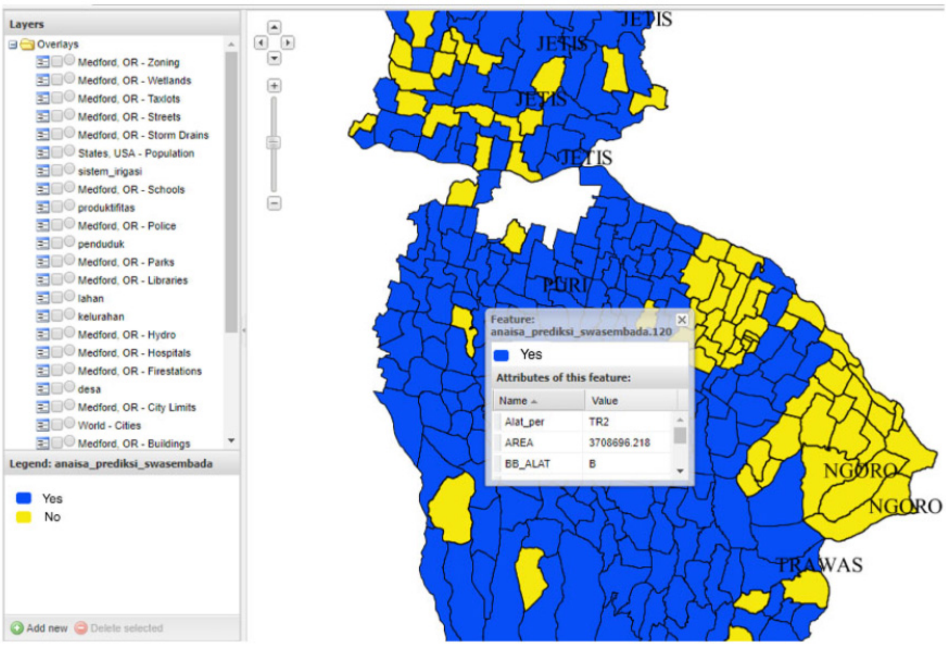


FIGURE 3. (color online) The result of mapping classification with the Naïve Bayes

TABLE 9. Food self-sufficient prediction datasets from 11 villages

Village No	Food self-sufficient prediction attributes					Status
	Types of seeds (x_1)	Type of fertilizer (x_2)	Irrigation system (x_3)	Agricultural land area (x_4)	Agricultural tools (x_5)	
1	Local	Organic	Technical Irrigation	100-200	TR2	Yes
2	Superior	Inorganic	Semi-Technical	0-100	RT/TRAY	Yes
3	Local	Organic	Rainfed	300-400	MIX	No
4	Local	Mix	Semi-Technical	> 400	TR2	Yes
5	Superior	Mix	Rainfed	300-400	RT/TRAY	No
6	Hybrid	Organic	Semi-Technical	200-300	TR2	Yes
7	Local	Inorganic	Rainfed	> 400	RT/TRAY	No
8	Hybrid	Organic	Technical Irrigation	300-400	MIX	Yes
9	Local	Organic	Semi-Technical	200-300	TR2	Yes
10	Hybrid	Mix	Technical Irrigation	> 400	MIX	Yes
11	Superior	Mix	Technical Irrigation	300-400	TR2	?

The testing of the spatial analysis for food self-sufficiency mapping application is performed by calculating the success rate of predictive analysis using the WP method. The correct predictions are 12 times out of 20 experiments. The Naïve Bayes method results in eight accurate predictions out of 15 experiments. The WP method is carried out to map food self-sufficiency using GIS. The validation of the predictive result shows 69% of precision, 85% of recall, and 75% of accuracy. Moreover, the Naïve Bayes method's precision, recall, and accuracy are 62%, 80%, and 70%, respectively.

4. **Conclusion.** This research examines the combination of WP and Naïve Bayes methods in classifying multi-attribute for spatial data modelling. The WP method on MADM allows comparative mapping results according to the priority level of importance of the parameters, weights, and priority rankings given to each multiparameter attribute in providing spatial sensitivity analysis. This paper considers quantitative data and computes

the Guttman scale classification parameter, and this research derives the V_i preference value from the WP approach and presents it. This is crucial in the decision-making process for selecting regions that are self-sufficient in terms of food production. While the Naïve Bayes method predicts the mapping of self-sufficient food areas, by maximizing the posterior probability, the method can quickly produce a structured result with a shorter processing time. The result of WP and Naïve Bayes methods combination unlocks new potential for further research in combining several different methods in spatial data modeling. Based on the test results, they have a good category agreement strength for GIS spatial data modeling to classify self-sufficient food areas. Kohen Kappa index is 0.78, and the analysis results determine the number of regions with abundant agricultural products and high self-sufficiency. The MADM method, classification method with optimization parameters, and datasets can be considered for further research for better accuracy.

Acknowledgments. This paper is collaborative research in GIS for spatial analysis and model estimation. The authors give the highest appreciation to the Department of Agriculture and Food Security, Mojokerto Regency, East Java Province, Indonesia, which has provided access to private data and multi-criteria parameters used for model testing.

REFERENCES

[1] Food and Agriculture Organization of the United Nations, *Farming Must Change to Feed the World*, http://www.fao.org/news/story/en/item/9962/icode/, 2009.

[2] X. Xiao et al., Mapping paddy rice agriculture in southern China using multi-temporal MODIS images, *Remote Sens. Environ.*, vol.95, no.4, pp.480-492, 2005.

[3] B. Zhu, Y. Fu, J. Liu and Y. Mao, Spatial distribution of 12 class B notifiable infectious diseases in China: A retrospective study, *PLoS One*, vol.13, no.4, pp.1-17, 2018.

[4] A. Vitianingsih, A. Choiron, D. Cahyono, A. Umam and S. Suyanto, Spatial data modeling on GIS for classification of measles-prone region using multiple attribute decision making, *Int. J. Intell. Eng. Syst.*, vol.12, no.3, pp.97-107, 2019.

[5] A. A. Mustafa, M. Singh, R. N. Sahoo, N. Ahmed, M. Khanna and A. Sarangi, Land suitability analysis for different crops: A multi criteria decision making approach using remote sensing and GIS, *Water Technol.*, vol.3, no.12, pp.61-84, 2011.

[6] T. Prakash, *Land Suitability Analysis for Agricultural Crops: A Fuzzy Multicriteria Decision Making Approach*, Master Thesis, 2003.

[7] M. Javari, M. Saghaei and F. F. Jazi, Analyzing the resilience of urban settlements using multiple-criteria decision-making (MCDM) models (case study: Malayer city), *Sustain. Environ.*, vol.7, no.1, 1889083, 2021.

[8] G. T. Ayehu and S. A. Besufekad, Land suitability analysis for rice production: A GIS based multi-criteria decision approach, *Am. J. Geogr. Inf. Syst.*, vol.4, no.3, pp.95-104, 2015.

[9] A. Ezra, A. A. Adebayo and A. S. Umar, Application of multi-criteria decision making (MCDM) model in agro-climatic zoning of Taraba state for Ofada rice production, *FUTY J. Environ.*, vol.14, no.1, pp.108-120, 2020.

[10] R. N. Parker and E. K. Asencio, *GIS and Spatial Analysis for the Social Sciences: Coding, Mapping and Modeling*, Taylor & Francis Group, 2009.

[11] G. P. M. Spagnuolo, *Heterogenous Spatial Data: Fusion, Modeling, and Analysis for GIS Applications*, Morgan & Claypool, 2016.

[12] X. Lei, W. Chen and B. T. Pham, Performance evaluation of GIS-based artificial intelligence approaches for landslide susceptibility modeling and spatial patterns analysis, *ISPRS Int. J. Geo-Information*, vol.9, no.7, pp.1-20, 2020.

[13] M. Otgonbayar, C. Atzberger, J. Chambers, D. Amarsaikhan, S. Böck and J. Tsogtbayar, Land suitability evaluation for agricultural cropland in Mongolia using the spatial MCDM method and AHP based GIS, *J. Geosci. Environ. Prot.*, vol.5, no.9, pp.238-263, 2017.

[14] B. Feizizadeh and T. Blaschke, Land suitability analysis for Tabriz County, Iran: A multi-criteria evaluation approach using GIS, *J. Environ. Plan. Manag.*, vol.56, no.1, pp.1-23, 2013.

[15] P. Zolfaghary, M. Zakerinia and H. Kazemi, A model for the use of urban treated wastewater in agriculture using multiple criteria decision making (MCDM) and geographic information system (GIS), *Agric. Water Manag.*, vol.243, 106490, 2021.

[16] H. R. Pourkhabbaz, S. Javanmardi and H. A. F. Sabokbar, Suitability analysis for determining potential agricultural land use by the multi-criteria decision making models SAW and VIKOR-AHP (case study: Takestan-Qazvin plain), *J. Agric. Sci. Technol.*, vol.16, no.5, pp.1005-1016, 2014.

[17] A. Nurkholis and I. S. Sitanggang, A spatial analysis of soybean land suitability using spatial decision tree algorithm, *Proceedings of SPIE*, p.65, 2019.

[18] K. T. Morrison, T. A. Nelson, F. S. Nathoo and A. S. Ostry, Application of Bayesian spatial smoothing models to assess agricultural self-sufficiency, *Int. J. Geogr. Inf. Sci.*, vol.26, no.7, pp.1213-1229, 2012.

[19] A. O. Onojeghuo, G. A. Blackburn, Q. Wang, P. M. Atkinson, D. Kindred and Y. Miao, Mapping paddy rice fields by applying machine learning algorithms to multi-temporal Sentinel-1A and Landsat data, *Int. J. Remote Sens.*, vol.39, no.4, pp.1042-1067, 2018.

[20] Z. Maddahi, A. Jalalian, M. M. K. Zarkesh and N. Honarjo, Land suitability analysis for rice cultivation using a GIS-based fuzzy multi-criteria decision making approach: Central part of Amol District, Iran, *Soil Water Res.*, vol.12, no.1, pp.29-38, 2017.

[21] A. Barakat, W. Ennaji, A. E. Jazouli, R. Amediaz and F. Touhami, Multivariate analysis and GIS-based soil suitability diagnosis for sustainable intensive agriculture in Beni-Moussa irrigated subperimeter (Tadla plain, Morocco), *Model. Earth Syst. Environ.*, vol.3, no.3, pp.1-8, 2017.

[22] J. Kihoro, N. J. Bosco and H. Murage, Suitability analysis for rice growing sites using a multicriteria evaluation and GIS approach in great Mwea region, Kenya, *Springerplus*, vol.2, no.1, pp.1-9, 2013.

[23] D. N. Kuria and E. Waithaka, Using geographic information systems (GIS) to determine land suitability for rice crop growing in the Tana delta, *J. Geogr. Reg. Plan.*, vol.4, no.9, pp.525-532, 2011.

[24] I. Nahib et al., Integration of logistic regression and multicriteria land evaluation to simulation establishment of sustainable paddy field zone in Indramayu Regency, West Java Province, Indonesia, *IOP Conference Series: Earth and Environmental Science*, vol.149, 012005, pp.1-12, 2018.

[25] D. W. Miller and M. K. Starr, *Executive Decisions and Operations Research*, 1st Edition, Prentice-Hall, 1960.

[26] Y. Li, Z. Sun and F. Accatino, Spatial distribution and driving factors determining local food and feed self-sufficiency in the eastern regions of China, *Food Energy Secur.*, pp.1-17, 2021.

[27] V. Talasila, C. Prasad, G. T. S. Reddy and A. Aparna, Analysis and prediction of crop production in Andhra region using deep convolutional regression network, *Int. J. Intell. Eng. Syst.*, vol.13, no.5, pp.1-9, 2020.

[28] E. Boltürk, A. Karaoan and C. Kahraman, Simple additive weighting and weighted product methods using neutrosophic sets, *Studies in Fuzziness and Soft Computing*, vol.369, pp.647-676, 2019.

[29] A. Kolios, V. Mytilinou, E. Lozano-Minguez and K. Salonitis, A comparative study of multiple-criteria decision-making methods under stochastic inputs, *Energies*, vol.9, no.7, 2016.

[30] D. M. Martin and M. Mazzotta, Non-monetary valuation using multi-criteria decision analysis: Sensitivity of additive aggregation methods to scaling and compensation assumptions, *Ecosyst. Serv.*, vol.29, pp.13-22, 2018.

[31] C. H. Yeh, A problem-based selection of multi-attribute decision-making methods, *Int. Trans. Oper. Res.*, vol.9, no.2, pp.169-181, 2002.

[32] R. Blanquero, E. Carrizosa and P. Ramrez-cobo, Variable selection for Naïve Bayes classification, *Comput. Oper. Res.*, vol.135, 105456, 2021.

[33] H. Zhang, L. Jiang and L. Yu, Attribute and instance weighted Naïve Bayes, *Pattern Recognit.*, vol.111, 107674, 2021.

[34] R. Liu et al., Assessing spatial likelihood of flooding hazard using Naïve Bayes and GIS: A case study in Bowen Basin, Australia, *Stoch. Environ. Res. Risk Assess.*, vol.30, no.6, pp.1575-1590, 2016.

[35] N. Friedman, D. Geiger and M. Goldszmidt, Bayesian network classifier, *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol.29, no.Part3, pp.131-163, 1997.

[36] M. Benjumeda, C. Bielza and P. Larrañaga, Tractability of most probable explanations in multidimensional Bayesian network classifiers, *Int. J. Approx. Reason.*, vol.93, pp.74-87, 2018.

[37] T. Gärtner and P. Flach, WBCSVM: Weighted Bayesian classification based on support vector machines, *Proc. of the 18th International Conference on Machine Learning*, pp.154-161, 2001.

[38] M. Miao, H. Zeng, A. Wang, C. Zhao and F. Liu, Discriminative spatial-frequency-temporal feature extraction and classification of motor imagery EEG: A sparse regression and weighted Naïve Bayesian classifier-based approach, *J. Neurosci. Methods*, vol.278, pp.13-24, 2017.

[39] C. Mao, L. Lu and B. Hu, Local probabilistic model for Bayesian classification: A generalized local classification model, *Appl. Soft Comput. J.*, vol.93, 106379, 2020.

[40] M. Kabak, M. Erbao, C. Çetinkaya and E. Özceylan, A GIS-based MCDM approach for the evaluation of bike-share stations, *J. Clean. Prod.*, vol.201, pp.49-60, 2018.

[41] Y. Qin, Y. Liu, Z. Hong and H. Jia, Hesitant fuzzy Muirhead mean operators and its application to multiple attribute decision making, *International Journal of Innovative Computing, Information and Control*, vol.14, no.4, pp.1223-1238, 2018.

[42] S. Drobne and A. Lisec, Multi-attribute decision analysis in GIS: Weighted linear combination and ordered weighted averaging, *Informatica*, vol.33, no.4, pp.459-474, 2009.

[43] A. V. Vitianingsih, I. Wisnubhadra, S. S. K. Baharin, R. Marco and A. L. Maukar, Classification of pertussis vulnerable area with location analytics using multiple attribute decision making, *International Journal of Innovative Computing, Information and Control*, vol.16, no.6, pp.1943-1957, 2020.

[44] M. J. Zhang and J. X. Nan, A compromise ratio ranking method of triangular intuitionistic fuzzy numbers and its application to MADM problems, *Iran. J. Fuzzy Syst.*, vol.10, no.6, pp.21-37, 2013.

[45] E. Triantaphyllou, *Multi-Criteria Decision Making Methods: A Comparative Study*, Springer, Boston, MA, 2000.

[46] H. Jafaryeganeh, M. Ventura and C. G. Soares, Effect of normalization techniques in multi-criteria decision making methods for the design of ship internal layout from a Pareto optimal set, *Struct. Multidiscip. Optim.*, vol.62, no.4, pp.1849-1863, 2020.

[47] K. Yoon, The propagation of errors in multiple-attribute decision analysis: A practical approach, *J. Oper. Res. Soc.*, vol.40, no.7, pp.681-686, 1989.

[48] Y. Yang and J. He, A novel method based on fixed point iteration and improved TOPSIS method for multi-attribute group decision making, *International Journal of Innovative Computing, Information and Control*, vol.17, no.1, pp.15-29, 2021.

[49] D. Soria, J. M. Garibaldi, F. Ambrogi, E. M. Biganzoli and I. O. Ellis, A 'non-parametric' version of the Naïve Bayes classifier, *Knowledge-Based Syst.*, vol.24, no.6, pp.775-784, 2011.

[50] P. Tsangaratos and I. Ilia, Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size, *Catena*, vol.145, pp.164-179, 2016.

[51] W. R. Kusumadewi Sri, H. Sri and H. Agus, *Fuzzy Multi-Attribute Decision Making (Fuzzy MADM)*, Graha Ilmu, Yogyakarta, Indonesia, 2006.

[52] L. Guttman, The determinacy of factor score matrices with implications for five other basic problems of common-factor theory, *Br. J. Stat. Psychol.*, vol.7, no.2, pp.65-81, 1955.

[53] L. Guttman, A basis for scaling qualitative data, *Am. Sociol. Rev.*, vol.9, no.2, p.139, 1944.

[54] R. E. Tractenberg, F. Yumoto, P. S. Aisen, J. A. Kaye and R. J. Mislevy, Using the Guttman scale to define and estimate measurement error in items over time: The case of cognitive decline and the meaning of 'points lost', *PLoS One*, vol.7, no.2, pp.1-8, 2012.

[55] A. Stegeman, A new method for simultaneous estimation of the factor model parameters, factor scores, and unique parts, *Comput. Stat. Data Anal.*, vol.99, pp.189-203, 2016.

[56] R. Azen and C. M. Walker, *Categorical Data Analysis for the Behavioral and Social Sciences*, Taylor & Francis Group, 2021.

[57] M. Sokolova and G. Lapalme, A systematic analysis of performance measures for classification tasks, *Inf. Process. Manag.*, vol.45, no.4, pp.427-437, 2009.

[58] M. Bramer, *Principles Data Mining*, Spinger, London, http://www.springer.com/series/7592, 2007.

[59] J. Davis and M. Goadrich, The relationship between precision-recall and ROC curves, *Proc. of the 23rd International Conference on Machine Learning (ICML'06)*, 2006.

[60] L. Torgo and R. Ribeiro, Utility-based regression, in *Lecture Notes in Computer Science Book Series*, Springer-Verlag Berlin Heidelberg, pp.597-604, 2007.

[61] A. Tripathy, A. Agrawal and S. K. Rath, Classification of sentiment reviews using n-gram machine learning approach, *Expert Syst. Appl.*, vol.57, pp.117-126, 2016.

Author Biography

Anik Vega Vitianingsih is Ph.D. candidate at the Faculty of Information and Communication Technology (FTMK), Universiti Teknikal Malaysia Melaka, Malaysia. Bachelor's degree in Informatics Engineering, Master's degree in Game Tech, her research interests are in spatial analysis, spatial data modeling, and artificial intelligence in geographical information systems. Being lecturer in the Informatics Department, she is also editor in chief of the International Journal of Artificial Intelligence & Robotics from Universitas Dr. Soetomo. With several papers in various Scopus Journal, she has been reviewer for various international journals.

Robert Marco is Ph.D. candidate at the Faculty of Information and Communication Technology (FTMK), Universiti Teknik Malaysia Melaka, Malaysia with earned Bachelor's degree in Electrical Engineering and Master's in Engineering. Experienced papers writer in various Scopus Journal, the author is a lecturer in the Department of Informatics, Amikom University, Yogyakarta with research interests in software engineering, machine learning and deep learning, systems analysis.

Anastasia Lidya Maukar is a lecturer in the Industrial Engineering Program at President University, Cikarang, West Java. The author received the Bachelor's degree in Industrial Engineering from the University of Surabaya, Master of Science in Information System Development from the University of Hertfordshire, UK and Master of Technology Management in Industrial Engineering from the Sepuluh Nopember Institute of Technology, Surabaya. Her research and teaching interests are facility layout, databases and information systems, industrial statistics, and production systems. The author is also active in managing national journals and a reviewer in several national journals.

Erri Wahyu Puspitarini is currently attending Ph.D. program in Information and Communication Technology at Universiti Teknologi Malaysia Melaka (UTeM), with Bachelor's degree in Information System and Master's in Information and Technology Management. Her research interest includes behaviour analysis, technology acceptance, system analysis, data science and eLearning Technology.

Seftin Fitri Ana Wati graduated with a bachelor's degree in informatics at UPN "Veteran" East Java and Master of Information Systems from the Sepuluh Nopember Institute of Technology, Surabaya. Interested in conducting research in the field of system dynamic modeling, forecasting, GIS, analysis systems, enterprise systems, she is currently a lecturer in Information Systems at UPN "Veteran" East Java.