

Spatial decision tree model for garlic land suitability evaluation

Andi Nurkholis¹, Imas Sukaesih Sitanggang², Annisa³, Sobir⁴

^{1,2,3}Department of Computer Science, IPB University, Indonesia

⁴Department of Agronomy & Horticulture, IPB University, Indonesia

Article Info

Article history:

Received Nov 4, 2020

Revised May 19, 2021

Accepted Jun 14, 2021

Keywords:

Garlic

Land suitability

Spatial decision tree

Spatial join relation

ABSTRACT

Predicting land and weather characteristics as indicators of land suitability is very important in increasing effectiveness in food production. This study aims to evaluate the suitability of garlic land using spatial decision tree algorithm. The algorithm is the improvement of the conventional decision tree algorithm in which spatial join relation is included to grow up spatial decision tree. The spatial dataset consists of a target layer that represents garlic land suitability and ten explanatory layers that represent land and weather characteristics in the study areas of Magetan and Solok district, Indonesia. This study generated the best spatial decision trees for each study area. On Magetan dataset, the best model has 33 rules with 94.34% accuracy and relief variable as the root node, whereas on Solok dataset, the best model has 66 rules with 60.29% accuracy and soil texture variable as the root node.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Andi Nurkholis

Department of Computer Science

IPB University

Wing 20, Meranti Campus IPB Street, Dramaga, Bogor, West Java, Indonesia 16680

Email: andi_n@ipb.ac.id

1. INTRODUCTION

The difficulty of Indonesia for garlic self-sufficiency is considered because of the lack of garlic farmland which causes the inability to fulfill the garlic consumption needs of Indonesian people so that is the main focus of the Indonesian government to be completed in 2019 [1]. The garlic farmland only reached 2,407 hectares (ha) in 2016 [2]. This figure even decreased by 6.09 percent compared to garlic farmland which was recorded as covering an area of 2,563 ha in 2015. Plant production land that did not grow significantly became one of the main causes of not being able to meet needs. This problem can be solved by increasing the effectiveness in food production by utilizing technological advances for land suitability evaluation modeling [3].

Previous research has discussed a lot about land suitability for various agricultural commodities. A knowledge-based system for evaluating physical land suitability for 45 cultivated plants based on fuzzy inference [4]. In its development, land-use suitability mapping and analysis is one of the most useful applications of geographic information systems (GIS) for spatial planning and management [5]. The conventional method of overlaying maps was widely used in the field of land suitability evaluation by integrating multi-criteria-decision-analysis/making (MCDA/MCDM) methods and GIS technology [6]-[9]. The artificial intelligence methods played an important role in the development of land suitability evaluations and can solve the problem of multi-index decision-making method that different multi-index analyzes can produce different evaluation results [10]. The GIS-MCDA/MCDM technique was combined with AHP as computational weights or level of influences of criteria to evaluate land suitability [10]-[18]. The limitations of those studies are inherent issues with the use of AHP method, namely the inconsistency of expert judgment [13].

Previous land suitability studies have not considered spatial data correlation for each variable/factor. Analysis of geographically referenced data was done in this case land suitability, it is essential to consider the correlation of spatial data (i.e., position, distance, and orientation) [19]. Land suitability evaluation can be accomplished using the classification method; this is due to the presence of existing garlic land suitability in spatial data based on assessment of land suitability uses provisions [20], namely S1 (highly suitable), S2 (moderately suitable), S3 (marginally suitable), and N (not suitable). Classification is a data extraction technique in which data stored in the database is analyzed to find rules that describe the partition of the database into a particular class set [21]. Spatial datasets for classification tasks are composed of some explanatory layers which in this study are ten garlic planting criteria and one target layer which is used to represent garlic land suitability class. Each layer represents a set of spatial objects which is characterized by several spatial (polygon, line, and point) and non-spatial attributes (label). One of non-spatial attributes in an explanatory layer is the explanatory attribute that identifies objects in the layer. The target layer has a target attribute that stores class labels of the target object. In a non-spatial dataset, target classes are discrete-valued and unordered (categorical) and explanatory attributes are categorical or numerical. In this study, spatial classification is used to extract rules that split a spatial dataset consisting of classified objects into a number of classes based on non-spatial and spatial properties, as well as spatial relations of the classified objects to other objects.

This study developed a classifier for evaluating garlic land suitability using spatial decision tree algorithm. The algorithm is a development by [19] an extended from ID3 algorithm [22] which has been used by [23]-[26] on the classification of spatial data to predict the occurrence of fires based on hotspots with fairly good results of accuracy, i.e., 74.72%, 87.69%, 75.66%, and 71.66%. In a recent study [26], the comparison between classification algorithms that involving spatial factor (spatial decision tree) and not involving spatial factor (ID3, C4.5, logistic regression) was carried out, showing spatial algorithm produces a model with better accuracy. The formula of entropy and information gain in the algorithm were modified by involving two types of spatial relationships namely metric and topological to relate two spatial objects. There are two spatial relationships used, namely 'in' with 'count' as spatial measure value and 'distance' with 'distance' as spatial measure value [23]-[27]. The relation 'in' is used when a target layer is represented by point feature and an explanatory layer represented by polygon feature, while the relation 'distance' is used when a target layer represented by point feature and an explanatory layer represented by point/line features. For example, when point feature as target layer is correlated to polygon feature as explanatory layer, then the way to get the spatial measure is calculated by counting the amount of target data in an explanatory layer. Instead of using point feature as target layer, both of target layer and explanatory layer in this study consist of polygon features. So that the spatial relation used will be different from previous research [23]-[27]. The spatial relationship proposed in this study is to measure the intersection area between the target layer area and the explanatory layer area.

2. RESEARCH METHOD

The study area is Magetan district, East Java province with an area of 70,143 ha [28] and Solok district, West Sumatra province with an area of 335,086.53 ha [29]. The two districts are predicted to be the center for producing garlic for Indonesia in the future [1]. The data used in this study are ten garlic planting criteria as explanatory layers and a garlic land suitability as target layer for each district. Seven spatial criteria are vector format collected from Indonesian Center for Agricultural Land Resources Research and Development (BBSDLP). Those criteria are drainage, relief (%), base saturation (%), cation exchange capacity (cmol), soil texture, soil pH ($^{\circ}$), and depth of soil mineral (cm). The three non-spatial criteria are rainfall (mm) and temperature ($^{\circ}$ c) obtained from meteorological, climatological, and geophysical agency (BMKG), while elevation (masl) in raster format acquired from United States geological survey (USGS). The non-spatial criteria need to be pre-processed before they can be integrated with other spatial data [24]. This study was conducted in several stages, i.e., preprocessing data, spatial decision tree classification, and classification evaluation. The following three preprocessing data stages were carried out in this study:

- The first step of preprocessing data is interpolation on rainfall and temperature data that produce two layers of rainfall and temperature in vector format. Interpolation is a mathematical method or function that predicts values in locations where data are not available or not obtained. The comparison of rainfall interpolation methods involving elevation factor (i.e., ordinary co-kriging) and does not involve elevation factor (i.e., ordinary kriging and kriging with external drift) [30]. The result of those study was ordinary co-kriging the best interpolation method in estimating the distribution of rainfall values with the lowest error; therefore, this method is used in this study. Variables included in spatial interpolation are rainfall/temperature as primary and elevation as secondary variables in the ordinary co-kriging method.
- The second step of preprocessing data is extracting topographic data contained in the digital elevation model (DEM) data to produce an elevation layer in vector format. The use of DEM as a source of

elevation data is possible because DEM is a quantitative, three-dimensional representation of the earth surface derived from elevation data [30]. A DEM is typically given in one of the three formats: the raster-based grid DEM, the vector-based triangular irregular network (TIN) and contour-based storage structure [31]. DEM type used in this study for making the elevation layer is the raster-based grid DEM acquired from USGS.

- The final step of preprocessing data is to verify the validity of each explanatory and target layer. The cause of invalid geometry in the polygon is self-intersection. Self-intersection status indicates that polygons do not meet the requirements of OpenGIS [32] so that polygons cannot be included in the spatial decision tree classification. The explanatory and target layer containing invalid geometry were repaired by deleting a small portion of the invalid polygon.

2.1. Spatial relationship, spatial entropy, and spatial information gain

Spatial data mining aims to discover hidden knowledge from spatial databases by combining spatial and non-spatial properties that accumulate in spatial systems such as geographic information systems [33]. Spatial data mining method is the development of those used in conventional data mining [34]. Spatial data mining has two functions [35]. The first function discusses spatial phenomena by exploring data. For example, this study is identifying land suitability by determining the spatial distribution of the location of land and weather characteristics. The second function explains or even predicts phenomena by discovering multiple relationships. For example, in this study land suitability can be ‘explained’ by land and weather characteristics at that location.

Spatial data represents real objects based on the earth geographical [26]. The objects are represented by using geometric such as point, line, polygon, and pixel. Objects in spatial data have spatial relationships with its neighbor which used in this study is topology. Topology is a spatial relation that deals with the various geometric shape which in this study is a polygon. A relation between spatial objects of two different layers is essential in spatial data mining systems [26]. Spatial relationships allow to include relations between two spatial objects in a dataset for a classification task. Spatial relations between two layers could produce quantitative values in the form of distance between points or area in the intersections of two polygons [23]. The explanatory layer and target layer used in this study are both represented by polygons so that the spatial relationship proposed are intersection between the target layer area and the explanatory layer area. We denote these quantitative values, i.e., area as spatial measure of spatial relationships between two objects. This spatial measure is used in the spatial entropy formula which replaces the number of tuples in a partition in the non-spatial entropy formula. Illustration of the intersection area between an explanatory layer and a target layer can be seen in Figure 1.

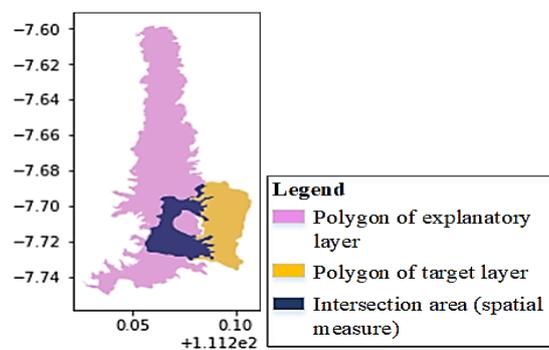


Figure 1. Illustration of the intersection area

Let L is a set of layers; L_i and L_j are two distinct layers in L . A spatial relationship applied to L_i and L_j is denoted $\text{SpatRel}(L_i, L_j)$ that can be topological relation or metric relation [23]. For example there are L_i (some explanatory layers) and L_j (a target layer), $i \neq j$, $i = 1, 2, \dots, p$ and p is number of layers in L_i , $j = 1, 2, \dots, q$ and q is number of layers in L_j which in this study is only one, for feature r_i with $R = \text{SpatRel}(L_i, L_j)$, spatial measure for r_i is denoted by $\text{SpatMes}(r_i)$. In this study, a new equation is formulated to measure $\text{SpatMes } r$ in (1).

$$\text{SpatMes}(r) = f\left(\text{SpatMes}(L_{i1} \cap L_{j1}), \text{SpatMes}(L_{i2} \cap L_{j2}), \dots, \text{SpatMes}(L_{im} \cap L_{jn})\right) \quad (1)$$

where

f : sum function

m : number of polygons in L_i

n : number of polygons in L_j

In spatial database, a layer is represented as a relation and applying a spatial relation between two layers results a new relation. A spatial relationship is applied to L_i and L_j in L results a new layer R . A spatial join relation (SJR) for all features p in L_i and q in L_j is formulated in (2) [23]:

$$\text{SJR} = \{(p, \text{SpatMes}(r), q) | p \text{ in layer } L_i, q \text{ in layer } L_j, \text{ and } r \text{ is feature in } R \text{ associated to } p \text{ and } q\} \quad (2)$$

Let a target attribute C in a target layer S has l distinct classes (i.e., c_1, c_2, \dots, c_l), spatial entropy for S represents the expected information needed to determine the class of tuples in the dataset which is formulated in (3) [23].

$$\text{Spatial Entropy } H(S) = - \sum_{i=1}^l \frac{\text{SpatMes}(S_{c_i})}{\text{SpatMes}(S)} \log_2 \frac{\text{SpatMes}(S_{c_i})}{\text{SpatMes}(S)} \quad (3)$$

Let an explanatory attribute V in an explanatory (non-target) layer L has q distinct values (i.e., v_1, v_2, \dots, v_q). We partition the objects in target layer S according to the layer L then we have a set of layers $L(v_i, S)$ for each possible value v_i in L . The expected entropy value for splitting is formulated in (4) [23].

$$\text{Split Information } H(S|L) = \frac{\text{SpatMes}(\neg L(S))}{\text{SpatMes}(S)} H(\neg L(S)) + \sum_{j=1}^q \frac{\text{SpatMes}(L(v_j, S))}{\text{SpatMes}(S)} H(L(v_j, S)) \quad (4)$$

The entropy value of a variable is denoted by $H(S)$, while the split information value of an attribute is denoted by $H(S|L)$. The spatial information gain value is formulated in (5) [23].

$$\text{Spatial Information Gain } (L) = H(S) - H(S|L) \quad (5)$$

The variable with highest spatial information gain is selected as the first node in the spatial decision tree known as root. The next node is sequentially filled with variables with lower gain value. Spatial decision tree will stop growing if they fulfill one of the following termination criteria [23]:

- Only one explanatory layer in L . In this condition, the algorithm returns a leaf node labeled with the majority class in the SJR for the best layer and the explanatory layer.
- The SJR for the best layer and explanatory layer contains the same class c . Then the algorithm returns a leaf node labeled with the class c .

2.2. Spatial decision tree

The spatial decision tree technique uses the basic concept of a decision tree which is a tree structure, where each node in a tree represents a variable, each branch represents the attribute value, and the leaf node represents a certain class [36]. Whereas the spatial decision tree is a rooted tree that meets the following criteria: i) each internal node is a decision node over a layer, ii) each branch denotes an outcome of the test and iii) each leaf represents one of the class values [19].

Figure 2 shows our proposed algorithm to generate a spatial decision tree that has been developed [23]. Algorithm inputs are divided into two groups: i) a set of layers containing some explanatory layers and one target layer that hold class labels for tuples in the dataset and ii) spatial join relations (SJRs) storing spatial measures for features resulted from spatial relations between two layers. The algorithm generates a tree by selecting the best layer to separate the dataset into smaller partitions as pure as possible meaning that all tuples in partitions belong to the same class.

The algorithm works on spatial data stored in a spatial database [26]. When the algorithm is applied to the data in the database, some new layers are produced as the result of spatial relations between two distinct layers. These new layers are created from existing explanatory layers and the value v_i of predictive attribute in the best splitting layer [26]. The value v_i is a selection criterion in the query to relate an explanatory layer and the best layer. The new layers are then used in calculating spatial information gain at the root branch which produce node internal/leave to compile spatial decision tree that illustrated by Figure 3.

```

Algorithm: Generate_SDT (Spatial Decision Tree)
Input:
  a Spatial dataset D, which is a set of training tuples and their associated class labels. These tuples are constructed from a set of layers, P, using spatial relations.
  b A target layer S ∈ P with a target attribute C.
  c A non empty set of explanatory layers L ⊆ P and L ∈ L has a predictive attribute V. P = S ∪ L.
  d Spatial Join Relation (SJR) on the set of layers P, SJR(P), as defined in (2).
Output: A Spatial Decision Tree
Method:
  1 Create a node N;
  2 If only one explanatory layer in L then
  3 return N as a leaf node labeled with the majority class in D; // majority voting
  4 endif
  5 If objects in D are all of the same class c then
  6 return N as a leaf node labeled with the class c;
  7 endif
  8 Apply layer_selection_method(D, L, SJR(P)) to find the "best" splitting layer, L*;
  9 Label node N with L*;
  10 Split D according to the best splitting layer L* in {D(v1), ..., D(vm)}. D(vi) is outcome i of splitting layer L* and v1, ..., vm are possible values of predictive attribute V in L*;
  11 L = L - {L*};
  12 for each D(vi), i = 1, 2, ..., m, do
  13 let Ni = Generate_SDT(D(vi), L, SJR(P));
  14 Attach node Ni to N and label the edge with a selected value of predictive attribute V in L*;
  15 endfor

```

Figure 2. Spatial decision tree algorithm [23]

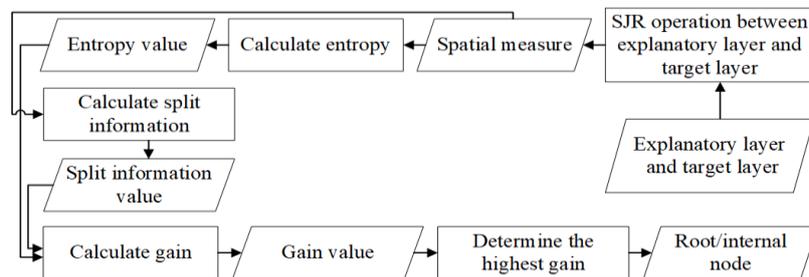


Figure 3. Step by step of spatial decision tree

2.3. Confusion matrix

Classification accuracy is done by testing the classification rules on the actual data so that they can be corrected in subsequent iterations. The higher level of accuracy, the lower classification error in the test data. Accuracy was obtained from Magetan and Solok data using confusion matrix in (6) [37].

$$\text{Accuracy} = \frac{tp+tn}{tp+fn+fp+tn} \times 100\% \quad (6)$$

where

- tp (true positive) : number of positive data that is correctly classified
- tn (true negative) : number of positive data that is incorrectly classified
- fp (false positive) : number of negative data that is correctly classified
- fn (false negative) : number of negative data that is incorrectly classified.

3. RESULTS AND DISCUSSION

The results of preprocessing data produce ten explanatory layers and one target layer, all of which are ready to be used for spatial decision tree classification. All explanatory layers and a target layer are stored

in a spatial database to be executed by the algorithm in Figure 2. List of layer names, number of polygons for each layer, and attribute names in a layer can be seen in Table 1.

Table 1. All layers and details

Layer's name	Number of polygons		Attributes
	Magetan	Solok	
Elevation (masl)	11	255	High (>1000), slightly high (851-1000), slightly low (601-850), low (≤ 600)
Drainage*	17	27	Slightly swift, good, slightly good, slightly hamper, hampered
Relief (%)	30	166	Flat (0), slightly flat (1-3), slightly slope (4-8), slope (9-15), slightly steep (16-25), steep (26-40), very steep (>40)
Base saturation (%)	19	86	Very low (<20), low (20-35), medium (36-60), high (61-80)
Cation exchange capacity (cmol)	15	53	Very low (<5), low (5-16), medium (17-24), high (24-40)
Soil texture*	18	53	Smooth, slightly smooth, medium, slightly rude
Soil pH (°)	16	33	Acid (4.5-5.5), slightly acid (5.6-6.5), neutral (6.6-7.5)
Depth of soil mineral (cm)	35	124	Very shallow (<25), shallow (25-50), medium (51-75), deep (76-100), very deep (>100)
Rainfall (mm)	3	5	High (>350), slightly high (301-350), slightly low (251-300), low (≤ 250)
Temperature (°c)	3	3	23, 24, 25, 26
Land suitability	78	307	Highly suitable, moderately suitable, marginally suitable

*Variables have no numeric value

3.1. Spatial decision tree for land suitability

In this study, 3 models were made for each of Magetan and Solok data with the aim of obtaining the best rules. The models created based on Magetan data is denoted by A, while the models made based on Solok data is denoted by B. The description of the differences in the variation of the Magetan and Solok models can be seen in Table 2. The model variations for Magetan can be seen in Table 3, while the model variations for Solok can be seen in Table 4.

Table 2. Model descriptions

Models	Descriptions
A0 and B0	Spatial decision tree model
A1 and B1	Optimization of A0 and B0 model adding the condition of spatial relation result ≥ 1 ha in SJR process, it is based on the smallest garlic farmland is 1 ha [38], so that we assume the area < 1 ha to be considered not representing the land suitability class which is related to explanatory factor
A2 and B2	Optimization of A1 and B1 model deleting planting area of <1 ha in new layer process, this is based on we assume that polygons that have an area below 1 ha will produce spatial relation result under 1 ha as well

Table 3. Spatial decision tree model of Magetan

Models	Number of intersection layer results	Number of rule results	Variables that are not involved	Node root
A0	887	56	-	Relief
A1	425	34 (2 none*)	Drainage	Relief
A2	425	33	Drainage	Relief

*Rules that have no class

Table 4. Spatial decision tree model of Solok

Models	Number of intersection layer results	Number of rule results	Variables that are not involved	Node root
B0	1746	131	-	Soil texture
B1	674	66 (4 none*)	Drainage	Soil texture
B2	616	66	Drainage	Soil texture

*Rules that have no class

Based on Tables 3 and 4, it can be concluded as: (i) Optimization of adding a condition to the SJR process has positive impact i.e., fewer layers and rules generated and provide information about the variable that is not involved in spatial decision tree result. Based on the model result, the variable that is not involved meaning that the variable is not so important in the garlic land suitability. That variable is drainage, due to the variable have variations data are very diverse, so the model cannot determine land suitability class of the variable. (ii) Optimization of deleting the planting area in the new layer process has positive impact i.e., fewer rule results and no rules that have no class.

3.2. Spatial decision tree evaluation

The evaluation was carried out on 6 models that have been produced using two testing datasets, namely Magetan and Solok. Evaluation was done by implementing confusion matrix in (6) on the results of applying the rules to the test data. The evaluation results can be seen in Table 5.

Table 5. Accuracy of spatial decision tree model

Models	Magetan testing dataset				Solok testing dataset			
	True*	False**	—***	Accuracy (%)	True*	False**	—***	Accuracy (%)
A0	46	0	7	86.79	33	17	86	24.26
A1	50	1	2	94.34	70	51	15	51.47
A2	50	1	2	94.34	70	51	15	51.47
B0	18	13	22	33.96	75	30	31	55.15
B1	23	18	12	43.4	82	44	10	60.29
B2	23	18	12	43.4	82	44	10	60.29

*Number of data that is correctly classified by the rules

**Number of data that is incorrectly classified by the rules

***Number of data that cannot be classified by rules

Based on Table 5, it can be concluded that optimization of adding a condition to the SJR processes affect the result of higher accuracy, while optimization of deleting the planting area in the new layer processes do not affect the result of accuracy. Overall, the Magetan and Solok models provide good result if tested using training data, for example, the Magetan model is tested with Magetan data and the Solok model is tested with Solok data. However, if the model is applied to other district data, the accuracy result is decreased. This is probably due to differences in characteristics that are quite significant in the two districts, so the result in one district can only represent the rules for the district itself. This is proven by:

- The amount of unclassified data results was higher than those classified when the Solok model is tested on Magetan data and vice versa. This can be seen in A0 and B0 model when tested with other district data, unclassified data results are higher than true and false data results.
- Variable resulted as root node differs between Magetan and Solok model. The entire Magetan model has a relief variable as a root node, while the entire Solok model chooses the soil texture variable as the root node.
- Some attribute values of a variable in Magetan data are not owned by the attributes in Solok data and vice versa. The temperature in Magetan only ranges from 23-25°C, while the temperature in Solok ranges from 25-26°C. It means that the Solok testing data containing temperatures variable with a value of 26°C cannot be classified by the Magetan model which only has a temperature range of 23-25°C.

Based on the model evaluation, it can be concluded that the best Magetan model is A2 model with a higher accuracy than A0 model and fewer rules than A1 model. Whereas the best Solok model is B2 model, with higher accuracy results than B0 model and fewer rules than B1 model. Here are some example results of A2 model rules:

- IF relief = steep AND elevation = slightly low AND soil pH = slightly acid AND depth of soil mineral = deep AND cation exchange capacity = medium THEN garlic land suitability class = S1 (highly suitable)
- IF relief = steep AND elevation = slightly low AND soil pH = slightly acid AND depth of soil mineral = medium THEN garlic land suitability class = S2 (moderately suitable)
- IF relief = slightly flat AND rainfall = slightly low THEN garlic land suitability class = S3 (marginally suitable)
- IF relief = steep AND elevation = slightly high AND temperature = 24°C AND cation exchange capacity = low AND rainfall = slightly high AND depth of soil mineral = deep THEN garlic land suitability class = S1 (highly suitable)
- IF relief = slightly flat AND rainfall = slightly high THEN garlic land suitability class = S2 (moderately suitable)
- IF relief = flat AND rainfall = slightly low AND depth of soil mineral = very steep THEN garlic land suitability class = S3 (marginally suitable)
- IF relief = slightly flat AND rainfall = slightly high THEN garlic land suitability class = S2 (moderately suitable)

Visualization is applied to the best spatial decision tree rules, where A2 model is implemented in Magetan data while B3 model is implemented in Solok data. The following is the land suitability visualization in Magetan and Solok district which can be seen in Figure 4.

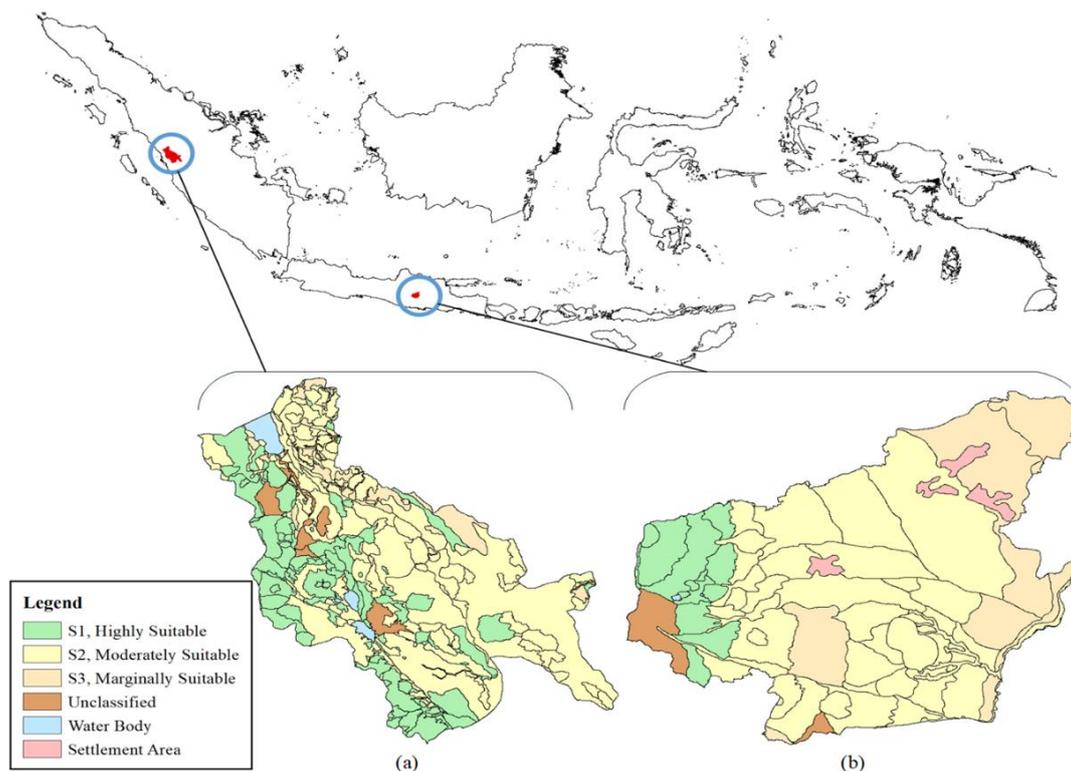


Figure 4. Garlic land suitability in: (a) Solok and (b) Magetan district

4. CONCLUSION

This work applied the spatial decision tree algorithm on the spatial garlic land suitability dataset in the study areas Magetan and Solok district, Indonesia. A spatial dataset is composed in a set of layers in which the layers are divided into two categories i.e., explanatory layers and a target layer. The explanatory layers are ten planting garlic criteria, i.e., elevation, drainage, relief, base saturation, cation exchange capacity, soil texture, soil acidity, mineral soil depth, rainfall, and temperature. A target layer is garlic land suitability that has three classes i.e., highly suitable, moderately suitable, and marginally suitable. The result is two best spatial decision trees for land suitability evaluation. Magetan model has 33 rules, accuracy of 94.34%, and relief variable as the root node, while Solok model has 66 rules, accuracy of 60.29%, and soil texture variable as the root node. The variable that is not involved in two best spatial decision trees is drainage, meaning the drainage variable is not so important in determining garlic land suitability. The two best spatial decision trees have weaknesses which when tested with other district data then the accuracy result will decrease, due to differences in characteristics of the two districts. Future works are expected: i) Development of land suitability geographic information system as an interactive map visualization, ii) Adding land cover factor to get land suitability rules with more specific land characteristics, so as not to damage the specified land order, such as protected forests.

REFERENCES

- [1] Directorate General of Horticulture, "National Garlic Development," Ministry of Agriculture Republic Indonesia, 2017. [Online]. Available at: <https://riph.pertanian.go.id/asset/media/download/file/547a6106025e209a3517aa07db2f27b7.pdf>.
- [2] BPS, "Production, Harvested Area and Vegetable Productivity in Indonesia," Statistics Indonesia, 2017, [Online]. Available at: <http://pangan.litbang.pertanian.go.id/files/Statistik/StatistikPertanian2017.pdf>.
- [3] D. Djaenudin, H. Marwan, H. Subagjo, and A. Hidayat, "Technical Guidelines for Land Evaluation for Agricultural Commodities," Bogor (ID): Indonesian Center for Agricultural Land Resources Research and Development, 2011, [Online]. Available at: <http://nasih.staff.ugm.ac.id/wp-content/uploads/Petunjuk-teknis-evaluasi-lahan-untuk-komoditas-pertanian-2011.pdf>.
- [4] S. Hartati and I. S. Sitanggang, "A fuzzy based decision support system for evaluating land suitability and selecting crops," *Journal of Computer Science*, vol. 6, no. 4, pp. 417-424, 2010, doi: 10.3844/jcssp.2010.417.
- [5] J. Malczewski, "GIS-based land-use suitability analysis: a critical overview," *Progress in planning*, vol. 62, no. 1, pp. 3-65, 2004, doi: 10.1016/j.progress.2003.09.002

- [6] I. W. Nuarsa, I. N. Dibia, K. Wikantika, D. Suwardhi, and I. N. Rai, "GIS Based Analysis of Agroclimate Land Suitability for Banana Plants in Bali Province, Indonesia," *HAYATI Journal of Biosciences*, vol. 25, no. 1, pp. 11-17, 2018, doi: 10.4308/hjb.25.1.11.
- [7] A. K. Nisyak and F. Ramdani, "Web-GIS development and analysis of land suitability for rice plant using GIS-MCDA method in Batu city," *International Symposium on Geoinformatics*, pp. 24-33, 2017, doi: 10.1109/ISYG.2017.8280667.
- [8] A. Ardiansyah, S. Sahirman, M. Rif'an, and E. H. Melmambessy, "Land Suitability Analysis for Paddy Field in Semangga District, Merauke Regency," *Journal of Regional and Rural Development Planning*, vol. 1, no. 1, pp. 87-97, 2017, doi: 10.29244/jp2wd.2017.1.1.87-97.
- [9] A.A. Baroudy, "Mapping and evaluating land suitability using a GIS-based model," *Catena*, vol. 140, pp. 96-104, 2016, doi: 10.1016/j.catena.2015.12.010.
- [10] L. Qu, Y. Shao, and L. Zhang, "Land suitability evaluation method based on GIS technology," *Second International Conference on Agro-Geoinformatics (Agro-Geoinformatics)*, 2013, pp. 7-12, doi: 10.1109/Argo-Geoinformatics.2013.6621869.
- [11] 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," *Journal of Geoscience and Environment Protection*, vol. 5 no. 9, pp. 238-263, 2017, doi: 10.4236/gep.2017.59017.
- [12] A. Bozdağ, F. Yavuz, and A. S. Günay, "AHP and GIS based land suitability analysis for Cihanbeyli (Turkey) County," *Environmental Earth Sciences*, vol. 75, no. 9, pp. 813-828, 2016, doi: 10.1007/s12665-016-5558-9.
- [13] S. G. Yalaw, A. Griensven, and P. Zaag, "AgriSuit: A web-based GIS-MCDA framework for agricultural land suitability assessment," *Computers and Electronics in Agriculture*, vol. 128, pp. 1-8, 2016, doi: 10.1016/j.compag.2016.08.008.
- [14] K. M. Ullah and A. Mansourian, "Evaluation of Land Suitability for Urban Land-Use Planning: Case Study D Haka City," *Transactions in GIS*, vol. 20, no. 1, pp. 20-37, 2016, doi: 10.1111/tgis.12137.
- [15] R. Mosadeghi, J. Warnken, R. Tomlinson, and H. Mirfenderesk, "Comparison of Fuzzy-AHP and AHP in a spatial multi-criteria decision-making model for urban land-use planning," *Computers, Environment and Urban Systems*, vol. 49, pp. 54-65, 2015, doi: 10.1016/j.compenvurbsys.2014.10.001.
- [16] J. Zhang, Y. Su, J. Wu, and H. Liang, "GIS based land suitability assessment for tobacco production using AHP and fuzzy set in Shandong province of China," *Computers and Electronics in Agriculture*, vol. 114, pp. 202-211, 2015, doi: 10.1016/j.compag.2015.04.004.
- [17] A. Hadipour, F. Vafaie, and V. Hadipour, "Land suitability evaluation for brackish water aquaculture development in coastal area of Hormozgan, Iran," *Aquaculture international*, vol. 23, no. 1, pp. 329-343, 2015, doi: 10.1007/s10499-014-9818-y.
- [18] G. Romano, P. Dal Sasso, G. T. Liuzzi, and F. Gentile, "Multi-criteria decision analysis for land suitability mapping in a rural area of Southern Italy," *Land Use Policy*, vol. 48, pp. 131-143, 2015, doi: 10.1016/j.landusepol.2015.05.013.
- [19] S. Rinzivillo and F. Turini, "Classification in geographical information systems," *Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases*, 2004, pp. 374-385, doi: 10.1007/978-3-540-30116-5_35.
- [20] FAO, "A Framework for Land Evaluation," Soil Resources Management and Conservation Service Land and Water Development Division. FAO Soil Bulletin No. 32. FAO-UNO, Rome, 1976.
- [21] K. Koperski, J. Han, and N. Stefanovic, "An efficient two-step method for classification of spatial data," *Proceedings of International Symposium on Spatial Data Handling*, 1998, pp. 45-54, doi: 10.1.1.12.2505
- [22] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81-106, 1986, doi: 10.1007/BF00116251.
- [23] I. S. Sitanggang, R. Yaakob, N. Mustapha, and A. A. B. Nuruddin, "An extended ID3 decision tree algorithm for spatial data," *International Conference on Spatial Data Mining and Geographical Knowledge Services*, pp. 48-53, 2011, doi: 10.1109/ICSDM.2011.5969003.
- [24] I. S. Sitanggang, R. Yaakob, N. Mustapha, and A. N. Ainuddin, "Classification model for hotspot occurrences using spatial decision tree algorithm," *Journal of Computer Science*, vol. 9, no. 2, pp. 244-251, 2013, doi: 10.3844/jcssp.2013.244.251.
- [25] Y.M. Khoiriyah and I.S. Sitanggang, "A spatial decision tree based on topological relationships for classifying hotspot occurrences in Bengkalis Riau Indonesia," *International Conference on Advanced Computer Science and Information System*, 2014, pp. 268-272, doi: 10.1109/ICACSIS.2014.7065844.
- [26] I. S. Sitanggang, R. Yaakob, N. Mustapha, and A. N. Ainuddin, "A decision tree based on spatial relationships for predicting hotspots in peatlands," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 12, no. 2, pp. 511-518, 2014, doi: 10.12928/TELKOMNIKA.v12i2.2036.
- [27] I. D. Nurpratami and I. S. Sitanggang, "Classification rules for hotspot occurrences using spatial entropy-based decision tree algorithm," *Procedia Environmental Sciences*, vol. 24, pp. 120-126, 2015, doi: 10.1016/j.proenv.2015.03.016.
- [28] Indonesian Center for Agricultural Land Resources Research and Development, "Semi Detailed Soil Map of Magetan District, Large Scale 1: 50,000," Bogor (ID): BBSDLP, 2017.
- [29] Indonesian Center for Agricultural Land Resources Research and Development, "Semi Detailed Soil Map of Solok District, Large Scale 1: 50,000," Bogor (ID): BBSDLP, 2017.

- [30] S. K. Adhikary, N. Muttil, and A. G. Yilmaz, "Cokriging for enhanced spatial interpolation of rainfall in two Australian catchments," *Hydrological Processes*, vol. 31, pp. 2143-2161, 2017, doi: 10.1002/hyp.11163.
- [31] S. Mukherjee, A. Mukhopadhyay, and A. B. A. M. Sananda, "Digital elevation model generation and retrieval of terrain attributes using CARTOSAT-1 stereo data," *International Journal of Science and Technology*, vol. 2, no. 5, pp. 265-271, 2012.
- [32] P. Oosterom, W. Quak, and T. Tijssen, "About invalid, valid and clean polygons," *Developments in Spatial Data Handling*, pp. 1-16, 2005, doi: 10.1007/3-540-26772-7_1.
- [33] S. Nagaprasad *et al.*, "Spatial data mining using novel neural networks for soil image classification and processing," *International Journal of Engineering Science and Technology*, vol. 2, no. 10, pp. 5621-5625, 2010.
- [34] N. Chelghoum, K. Zeitouni, and A. Boulmakoul, "A decision tree for multi-layered spatial data," *Advances in Spatial Data Handling*, pp. 1-10, 2002, doi: 10.1007/978-3-642-56094-1_1.
- [35] K. Zeitouni, "A survey of spatial data mining methods databases and statistics point of views," *Data warehousing and web engineering*, pp. 229-242, 2002, doi: 10.4018/978-1-931777-02-5.ch013.
- [36] J. Han, M. Kamber, and J. Pei. "Data Mining Concepts and Techniques Third Edition," San Francisco (US): Morgan Kaufman Publisher, 2012.
- [37] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing and Management*, vol. 45, no. 4, pp. 427-437, 2009, doi: 10.1016/j.ipm.2009.03.002.
- [38] Ministry of Agriculture Republic Indonesia, "Production, Harvested Area and Vegetable Productivity in Indonesia," 2016. [Online]. Available at: <https://aplikasi2.pertanian.go.id/bdsp>.

BIOGRAPHIES OF AUTHORS



Andi Nurkholis is a Lecturer of Informatics in Teknokrat Indonesia University, Indonesia. His research areas are focused on the development of computer science (artificial intelligence, data mining, and expert system) in agriculture, specifically related to Geographical Information System.



Imas Sukaesih Sitanggang is a Professor of Computer Science in IPB University, Indonesia. Her research areas are focused on the development of computer science (artificial intelligence, data mining, and expert system) in agriculture, specifically related to Geographical Information System.



Annisa is a Lecturer of Computer Science in IPB University, Indonesia. Her research areas are focused on the development of computer science (artificial intelligence, data mining, and expert system) in agriculture.



Sobir is a Professor of Agronomy and Horticulture in IPB University, Indonesia. His research areas are focused on development of horticultural agricultural commodity varieties for food security.