IT Journal Research and Development (ITJRD)
Vol.9, No.1, August 2024, E-ISSN : 2528-4053 | P-ISSN : 2528-4061
DOI : 10.25299/itjrd.2024.13595

Classification of Land Suitability For Soybean Crops Using The Cart Method and Feature Selection Using an Algorithm ABC

Rusdi Efendi[1], Mochammad Yusa[2], Stefani Tasya Hallatu[3]
Department of Informatics Faculty of Engineering, University of Bengkulu [1,2,3]
Rusdi.efendi@unib.ac.id[1], mochammad.yusa@unib.ac.id[2], fhallatu08@gmail.com[3]

Article Info

Article history:

Received Aug 6, 2023
Revised Sep 8, 2023
Accepted Jun 5, 2024

Keyword:

Soyabean
Classification
CART
Feature selection
ABC

ABSTRACT

The allocated area for soybean cultivation has been gradually decreasing, leading to a decline in both production and productivity. Consequently, the current level of soybean production and productivity falls short of meeting the demand within the community. One potential solution to augment soybean output and efficiency involves allocating specific parcels of land for soybean cultivation. It is essential to conduct land evaluations tailored to soybean cultivation, accounting for the land's inherent potential, in order to optimize land utilization. Thus, a comprehensive system is required to assess land suitability, particularly for soybean cultivation, and employ the results of this classification as recommendations for land allocation. This research employess combination the Classification and Regression Tree (CART) method and the Artificial Bee Colony (ABC) algorithm to classify suitable land for soybean cultivation. CART is used for classification and ABC is utilized for feature selection to identify the most relevant attributes in case of the algorithm improvement. Through a series of iterative experiments involving 5, 10, 25, 50, 75, and 100 iterations, the best attribute was determined following three attempts at each iteration. The Confusion Matrix test yielded an accuracy rate of 94.22% for the CART method in the second experiment, while the combined use of the best ABC and CART combination resulted in an accuracy rate of 97.11%. Therefore, it can be concluded that the integration of the artificial bee colony (ABC) algorithm with the classification and regression tree (CART) method outperforms the sole use of the CART method in terms of accuracy.

Corresponding Author:

Rusdi Effendi
Departement of Informatics
Faculty of Engineering, University of Bengkulu
Email: Rusdi.efendi@unib.ac.id

1. INTRODUCTION

Soybeans hold a prominent position as a priority food crop, following rice and corn. They are considered one of the most crucial commodities, with a consistently increasing demand within the country. According to the publication "Statistik Konsumsi Pangan Tahun 2020," the per capita consumption of soybean-containing food items exhibited an average annual growth rate of 8.21%

during the period of 2016-2020 [1]. In order to meet the escalating demand for soybean consumption, it is imperative to continually augment soybean production and productivity. However, data from the book "*Analisis Produktivitas Jagung dan Kedelai di Indonesia 2020*" reveals that the average national soybean productivity in 2020 was merely 15.69 kilograms per hectare [2]. Furthermore, statistics provided by the Central Bureau of Statistics indicate a continuous decline in soybean production from 1993 to 2015. In 1993, soybean production reached 1,707,126 tons, whereas by 2015, it had plummeted to a mere 963,183 tons [2]. Consequently, the domestic production and productivity of soybeans have not been able to meet the populace's demand, leading to heavy reliance on soybean imports. According to "Food Consumption Statistics for 2020," the average annual growth rate of soybean imports between 2016 and 2020 stood at 2.06% [1].

The provision of dedicated land for soybean cultivation emerges as a crucial factor in driving an increase in soybean production [3]. According to the Food and Agriculture Organization (FAO) (1976), land constitutes an integral component of the broader landscape, encompassing the physical environment, climate, topography, soil characteristics, hydrology, and natural vegetation, all of which potentially influence land utilization. A report from the Central Bureau of Statistics (BPS) reveals a notable decline in the soybean harvest area in Indonesia between 1993 and 2015. Specifically, the total area allocated for soybean cultivation stood at 1,468,316 hectares in 1993, whereas by 2015, this specific soybean harvest area had dwindled to 614,095 hectares [4]. It is worth noting that Bengkulu Province is also a soybean-producing region. Data provided by the Central Bureau of Statistics indicates a relatively small soybean harvest area in Bengkulu, which experienced a continuous decrease from 1993 to 2015. In 1993, the dedicated soybean harvest area in Bengkulu amounted to merely 15,417 hectares, while by 2015, it had further declined to 4,235 hectares [4].

The allocation of specific land for soybean cultivation plays a pivotal role in enhancing soybean production and productivity, as it allows for the cultivation of soybeans on a larger scale. This can be achieved through a systematic land evaluation process that involves assessing and evaluating the land based on specific criteria and characteristics suitable for soybean cultivation. We can identify areas with genuine potential for successful soybean cultivation by thoroughly evaluating land suitability. Several physical environmental factors, such as climate, soil composition, topography, hydrology, drainage, and others, determine the land's inherent potential. Adhering to land use practices that align with the actual potential of the land enables the preservation of soil quality, leading to improved crop production and enhanced productivity.

Technological advancements offer new opportunities to address the challenge of determining land suitability for soybean cultivation through a systematic classification system that provides valuable land-use decision-making recommendations. This study employs the Classification and Regression Tree (CART) method for land classification. The CART method entails constructing decision trees that group data and describe the relationship between the dependent and independent variables [5]. Notably, the CART method is versatile as it can handle numerical and categorical data, ensuring flexibility and adaptability to generate optimal classification outcomes. Compared to alternative approaches, the CART method offers several advantages. It is relatively straightforward to formulate, performs calculations more efficiently and accurately, and can be applied to large datasets [6]. Despite these strengths, the CART method is not without limitations. Its complex data structure, involving multiple attributes, may lead to the formation of unstable trees that fail to classify accurately.

To address the limitations of the CART (Classification and Regression Tree) method, specifically its suboptimal performance when dealing with many attributes, the authors of this study employed a feature selection process as an optimization technique. Feature selection is a crucial preprocessing stage in machine learning that aims to eliminate irrelevant data, thereby enhancing efficiency, performance, and accuracy [7]. In this particular study, the feature selection process was carried out using the Artificial Bee Colony (ABC) algorithm, which mimics the foraging behavior of bees [8]. By leveraging the ABC algorithm, the researchers were able to identify the most relevant attribute among the entire set of attributes, improving the overall effectiveness of the classification process.

In a study conducted by Indra Irawan et al [9] the influence of the Artificial Bee Colony (ABC) algorithm on enhancing the performance of the CART (Classification and Regression Tree) classification method was investigated. The results demonstrated a notable performance improvement, with an achieved accuracy of 82%. This accuracy surpassed that of alternative methods, indicating the superiority of the combined ABC algorithm and CART classification approach. The integration of these techniques occurred after the data preparation stage. Following data preparation, the feature selection process was performed using the ABC algorithm to identify the most relevant attributes for the classification task. This ensured that the classification process utilized the optimal attributes selected through the feature selection stage. Ultimately, the study successfully employed the CART method with feature selection using the ABC algorithm, improving classification performance.

Consequently, developing a classification system becomes imperative to address the challenge of determining land suitability for soybean cultivation. Such a system would facilitate the classification of land based on its suitability for soybean plants, thereby providing valuable recommendations for determining appropriate land use and effectively preserving land quality. Ultimately, the implementation of this classification system would contribute to the cultivation and productivity enhancement of soybean plants.

2. RESEARCH METHOD

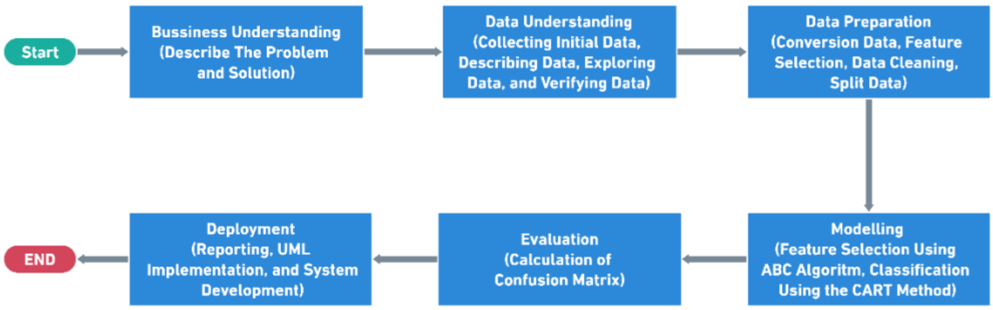

Figure 1. System Development Flowchart

Figure 1 above is a flowchart of the system development method in this study, where the system development method used is the CRISP-DM (Cross-Industry Standard Process for Data Mining) method. The stages of the system development method consist of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment which are described below:

2.1. Bussiness Understanding

The business understanding stage is the stage in determining goals based on the business situation, eventually resulting in the correct problems and solutions in dealing with these problems [10]. The issues raised and identified in this study are land use that is still not by the actual land potential; it can result in the incompatibility of planting land used in soybean planting so that land quality decreases, and lack of soybean cultivation due to lack of soybean land which results in production and productivity soybeans that have not been able to meet the needs of the community.

The offered solution is implementing a soybean land suitability classification system using the CART method and feature selection using the ABC algorithm to determine land suitable for soybean plants, and it is hoped that the resulting classification results can provide recommendations in determining and expanding soybean land use according to the actual land potential.

2.2. Data Understanding

Several things were done to understand data: collecting initial data, describing and exploring data, and verifying data. The following are the stages in data analysis :

1. Initial Data Collection

This study to collect data carried out by several methods, namely :

a. Library Studies

The data collection process for this study involved a comprehensive literature review. The finding was conducted to identify relevant criteria that underpin the evaluation of soybean land suitability. To obtain the necessary information, authoritative sources were consulted, including the book authored by Wahyunto et al [11] and the book authored by Ritung et al [12]. Furthermore, a book focused on soybeans authored by Aidah and the KBM Indonesia Publishing Team [13] was utilized to deepen the understanding of soybeans. Additionally, scientific journals pertaining to soybean land suitability and the methods employed in this study were thoroughly examined to gather pertinent data.

b. Interview technique

In this study, data was collected by interviewing from the Bengkulu Agricultural Technology Assessment Center (BPTP) staff regarding land information and validating crop suitability criteria.

c. Collection of data with documents

Preliminary data collection with documents obtained through the Bengkulu Agricultural Technology Assessment Center (BPTP) Based on four districts in Bengkulu Province, namely Central Bengkulu, Kepahiang, Lebong, and Mukomuko Regencies as many as 111 data shown in Table 1 :

Table 1. Land Dataset Agricultural Technology Assessment Center (BPTP) Bengkulu

Num.	Dataset	Year	Number of Data
1.	Land suitability data for Central Bengkulu Regency	2013	29
2.	Land suitability data for Mukomuko Regency	2014	31
3.	Land suitability data for Lebong Regency	2015	32
4.	Land suitability data for Kepahiang Regency	2015	19

2. Data Description and Exploration

Data description and exploration were conducted to analyze and comprehend the structure of the initially collected data, aiming to generate tables that adhere to predetermined criteria for soybean land suitability. At the initial data collection stage, 111 data points were acquired from Central Bengkulu, Kepahiang, Lebong, and Mukomuko Regencies. These data encompassed 33 attributes along with one target attribute.

While some attributes corresponded to the criteria for soybean land suitability and were included in this study, not all attributes were utilized. This is due to certain factors, such as the requirement for further laboratory testing for several attributes and the absence of data for certain attributes. Consequently, these unused attributes were excluded from the study. Additionally, location attributes, such as mapping units, sub-districts, districts, and provinces, were incorporated, resulting in 20 attributes and one target attribute. Subsequently, the collected data was divided based on sub-districts within each respective mapping unit found on the land map. This division yielded an additional 135 data points, increasing the total dataset from the initial 111 data points to 247.

Furthermore, data exploration was conducted to analyze and assess various aspects of the data, including data types, volume, and other relevant characteristics. This exploration was performed after amalgamating the data obtained from Central Bengkulu, Kepahiang, Lebong, and Mukomuko Regencies, creating a unified dataset for further analysis.

3. Verify

Data verification is done to re-check the data processed at the data exploration stage. This verification stage is the stage after conducting data exploration, where data is obtained that will be used in the data mining process. In this study, the attributes used in the implementation of feature selection using the Artificial Bee Colony (ABC) algorithm and classification using the Classification And Regression Tree (CART) method are shown in Table 2 :

Table 2. Attributes in the Implementation of Feature Selection and Classification

Num.	**Requirements for land use/characteristics**	**Variable**
1.	Average temperature (°C)	attribute
2.	Annual rainfall (mm)	attribute
3.	Humidity (%)	attribute
4.	Number of wet months (>200 mm/bl)	attribute
5.	Drainage	attribute
6.	Texture	attribute
7.	Soil depth (cm)	attribute
8.	KTK land (cmol)	attribute
9.	Base Saturation (%)	attribute
10.	pH H2O	attribute
11.	C-organic (%)	attribute
12.	N total (%)	attribute
13.	P2O5 (mg/100 g)	attribute
14.	K2O (mg/100 g)	attribute
15.	Slope (%)	attribute
16.	Erosion hazard	attribute
17.	Evaluation Labels	targeted attribute

2.3. Data Preparation

The stages of data preparation that will be carried out in this study include several steps, namely data cleaning and data split :

1. Conversion Data

At the data conversion stage, several data conversions were carried out on attributes that have non-numeric values, namely drainage, texture, N Total, P2O5, dan K2O, and erosion hazard into numerical data represented by the numbers 0, 1, 2, etc.

2. Feature Selection

At the feature selection stage, feature selection is made from the existing attributes. For the mapping unit, sub-district, district, and province attributes, a correlation value is calculated where a negative correlation value is obtained for this attribute, which means that it does not affect the evaluation label attribute. The value for the mapping unit attribute is -0.190709; for the sub-district, the value is -0.221255; for the regency, the value is -0.074974; and for the province attribute, the value is NaN.

3. Data Cleaning

In dealing with the missing values contained in the data, it is handled by changing the value to "Unknown" so as not to cause conflicts in the data type used so that when checking the missing values, no NaN or empty values are found.

4. Split Data

In the data split stage, the data is divided into training and testing data. The distribution of data for each training data and testing data will be divided into two parts, namely 70% training data with 173 data and 30% testing data with 74 data.

2.4. Modelling

System flowcharts, feature selection flowcharts with the ABC algorithm, and classification flowcharts using the CART method are as follows :

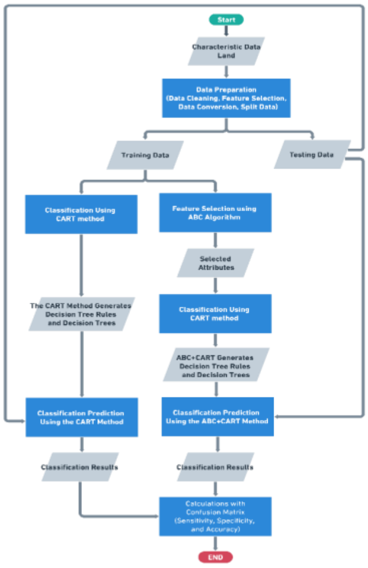

Figure 2. System Flowcharts

First, the input data is permit location information: provinces, districts/cities, and sub-districts. As well as a dataset of several attributes of the needs/characteristics of soybean plant permits. Furthermore, the data preparation stage includes cleaning the data to recover lost or "unknown" values and obtaining inconsistent data. Furthermore, then split the data where the data is divided into 70% training data and 30% testing data. Then the training data can be directly classified using the CART (Classification And Regression Tree) method or perform feature selection using the ABC (Artificial Bee Colony) algorithm.

Furthermore, the best attribute method is obtained after feature selection, which is used in the classification process using CART (Classification And Regression Tree). After that, the model will be accepted as decision tree rules and decision trees. After that, the model is implemented on training and testing data to make predictions in the classification process, and the classification results will be obtained in the form of S1 (very suitable class), S2 (reasonably suitable class), S3 (suitable marginal class), and N (not according to class). After that, an evaluation is carried out by calculating the values of Sensitivity, Specificity, and $Accurac$y and calculating the Confusion Matrix.

2.5. Evaluation

This stage is where an evaluation is carried out and ensures that the resulting model is by the initial objectives that have been determined. If appropriate, it can proceed to the deployment stage; if it is not suitable, it can return to the business understanding stage [10]. The calculation of the confusion matrix is an evaluation stage in this study by calculating the values of Sensitivity, Specificity, and $Accurac$y.

2.5. Deployment

At this stage, the process of making reports and journal articles is carried out by implementing the results of the research that has been done. It is also done at this stage by making UML (Unified Modeling Language). UML (Unified Modeling Language) is a language that has become an industry standard used to design a software system to provide a model and model visualization, and a modeling language without a programming language used for developing software systems [14]. Some UML modeling such as usecase diagrams and class diagrams [15]. As well as in this research, interface design, interface creation, and system development were carried out based on previously designed arrangements.

3. RESULTS AND ANALYSIS

3.1. Artificial Bee Colony (ABC)

Feature selection is selecting several features that affect accuracy the most. Also, feature selection is the best solution to improve the performance of classification methods by reducing irrelevant features [16]. Feature selection in this study using the artificial bee colony (ABC) algorithm. The ABC (Artificial Bee Colony) algorithm was introduced by Karaboga and Basturk [17]. The following are the steps in the optimization process for the ABC (Artificial Bee Colony) algorithm :

1. Initialization Phase, Initialization Phase is the phase of randomly determining the position of the food source using the following equation (1):
$$x_{ij} = x_{\min j} + rand[0,1]\ x\ (x_{\max j} - x_{\min j}) \tag{1}$$
2. Employee Bee Phase, in this phase, the employee bee will explore the neighborhood of food sources related to the employee bee. The equation (2) of the neighborhood exploration is as follows :
$$v_{ij} = x_{ij} + \Phi_{ij}(x_{ij} - x_{kj}) \tag{2}$$
After v_ij is generated, the fitness value is generated using the following equation (3):
$$fitness_i = \begin{cases} \dfrac{1}{1 + f_i}, & if\ f_i \geq 0 \\ 1 + abs(f_i), & if\ f_i < 0 \end{cases} \tag{3}$$
After the employee bee has searched, the employee bee provides information regarding food sources to the onlooker bee. The probability value helps the onlooker bee to choose a food source to be explored next with the following equation (4):
$$p_i = \frac{fitness_i}{\sum_{n=1}^{F} fitness_i} \tag{4}$$
3. Onlooker bee phase. In this phase, the food source with the best probability value will be selected by the onlooker bee to become an employee bee by using a roulette wheel (RW) for neighborhood exploration of preferred food sources described in the employee bee phase.
4. In the scout bee phase, in this phase, the scout bee determines whether to renew the food source or not based on the LIMIT variable, which is updated on each exploration, which is if the LIMIT value > MAX LIMIT, then the food source will be updated with a new food source randomly by the scout bee. And if LIMIT < MAX LIMIT, then the food source will not be renewed. [18]

Several iterations were carried out in the artificial bee colony (ABC) algorithm. From each number of iterations, three trials were carried out. Where the best attribute is taken from rank 8 with the highest number of occurrences, and if the following sequence has the same number as rank 8, then it will still be taken as the best attribute. With ten iterations, the best green attributes were produced as follows:

Table 3. The Number of Attributes Appears 10 Iterations

#	Trial 1		Trial 2		Trial 3	
	Attribute	Number of Appearances	Attribute	Number of Appearances	Attribute	Number of Appearances
1.	Erosion Hazard	7	Humidity	7	Erosion Hazard	8
2.	Slope	7	Drainage	6	KTK Land	6
3.	Texture	6	Average Temperature	5	Average Temperature	6
4.	K_2O	6	pH H2O	5	Number of Wet Months	6
5.	Number of Wet Months	6	N Total	5	Base Saturation	5
6.	Rainfall	5	Number of Wet Months	5	N Total	5
7.	C Organic	5	P2O5	5	Humidity	5
8.	Humidity	5	Slope	5	Texture	5
9.	Base Saturation	4	Base Saturation	4	pH H_2O	4
10.	pH H_2O	4	K2O	4	P_2O_5	4
11.	Drainage	4	Erosion Hazard	4	Rainfall	3
12.	P_2O_5	4	C Organic	4	Soil Depth	3
13.	Average Temperature	3	Rainfall	3	K2O	3
14.	Soil Depth	3	Texture	3	Drainage	3
15.	N Total	2	KTK Land	3	C Organic	3
16.	KTK Land	2	Soil Depth	2	Slope	1

Table 3 displays the feature selection results using the ABC algorithm with ten iterations and three trials because ten iterations produced the highest level of accuracy, namely 97.11%, compared to other iterations. The dataset used has 16 attributes subjected to feature selection using the ABC algorithm to produce several attributes that frequently appear in iterations. The attributes that appear are then added up and ranked to get the most influential attributes, namely the top 8 in green in Table 3, which are the attributes that will be used in the ABC+CART method.

3.2. Classification And Regression Tree (CART)

The decision tree technique known as CART (Classification And Regression Tree) aims to group data and also describe the dependent variable with the independent variable [5]. The CART algorithm has several advantages, which are as follows:

1. Because it is a non-parametric method, the CART algorithm does not require a specification of any functional form.
2. Since it is not required to determine variables at the start of the process, the CART algorithm selects the most significant variables and eliminates the insignificant variables.
3. Outliers are data with different characteristics that appear as extreme values in either a single variable or a combination. This CART algorithm easily handles outliers.
4. The CART algorithm is computationally fast and makes no assumptions.
5. The CART algorithm is a flexible and can adjust according to needs.

The CART algorithm has several drawbacks, which are as follows:

1. The CART algorithm can generate a hazardous decision tree.
2. Data is divided by the CART algorithm based on only one variable. [19]

According to Steinberg [9] in implementing the CART algorithm, there are several stages in forming a classification tree using the CART algorithm, which is described below:

1. Define root
2. Calculate the Gini index for each prospective branch
3. Choose the most extensive Gini index as the splitting attribute used as a branch.
4. Repeat steps 2 and 3 until the leaves are close to pure.

The attribute selection measure is crucial in determining the most suitable criteria for classifying training data in size. It is a heuristic method to select criteria that effectively partition or restrict the data, resulting in close-to-pure divisions. The attribute selection measure determines the order and ranking of attributes within the training data, with the attribute possessing the highest rank selected as the splitting attribute. In cases where the splitting attribute exhibits a constant value, a split point is determined. However, if the splitting attribute is discrete, the decision tree formed must adhere to a binary structure.

The Gini index serves as an attribute selection measure within the CART (Classification and Regression Tree) algorithm. It employs the binary properties of each attribute to generate divisions. The Gini index quantifies the impurity of a particular partition, denoted as D, by applying a formula (5).

$$Gini\ (D) = 1 - \sum_{i=1}^{m} p_{i^2} \tag{5}$$

The impurity of each resulting partition can be summed up for checking binary division. The formula (6) can be used to determine the Gini index D :

$$Gini_A\ (D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \tag{6}$$

Attribute A was chosen as the splitting subset because the attribute has discrete values so that the gini index value with the smallest value close to 0 is given by the subgroup. Each split point must be checked on continuous-valued attributes. The attribute values are sorted, and the midpoint between each attribute value pair can be used as the split point. The split point will then be the attribute with the smallest Gini index. The following formula can be used to determine the decrease in the level of impurity resulting from the binary division on attribute A :

$$\Delta Gini(A) = Gini(D) - Gini_A\ (D) \tag{7}$$

Splitting attributes will be chosen based on how maximally the attribute can reduce impurity. If the attribute has a discrete value or a split point, it will form the sharing criteria for splitting subsets. If the attribute has a continuous value, it will create a division criterion [20].

In this study, the CART method produces decision trees and decision tree rules that will be used as models in the classification process in data testing. Where the decision tree is shown as follows :

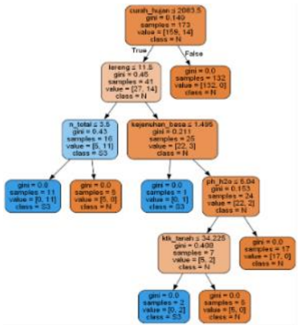

Figure 3. CART Decision Tree

The decision tree rules generated from the CART method are shown in the following table:

Table 4. Exploration of the CART Decision Tree Rules from Figure 3

No.	Id Aturan	Aturan
1	1	if (curah_hujan > 2083.5) then N
2	2	if (curah_hujan <= 2083.5) and (lereng > 11.5) and (kejenuhan_basa > 1.495) and (ph_h2o > 5.04) then N
3	3	if (curah_hujan <= 2083.5) and (lereng <= 11.5) and (n_total <= 3.5) then S3
4	4	if (curah_hujan <= 2083.5) and (lereng > 11.5) and (kejenuhan_basa > 1.495) and (ph_h2o <= 5.04) and (ktk_tanah > 34.225) then N
5	5	if (curah_hujan <= 2083.5) and (lereng <= 11.5) and (n_total > 3.5) then N
6	6	if (curah_hujan <= 2083.5) and (lereng > 11.5) and (kejenuhan_basa > 1.495) and (ph_h2o <= 5.04) and (ktk_tanah <= 34.225) then S3
7	7	if (curah_hujan <= 2083.5) and (lereng > 11.5) and (kejenuhan_basa <= 1.495) then S3

3.3. ABC+CART

This study also carried out the combination calculation process between the CART method and the ABC algorithm. In the calculation process using the combination method, an experiment was carried out with the attributes obtained at each iteration used in this study, namely iterations 5, 10, 25, 50, 75, and 100. So that the resulting decision tree is as follows:

Figure 4. Decision Tree in 5 Iterations

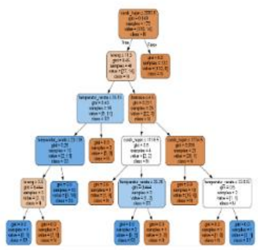

Figure 6. Decision Tree in 25 Iterations

Figure 8. Decision Tree in 50 Iterations

Figure 5. Decision Tree in 10 Iterations

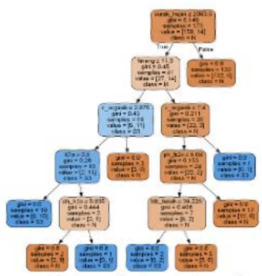

Figure 7. Decision Tree in 25 Iterations

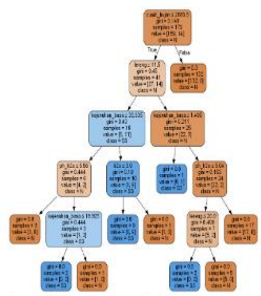

Figure 9. Decision Tree in 100 Iteratio

3.3. Confusion Matrix

In the testing stage, the performance of the classification method is evaluated by comparing the original labels of the data with the labels obtained from the analysis. This evaluation is carried out using a confusion matrix, which is a data mining concept that provides values such as Sensitivity, Specificity, and Accuracy. Accuracy represents the percentage of correctly classified data, Specificity measures the ratio of correctly classified harmful data to the total harmful data, and Sensitivity represents the ratio of correctly classified positive data to the total positive data [19]. The calculation of Accuracy, Specificity, and Sensitivity requires the use of several equations (8,9,10), as follows:

1. Sensitivity [21]

$$Sensitivity = \frac{TP}{P} \tag{6}$$

2. Specificity [21]

$$Specificity = \frac{TN}{N} \tag{7}$$

3. Accuracy [21]

$$Accuracy = \frac{TP + TN}{P + N} \tag{8}$$

The evaluation of the machine learning model in this study involved a train/test split process. The data was divided into training data, comprising 70% of the total data (173 instances), and testing data, comprising 30% of the total data (74 instances). After the data division, the model was implemented, and the labels predicted by the model were compared with the original labels. Two methods were employed in this study: the classification and regression tree (CART) method and a combination of the CART method with the artificial bee colony (ABC) algorithm. The confusion matrix was then calculated for both methods, and the results were compared as follows in Table 5:

Table 5. Confusion Matrix Results

Num.	Metrics	*Iteration*	*Trial*	*Accuracy* (%)	*Sensitivity* (%)	*Specificity* (%)	*Error Rate* (%)
1.	*CART*	-	-	94.22 %	98.74 %	42.86 %	5.78 %
2.	*ABC+CART*	5	1	94.22 %	98.11 %	50.0 %	5.78 %
			2	94.22 %	98.74 %	42.86 %	5.78 %
			3	96.53 %	96.23 %	100.0 %	3.47 %
3.	*ABC+CART*	10	1	94.22 %	98.11 %	50.0 %	5.78 %
			2	**97.11 %**	**100.0 %**	**64.29 %**	**2.89 %**
			3	95.38 %	96.86 %	78.57 %	4.62 %
4.	*ABC+CART*	25	1	94.8 %	98.11 %	57.14 %	5.2 %
			2	94.22 %	98.11 %	50.0 %	5.78 %
			3	94.8 %	98.11 %	57.14 %	5.2 %
5.	*ABC+CART*	50	1	93.64 %	98.74 %	35.71 %	6.36 %
			2	94.22 %	98.74 %	42.86 %	5.78 %
			3	93.64 %	98.74 %	35.71 %	6.36 %
6.	*ABC+CART*	75	1	94.22 %	99.37 %	35.71 %	5.78 %
			2	94.22 %	93.71 %	100.0 %	5.78 %
			3	94.22 %	98.11 %	50.0 %	5.78 %
7.	*ABC+CART*	100	1	94.22 %	99.37 %	35.71 %	5.78 %
			2	93.64 %	98.74 %	35.71 %	6.36 %
			3	94.22 %	93.71 %	100.0 %	5.78 %

Table 5 above shows the results of the model's success in predicting data. There was the best improvement in the use of the ABC+CART method with ten iterations in the 2nd experiment, which used the attributes, namely Soil Depth, KTK Land, Drainage, Average Temperature, Erosion Hazard, Number of Wet Months, Rainfall, Humidity, pH H_2O, N Total, Slope, and C Organic. The increase occurred when the accuracy with the CART method was 94.22%, while the accuracy with the ABC+CART method was 97.11%. The specificity value with the CART method produces a value of 42.86%, and the ABC+CART method produces a value of 64.29%. The sensitivity value and error rate are inversely proportional to the accuracy and specificity, where the sensitivity value for the CART method is 98.74%, and the ABC+CART method is 100.0%. The error rate produced by the CART method is 5.78%, and the ABC+CART method is 2.89% which is the error rate of the classification model using the ABC+CART method with the attributes Soil Depth, KTK Land, Drainage, Average Temperature, Erosion Hazard, Number of Wet Months, Rainfall, Humidity, pH H_2O, N Total, Slope, and C Organic are more accurate than the CART method.

4. CONCLUSION

The attributes generated through multiple iterations (5, 10, 25, 50, 75, and 100) using the artificial bee colony (ABC) algorithm were evaluated three times. It was observed that the attributes obtained from the ten iterations performed the best. In the second experiment, these attributes were identified as follows: Soil Depth, KTK Land, Drainage, Average Temperature, Erosion Hazard, Number of Wet Months, Rainfall, Humidity, pH H_2O, N Total, Slope, and C Organic.

By comparing the accuracy results of the two methods, an increase of 2.89% was observed. This finding indicates that the combination of the artificial bee colony (ABC) algorithm and the classification and regression tree (CART) method outperformed the sole use of the CART method. The selected attributes from the ten iterations in the second experiment using the ABC algorithm contributed to the improved accuracy of the model.

REFERENCES

[1] Kementerian Pertanian, “Statistics of Food Consumption 2020,” *Cent. Agric. Data Inf. Syst.*, pp. 1–103, 2021, [Online]. Available: http://epublikasi.setjen.pertanian.go.id/arsip-perstatistikan/163-statistik/statistik-konsumsi/751-statistik-konsumsi-pangan-tahun-2020.

[2] Badan Pusat Statistik, “Produksi (Ton), 1993-2015,” *bps.go.id*, 2016. https://www.bps.go.id/indicator/53/23/1/produksi.html (accessed May 29, 2023).

[3] Sumarno and M. M. Adie, “Strategi pengembangan produksi menuju swasembada kedelai berkelanjutan,” *Iptek Tanam. Pangan*, vol. 5, no. 1, pp. 49–63, 2010.

[4] Badan Pusat Statistik, “Luas Panen (Hektar), 1993-2015,” *bps.go.id*, 2016. https://www.bps.go.id/indicator/53/21/1/luas-panen.html (accessed May 29, 2023).

[5] A. Andarama, Nurfalinda, and N. Ritha, “Implentasi Algoritma Classification And Regression Trees (Cart) Dalam Klasifikasi Penerima Bantuan Sosial Beras Sejahtera (Studi Kasus : Desa Kuala Sempang Kabupaten Bintan),” vol. 4, no. 1, pp. 1–23, 2016.

[6] F. E. Pratiwi and I. Zain, “Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara,” *J. Sains Dan Seni Pomits*, vol. 3, no. 1, pp. D54–D59, 2014, [Online]. Available: http://www.ejurnal.its.ac.id/index.php/sains_seni/article/view/6129.

[7] N. A. Sugianto, I. Cholissodin, and A. W. Widodo, “Klasifikasi Keminatan Menggunakan Algoritme Extreme Learning Machine dan Particle Swarm Optimization untuk Seleksi Fitur (Studi Kasus : Program Studi Teknik Informatika FILKOM UB),” vol. 2, no. 5, pp. 1856–1865, 2018.

[8] A. Nurdiansyah, M. T. Furqon, and B. Rahayudi, “Prediksi Harga Bitcoin Menggunakan Metode Extreme Learning Machine (ELM) dengan Optimasi Artificial Bee Colony (ABC),”

J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 3, no. 6, pp. 5531–5539, 2019, [Online]. Available: http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/5507.

[9] I. Irawan, "Peningkatan Performa Algoritma CART dengan Seleksi Fitur Menggunakan ABC untuk Penilaian Kredit," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 199–208, 2021, doi: 10.35957/jatisi.v8i1.553.

[10] A. R. Wibowo and A. Jananto, "Implementasi Data Mining Metode Asosiasi Algoritma FP-Growth Pada Perusahaan Ritel," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 2, p. 200, 2020, doi: 10.35585/inspir.v10i2.2585.

[11] Wahyunto *et al.*, *Technical Guidance Guidelines for Land Suitability Assessment for Strategic Agricultural Commodities Semi-Detailed Scale 1:50.000*. 2016.

[12] S. Ritung, K. Nugroho, A. Mulyani, and E. Suryani, *Petunjuk Teknis Evaluasi Lahan untuk Komoditas Pertanian (Edisi Revisi)*. 2011.

[13] T. P. K. Indonesia, *Ensiklopedia Kedelai (Deskripsi, Filosofi, Manfaat, Budidaya, dan Peluang Bisnisnya)*. Bojonegoro: Penerbit Karya Bakti Makmur (KBM) Indonesia, 2020.

[14] D. Etika Profesi and Henderi, "Analisis Dan Perancangan Sistem Informasi Kepegawaian Menggunakan Unified Modeling Language (UML) Analysis And Design Of Employee Information System Use Unified Modeling Language (UML)," *Ijccs*, vol. x, No.x, no. 1, pp. 22–33, 2018.

[15] Muhamad Syarif and Wahyu Nugraha, "Pemodelan Diagram Uml Sistem Pembayaran Tunai Pada Transaksi E-Commerce," *J. Tek. Inform. Kaputama*, vol. 4, no. 1, pp. 64–70, 2020.

[16] A. K. B. Ginting, M. S. Lydia, and E. M. Zamzami, "Peningkatan Akurasi Metode K-Nearest Neighbor dengan Seleksi Fitur Symmetrical Uncertainty," *J. Media Inform. Budidarma*, vol. 5, no. 4, pp. 1714–1719, 2021, doi: 10.30865/mib.v5i4.3254.

[17] M. D. Arifin and A. D. Laksito, "Implementasi Algoritma Bee Colony Untuk Optimasi Rute Distribusi Carica Nida Food Wonosobo," *Sistemasi*, vol. 8, no. 2, p. 243, 2019, doi: 10.32520/stmsi.v8i2.470.

[18] M. Schiezaro and H. Pedrini, "Feature Selection Based On Artificial Bee Colony Algorithm," *EURASIP J. Image Video Process.*, p. 47, 2013, [Online]. Available: http://jivp.eurasipjournals.com/content/2013/1/47.

[19] N. Indah Prabawati, Widodo, and H. Ajie, "Kinerja Algoritma Classification And Regression Tree (Cart) dalam Mengklasifikasikan Lama Masa Studi Mahasiswa yang Mengikuti Organisasi di Universitas Negeri Jakarta," *PINTER J. Pendidik. Tek. Inform. dan Komput.*, vol. 3, no. 2, pp. 139–145, 2019, doi: 10.21009/pinter.3.2.9.

[20] J. Wijaya, "Implementasi Algoritma Pohon Keputusan CART untuk Menentukan Klasifikasi Data Evaluasi Mobil," Yogyakarta, 2019.

[21] J. Han, M. Kamber, and J. Pei, *Data mining: Data mining concepts and techniques (Third Edition)*, Third Edit. Morgan Kaufmann: Morgan Kaufmann, 2011.