Spatial Data Modelling for Irrigation Canal Development using Decision Tree Algorithm C4.5 Method

Anastasia Lidya Maukar
Industrial Engineering,
Faculty of Engineering
President University
Bekasi, Indonesia
almaukar@president.ac.id

Erri Wahyu Puspitarini
Department of Informatic Engineering
STMIK Yadika Bangil
Pasuruan, Indonesia
www.erri@stmik-yadika.ac.id

Anik Vega Vitianingsih*
Department of Informatics Engineering
Dr. Soetomo University
Surabaya, Indonesia
vega@unitomo.ac.id

Morinda Rori Rosalina
Department of Informatics Engineering
Dr. Soetomo University
Surabaya, Indonesia
morinda.rori@gmail.com

Fitri Marissa
Department of Informatics,
Faculty of Information and
Communication Technology
Widyagama University of Malang
Malang, Indonesia
fitrimarisa@widyagama.ac.id

***Abstract*— The development of irrigation canals is one of the optimal ways to increase food production. These efforts could assist farmers in using water to increase the production of agricultural products, especially rice. The purpose of the discussion of this paper is to determine the classification of areas that are necessary or not in the development of irrigation canals based on Web-GIS technology. The Decision Tree Algorithm C4.5 method is used in the spatial data modelling process based on land type, rice productivity, water availability, water demand, and rainfall parameters. The results of spatial data modelling with the Decision Tree Algorithm C4.5 method get an accuracy value of 83%, which states that this method is recommended for further research with the same data behaviour. The benefits of this research can be used as a policymaker to determine the priority of irrigation canal development based on the danger of drought level with a high category.**

Keywords—spatial data modelling, irrigation canal development, decision tree, C4.5 method, Web-GIS

I. INTRODUCTION

East Java is one of the rice barns and acts as a national food buffer, being in the first position from 2001 to 2015 with an increase in production of 0.52%, an average of 10,460,278 tons per year [1]. Rice production in 2020 is expected to reach 55.16 million tons of GKG, up from 54.60 million tons in 2019, an increase of 556.51 thousand tons or 1.02 per cent [2][3]. Agricultural production in developing countries such as Indonesia will increase by 1.4% for 2015-2030 [4]. Rice is the one of staple food for Indonesian. The government must seek to improve food security, especially by increasing production. Food security cannot be separated from the nature of food commodity production, but it is seasonal and fluctuating due to climate/weather’s easy influence on it [2]. The production behaviour, which is strongly influenced by the climate, significantly affects the national food availability. Drought will become a disaster that routinely occurs when the dry season comes. Hydrometeorological disasters, one of which is drought. In this condition, Indonesia is located in a tropical climate with two seasons, namely summer and rainy season, characterized by changes in weather, temperature and direction. According to BPBD (Regional Disaster Management Agency), drought is a potential disaster threat [5] [6]. One condition that must be taken seriously is in terms of efforts to maintain the quality of agricultural products and meet water requirements for other activities. Thus, innovation is needed in water management during the dry season, one of which is the development of irrigation canals, as drought mitigation [7]. Irrigation has a vital role for the nation and the State of the Republic of Indonesia, following the mandate of the 1945 Indonesian Constitution, article 5 paragraph 2 [8]. The purpose of irrigation in an area is to regulate and support agriculture and distribute water sources technically and systematically to areas of need [8].

The development of irrigation canals is one of the optimal ways to increase food production. The case studies in this paper are the districts, Lamongan and Tuban. The condition of agricultural land in Bojonegoro, East Java, currently only relies on reservoirs and river flows. Therefore, water conditions are not able to provide maximum water supply, so irrigation needs serious attention. [9]. Village reservoirs with small water reservoirs are located in Lamongan; their utilization cannot meet the needs in agricultural locations because the volume of water decreases every year and some have experienced siltation which causes sediment [10]. Meanwhile, there is a lot of damage to irrigation canals in Tuban; uneven distribution of water and suboptimal cropping patterns are the main factors, so that not all areas can be planted. Therefore, speedy maintenance and repairs are needed to meet food production targets [11]. The cause of all these cases is the inefficient fulfilment of water needs for agriculture, irrigation channels that have not met the needs of the area. During the dry season, many water sources dry up, so they cannot irrigate rice fields. For optimal plant growth to increase high productivity, it is essential to improve irrigation canals based on spatial data modelling..

GIS spatial data modelling is data modelling by performing geospatial analysis based on the model used or proposed to assess the accuracy of the model. [12][13]. GIS software integrates spatial data sets to visualize spatial information-based decision making [14].

Previous research discussed the optimization of land development based on irrigation areas to determine development priorities using geospatial Multi-Criteria Decision Analysis (GIS-MCDA) with the Analytical Hierarchy Process (AHP), Knowledge and data-based criteria and sub-criteria have an important role in geographic-based decision making [15]. Multi-Criteria Evaluation (MCE) is used to evaluate the land suitability for the irrigation canal development as an effort to plan sustainable agricultural systems [12]. The zoning of the water resource utilization area for irrigation development using the AHP method, the final result is that three aspects that become criteria have the same level of importance, namely with a weight of 33.3%. Planning, institutions, and participation are prerequisites in the aspect of water supply with equal importance [16]. Fuzzy analytical hierarchical process (FAHP) multi-criteria decision making technique based on GIS is used for land suitability analysis in rice cultivation, FAHP in this study can increase the accuracy of the criteria weights [17]. This research has actually been widely used, but there has never been a study using the Decision Tree Algorithm C4.5 method, however, this algorithm has been used for research on the subject of spatial analysis for soybean land suitability [18].

The Decision Tree method classifies priority areas in the development of irrigation canals using the parameters of rice productivity, rainfall, water availability, land types, and water requirements. The parameter data is taken from a 5-year range, namely 2015-2019, from the East Java Agriculture and Food Crops Service, East Java Provincial Public Works Service. The results of this study can be used as drought mitigation in Bojonegoro, Tuban, and Lamongan regencies with a high level of drought hazard. The use of Decision Tree Algorithm C4.5 in this study proves the statement in previous research[19] which states that the decision tree will produce the best decisions if it is optimized based on examining coexisting factors multi-criteria parameters. Expert heuristics on the impacts of various irrigation techniques are represented by the decision tree[20], wherewith the decision tree, it will be easy to model the behaviour of farmers in improving the sustainable management of irrigation areas [21].

The classification results have two classes: the yes class if it has dry land, less rainfall, less and sufficient water needs, insufficient and sufficient water availability, and sufficient and very good rice productivity, then carry out irrigation development. If the data is other than that, it has a class that does not develop irrigation canals. The accuracy value in the C4.5 algorithm method reaches 83%.

II. Spatial Datasets

A spatial dataset consists of two components: spatial data and attribute data. Establishing knowledge-based criteria and sub-criteria in decision-making based on a geospatial multi-criteria decision-making framework. [15]. This parameter refers to the previous research in Table I.

Depending on their physical form and ecosystem, the parameters of land types can be divided into two, namely wetland and dry land [8]; the data source refers to the land type of the 2018 physical for, Department of Agriculture and Food Security, East Java Province. The rainfall parameter is the amount of rainwater during a specific period and can be calculated in millimetres (mm) above horizontal ground level [9]; data source refers to rainfall data in the last ten years in East Java Province taken from Central Bureau of Statistics of East Java Province.

TABLE I. Spatial Datasets

Parameters	Range	Category
Land type		Wet
		Dry
Rainfall (R) (millimeter)	$50 \leq CH < 130$	Deficient
	$CH \geq 130$	Good
Water needs (W) (discharge per liter)	$W < 40.000$	Deficient
	$40.000 \leq W < 80.000$	Adequate
	$W \geq 80.000$	Very good
Water availability (A) (discharge per liter)	$A < 40$	Deficient
	$40 \leq A < 50$	Adequate
	$A \geq 50$	More than enough
Rice productivity (P) (quintals per hectare)	$P < 60$	Adequate
	$60 \leq P < 90$	Good
	$P \geq 90$	Very Good

The required water parameter is the volume of water needed to meet the very low water needs, beneficial for plants by considering the amount of water provided by nature through rain and soil-water[10]. The data were taken from the minimum service standard value for the reliability of the water needs of the irrigation area under the authority of the Province in 2017/2018 Public Works Department of East Java Province. Parameters of water availability come from rainwater (atmospheric), where surface water and groundwater will partially evaporate according to the climate [10]. The data were taken from the minimum service standard value for the reliability of the irrigation area's water needs under the Province's authority in the 2017/2018 Public Works Department of East Java Province. Rice production is the yield of rice in an area of land where per unit productivity can be measured in units of tons per hectare (tons/ha) [11], refers to rice productivity data according to the 2018 area sample framework of the East Java Province Agriculture and Food Security Service.

III. Methodology

The spatial data modelling of this sub-system determines the information or data sets that the GIS can generate [22]–[25]. The spatial data collection is used as the basis for determining the classification of irrigation canal development in East Java with case studies of 3 districts, namely Lamongan, Tuban, and Bojonegoro regencies. The stages for the spatial data modelling process for the classification of irrigation canal development can be seen in Fig. 1.

This stage describes all the working systems, first, from inputting all data requirements, using the C4.5 Algorithm method [26], then calculating all of the equations from the method, and finally the decision tree. To facilitate the calculation of the initialization values for each parameter, the data are classified as shown in Table II.

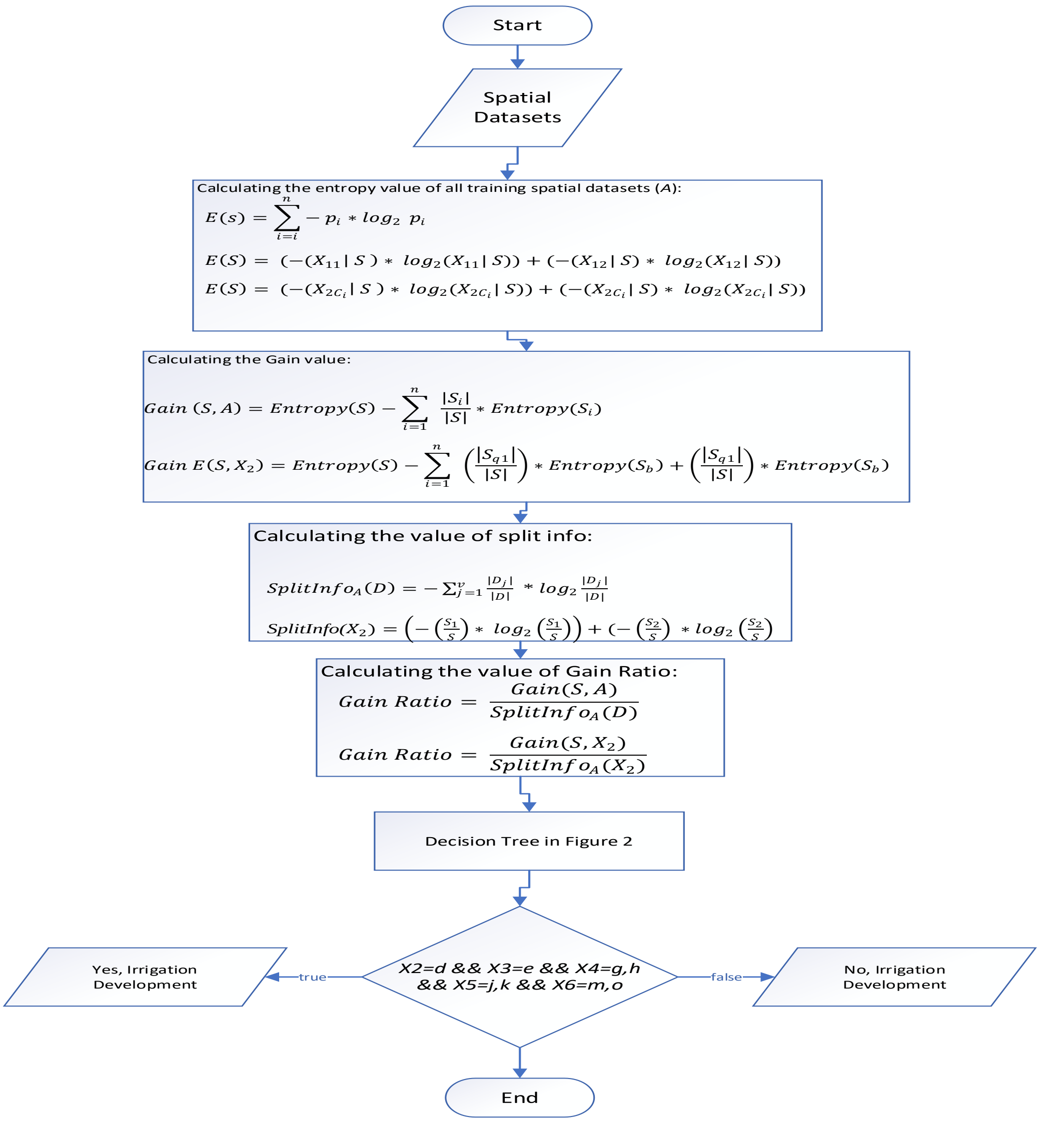

Fig. 1. Flowchart of spatial data modelling using C4.5 algorithm method

TABLE II. PARAMETERS CATEGORY

Parameters	**Initialization**		
Irrigation development	X_1	Yes	a
		No	b
Land type	X_2	Wet	c
		Dry	d
Rainfall	X_3	Deficient	e
		Good	f
Water needs	X_4	Deficient	g
		Adequate	h
		Very good	i
Water availability	X_5	Deficient	j
		Adequate	k
		Good	l
Rice productivity	X_6	Adequate	m
		Good	n
		Very good	o

Referring to Fig. 1, the flowchart of the C4.5 algorithm method can be carried out in several stages, as follows:

1. Calculate the entropy(s) value for the root node (all data) against the class composition. $(p_i | s)$ the proportion of *i*-th class in all training data X_1, X_2, which is processed at node s. $(p_i | s)$ is obtained from the number of all data rows with the class *i* table divided by the number of rows of all data.
2. Calculate the gain value used for choosing the test attribute on each node in (S, X_2) tree. X_2: land type, X_3: rainfall, X_4: water needs, X_5: water availability, X_6: rice productivity, $|S|$ is the total cases in S, and S_b entropy(s) used for the sample having the value of the class. Perform calculations to all attributes with the highest gain information picked as the test attribute of a node.
3. Calculate the *splitInfo* (S, X_2). S_i is the number of samples in attribute *I*, and S is the set of all cases.
4. Calculate the *gainRatio* used for calculating the most widely used criteria for selecting features as solvers in the C4.5 algorithm stating the gain ratio value for the

A-th feature. S is case set, A: attribute, D sum of values of splitInfo.

5. Form the decision tree with the highest gainRatio value as the branch node. After calculating the C4.5 algorithm, check the condition where if X_{2d}: is dry land type, X_{3e}:less rainfall, X_{4g} and X_{4h}: less water demand and sufficient, X_{5j}, X_{5k} is insufficient and sufficient water availability and X_{6m}, X_{6o}: rice productivity is sufficient and very good, so carry out irrigation development. If the data is other than that, then do not carry out irrigation development.

The C4.5 algorithm is one of the classification algorithms in data mining that uses a decision tree [12]. The decision tree is a tree used as a reasoning procedure to get answers to problems that have been entered. The decision tree is a powerful classification method that converts big facts into decision trees and represents a rule.

The C4.5 algorithm uses entropy(s), gain, split info and gain ratio calculations to select attributes to be nodes [26] [27]. The first step to calculate entropy(s) can be seen (1). Where, *S* variable is set of cases, *n* variable is the number of *S* partition, and p_i variable is the sum of the proportions of the *i*-th case set.

$$Entropy(s) = \sum_{i=i}^{n} - p_i * log_2\ p_i \quad (1)$$

The formula for calculating information gain can bee seen in (2). Where *A* variable is attribute, *n* variable is sum of attribute towards *A*, $|S|$ variable is total cases in S, $|S_i|$ variable is the proportion of S_i towards S, and $E(S_i)$ variable is entropy(s) used for samples that have i-th value.

$$Gain\ (S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * Entropy\ (S_i) \quad (2)$$

The next step to calculate the split info and gain ratio can be seen in (3) and (4). The D variable is the sample data space used for training, and the D_j variable is the number of the *j*-th attribute [13].

$$SplitInfo_A(D) = -\sum_{j=1}^{v} \frac{|D_j|}{|D|} * log_2 \frac{|D_j|}{|D|} \quad (3)$$

$$gainRatio = \frac{Gain(S,A)}{SplitInfo_A(D)} \quad (4)$$

The stages for making the decision tree algorithm C4.5 are as follows:

1. Calculate the value of entropy(s) by using (1). All data are grouped into certain classes to make it easier to calculate the initialization values for each parameter can be seen in Table II.
2. Calculating the Gain value to show the formation of the root node, the Gain of the *A*-th feature is calculated using (2).

$$G = E(S) + \left(\left(-\frac{|S_1|}{|s|}\right) + \left(\frac{|S_2|}{|s|}\right)\right) * \ E_a)$$

Where the *S* variable is set of all cases, the X_2 variable is the proportion of attribute $|s_1|$ and $|s_2|$ in *S*, the $|S|$ variable is total cases in *S*, and the E_a variable is entropy(s) used for samples with a class value from one of the classes in the *a*-th data.

3. The next step is to calculate the *splitInfo* value for node selection as an attribute usimg (3).

$$splitInfo\ (X_2) = \left(-\left(\frac{S_1}{S}\right) * \ log_2\left(\frac{S_1}{S}\right)\right) - \left(\frac{S_2}{S}\right) * log_2\left(\frac{S_2}{S}\right)$$

Where X_2, X_3, X_4, X_5, X_6 variable are data used for training and S_i is the sum of samples at *i* attribute.

4. Calculating the *gainRatio* to determine the root node with (4).

$$gainRatio = \frac{Gain(S, X_2)}{SplitInfo_A(X_2)}$$

The validation test in this paper uses a confusion matrix to determine the accuracy value [28]–[30] The confusion matrix is used to describe the evaluation of the classification by calculating the accuracy value, Precision-Recall on all correct prediction values divided by the entire data and computed using (5), (6), and (7). The details of the parameters are as follows:

a. *TP*: True positive prediction results are under the actual value, which is both true.
b. *TN*: True negative prediction results follow the actual value, which is both false.
c. *FP*: False-positive results predict correct, but the actual result is wrong.
d. *FN*: False-negative results predict wrong, but the actual result is true.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \dots \dots \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \dots \dots \quad (7)$$

IV. Results and Discussions

The test results to determine the development of irrigation canal based on the parameters of rice productivity, rainfall, water availability, land type, water demand using the decision tree method Algorithm C4.5. The process of the Decision Tree (Fig. 2) Algorithm C4.5 method, as follows:

X_1 = Irrigation Canal Development
X_2 = Soil Type
X_3 = Rainfall
X_4 = Water Requirements
X_5 = Water Availability
X_6 = Rice Productivity

The dataset used in the calculation of the C4.5 algorithm is the irrigation canal development. The irrigation canal development dataset comprises 70 data from three districts. The process of calculating the decision tree method is to calculate the entropy(s), Gain, splitInfo and gainRatio values, where the highest gainRatio value will be the first branch node. The results of the calculation of all attributes can be seen in Table III. The next step is to determine the results of the decision tree as seen in Fig. 2.

a. Calculate the entropy(s) value using (1).

entropy(s) $(X_1) = (-\frac{35}{70} * \ log\,2(\frac{35}{70})) + (-\left(\frac{35}{70}\right) * log\,2\,(\frac{35}{70})) = 1$

b. Obtain the *Gain* (X_2) from (2).

$$Gain\ (X_2) = 1 - \left(\left(\frac{42}{70} * 0.749\right) + \left(\frac{28}{70} * 0.371\right)\right) = 1.301$$

c. Calculate the *splitInfo* (X_2) using (3).

$$splitInfo\ (X_2) = \left(-\left(\frac{42}{70}\right) * log\ 2\left(\frac{42}{70}\right)\right) + \left(-\left(\frac{28}{70}\right) * log\ 2\left(\frac{28}{70}\right)\right) = 0.970$$

d. Calculate the *gainRatio* (X_2) using (4)

$$gainRatio\ (X_2) = \frac{1.301}{0.970} = 1.340$$

TABLE III. CALCULATION RESULTS OF ALL ATTRIBUTES

Att.	Category	Total	Yes	No	Entr.	Gain	Split Info	Gain Ratio
X_1		70	35	35	1			
X_2	Wet	42	9	33	0.50	1.30	0.970	1.34
	Dry	28	26	2	0.37			
X_3	Deficient	36	29	7	0.71	0.30	0.999	0.31
	Good	34	6	26	0.67			
X_4	Deficient	19	14	5	0.83	0.60	1.497	0.04
	Adequate	35	14	21	0.97			
	Better	16	7	9	0.99			
X_5	Deficient	28	22	6	0.75	0.17	1.51	0.11
	Adequate	29	10	19	0.93			
	Better	13	3	10	0.79			
X_6	Deficient	16	9	7	0.99	0.05	1.39	0.00
	Good	13	7	6	1.00			
	Verry Good	41	19	22	1.00			

The final result of the irrigation canal development decision tree can be seen in Fig. 2. No more nodes are processed, then the decision tree is declared complete. The final result of the decision tree is presented in Fig. 2.

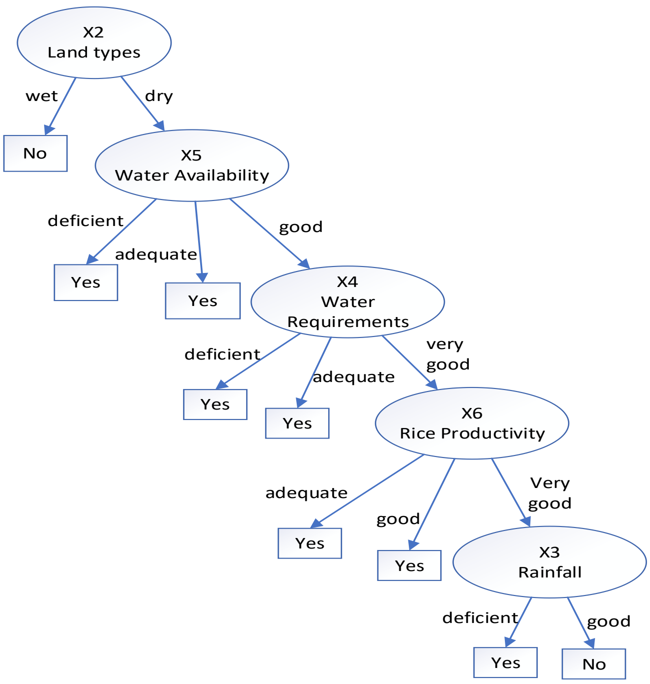

Fig. 2. The decision tree for Irrigation canal development

The form of the IF-THEN rule for the decision tree (see Fig, 2) is as follows:

- If the land type is **wet**, then the irrigation development is **no.**
- If the land type is **dry** and the water availability is **deficient**, then the irrigation development is **yes**.
- If the land type is **dry** and the water availability is **adequate**, then the irrigation development is **yes**.
- If the land type is **dry**, the water availability is **more than enough,** and the water requirements are **enough**, then the irrigation development is **yes**.
- If the land type is **dry**, the water availability is **more than enough,** and the water requirements are **deficient**, then the irrigation development is **yes**.
- If the land type is **dry**, the water availability and requirements are **more than enough,** and the rice productivity is **very good**, then the irrigation development is **yes**.
- If the land type is **dry**, the water availability and requirements are **more than enough,** and the rice productivity is **adequate**, then the irrigation development is **yes.**
- If the land type is **dry**, the water availability and water requirements are **more than enough,** and the rice productivity is **good,** and the rainfall is **deficient**, then the irrigation development is **yes**.
- If the land type is **dry**, the water availability and requirements are more than enough, the rice productivity is good, and the rainfall is **good**, then the irrigation development is **no**.

The Decision Tree Algorithm C4.5 method calculation results to determine areas that require irrigation canal development are shown in Fig.3.

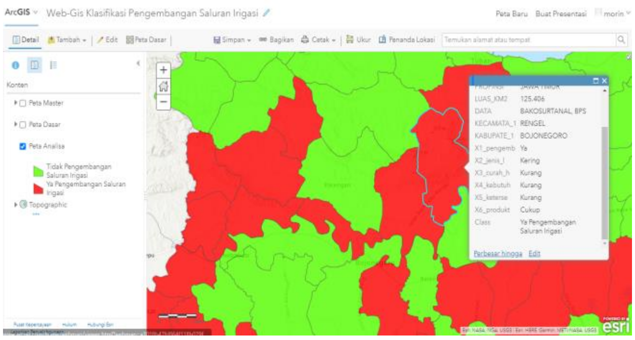

Fig. 3. Web-GIS with **yes** classification for irrigation development

Meanwhile, classes that do not develop irrigation canals can be seen in Fig. 4. There are Attribute Data of five parameters for all regions; namely there are three districts and have two class results, which are yes, which has a red Legend, and one that does not have a green Legend in each District in three districts.

There are 70 data (see Table IV) used for testing. Confusion Matrix Calculation Test Value Method Decision Tree Algorithm C.45 results are presented in Table V.

Based on Table V, using the C4.5 Algorithm method and referring to (5) – (7), the accuracy values obtained, Precision and Recall, reached 83%, 96% and 69%, respectively, of the data that had been accumulated with real data. These results can be used as an information reference for the development of irrigation canals in East Java.

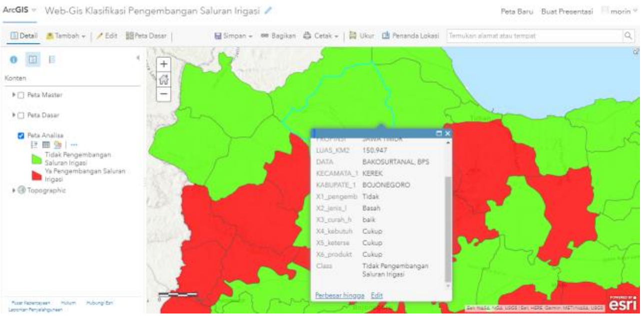

Fig. 4. Web-GIS with **no** classification for irrigation development

TABLE IV. REAL DATA RESULTS AND CALCULATIONS OF DECISION TREE ALGORITHM C.45

Data Real for classification for irrigation development		**Data Metode Algoritma C4.5 for classification for irrigation development**	
Yes	***No***	***Yes***	***No***
35	35	25	45

TABLE V. CONFUSION MATRIX CALCULATION TEST VALUE METHOD DECISION TREE ALGORITHM C.45

		Predict	
		True	***False***
Actual	True	24	11
	False	1	34

CONCLUSION

Based on the classification analysis of irrigation canal development results in East Java using the parameters of land type, rainfall, rice productivity, water availability, and water demand, it is much accurate using the Decision Tree Algorithm C4.5 method. The accuracy value of the C4.5 method reaches 83%, Precision 96% and Recall 69%. Thus, it can be recommended to use the analysis process of irrigation canal development areas in the data series the following year. It is hoped that when there is new data, further research is more developed in carrying out automatic updates. It is necessary to compare other similar methods such as ID3 or CART to compare the accuracy values of the methods used, add different and more specific parameter data.

REFERENCES

[1] I. M. Yoga Prasada, A. Dhamira, and A. Dwi Nugroho, "Supply response of paddy in east java: policy implications to increase rice production," *Agrar. J. Agribus. Rural Dev. Res.*, vol. 4, no. 2, 2018.

[2] Statistics Indonesia, "The 2020 harvested area and production of paddy in Indonesia", 2020. [Online]. Available: https://www.bps.go.id/publication/2021/07/12/b21ea2ed9524b784187be1ed/luas-panen-dan-produksi-padi-di-indonesia-2020.html#:~:text=Sementara%20itu%2C%20produksi%20padi%20pada,dengan%20produksi%20beras%20tahun%202019.

[3] ADB, *Policies to Support Investment Requirements of Indonesia's Food and Agriculture Development During 2020-2045*. 2019.

[4] C. B. Luttrell, *Outlook for Agriculture*, vol. 58. 1976.

[5] N. A. for D. Management, "Strengthening National Disaster Databases in Asia Study in Indonesia," 2013.

[6] N. A. for D. Management, *Indonesia National Disaster Management Plan 2010-2014*. 2010.

[7] J. Liu *et al.*, "Water quality in irrigated paddy systems," *Irrig. Agroecosystems*, pp. 1–17, 2019.

[8] P. R. Indonesia, *Undang-Undang RI No 17 Thn 2019 Tentang Sumber Daya air*. 2019.

[9] Suparwoto, "Pembangunan dan pengaruh waduk pacal terhadap pertanian masyarakat kabupaten Bojonegoro," vol. 2, no. 3, pp. 90–103, 2014.

[10] C. Anam and A. Amiroh, "Kajian Lahan Kritis dan Sumber-Sumber Air di Lamongan," *J. Kelitbangan*,vol. 2, no.4, pp. 66–74, 2019.

[11] "https://tubankab.go.id/entry/pemkab-perbaiki-saluran-irigasi-di-12-kecamatan." .

[12] A A. W. Worqlul, Y. T. Dile, J. Jeong, Z. Adimassu, N. Lefore, T. Gerik, R. Srinivasan, and N. Clarke, "Effect of climate change on land suitability for surface irrigation and irrigation potential of the shallow groundwater in Ghana," *Computers and Electronics in Agriculture*, vol. 157, pp. 110–125, 2019.

[13] N. Tryfona and C. S. Jensen, "Conceptual data modeling for spatiotemporal applications," *Geoinformatica*, vol. 3, no. 3, pp. 245–268, 1999.

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

[15] M. Paul, M. Negahban-Azar, A. Shirmohammadi, and H. Montas, "Assessment of agricultural land suitability for irrigation with reclaimed water using geospatial multi-criteria decision analysis," *Agric. Water Manag.*, vol. 231, pp. 1–15, Dec. 2020.

[16] H. Rengganis, "Zonasi Wilayah Pendayagunaan Sumber Daya Air untuk Pembangunan Irigasi di Pulau Sumba, Nusa Tenggara Timur," *Anal. Kebijak. Pertan.*, vol. 14, no. 1, p. 17-33, 2018.

[17] 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.

[18] A. Nurkholis and I. S. Sitanggang, "A spatial analysis of soybean land suitability using spatial decision tree algorithm," *Proc.of SPIE*, Dec. 2019, p.1137201-1-10.

[19] P. Christias, I. N. Daliakopoulos, T. Manios, and M. Mocanu, "Comparison of three computational approaches for tree crop irrigation decision support," *Mathematics*, vol. 8, no. 5, pp. 1–26, 2020.

[20] L. Lilburne, J. Watt, and K. Vincent, "A prototype DSS to evaluate irrigation management plans," *Comput. Electron. Agric.*, vol. 21, no. 3, pp. 195–205, 1998.

[21] R. González Perea, E. Camacho Poyato, P. Montesinos, and J. A. Rodríguez Díaz, "Prediction of irrigation event occurrence at farm level using optimal decision trees," *Comput. Electron. Agric.*, vol. 157, no. June 2018, pp. 173–180, 2019.

[22] 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.

[23] A. V. Vitianingsih, D. Cahyono, and A. Choiron, "Analysis and design of web-geographic information system for tropical diseases-prone areas: a case study of East Java Province, Indonesia," in *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2017, pp. 255–260.

[24] 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," *Int. J. Innov. Comput. Inf. Control*, vol. 16, no. 6, pp. 1943–1957, 2020.

[25] A. V. Vitianingsih and D. Cahyono, "Geographical information system for mapping road using multi-attribute utility method," in *International Conference on Science and Technology-Computer (ICST)*, 2016, pp. 0–4.

[26] G. Sharma, R. Bhargava, and M. Mathuria, "Decision tree analysis on j48 algorithm," *Int. J. Adv. Res. inComputer Sci. Softw. Eng.*, vol. 3, no. 6, pp. 1114–1119, 2013.

[27] A. A. Aprilia Lestari, "Increasing accuracy of c4.5 algorithm using information gain ratio and adaboost for classification of chronic kidney disease," *J. Soft Comput. Explor.*, vol. 1, no. 1, pp. 32–38, 2020.

[28] 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.

[29] W. Song and H. Li, "A new method for acquiring long-term high-precision spatial data on rural settlements," *MethodsX*, vol. 8, no. 101249, pp. 1–7, 2021.

[30] G. M. Foody, "Explaining the unsuitability of the kappa coefficient in the assessment and comparison of the accuracy of thematic maps obtained by image classification," *Remote Sens. Environ.*, vol. 239, pp. 1-18, Mar. 2020.