

C.45 Algorithm for Classification of Causes of Landslides

Yopi Handrianto¹⁾
Universitas Bina Sarana Informatika
Jakarta, Indonesia
yopi.handrianto@gmail.com

Muhammad Farhan²⁾
STMIK Nusa Mandiri
Jakarta, Indonesia
mfarhan_207@yahoo.com

Abstract— Natural disasters are disasters caused by natural events and cannot be avoided including earthquakes, tsunamis, volcanic eruptions, floods, hurricanes, droughts and landslides. One of the natural disasters that often occurs in Indonesia is a landslide disaster. One of the regencies in West Java Province that had experienced a landslide was a Purwakarta district area. Landslide is one type of mass or rock mass movement, or a mixture of both, down or out of the slope due to the disruption of the stability of the soil or rocks making up the slope. With a data mining approach that uses the decision tree method or C4.5 Algorithm, a classification model will be made where the model functions as a classification of the causes of landslides in Purwakarta district.

Keywords— Natural Disasters, Landslides, Decision Trees, Algorithm C4.5.

I. INTRODUCTION

According to (Undang-Undang Republik Indonesia Nomor 24 Tahun 2007 Tentang Penanggulangan Bencana, 2007) concluded that : a disaster is an event or a series of events that threaten and disrupt people's lives and livelihoods caused, either by natural or non-natural factors or human factors, resulting in human casualties, environmental damage, property losses and psychological impacts. Natural disasters are disasters caused by natural events and cannot be avoided including earthquakes, tsunamis, volcanic eruptions, floods, hurricanes, droughts and landslides. One of the natural disasters that often occur in Indonesia is landslides with 496 times in 2019. In the Province of West Java during 2019 there have been 178 landslides. Of the total landslide events in 2019, West Java is the second most vulnerable area to landslides in Indonesia after Central Java with 217 events (BNPB, 2017). One of the regencies in West Java that had experienced a landslide was Purwakarta. Landslides are one type of mass or rock mass movement, or a mixture of both, down or out of slopes due to disturbance in the stability of the soil or rocks making up the slope (BNPB, 2017). The causes of landslides include the addition of loads on the slope, excavation of the soil

at the foot of the slope, an increase in soil pressure by water, a decrease in the sliding resistance of the slope due to increased water pressure, and a vibration or earthquake (Hardiyatmo, 2012). From the explanation above, some of the causes of landslides are caused by humans, for example excavation of slopes for road construction, construction of railroad tracks, construction of housing and mining of sand around rivers.

With a data mining approach that uses the Decision Tree Method or C4.5 Algorithm, a classification model is created where the model functions as a classification of the causes of landslides in Purwakarta district. The purpose of making a classification of the causes of landslides in Purwakarta district area is by using the C4.5 algorithm to provide information specifically for landslide researchers and generally for the public about the level of potential hazards from landslides. Data mining has attracted much attention in the world of information systems and society in general in recent years, due to the wide availability of large amounts of data and the immediate need to turn that data into useful information and knowledge (Septiani, 2017). According to Giudici & Figini in (Sunge, 2018) the use of data mining is used to analyze a behavior or prediction, also not only used

in computer science but other fields such as business and industry. Classification is one of the techniques of data mining that can be used to help predictions. Classification can be done by Decision Tree using C4.5 Algorithm (Fatmawati, 2016). Data mining has attracted much attention in the world of information systems and society in general in recent years, due to the wide availability of large amounts of data and the immediate need to turn that data into useful information and knowledge (Septiani, 2017). According to Giudici & Figini in (Sunge, 2018) the use of data mining is used to analyze a behavior or prediction, also not only used in computer science but other fields such as business and industry. Classification is one of the techniques of data mining that can be used to help predictions. Classification can be done by Decision Tree using C4.5 Algorithm (Fatmawati, 2016).

II. LITERATURE REVIEW

A. Data Mining

According to Turban in (Nofriansyah, 2015) concluded that : data mining is a term used to describe the discovery of knowledge in a database. Data Mining is a process that uses statistical, mathematical, artificial intelligence and machine learning techniques to extract and identify useful information and related knowledge from various large databases. According to Mulyadi in (Sudiby et al., 2018) data mining is a technique of analyzing data (for large-scale data), so finding clear relationships and concluding unknowns in the latest way so that they are easy to understand and useful for data owners. Classification (taxonomy) is the process of placing certain objects (concepts) in a set of categories, based on each object (concept) property, according to Gorunescu in (Septiani, 2017). The classification process is based on four basic components namely: class, predictor, training set, and data set testing. Classification is one of the techniques of data mining that can be used to help predictions. Classification can be done with the Decision Tree using the C4.5 algorithm (Fatmawati, 2016). According to Moghimipour in (Puspita & Wahyudi, 2015) concluded that : Classification is part of the prediction, where the predicted value is in the form of a label. Classification determines the class or group for each data sample, the input of the classification model is an attribute of the sample data and the output is the class of the data samples themselves, in machine learning to build a classification model using the supervised learning method.

B. Decision Tree

Decision tree is a structure that can be used to convert data into decision trees that will produce large decision rules into smaller record sets by applying a series of decision rules (Sukma, Halfis, & Hermawan, 2019). According to Han & Kamber in (Septiani, 2017) decision tree is one of the classification methods that use tree structure representations where each node represents an attribute and leaves represent a class. The topmost node of the decision tree is root. According to Han & Kamber in (Tabrani, 2016), the most popular classification technique used is the decision tree. While the algorithm that can be used in a decision tree is the C4.5 algorithm.

C. Algorithm C.45

According to (Prasetyo, 2014) explained that, "C4.5 algorithm was introduced by J. Ross Quinlan (1996) as an improved version of ID3 (Iterative Dichotomiser 3). The improvements that differentiate the C4.5 algorithm from ID3 are that it can handle numeric type features, pruning decision tree, and deriving rule set. The C4.5 algorithm is used for the decision tree. According to Larose in (Tabrani, 2016) the steps in making a decision tree in C4.5 algorithm:

1. Prepare training data. Training data is usually taken from historical data that has happened before or is called past data and has been grouped in certain classes.
2. Count the roots of the tree. The root will be taken from the attribute to be selected, by calculating the gain value of each attribute, the highest gain value that will be the first root. Before calculating the gain value of the attribute, first calculate the entropy value. To calculate the entropy value the formula is used:

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i \log_2 p_i$$

Note :

S = Case set

n = Number of partitions S

Pi = Proportion of Si to S

Then calculate the gain using the formula:

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

Note :

S = Case Set

A = Feature

n = Number of attribute attributes A

| Si | = Proportion of Si to S

| S | = Number of cases in S

3. Repeat step 2 and step 3 until all records are closed.

4. The decision tree partition process will stop when:
 - a. All records in node N have the same class.
 - b. There are no attributes in the partitioned record anymore.
 - c. There are no records in the empty branch.

D. Landslide

According to the Bejis Project Team in (Priyono, 2015), a landslide is one of the results of a mass movement along its critical slide area. Mass movement is the transfer of rock mass and soil from high to low due to the influence of gravity. Landslides are one type of mass or rock mass movement, or a mixture of both, down or out of slopes due to disturbance in the stability of the soil or rocks making up the slope (BNPB, 2017). Landslides are one type of mass or rock mass movement, or a mixture of both, down or out of slopes due to disturbance in the stability of the soil or rocks making up the slope (BNPB, 2017). According to (Hardiyatmo, 2012) concluded that: Avalanche is a mass movement of soil or rock in a potential landslide field. This ground mass movement is a sagging downward movement of slope-forming material, which can be soil, rock, pile or mixed with other materials. If the land mass movement is very excessive, it is called a landslide. According to (Hardiyatmo, 2012) concluded that : landslide is a mass movement of soil or rock in a potential landslide field. This ground mass movement is a sagging downward movement of slope-forming material, which can be soil, rock, pile or mixed with other materials. If the land mass movement is very excessive, it is called a landslide. Many factors cause landslides, such as geological and hydrological conditions, topology, climate, and weather changes that affect the stability of slopes that cause landslides.

E. RapidMiner

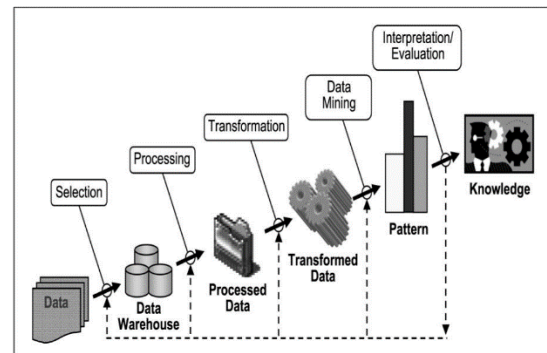
RapidMiner is opensource software. RapidMiner is a solution for analyzing data mining, text mining and predictive analysis. RapidMiner uses a variety of descriptive and predictive techniques in providing insights to users so they can make the best decisions. RapidMiner has approximately 500 data mining operators, including operators for input, output, data preprocessing and prediction visualization. RapidMiner provides a GUI (Graphic User Interface) for designing an analytical pipeline. This GUI will produce an XML (Extensible Markup Language) file that defines the user's analytical process to be applied to predictive data (Aprillia, Baskoro, Ambarwati, & Wicaksana, 2013).

F. Decision Tree on RapidMiner

RapidMiner as a data mining processing software provides tools for making decision trees. This will certainly make it easier for us to make a decision tree using RapidMiner compared to making a decision tree manually by performing calculations using the C4.5 algorithm (Aprillia et al., 2013).

III. PROPOSED METHOD

In this study, using the concept of data mining, Knowledge Discovery in Databases, which is a framework as a reference in assessing data mining. The stages of Knowledge Discovery in Databases, will be explained in (Figure III.1),



Sources: Fayyad in (Sukma et al., 2019)

Figure 1. Knowledge Discovery in Databases

In conducting this research, research steps are needed so that the objectives of the research can be achieved. The research steps were carried out by the author using the stages of knowledge discovery in databases :

1. Data Selection. Selection (selection) of data from a set of operational data needs to be done before the information gathering stage in Knowledge Discovery in Databases begins. Data from the selection results used for the data mining process are stored in a separate file from the operational database.
2. Pre-processing (Cleaning). Before the data mining process can be carried out, it is necessary to clean up the data that is the focus of Knowledge Discovery in Databases. The cleaning process includes removing duplicate data, checking inconsistent data, and correcting errors in the data.
3. Transformation. Coding is the process of transformation of data that has been selected, so that the data is suitable for the data mining process. The coding process in Knowledge Discovery in Databases is a creative process and

- is very dependent on the type or pattern of information that will be searched in the database.
4. Data mining. Data mining is the process of finding interesting patterns or information in selected data using certain techniques or methods. Techniques, methods, or algorithms in data mining vary greatly. The choice of the right method or algorithm is very dependent on the objectives and the process of Knowledge Discovery in Databases as a whole.
 5. Interpretation (Evaluation). The pattern of information generated from the data mining process needs to be displayed in a form that is easily understood by interested parties. This stage is part of the Knowledge Discovery in Databases process called interpretation. This stage involves checking whether the pattern or information found is contrary to the facts or hypotheses that existed before.

IV. RESULT AND DISCUSSION

There are 4 stages in the decision tree modeling using the C4.5 algorithm :

1. Determine Training Data

Training Data is determined from historical data that has happened before. The data used in this study are 17 landslide events over the past 10 years.

Table 1. Training Data That Cause of Landslides

Landslide Hazard Zone	Slope	Weather	Type of Soil	Causes of Landslides
High	Steep	Rain	Latosol	Yes
Middle	Sloping	Rain	Clay	Yes
High	Rather Steep	Rain	Latosol	Yes
Low	Sloping	Bright	Latosol	No
Middle	Steep	Rain	Clay	Yes
Middle	Steep	Bright	Clay	No
High	Rather Steep	Bright	Latosol	No
Middle	Steep	Rain	Clay	Yes
Middle	Sloping	Rain	Clay	No
High	Steep	Rain	Latosol	Yes
Middle	Sloping	Rain	Clay	Yes
Middle	Sloping	Rain	Latosol	No
Middle	Rather Steep	Rain	Clay	No
High	Steep	Rain	Latosol	Yes
Low	Rather Steep	Rain	Clay	No
Middle	Sloping	Rain	Latosol	No
High	Steep	Rain	Clay	Yes

Note :

Sloping : 8 – 15%, **Rather Steep** : 15 – 25%, **Steep** : 25 – 45%.

High : Tegalwaru Subdist., Pondok Salam Subdist., Bojong Subdist., Sukatani Subdist., Wanayasa Subdist.

Middle : Bojong Subdist., Kiarapedes Subdist., Tegalwaru Subdist., Sukatani Subdist., Wanayasa Subdist., Mani'is Subdist. **Low** : Darangdan Subdist.

2. Calculating the Root Value

The root will be taken from the attributes that will be selected by calculating the gain value of each attribute, the highest gain value that will be the first root. Before calculating the gain value of the attribute, first calculate the entropy value. To calculate the entropy value the formula is used, $Entropy(S) = \sum_{i=1}^n - p_i \log_2 p_i$

Note :

S = Case set

n = Number of Partitions S

Pi = Proportion of Si to S

Then,

$$Entropy(S) = \left(- \left(\frac{Sum(No)}{Total} \right) \times \log_2 \left(\frac{Sum(No)}{Total} \right) \right) + \left(- \left(\frac{Sum(No)}{Total} \right) \times \log_2 \left(\frac{Sum(No)}{Total} \right) \right)$$

$$Entropy(\text{all cases}) = \left(- \left(\frac{9}{17} \right) \times \log_2 \left(\frac{9}{17} \right) \right) + \left(- \left(\frac{8}{17} \right) \times \log_2 \left(\frac{8}{17} \right) \right) = 0,997502546$$

Then calculate the gain value using the formula,

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

Note :

S = Case Set

A = Feature

n = number of attribute attributes A

| Si | = Proportion of Si to S

| S | = number of cases in S

Then,

$$Gain(S,A) = entropy(total) - \left(\left(\frac{Sum(total)}{number\ of\ case} \right) * \right)$$

$$Entropy(atribut)$$

$$Gain(\text{regional zone}) = 0,997502546 - \left(\left(\frac{6}{17} \right) * 0,650022422 + \left(\frac{9}{17} \right) * 0,991076060 + \left(\frac{2}{17} \right) * 0 \right) = 0,243395542.$$

3. Entropy and Gain Value Calculation Results

After calculating the entropy value of all attributes, it will appear in the table below,

Table 2. Entropy Calculation

Attribute	Attribute Value	Sum Value	Yes	No	Entropy
Total	Total	17	9	8	0,997502546
Landslide Hazard Zone	High	6	5	1	0,650022422
	Middle	9	4	5	0,991076060
	Low	2	0	2	0
Slope	Steep	7	6	1	0,591672779
	Rather Steep	4	1	3	0,811278124
	Sloping	6	2	4	0,918295834
Wheater	Rain	14	9	5	0,940285959
	Bright	3	0	3	0
Type of Soil	Latosol	8	4	4	1
	Clay	9	5	4	0,991076060

The next step is to calculate the gain value for each attribute, it will appear in the table below,

Table 3. Calculation of Gain

Attribute	Attribute Value	Sum Value	Yes	No	Entropy	Gain
Total	Total	17	9	8	0,997502546	
Landslide Hazard Zone	High	6	5	1	0,650022422	0,243395542
	Middle	9	4	5	0,991076060	
	Low	2	0	2	0	
Slope	Steep	7	6	1	0,591672779	0,238879196
	Rather Steep	4	1	3	0,811278124	
	Sloping	6	2	4	0,918295834	
Wheater	Rain	14	9	5	0,940285959	0,610325975
	Bright	3	0	3	0	
Type of Soil	Latosol	8	4	4	1	0,002226985
	Clay	9	5	4	0,991076060	

4. Determine the Decision Tree.

Based on table 3, the weather attribute has the largest gain value, 0.610325975, then the weather attribute becomes the root node. If drawn in a decision tree, the picture will look like in the picture below.

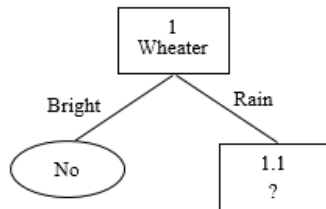


Figure 2. Decision Tree Node 1

Based on the decision tree node 1 (root Node) that has been formed, it will proceed to the next stage. In the training data filtered by just taking data on rainy weather. Then the Entropy calculation is repeated for each attribute value and also calculates the Gain for each attribute. So the results are obtained as in table 4, then determine the attribute that has the highest Gain value to form the next decision tree node.

Table 4. Node Analysis Results 1.1

Attribute	Attribute Value	Sum Value	Yes	No	Entropy	Gain
Total	Total	14	9	5	0,940285959	
Landslide Hazard Zone	High	5	5	0	0	0,368857388
	Middle	8	4	4	1	
	Low	1	0	1	0	
Slope	Steep	6	6	0	0	0,396740211
	Rather Steep	3	1	2	0,918295834	
	Sloping	5	2	3	0,970950594	
Type of Soil	Latosol	6	4	2	0,918295834	0,001339743
	Clay	8	5	3	0,954434003	

In the table above the highest Gain value is owned by the slope attribute of 0.396740211, then the slope is used as Node as the root node 1.1. The value used as a leaf (branch) is Steep, Rather Steep and Sloping, then the decision tree will look like in Figure 2.

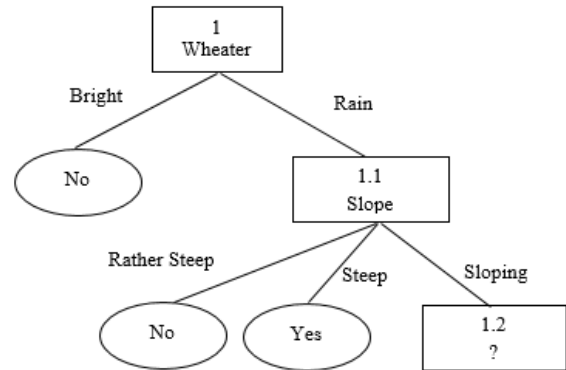


Figure 3. Decision Tree Node 1.1

Based on the decision tree Node 1.1, it will proceed to the next stage. In the Training Data filter is performed, to select the data of the slope, and the selected data is sloping. Then the Entropy calculation is repeated for each attribute value and also calculates the Gain for each attribute. So the results are as in table 5, then determine which attribute has the highest Gain value to form the next decision tree node.

Table 5. Node Analysis Results 1.2

Attribute	Attribute Value	Sum Value	Yes	No	Entropy	Gain
Total	Total	5	2	3	0,970950594	
Landslide Hazard Zone	High	0	0	0	0	0
	Middle	5	2	3	0,970950594	
	Low	0	0	0	0	
Type of Soil	Latosol	2	0	2	0	0,419973094
	Clay	3	2	1	0,918295834	

In the table above the highest Gain value is owned by the soil type attribute, 0.419973094, then the soil type is used as a node as the root node 1.2. The values used as leaf (branch) are Latosol, and Clay, then the decision tree will look like Figure 3.

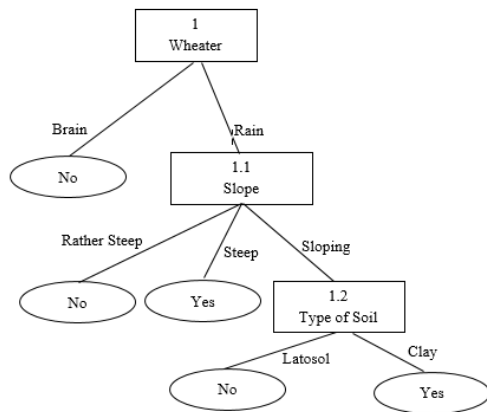


Figure 4. Decision Tree Node 1.2

Based on the decision tree node 1.2 where the soil type is the result of node 1.2, then no attributes are left. The soil type attribute becomes the last node and the values used as leaf (branch) are Latosol and Clay. Then from table 5 a decision can be made from each type of soil : Latosol = No, Clay = Yes. So that these results can be described in the whole decision tree as follows in Figure 4.

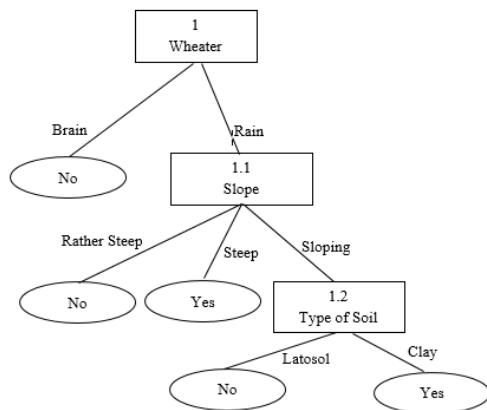


Figure 5. The Last Decision Tree

From the decision tree, the following rules are obtained,

1. IF (Weather = Bright) THEN No.
2. IF (Weather = Rain AND Slope = Steep) THEN Yes.
3. IF (Weather = Rain AND Slope = Steep AND Soil Type = Clay) THEN Yes.
4. IF (Weather = Bright AND Slope = Rather Steep) THEN No.
5. IF (Weather = Rain AND Slope = Sloping AND Land Type = Clay) THEN Yes.
6. IF (Weather = Bright AND Slope = Rather Steep AND Soil Type = Latosol) THEN No.

Based on the results of the final decision tree and the training data used to determine the decision tree, the following conclusions,

1. Causes of landslides in Purwakarta district area are based on decision trees: weather, slope and soil type.
2. Based on the decision tree the weather attribute becomes the root node with the rain branch as the main cause of landslides. The slope attribute with steep and sloping branches is the second cause of landslides. Attributes of soil type with clay branches as the third cause of landslides.
3. Based on the attributes of the landslide hazard zone contained in the training data, it can be concluded that the area with a high level of landslide hazard is Tegalwaru sub-district, pondok salam sub-district, Bojong sub-district, sukutani sub-district, wanayasa sub-district.

From the test results using the RapidMiner application, the results obtained by the number of True Positive (TP) are as many as 8 records classified as causing landslides and False Negative (FN) as many as 7 records classified as non-causative categories, but there are possible causes. Next to the number of True Negative (TN) of 1 record is categorized as not causing landslides and the number of False Positives (FP) is categorized as a cause but has a risk of not causing a total of 1 record.

Table View Plot View

accuracy: 50.00% +/- 40.82% (micro average: 52.94%)

	true Ya	true Tidak	class precision
pred. Ya	8	7	53.33%
pred. Tidak	1	1	50.00%
class recall	88.89%	12.50%	

Figure 6. Output Accuracy Data on RapidMiner 9.3

Based on the data above, we can calculate it to find the accuracy, Sensitivity, Specificity, PPV and NPV values:

$$\text{Sensitivity} = \frac{TP}{TP+FN} = \frac{8}{8+7} = 0,5333$$

$$\text{Spesificity} = \frac{TN}{TN+FP} = \frac{1}{1+1} = 0,5$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{8+1}{8+1+1+7} = 0,52$$

$$\text{PPV} = \frac{TP}{TP+FP} = \frac{8}{8+1} = 0,8888$$

$$\text{NPV} = \frac{TN}{TN+FN} = \frac{1}{1+7} = 0,125$$

From the above calculation the value of Accuracy, Sensitivity, Specificity, PPV and NPV can be seen in the table below which has been converted into a percentage.

Table 6. The Calculation Results

Calculation results for Accuracy, Sensitivity, Specifity, PPV and NPV in (%)	
Accuracy	52
Sensitivity	53,33
Specifity	50
PPV	88,88
NPV	12,5

Based on the table above shows that the level of accuracy using the C4.5 algorithm is 52%. The calculation results are visualized with the ROC curve. Comparison of the two classes can be seen in Figure 7 which is the ROC curve for the C45 algorithm. The ROC curve in Figure 7 expresses the confusion matrix of Figure 6. Horizontal lines are false positives and vertical lines are true positives. Generates an AUC (Area Under Curve) value of

V. CONCLUSION AND SUGGESTION

Based on the results of research and discussion that has been outlined, it can be concluded:

1. The cause of landslides in Purwakarta regency is based on the decision tree, namely weather, slope and soil type.
2. Based on the zone zone attributes, the landslide hazard contained in the training data can be concluded that the areas with high landslide hazard level are Tegalwaru District, Pondok Salam District, Bojong Subdistrict, Sukatani Subdistrict, Wanayasa Subdistrict.
3. Implementation of data mining with the C4.5 algorithm can produce information in the form of a classification of the causes of landslides where from the training data used with the number 17 a decision tree that can produce 6 rules can be built.
4. Based on the test results using RapidMiner which measured its accuracy using the confusion matrix test and the ROC curve produces an accuracy value of 52 percent and produces an AUC (Area Under Curve) value of 0.500 with an incorrect classification accuracy value (Incorrect Classification).

Some suggestions that can be submitted for further research:

1. The Decision Tree method using the C4.5 Algorithm should be implemented in the form of a useful application program for analyzing to suppress the numbers to the smallest possible.
2. To improve the quality of research results, it is better that further research use more training data.
3. For further research, other methods such as Naive Bayes can be tested for use in classification so that we can compare methods previously used that have a higher level of accuracy.

0.500 with an incorrect classification accuracy value (Incorrect Classification).

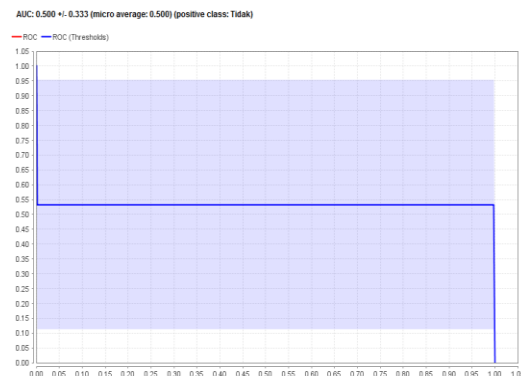


Figure 7. AUC Results in ROC Graph C4.5 Algorithm

VI. REFERENCES

- Aprillia, D. C., Baskoro, D. A., Ambarwati, L., & Wicaksana, I. W. S. (2013). *Belajar Data Mining Dengan Rapid Miner*. Retrieved from https://www.academia.edu/7712860/Belajar_Da ta_Mining_dengan_RapidMiner
- BNPB. (2017). *Definisi Bencana*. p. 1. Retrieved from <https://bnpb.go.id/home/definisi>
- Fatmawati. (2016). *Perbandingan Algoritma Klasifikasi Data Mining Model C.45 Dan Naive Bayes Untuk Prediksi Penyakit Diabetes*. XIII(1), 50–59.
- Hardiyatmo, H. C. (2012). *Tanah Longsor Dan Erosi* (2nd ed.). Yogyakarta: Gajah Mada University Press.
- Nofriansyah, D. (2015). *Konsep Data Mining VS Sistem Pendukung Keputusan*. Yogyakarta: Deepublish.
- Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab* (1st ed.). Yogyakarta: Andi Offset.
- Priyono. (2015). *Hubungan Klasifikasi Longsor, Klasifikasi Tanah Rawan longsor dan Klasifikasi Tanah Pertanian Rawan Longsor*.
- Puspita, A., & Wahyudi, M. (2015). *Algoritma C4.5 Berbasis Decision Tree Untuk Prediksi Kelahiran Bayi Prematur*. 97–102.
- Septiani, W. D. (2017). *Komparasi Metode Klasifikasi Data Mining Algoritma C.45 Dan Naive Bayes Untuk Prediksi Penyakit Hepatitis*. 13(1), 76–84.

- Sudibyo, A., Asra, T., Rifai, B., Informasi, S. S., Informatika, B. S., & Informatika, S. T. (2018). *Klasifikasi Seleksi Atribut Pada Serangan Spam Menggunakan Metode Algoritma Decision Tree*. 14(2), 145–150.
- Sukma, A. R., Halfis, R., & Hermawan, A. (2019). *Klasifikasi Channel Youtube Indonesia Menggunakan Algoritma C4 . 5*. V(1), 21–28. <https://doi.org/10.31294/jtk.v4i2>
- Sunge, A. S. (2018). *Optimasi Algoritma C4 . 5 Dalam Prediksi Web Phishing Menggunakan Seleksi Fitur Genetic Algoritma*. XX(2), 27–32. <https://doi.org/10.31294/p.v>
- Tabrani, M. (2016). *Klasifikasi Penerima Beasiswa Kopertis Dengan Menggunakan Algoritma C.45*. (1), 72–80.
- Undang-Undang Republik Indonesia Nomor 24 Tahun 2007 Tentang Penanggulangan Bencana. , (2007).