*Jurnal Mantik, 6 (3) (2022), ISSN 2685-4236 (Online)*

IOCSCIENCE

Published by: Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

# Classification of Tsunami Potential Based on Earthquakes in Indonesia Using the C4.5 Algorithm

Eni Irfiani[1], Arnoldus Pius Purnomo Raben Galla[2], Vivien Sufi Hadi[3], Muhammad Iqbal Syaputra[4], Sriyadi5, Haryani[6], Dio Okta Rovelino[7]

[1,2,3,4,5,6,7]Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, Jl. Kramat Raya No. 98, Senen District, Central Jakarta 10450

ARTICLE INFO

*Article history:*

Received Oct 26, 2022
Revised Nov 02, 2022
Accepted Nov 23, 2022

*Keywords:*

Tsunami Potential,
Earthquake,
Data Mining,
Classification,
C4.5 Algorithm

ABSTRACT

Natural disasters are natural phenomena that are difficult to prevent or avoid. The tsunami disaster caused by the earthquake was a natural disaster with a large impact with many casualties and material and non-material losses. One of the efforts to help deal with the tsunami natural disaster with the help of the Data Mining method is to classify the potential for a tsunami caused by an earthquake in Indonesia based on the BMKG dataset so that the community and government are alert and able to reduce the impact. Classification technique to predict the potential occurrence of tsunami waves generated by earthquakes in Indonesia by applying the C4.5 algorithm. The results of data processing obtained a prediction of the potential for a tsunami based on the classification of the attribute magnitude of the earthquake strength at sea (SR) and the area where the earthquake occurred (KM). The results of model testing using the confusion matrix to classify the potential for a tsunami show an accuracy value of 99.96%.



*Corresponding Author:*

Eni Irfiani,
Faculty of Engineering and Informatics,
Universitas Bina Sarana Informatika,
Jl. Kramat Raya No. 98, Senen District, Central Jakarta 10450.
Email: eni.enf@bsi.ac.id

## 1. INTRODUCTION

Earthquakes are shocks that occur under the surface of the earth, and are caused by the movement of the earth's plates, sometimes earthquakes can also occur due to volcanic eruptions. The consequences of an earthquake can cause a devastating tsunami that takes many lives(BPBD (Regional Disaster Management Agency) City of Aceh, 2018).

Natural disasters that occur in Indonesia require the government to make preparations in dealing with these disasters. The tsunami that occurred in Indonesia is one of the natural disasters that claimed many lives, there have been 3 tsunami incidents that coincided with the largest earthquake that ever hit Indonesia which resulted in many casualties, the first was the Flores Tsunami in 1992 at 13.29 WITA, an earthquake measuring 6.8 on the Richter scale rocked the Flores Sea, with a depth of the epicenter 35 Km below sea level to the southwest of the city of Maumere, this 30 meter high tsunami hit for 15 minutes, as a result of this incident killing more than 3000 people, and many other unexpected damages, the second was the Aceh earthquake and tsunami in 2004 at 07.58 WIB,(BNPB, 2022)

In several tsunami events that hit Indonesia, there were several tsunamis that were not detected, this is because most potential tsunamis that occur are usually caused by tectonic earthquakes, while there are tsunamis that occur due to volcanic earthquakes. An example that can be taken is the tsunami in Banten which was caused by the Anak volcanic activity. Mount Krakatau, this incident resulted in 176 people died, 509 people were injured, and 30 people were missing(Banten, 2018).

Earthquake disasters cannot be predicted, but the impact of disasters can be estimated based on previous events. One method that can be used to extract information is the C4.5 algorithm. The output of this method can predict the impact of the earthquake. This prediction is expected to reduce the number of fatalities due to the earthquake(Buulolo et al., 2017).

The post-disaster rehabilitation program database is an asset for data mining research where data patterns can be grouped or classified. This pattern is useful for classifying the occupant status of houses for victims of Mount Merapi(Wijaya et al., 2020). Through data mining techniques using the Naive Bayes Classifier method, earthquake data that occurs in Indonesia can be predicted for the probability that they will occur and the possible effects that can be caused by earthquakes can be identified.(Utomo & Purba, 2019).

Damage caused by flooding in Indonesian territory can disrupt community activities. The pattern of flood disasters that occur in Indonesia occurs repeatedly and requires handling from the Government. In order to help reduce the impact of flooding, an intelligent decision pattern is made from historical data patterns so that natural disasters can predict future rainfall and flooding.(Faiza et al., 2022).

The magnitude of the impact generated by the tsunami disaster requires efforts to reduce the number of victims and damage caused by the tsunami. Data Mining is a research technique that can help researchers make predictions using a decision tree model that can be applied to classify the causes of potential tsunamis based on earthquakes in Indonesia.

## 2. RESEARCH METHOD

Data mining research using the C4.5 Algorithm method was then tested using the decision tree method regarding the Earthquake natural disaster that triggered the Tsunami that hit Indonesia. Data that is focused on the time, location, depth, and magnitude that causes an earthquake and has the potential to cause a tsunami. Earthquake data sourced from BMKG with a timeframe from 2018-2022 with a total of 50,116 data(Repository, 2022).

a. Data Mining
   It is an activity of searching for available data to create a model, then the model is used to identify other data patterns that are not available in the stored database(Jamhur, 2016).
b. Decision Tree
   Is a commonly used classification algorithm with a simple structure and easy interpretation(Novianti et al., 2016).
c. Tsunamis
   Tsunamis are sea waves that are generated vertically from the surface of the sea. These sea waves are caused by underwater earthquakes, volcanoes that erupt under the sea, underwater landslides, tsunami waves can move in all directions.(Simamora et al., 2017).
d. C4.5 Algorithm

Is an algorithm used to build a decision tree. The C4.5 algorithm is a development of the ID3 algorithm.

The method used to classify potential tsunamis in Indonesia based on earthquakes uses the C4.5 Algorithm with the data mining method(Irawan et al., 2020).

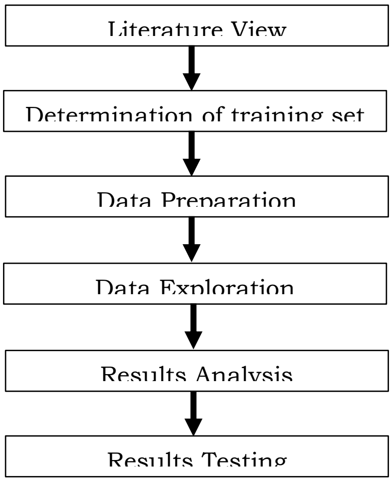


Figure 1. Research Stages

a. Data Training Set, in order to solve the problems in this study, the dataset was taken from the Meteorology, Climatology and Geophysics Agency's (BMKG) earthquake repository from 2018 to 2022, the authors performed good data processing and then proceeded to the data preparation stage. The author gets a total of 50,116 data records(Repository, 2022).
b. Data Preparation, in the next stage, the writer performs several data preparation steps. This step consists of 3 main parts, namely data cleaning, data integration, and data reduction.
c. Data Validation, this stage is carried out by the author to ensure that the training dataset is no longer a missing value and the existing data can be processed towards the next stage.
d. Data Transformation, at this stage the author transforms data or creates a categorical data that aims to ensure the accuracy of existing data. After the data transformation process, a more accurate dataset is obtained, then proceed to the next stage(Rismayanti, 2016).
e. Data Exploration, is an important stage in the classification process, at this stage the dataset that the author has obtained is then processed appropriately using the C4.5 algorithm. In general, the C4.5 algorithm for building a decision tree is as follows:
   a) choose attribute to be root
   b) create a branch for each value
   c) split cases into branches
   d) repeating the process on each branch until all cases on the branch have the same class. To choose an attribute as the root, it is based on the highest gain value of the existing attributes. To calculate the gain, the following formula is used(Mardy, 2017):

   *Gains S, A*= Entropy S$\frac{n}{i=1}$* Entropy Si$\frac{|Si|}{|S|}$

   Information :
   S = set of cases
   A = attribute

n = number of partitions attribute A
|Si| = number of cases on the i-th partition
|S| = number of cases in S

f. Results Analysis
Based on the image from the decision tree then an analysis is carried out

g. Results Testing
he training set data that has been processed will then be tested for validation.

## 3. RESULTS AND DISCUSSIONS

### 3.1 Data Training Sets

Following are some of the training data sets used by the author in the research.

Table 1. Indonesia Earthquake Dataset

| Time | Latitudes | Longitude | depth | Magnitude | Remarks | Potency |
|---|---|---|---|---|---|---|
| Night | 0.33 | 123.73 | 112 | 5,1 | Minahasa Peninsula - Sulawesi | No Potential |
| Morning | 0.54 | 128,1 | 17 | 7,2 | Halmahera - Indonesia | Tsunamis |
| Morning | 6,46 | 101.7 | 10 | 4,9 | Southwest of Sumatra - Indonesia | No Potential |
| Morning | 4,41 | 102,14 | 43 | 4,8 | Southern Sumatra - Indonesia | No Potential |
| Night | 3,42 | 128.45 | 10 | 6,6 | Spooky - Indonesia | Landslide B Sea |
| Morning | 4,41 | 102,14 | 43 | 4,8 | Southern Sumatra - Indonesia | No Potential |
| Afternoon | 8,49 | 121.3 | 181 | 4,5 | Flores Region - Indonesia | No Potential |
| Afternoon | 2.34 | 126.75 | 56 | 6,6 | Northern Molucca Sea | Landslide B Sea |
| Night | 1.89 | 122.53 | 10 | 6,9 | Sulawesi - Indonesia | Landslide B Sea |
| Afternoon | 7.96 | 116.84 | 288 | 4,5 | Bali Sea | No Potential |
| Night | 6,11 | 130.34 | 157 | 4,3 | Banda Sea | No Potential |
| Morning | 0.28 | 125.35 | 10 | 4 | Southern Molucca Sea | No Potential |
| Afternoon | 7,66 | 119,28 | 19 | 3,5 | Flores Sea | No Potential |
| Night | 3,24 | 128.84 | 10 | 3,5 | Spooky - Indonesia | No Potential |
| Afternoon | 0.2 | 125,41 | 10 | 3,3 | Southern Molucca Sea | No Potential |
| Afternoon | 1.01 | 132.52 | 10 | 3,3 | Irian Jaya Region - Indonesia | No Potential |
| Morning | 0.22 | 119.85 | 10 | 7,5 | Minahasa Peninsula - Sulawesi | Tsunamis |
| morning | 0.22 | 119.85 | 10 | 7,5 | Minahasa Peninsula - Sulawesi | Tsunamis |
| afternoon | 1.66 | 126,37 | 62 | 7,1 | Northern Molucca Sea | Tsunamis |
| morning | 6,9 | 106,83 | 10 | 3,3 | Java-Indonesia | No Potential |
| afternoon | 9,82 | 124.25 | 10 | 3,1 | Timor Region | No Potential |
| afternoon | 9,48 | 118,11 | 22 | 3,1 | Sumbawa Region - Indonesia | No Potential |
| night | 7.59 | 122,23 | 10 | 7,3 | Flores Sea | Tsunamis |
| night | 2,31 | 97.06 | 59 | 2,9 | Northern Sumatra - Indonesia | No Potential |
| morning | 0.23 | 122.94 | 12 | 2,7 | Minahasa Peninsula - Sulawesi | No Potential |

The dataset table above consists of 6 attributes, namely: Time (morning, afternoon, evening, night), Latitude, Longitude, Depth, Earthquake Effect, Remark. The resulting targets or labels include tsunami potential, underwater landslides, and no tsunami potential

### 3.2 Data Preparation

The author in carrying out this stage, the author uses the RapidMinner application. The stages consist of:

a. Data Cleaning to remove empty values.

b. Data Integration is done to unify different storages into 1 data.
c. Data Reduction, is done when many attributes are used but not all attributes are conditional on the determining attribute(Iriadi & Nuraeni, 2016).

3.3 Data Validation

From this stage it will produce a training data set that is in normal condition and no longer has data with missing values.

3.4 Data Transformation

From the existing dataset, a categorical form for earthquake effect nodes and time division is formed, as shown below:

Tabel 2. Earthquake Effect Categories

| Richter scale | Earthquake Effect |
|---|---|
| <2.0 | Not felt, small earthquake |
| 2.0 - 2.9 | Not felt, recorded tool |
| 3.0 - 3.9 | Often felt, but rarely damage |
| 4.0 - 4.9 | The vibration can be known, the damage is not significant |
| 5.0 - 5.9 | Causes damage to a small area |
| 6.0 - 6.9 | Can damage areas up to a distance of about 160 Km |
| 7.0 - 7.9 | Causes damage over a wide area |

Tabel 3. Earthquake Effect Categories

| O'clock | Time |
|---|---|
| 05:00 - 10:00 | Morning |
| 11:00 - 15:00 | Afternoon |
| 15:00 - 18:00 | Afternoon |
| 18:00 - 05:00 | Evening |

3.5 Data Exploration

Starting from data calculations to analyzing results. In this study, the data was processed and then calculated using the RapidMiner application. To see the classification results, the C4.5 algorithm method is used and produces the results of the Decision Tree model. The following data is in the form of images and results from dataset testing using the RapidMiner application, then the method used is the C4.5 Algorithm or Decission Tree.

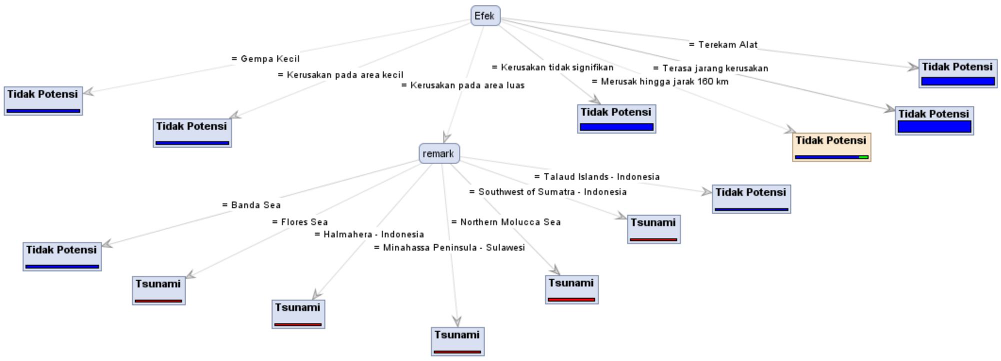


Figure 2. Tsunami Potential Prediction Decision Tree Results

3.6 Results Analysis

From the Decision Tree obtained in the form of a pattern which is the highest attribute is the effect of the earthquake. If the effect of the earthquake is small, the damage is in a small area, the damage is insignificant, the damage is rarely felt, and it is recorded by tools, then there is no potential for a tsunami to occur. If the damaging effect extends to a distance of 160 km there is a potential for an underwater landslide. Then another attribute which is a causative factor which becomes the node of the root effect is the remark (location of the earthquake) in the Banda Sea area (Banda Sea), and the Talaud Islands have no potential for a tsunami, while the Flores sea area, Halmahera, Minahasa Peninsula, North Maluku Sea, west Sumatra, Indonesia has the potential for a tsunami.

## 3.7 Results Testing

Testing was carried out using the 10-fold Cross validation method. Then the results of data processing will be analyzed using the Algortima C4.5 method, along with the results of the accuracy:

Table 4. Accuracy Testing Results

accuracy: 99.96% +/- 0.01% (mikro: 99.96%)

| | true Tidak Potensi | true Longsor Bawah Laut | true Tsunami | class precision |
|---|---|---|---|---|
| pred. Tidak Potensi | 50084 | 18 | 2 | 99.96% |
| pred. Longsor Bawah Laut | 0 | 0 | 0 | 0.00% |
| pred. Tsunami | 2 | 0 | 10 | 83.33% |
| class recall | 100.00% | 0.00% | 83.33% | |

From table 4 it can be seen that the test results obtained an accuracy value of 99.96%. Based on the results of the analysis above, a prototype for predicting the potential for a tsunami in Indonesia was designed:

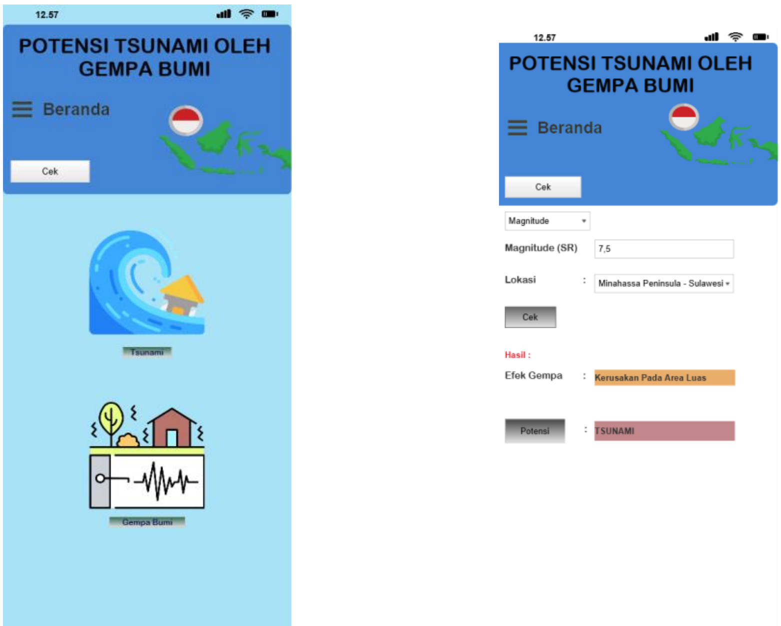


Figure 3. Prototype Design for Prediction of Tsunami Occurrence Based on Earthquakes

## 4 CONCLUSION

From the research that has been made, it can be concluded that from the training data set the attributes that are the highest attributes on the decision tree are the earthquake effect and remark (the location of the earthquake). This can be used by the BMKG to predict the potential for a tsunami, based on the C4.5 algorithm which generated in this study, the possibility of a tsunami occurring in the Indonesian seas based on measurements of the magnitude (Ritcher Scale) of the earthquake has a damaging effect, including: if the Magnitude < 5.9 SR, then there is no potential for a tsunami, if the Magnitude is 6.0 – 6.9 SR, then there is the potential for landslides under the sea causes a secondary tsunami, if magnitude > 7.0 SR, seen from the remark (location of the earthquake incident) in several areas such as the Banda Sea (Banda Sea), and the Talaud Islands have no potential for a tsunami,while the marine areas of Flores, Halmahera, Minahasa Peninsula, North Maluku Sea, southwest Sumatra Indonesia have the potential for a tsunami while based on the validation test using the confusion matrix, the accuracy value of the test using the C4.5 algorithm is 99.96%.

REFERENCES

Banten, P. (2018). The Provincial Government of Banten Announces the Number of Victims who Died Due to the Tsunami Reaches 176 and 30 Disappeared. Banten Government Bureau. https://biroGovernment.bantenprov.go.id/post/Pemprov-Banten-Umumkan-Jumlah-Korban-Meninggal-Akibat-Tsunami-Capai-176-dan-30-Hilang

BNPB, SBNPB (2022). 10 Biggest Natural Disasters in Indonesia, Have Killed Most of the Earth's Population. BPBD Bogor Regency. https://bpbd.bogorkab.go.id/10-bencana-alam-terbesar-di-indonesia-pernah-tewaskan-sebagian-large-penduduk-bumi/

BPBD (Regional Disaster Management Agency) Aceh City. (2018). Understanding Earthquakes, Types, Causes, Effects, and How to Deal with Earthquakes. https://bpbd.bandaacehkota.go.id/2018/08/05/pengertian-gempa-bumi-tipe-Jadibab-akibat-dan-cara-menghadapi-gempa-bumi/

Buulolo, E., Silalahi, N., Fadlina, & Rahim, R. (2017). C4.5 Algorithm To Predict the Impact of the Earthquake. International Journal of Engineering Research & Technology (IJERT), 6(2), 10–15.

Faiza, IM, Gunawan, G., & Andriani, W. (2022). Systematic Literature Review: Application of Machine Learning Methods for Flood Disaster Detection. Polgan Minfo Journal, 11(2), 59–63. https://doi.org/10.33395/jmp.v11i2.11657

Irawan, L., Hermawan Hasibuan, L., & Fauzi. (2020). Predicted Analysis of Earthquake Damage Effects from Magnitude (Richter Scale) Using the ID3 Algorithm Using the Orange Data Mining Application. Information Technology, 14(2), 189. https://doi.org/10.47111/JTI

Iriadi, N., & Nuraeni, N. (2016). Study of the Application of the C4.5 Algorithm Data Mining Classification Method for Predicting Creditworthiness at Bank Mayapada Jakarta. Computer Engineering AMIK BSI, II(1), 132.

Jamhur, AI (2016). Application of Data Mining To Analyze Quantity. 23(2), 12–20.

Mardi, Y. (2017). Data Mining : Classification Using Algorithm C4.5. Informatics Edik, 2(2), 213–219. https://doi.org/10.22202/ei.2016.v2i2.1465

Novianti, B., Rismawan, T., & Bahri, S. (2016). Data Mining Implementation With C4 Algorithm . 5 For Student Majors (Case Study: Sma Negeri 1 Pontianak). 04(3).

Reposity, BE (2022). Earthquakes In Indonesia. Kaggle. https://www.kaggle.com/datasets/kekavigi/earthquakes-in-indonesia

Rismayanti. (2016). Implementation of the C4.5 Algorithm to Determine Scholarship Recipients at Stt Harapan Medan. Infotama Media Journal, 12(2), 116–120.

Simamora, DAS, Furqon, MT, & Priyambadha, B. (2017). Clustering Data on Tsunami Events Caused by Earthquakes Using the K-Medoids Algorithm. Journal of Information Technology Development and Computer Science, 1(8), 635–640.

Utomo, DP, & Purba, B. (2019). Application of Datamining on Earthquake Data on Tsunami Potential in Indonesia. September, 846–853.

Wijaya, N., Endah, M., Feliati, M., Study, P., Program, I., & Correspondence, P. (2020). Application of Decision Tree Algorithm C.45 for Classification of Occupied Status Data of Post Merapi Eruption Rehabilitation Houses. 424–430.