Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

# C4.5 Algorithm to Predict the Impact of the Earthquake

Efori Buulolo[1]
Departement of Computer Engineering
STMIK Budi Darma
Medan, Indonesia
Jl. Sisingamangaraja XII No. 338, Siti Rejo I, Medan Kota,
Kota Medan, Sumatera Utara, 20216

Fadlina[3]
Departement of Informatics Management
AMIK STIEKOM SUMUT
Medan, Indonesia
Jl. Abdul Haris Nasution No.19, Kwala berkala,
Kota Medan, Sumatera Utara, 20142

Natalia Silalahi[2]
Departement of Informatics Management
AMIK STIEKOM SUMUT
Medan, Indonesia
Jl. Abdul Haris Nasution No.19, Kwala berkala,
Kota Medan, Sumatera Utara, 20142

Robbi Rahim[4]
Departement of Computer Engineering
Medan Institute of Technology
Medan, Indonesia
Jl. Gedung Arca No.52 Kota Medan, Sumatera Utara,

***Abstract*: One of the impacts of the quake was heavily damaged, the even tsunami killed at no less. One cause many deaths is because many can not predict the impact of earthquakes. Data earthquakes that occurred earlier can be used to predict the incidence of the quake will probably happen someday. One algorithm that can be used to predict is the algorithm C4.5. The results of the algorithm C4.5 decision tree form, decision trees characteristic or condition of the earthquake and the decision, where the decision is a fruit of the earthquake that occurred modeling**

***Keywords—Earthquake; Impact; The Algorithm C4.5***

## I. INTRODUCTION

Earthquakes often cause massive damage, and human casualties are not small, one reason is that many people can not predict the incidence of the quake which occurred mainly in the earthquake-ravaged region.

The earthquake can not predict when it would happen, but the expected impact of the quake based on seismic data that never happened before[1][2]. One of the methods used to dig or search for information on old data is data mining algorithm C4.5. The output of the algorithm C4.5 in predicting the impact of the quake is divided into three parts[3][4]. Namely, there is no impact / minor damage, severe damage, and the damage and tsunami. With predictions of the implications of the earthquake is expected to be minimized as a result of the quake victims.

## II. THEORY

### A. *Earthquake*

An earthquake is a vibration or shock caused by the release of energy from the earth suddenly and creates seismic waves. Usually, earthquakes caused by the movement of the earth's crust or plates.

Several theories have been making the quake is the collapse of caverns below the surface of the Earth, meteor impact on Earth's surface, due to the volcanic eruption magma activity that occurred before the volcanoes and tectonic activity. Damage caused by earthquakes is death and disability living beings, and the environmental damage and the collapse of the construction of buildings and tsunami waves[5].

### B. *Algorithms C4.5*

The c4.5 algorithm is one of the data mining algorithms that included in the classification groups. C4.5 algorithms are used to form a decision tree. The resulting decision tree is the result of the algorithm C4.5 and can represent and model the results of the exploration of significant data, so the knowledge or information from these data more easily identified [6][7].

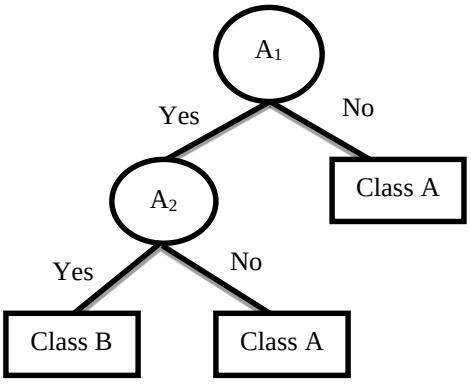


Fig 1. Decision tree example C.45

C4.5 algorithm formula in the form of a decision tree as follows:

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

With:

S: Set Case
A: Attributes
n: number of partitions attribute A
|Si| : Number of cases in the partition to-i


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

|S| : Number of cases in S

To find the value of Entropy is

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

With:

S: Set Case

A: Features

n: number of partitions S

pi: a proportion of Si to S

The steps of the algorithm C4.5 is

1. Calculate the value Entropy (S) and Gain (S, A) to seek early roots. Old sources taken from one of the attributes table and the value of Gain (S, A) is the highest.
2. Create a branch for each value
3. For cases in branch
4. Repeat the process for each branch, until all the cases to the branches have the same class[8]

## III. ANALYSIS AND DISCUSSION

To predict the impact of earthquakes with C4.5 algorithm then takes the old data of the earthquake never happened before. Below are the seismic data that never happened[9]

TABLE I. EARTHQUAKE DATA

| No | Region earthquake | The epicenter | Distance from the beach (km) | Depth (km) | Scale | Duration (second) | Effect |
|---|---|---|---|---|---|---|---|
| 1 | Deli Serdang Medan I | Land | 0 | 10 | 3,9 | 6 | No effect |
| 2 | Deli Serdang Medan II | Land | 0 | 10 | 5,6 | 15 | No effect |
| 3 | Aceh Pidie | Land | 0 | 15 | 6,5 | 59 | Broken |
| 4 | Nias | Sea | 96 | 30 | 8,2 | 60 | Broken and Tsunami |
| 5 | Aceh | Sea | 160 | 30 | 9,1 | 600 | Broken and Tsunami |
| 6 | Padang | Sea | 50 | 87 | 7,6 | 60 | Broken |
| 7 | Mentawai | Sea | 682 | 10 | 7,8 | 65 | Broken and Tsunami |
| 8 | Yogyakarta | Land | 0 | 17,1 | 5,9 | 57 | Broken |
| 9 | Sendai, Jepang | Sea | 130 | 24,4 | 9 | 300 | Broken and Tsunami |
| 10 | Illapel, Chile | Sea | 46 | 25 | 8,3 | 180 | Broken and Tsunami |
| 11 | Nepal | Land | 0 | 15 | 7,8 | 25 | Broken |
| 12 | Afghanistan | Land | 0 | 196 | 7,5 | 30 | Broken |
| 13 | West Southeast Maluku | Sea | 179 | 184 | 5 | 9 | No effect |
| 14 | Morotai | Sea | 122 | 10 | 5 | 8 | No effect |
| 15 | Karo | Land | 0 | 10 | 2,8 | 4 | No effect |

Attributes distance from shore, depth, scale and duration molded into the form of the categories of data, based on the value of each attribute.

TABLE II. CATEGORY DISTANCE FROM THE BEACH

| Distance from the beach *(km)* | Categories |
|---|---|
| 0 | No |
| <= 100 | Far |
| > 100 | Very far |

TABLE III. CATEGORY DEPTH

| **Depth*(km)*** | **Categories** |
|---|---|
| <= 10 | Deep |
| > 10 | Deeper |

TABLE IV. CATEGORY SCALE

| **Scale** | **Categories** |
|---|---|
| <= 5 | Low |
| 5,1 – 7 | Medium |
| >7,1 | High |

TABLE V. CATEGORY DURATION

| **Duration** | **Categories** |
|---|---|
| <=20 second | Short |
| > 20 second | long |



(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

TABLE VI. EARTHQUAKE DATA THAT HAS CATEGORIZE

| *No* | *Region earthquake* | *The epicenter* | Distance from the beach *(km)* | *Depth (km)* | *Scale* | *Duration (second)* | *Effect* |
|---|---|---|---|---|---|---|---|
| 1 | Deli Serdang Medan I | Land | No | Deep | Low | Short | No effect |
| 2 | Deli Serdang Medan II | Land | No | Deep | Medium | Short | No effect |
| 3 | Aceh Pidie | Land | No | Deepen | Medium | Long | Broken |
| 4 | Nias | Sea | Far | Deeper | High | Long | Broken and Tsunami |
| 5 | Aceh | Sea | Very far | Deeper | High | Long | Broken and Tsunami |
| 6 | Padang | Sea | Far | Deeper | high | Long | Broken |
| 7 | Mentawai | Sea | Very far | Deep | High | Long | Broken and Tsunami |
| 8 | Yogyakarta | Land | No | Deeper | Medium | Long | Broken |
| 9 | Sendai, Jepang | Sea | Very far | Deeper | High | Long | Broken and Tsunami |
| 10 | Illapel, Chile | Sea | Length | Deeper | High | Long | Broken and Tsunami |
| 11 | Nepal | Land | No | Deeper | High | Long | Broken |
| 12 | Afghanistan | Land | No | Deeper | High | Long | Broken |
| 13 | Maluku Tenggara Barat | Sea | Very far | Deeper | Low | Short | No effect |
| 14 | Morotai | Sea | Very far | Deep | Low | Short | No effect |
| 15 | Karo | Land | No | Deep | Low | Short | No effect |

The next step is to calculate the number of cases($S$), the number of declared cases of non-effect($S_1$), the number of cases for decision broken($S_2$) and the number of cases reported broken and tsunami($S_3$). After that calculating the gain for each attribute. The results show in the following table.

TABLE VII. CALCULATION NODES 1

| Node | | | S | $S_1$ | $S_2$ | $S_3$ | Entropy | Gain |
|---|---|---|---|---|---|---|---|---|
| 1 | Total | | 15 | 5 | 5 | 5 | 1,584962501 | |
| | The epicenter | | | | | | | 0,432498736 |
| | | Land | 7 | 3 | 4 | 0 | 0,985228136 | |
| | | Sea | 8 | 2 | 1 | 5 | 1,298794941 | |
| | Distance from the beach | | | | | | | 0,617880006 |
| | | No | 7 | 3 | 4 | 0 | 0,985228136 | |
| | | Far | 3 | 0 | 1 | 2 | 0,918295834 | |
| | | Very far | 5 | 2 | 0 | 3 | 0,970950594 | |
| | Depth | | | | | | | 0,2490225 |
| | | Deep | 6 | 4 | 1 | 1 | 1,251629167 | |
| | | Deepen | 9 | 1 | 4 | 4 | 1,392147224 | |
| | Scale | | | | | | | 0,892271866 |
| | | Low | 4 | 4 | 0 | 0 | 0 | |
| | | Medium | 3 | 1 | 2 | 0 | 0,918295834 | |
| | | High | 8 | 0 | 3 | 5 | 0,954434003 | |
| | Duration | | | | | | | 0,880467701 |
| | | Short | 5 | 5 | 0 | 0 | 0 | |
| | | Long | 10 | 0 | 5 | 5 | 1,0567422 | |


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

From the Table VII, the calculation could see that the highest attribute is a scale that is equal to 0.892271866. Thus the scale can be the root node. There is three attributes value, low, medium and high. Due to Low Entropy value of 0 means, the case has classified into (S1) indicates the decision to no effect. While Medium and High does not have decision-making needs to be calculated again.

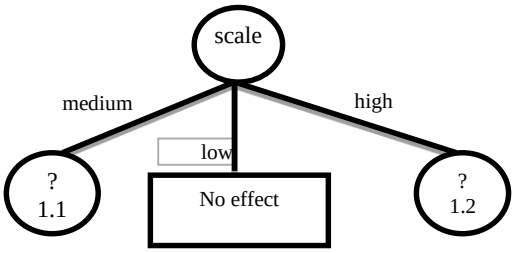


Fig 2. Decision tree calculation results

The next step is to calculate the 1.1 branch nodes of medium and branch nodes of high 2.1

TABLE VIII. CALCULATION NODES 1.1

| Node | | | S | S1 | S2 | S3 | Entropy | Gain |
|---|---|---|---|---|---|---|---|---|
| 1.1 | Scale-medium | | 3 | 1 | 2 | 0 | 0,918295834 | |
| | The epicenter | | | | | | | 0 |
| | | Land | 3 | 1 | 2 | 0 | 0,918295834 | |
| | | Sea | 0 | 0 | 0 | 0 | 0 | |
| | Distance from the beach | | | | | | | |
| | | No | 3 | 1 | 2 | 0 | 0,918295834 | 0 |
| | | Far | 0 | 0 | 0 | 0 | 0 | |
| | | Very far | 0 | 0 | 0 | 0 | 0 | |
| | Depth | | | | | | | 0,251629167 |
| | | Deep | 2 | 1 | 1 | 0 | 1 | |
| | | Deeper | 1 | 0 | 1 | 0 | 0 | |
| | Duration | | | | | | | 0,918295834 |
| | | Short | 1 | 1 | 0 | 0 | 0 | |
| | | Long | 2 | 0 | 2 | 0 | 0 | |

From Table VIII is the highest gain value with the value 0.918295834 duration, the duration becomes a branch node of the medium. The duration has two branches, namely short and long, the two branches already have a decision for entropy value of 0, as shown below

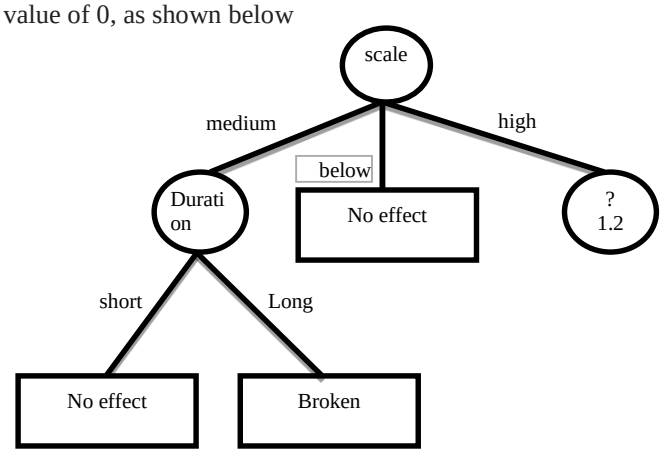


Fig 3. Decision tree node calculation in 1.1

Duration branch already has branched decision means the process stops. The next step is to form a branch node of 2.1 out of high.


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

TABLE IX. CALCULATION NODES 1.2

| Node | | | S | S1 | S2 | S3 | Entropy | Gain |
|---|---|---|---|---|---|---|---|---|
| 1.2 | Scale-high | | 8 | 0 | 3 | 5 | 0,954434003 | |
| | The epicenter | | | | | | | 0,466917187 |
| | | Land | 2 | 0 | 2 | 0 | 0 | |
| | | Sea | 6 | 0 | 1 | 5 | 0,650022422 | |
| | Distance from the beach | | | | | | | 0,610073065 |
| | | No | 2 | 0 | 2 | 0 | 0 | |
| | | Far | 3 | 0 | 1 | 2 | 0,918295834 | |
| | | Very far | 3 | 0 | 0 | 3 | 0 | |
| | Depth | | | | | | | 0,092359384 |
| | | Deep | 1 | 0 | 0 | 1 | 0 | |
| | | Deepen | 7 | 0 | 3 | 4 | 0,985228136 | |

From Table IX, the highest value gain distance from the beach is 0.610073065, the distance from the coast to the high scale branch node. Distance from the beach is owned by the three branches namely No. and very much with entropy values 0, during the length because the decision did not have entropy value is not 0, then continued the following process.

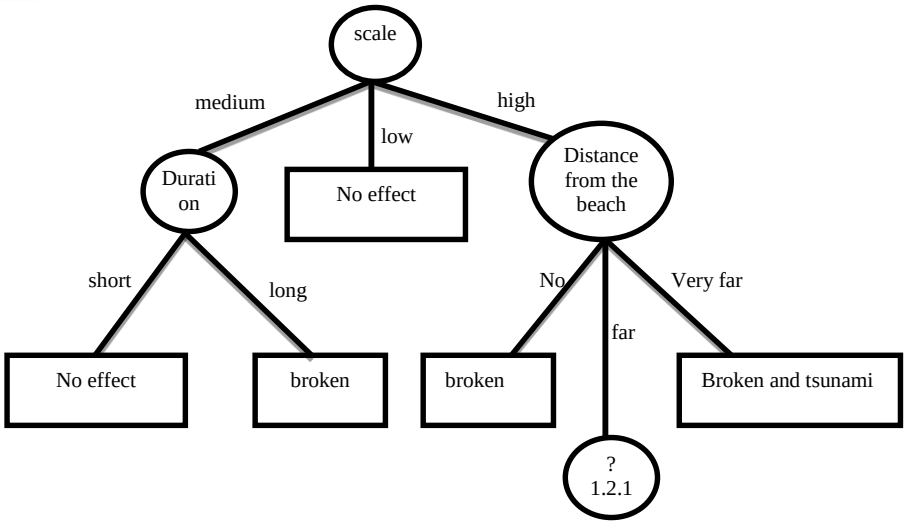


Fig 4. Decision tree node results in 1.2

To search for a branch node of the from calculation table X, like the following.

TABLE X. CALCULATION NODES 1.2.1

| Node | | | S | S1 | S2 | S3 | Entropy | Gain |
|---|---|---|---|---|---|---|---|---|
| 1.2.1 | Scale-high-distance from the beach far | | 3 | 0 | 1 | 2 | 0,918295834 | |
| | The epicenter | | | | | | | 0 |
| | | Land | 0 | 0 | 0 | 0 | 0 | |
| | | Sea | 3 | 0 | 1 | 2 | 0,918295834 | |
| | Depth | | | | | | | 0 |
| | | Deep | 0 | 0 | 0 | 0 | 0 | |
| | | Deepen | 3 | 0 | 1 | 2 | 0,918295834 | |


(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 02, February-2017

From the calculation table X, The epicenter and depth have the same gain value. The epicenter and depth mean a similar position to be a remote branch node. In this case is more likely to influence the impact of the earthquake is the epicenter

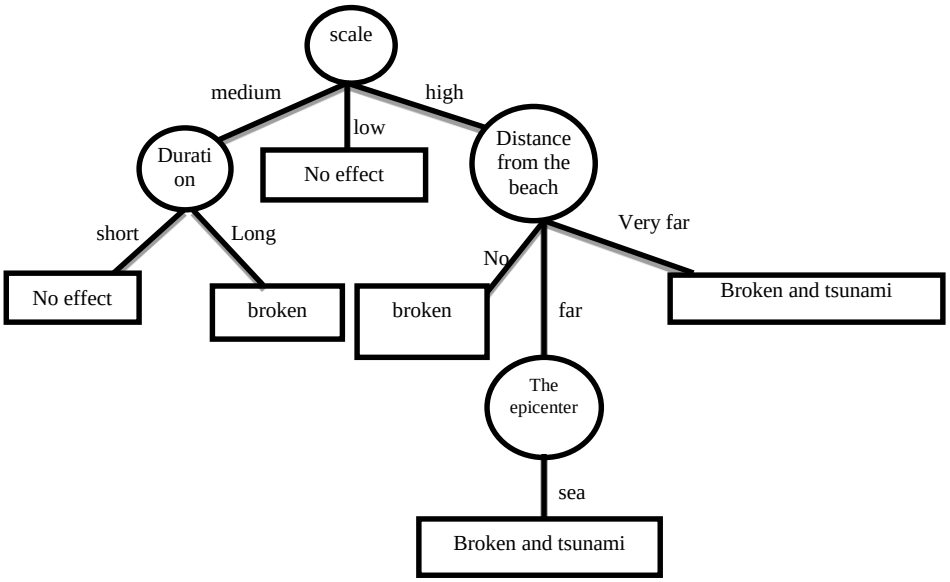


Fig 5. Decision tree node calculation 1.2.1

The decision tree above is the product of the algorithm C4.5. A decision tree can be used to predict the impact of the earthquake based on the characteristics and condition of the quake. The explanation of the decision tree above are as follows:

a. If the scale is low, does not cause any effect
b. If the scale of medium and short duration then no effect
c. If the scale of medium and long duration then cause broken
d. If the scale height and distance from the coast 0 / happened on land, it causes broken
e. If the scale height and distance from the coast very far then cause broken and tsunami
f. If the scale of height and distance from the coast far and The epicenter sea it causes broken and tsunami

## IV. CONCLUSION

Based on the description above can be summarized as follows:

a. The data of earthquakes that has ever happened can provide useful information or knowledge
b. Data mining algorithms can be used to predict C4.5
c. Algorithms C4.5 can predict the impact of the quake based on seismic data that has ever happened which modeled in the form of a decision tree
d.
e. An impact of earthquake affected by some characteristics or conditions of an earthquake that is the scale, duration, distance from the beach and The epicenter.

## V. REFERENCES

[1] Ruxandra and S. Petre, "Data mining in Cloud Computing," *Database Systems Journal*, vol. III, pp. 67-71, 2012.
[2] F. Chen, P. Deng, J. Wan, D. Zhan, V. A. Vasilakos and X. Rong, "Data mining for the internet of things: Literature Review and Challenges," *Hindawi Publishing Corporation Internasional Journal of Distributed Sensor Networks*, vol. 2015, pp. 1-5, 2015.
[3] L. Marlina, Muslim, and A. P. Utama Siahaan, "Data Mining Classification Comparison (Naïve Bayes and C4.5 Algorithms)," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 38, pp. 380-383, 2016.
[4] H. Chauhan and A. Chauhan, "Implementation of decision tree algorithm c4.5," *International Journal of Scientific and Research Publications*, Vols. 1-3, p. III, 2013.
[5] Y. MARUYAMA, M. SAKAYA, and F. YAMAZAKI, "AFFECTS OF EARTHQUAKE EARLY WARNING TO EXPRESSWAY DRIVERS BASED ON DRIVING SIMULATOR EXPERIMENTS," *Journal of Earthquake and Tsunami*, vol. III, pp. 1-11, 2009.
[6] M. Purnamasari and Sulistiyono, "Decision Support System for Classification of Child Intelligence Using C4.5 Algorithm," *International Journal of Advanced Research in Computer Science*, vol. 5, pp. 16-20, 2014.
[7] B. Hssina, A. Merbouha, H. Ezzikouri and M. Erritali, "A comparative study of decision tree ID3 and C4.5," *(IJACSA) International Journal of Advanced Computer Science and Applications*, pp. 13-19.
[8] K. Adhatrao, A. Gaykar, A. Dhawan, R. Jha and V. Honrao , "PREDICTING STUDENTS' PERFORMANCE USING ID3 AND C4.5 CLASSIFICATION ALGORITHMS," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. III, pp. 39-52, 2013.
[9] BMKG, "BADAN METEOROLOGY, KLIMATOLOGI, DAN GEOFISIKA," [Online]. Available: http://www.bmkg.go.id/gempabumi/gempabumi-dirasakan.bmkg. [Accessed 18 1 2017].


(This work is licensed under a Creative Commons Attribution 4.0 International License.)