Decision Rules on Damaged House Data for Natural Disasters Using The If–Then Method Of The Rough Set Theory

Siti Arni Wulandya¹, Wiwit Pura Nurmayanti²,  
¹²Universitas Hamzanwadi  
*sitiarniwulandya@gmail.com

Abstract. Natural disasters remain a real threat for Indonesia. One province in Indonesia which has a high disaster risk is Nusa Tenggara Barat. The impact of disasters can be understood better if the data shape systematically. Data mining techniques that are used in this research is the Rough Set theory. It is an alternative analysis method for analyzing the imperfect data. Disaster that cause damaged houses in Nusa Tenggara Barat, including inconsistent data (rough). Researchers want to know the rules of data patterns (decision rules) using if–then rules from rough set theory. Based on the analysis, gained some rules of data patterns cause heavy damaged houses, minor and minor and heavy damaged houses. 32.75% of the houses were heavily damaged happens when a tornado disaster-stricken area with rain intensity is very low, flat topography and medium density of population. 89.93% of homes with minor damage occurs when an area flooded with low rainfall intensity, undulating topography of the area and the condition of the population is very dense. 38.45% of homes with minor and heavy damage occurred when a severe flood-stricken area with low rainfall intensity, topography of the area is very steep and very dense population conditions.

Keywords: Natural Disaster, Damaged House, Rough Set, Decision Rules, NTB

1. Introduction
Indonesia has geographical, geological, hydrological and demographic conditions that potentially cause disaster. One province in Indonesia which has a high disaster risk is Nusa Tenggara Barat. Nusa Tenggara Barat is a province with two main islands, Lombok and Sumbawa, bordering the Indonesian Ocean in the south and the Flores sea in the north. All districts / cities in the province of Nusa Tenggara Barat are in a high risk class with an index score between 149 - 209. The population of NTB in 2010 amounted to 4,500,212 people [1]. It includes the region of Nusa Tenggara Barat in the volcanic belt (volcanic arc) which extends from the islands of Sumatera - Jawa - Nusa Tenggara - Sulawesi. While the sides are in the form of old volcanic mountains and lowlands which are partly dominated by swamps.

Potential losses that may arise for disaster events include fatalities, missing victims, injured victims, suffering and displaced, property damage, damage to heavy and light houses, damage to land, livestock and other material losses. Based on data from the Badan Penanggulangan Bencana Daerah (BPBD) of Nusa Tenggara Barat province, natural disasters in the province of Nusa Tenggara Barat in 2013 resulted in damage to 26,120 housing units, including 4,453 heavily damaged and 21,667 slightly damaged. One indicator of a prosperous community is the fulfillment of safe or comfortable housing or housing needs as a place for the formation of a prosperous family. Not fulfilling this need implies the condition of a people that still below the poverty line. In 2007 there were 250,000 households in...
the province of Nusa Tenggara Barat still inhabiting uninhabitable homes. This condition is worsened by the increasing number of uninhabitable houses damaged by natural disasters. The risk of damaged house for natural disasters can be reduced if the factors causing damaged house can be detected earlier. By using the method of observation and interpretation of maps of the earth and satellite imagery found that the factors that affected the damage to houses in Pangalengan sub-district of Bandung regency for earthquakes included being close to a fault zone, steep slope, high earthquake intensity, earthquake resistant housing conditions, clustered settlement patterns, high levels of population density and community capacity regarding disaster knowledge are still lacking [2]. In addition, a study also conducted observations on 30 respondents who suffered to damaged house because of the earthquake and found that one factor affecting the level of damaged house was the quality of construction, implementation and strengthening of building walls [3].

The data used in this study is the data of natural disaster events in the province of Nusa Tenggara Barat in 2013. Some parts of the data were found inconsistent (rough) so we cannot draw the conclusions of the pattern of causes of the level of damaged house. Disaster data is a big data, therefore it needs data mining in the analysis process. Data mining techniques that are used in this research is the Rough Set theory [4] [5]. The pattern (rules) of decisions based on a set of inconsistent data (rough) can be found with the help of the rough set theory proposed by Pawlak [6]. A study used a decision rule as a quantitative measure of the rough set theory strengthened that to find hidden data patterns of factors that cause Medirose disease [7]. Besides another study also used decision rules from the rough set theory to classify types and causes of cow disease [8]. Therefore, in this study the author wants to find out what factors influence damaged house then based on these factors find a decision rule (decision rule) that cause the damaged house because of disaster events. In this study, the authors wanted to know whether the factors of the disaster events, the intensity of the rain, the slope, the level of regional risk and population density affect the level of damaged house because of natural disasters.

2. Method

The data used in this study is the data of natural disaster events and risks caused in Nusa Tenggara Barat province in 2013. The decision variable used in this study is the type of damaged house because of natural disasters. Condition variables used in this study include disaster types, rainfall intensity, slope area, level of disaster risk and population density [9] [10].

Data analysis methods used in this study are Descriptive Analysis, Chi Square analysis and Rough Set analysis. The stages of analysis in this research are:

1. Descriptive analysis is used to explain the general description and characteristics [11] [12] of natural disasters in 2013 and their impact in the province of Nusa Tenggara Barat based on various causes.

2. Chi-square test is used to see the dependency between the independent variable and the dependent variable on a nominal or ordinal scale. Chi-square test procedures tabulate one or more variables into categories and calculate chi-square statistics. If it comprises two variables or better known as an independence test that serves to test the relationship of two variables [13] [14] [15]. Chi Square is used to test H0 that the row and column variables are independent of each other. Large samples $\chi^2$ have an approximated Chi Square distribution with (b-1) (k-1) degree of freedom [16].

3. The Roughset method is a new mathematical approach to analyzing vague and uncertain data patterns [17] [18] [19]. In the process of vague and uncertain data analysis, the rough set has several terms explained as follows:
   - Information System
   - Indiscernibility Relation
   - Set Approximation
   - Attribute Dependency
In a data set, the most important thing to look for is dependency between attributes. Intuitively, a $D$ attribute set totally depends on the $C$ attribute set, denoted by $C \rightarrow D$, if all the values of the $D$ attribute are uniquely determined by the values of the $C$ attribute [20].

- Attribute reduction
- Quantitative measure is a measure that can be expressed in a number or size consisting of several measurements, namely: support, strength, certainty, and coverage.
- Decision rules are rules consisting of if then or if f then g which can be represented as $f \rightarrow g$. Section $f$ in rule $f \rightarrow g$ is called antecedent and part $g$ is called conclusion. In a rough set, we can draw decision rules from the reduce attributes that have been obtained.

3. Results and Discussion

3.1. Descriptive Statistics:
Natural disasters that dominated in West Nusa Tenggara Province in 2013 were whirlwind disasters that were 18 events, floods 17 times, tidal waves and earthquakes 5 events, strong winds and landslides 4 events and fires 3 events.

The pattern of frequency of natural disaster events in Figure 1 is high in January, February and then slowly decreases in March to August and the frequency of disaster events will increase again in September, October and so on. Many of the disasters that occur in these dominant months are hydrometeorological disasters such as floods, tornadoes, strong winds, tidal waves and landslides. This hydrometeorological disaster is a disaster that usually occurs every year in Indonesia including Nusa Tenggara Barat. Most regions in Indonesia in September - March experience the rainy season, ranging from mild to severe intensity. This rain intensity is one cause of many hydrometeorological disasters occurring in September to March.

All districts in the province of Nusa Tenggara Barat are in the high disaster risk class with a score between 149 and 209. The highest risk score of 209 is in Bima Regency while the lowest score of 149 is in the city of Mataram.
Natural disasters in 2013 resulted in 28 livestock lost/dead, 34.4 hectares of gardens and 1,157.2 hectares of damaged rice fields, 1.82 kilometers of dykes damaged, 6,757 kilometers of damaged roads, 18 units of bridges damaged, 10 units of schools damaged, 26,120 units the house was damaged and resulted in the death/loss of 9 people and caused 27,720 other people to suffer injuries and displaced. One of the very serious impact of the disaster was damage to houses both in the category of minor damage and severe damage.

3.2. Chi Square Analysis
Further analysis will be performed to see the relationship between condition variables and decision variables. Condition variables that used in this study include disaster variables (X1), rainfall intensity (X2), area slope (X3), disaster risk level (X4) and population density (X5). While the decision variable used is the damaged house variable (Y).

| Interaction                        | df | $\chi^2$  | p-value | Decision     |
|-----------------------------------|----|-----------|---------|--------------|
| rainfall intensity * damaged house | 8  | 474.431   | .000    | Reject $H_0$ |
| area slope * damaged house         | 6  | 760.865   | .000    | Reject $H_0$ |
| population density * damaged house | 6  | 65.792    | .000    | Reject $H_0$ |
| damaged house * disaster           | 12 | 331.605   | .000    | Reject $H_0$ |
| damaged house * disaster risk level| 6  | 3.220     | .781    | Failed to reject $H_0$ |

The initial hypothesis used is that there is no relationship between the condition variable (Xi) and the decision variable (Y) the type of damaged house because of natural disasters, where i = 1,2,3,4,5. By using alpha 0.05 it was decided to reject the initial hypothesis for the p-value of each interaction which was smaller than the alpha that was determined. So it was concluded that there is a relationship between the variable conditions Disaster Type (X1), Rain Intensity (X2), Regional Slope (X3) and Population Density (X5) with the decision variable the level of house damage because of natural disasters.

3.3. Rough Set Analysis
The discussion for analyzing data patterns with rough sets uses the order below:

1) Information System is a table that contains condition attributes in the form of variable types of disaster events, rainfall intensity, regional slope, level of regional disaster risk and population density in the form of categorical data. It also comes with a decision variable of the level of damaged house in the form of categorical data as well as the frequency of the number of damage to houses for each category of damage level. Based on the information system, information was obtained that the most light and heavy damaged homes of 9,836 units occurred if of a flood disaster with low rainfall intensity, the slope of the region was very steep, the level of disaster risk was not risk and the population density was very dense; the most severely damaged houses as many as 131 units occurred in a whirlwind disaster condition with very low rainfall intensity, slope of a flat area, moderate level of disaster risk and very dense population density; In addition, 125 units of lightly damaged houses occurred during the flood disaster conditions with low rainfall intensity, the slope of the area is bumpy, the level of disaster risk is not at risk and the population density is very dense.

2) The approximation set contains a set of decision rule approaches which causes some damage to the house. Lower approximation contains a set of rule approaches that definitely results in a level of house damage, upper approximation contains a set of rule approaches that can be categorized generally as causing a level of damage, whereas boundary approximation contains a set of gray rule approaches, which can be classified as a level of house damage or the level of damage to other houses. A non-empty Boundary Approximation shows rough data. In this study, there is a boundary approximation for all three levels of damage to the house.
3) Excessive data can be moved from the information system table to simplify decision rules using reduced table data. In the process of data reduction, it must maintain data consistency without changing the data information system.

4) Quantitative measure is the calculation of the probability of a decision based on various conditions. This quantitative measure can improve the accuracy of the decision rule got. Quantitative measures calculated include support, strength, certainty and coverage.

5) It bases the decision rule obtained on certainty and coverage calculations.
   a) According to certainty values, it is obtained a rule that fire disasters always cause severely damaged houses; Earthquake disasters always cause the house to be lightly and severely damaged; landslides always cause the house to be slightly damaged and severely damaged; catastrophic floods with very low rainfall intensity always result in houses being slightly damaged to severely damaged; high tide disasters with low and high rainfall always cause heavy damage to houses; and a tornado disaster with moderate rainfall intensity always results in heavily damaged houses.

   b) According to the coverage value, a rule is obtained that 32.75% of houses were severely damaged when an area was hit by a tornado with very low rainfall intensity, flat topography and moderate population density; 89.93% of the houses were slightly damaged when an area was flooded with low rainfall intensity, the topography of the area was bumpy and the condition of the population was very dense; 38.45% of the houses were lightly and severely damaged when an area was flooded with low rainfall intensity, the topography of the area was very steep and the condition of the population was very dense.

4. Conclusion
A tornado disaster caused the most heavily damaged houses with very low rainfall intensity, flat topography and moderate population density. The most minor damage to the house occurred because of flooding with low rainfall intensity, undulating topography and very dense population conditions; meanwhile houses were both slightly damaged and heavily damaged because of floods with low rainfall intensity, the topography of the area was very steep and the population conditions were very dense. The researcher can then add other attributes that influence the level of house damage because of natural disasters, for example, the building (permanent or not) and the presence or absence of government regulations that have been applied to the area.

5. References
[1] S. Indonesia, “Badan pusat statistik,” Diakses dari http://www.bps.go.id pada, vol. 29, 2016.
[2] Y. Malik, “Penentuan Tipologi Kawasan Rawan Gempabumi Untuk Mitigasi Bencana Di Kecamatan Pangalengan Kabupaten Bandung,” J. Geogr. GEA, vol. 10, no. 1, pp. 59–70, 2010.
[3] V. N. Hasanah and I. Sanaky, “Analisis Pengaruh Kualitas Pelaksanaan Tenbok dan Perkuatannya terhadap Tingkat Kerusakan Bangunan Rumah Tinggal Sederhana Tenbokan (BRTST) Akibat Gempa,” 2006.
[4] W. I. D. Mining, “Data mining: Concepts and techniques,” Morgan Kaufmann, 2006.
[5] G. S. Linoff and M. J. A. Berry, Data mining techniques: for marketing, sales, and customer relationship management. John Wiley & Sons, 2011.
[6] Z. Pawlak, “Rough set theory and its applications to data analysis,” Cybern. Syst., vol. 29, no. 7, pp. 661–688, 1998.
[7] R. Soelaiman, W. Anggraeni, and E. Setiawan, “Penerapan Rough Set Quantitative Measure Pada Aplikasi Pendukung keputusan,” J. Fak. Huk. UII, 2008.
[8] N. Listiana, W. Anggraeni, and A. Muhlason, “Implementasi algoritma rough set untuk deteksi dan penanganan dini penyakit sapi,” ITS, Surabaya, 2011.
[9] B. N. P. Bencana, “Badan Nasional Penanggulangan Bencana,” Dipetik April, vol. 20, p. 2017, 2013.
[10] P. R. Indonesia, “Undang-undang republik indonesia nomor 24 tahun 2007 tentang penanggulangan bencana,” Jakarta. Diakses dari https://doi.org/10.1007/s13398-014-0173-7.
2007.

[11] H. T. Lawless, “Descriptive analysis of complex odors: reality, model or illusion?,” *Food Qual. Prefer.*, vol. 10, no. 4–5, pp. 325–332, 1999.

[12] A. Muhson, “Teknik Analisis Kuantitatif,” *Univ. Negeri Yogyakarta. Yogyakarta*, 2006.

[13] M. L. McHugh, “The chi-square test of independence,” *Biochem. medica Biochem. medica*, vol. 23, no. 2, pp. 143–149, 2013.

[14] R. Azen and C. M. Walker, *Categorical data analysis for the behavioral and social sciences.* Routledge, 2011.

[15] A. Agresti and M. Kateri, *Categorical data analysis.* Springer, 2011.

[16] J. Nugraha, *Pengantar Analisis Data Kategorik: Metode dan Aplikasi Menggunakan Program R.* Deepublish, 2014.

[17] Z. Pawlak, “A primer on rough sets: a new approach to drawing conclusions from data,” in *The Dynamics of Judicial Proof*, Springer, 2002, pp. 135–142.

[18] Z. Pawlak and A. Skowron, “Rough sets: some extensions,” *Inf. Sci. (Ny).* vol. 177, no. 1, pp. 28–40, 2007.

[19] Z. Pawlak, “Rough sets and intelligent data analysis,” *Inf. Sci. (Ny).* vol. 147, no. 1–4, pp. 1–12, 2002.

[20] Z. Suraj, “An introduction to rough set theory and its applications,” *ICENCO, Cairo, Egypt*, vol. 3, p. 80, 2004.