ANALYSIS OF RAINY DAYS AND RAINFALL TO LANDSLIDE OCCURRENCE USING LOGISTIC REGRESSION IN PONOROGO EAST JAVA

Dihin Muriyatmoko[1] , Sisca Mayang Phuspa[2]
[1]Department of Informatics, University of Darussalam Gontor, Indonesia
[2]Department of Occupational Safety and Health, University of Darussalam Gontor, Indonesia
Email : dihin@unida.gontor.ac.id, siscamayang@unida.gontor.ac.id

Abstract

Referred to data of Badan Nasional Penanggulangan Bencana (BNPB) and Kementerian Kesehatan Republik Indonesia (Kemenkes RI), almost landslide occurrence in Ponorogo always starts with high-intensity rain. This research aimed to determine simultaneously correlation and partial assessment impact of rainy days every month and monthly rainfall toward landslide occurrence in Ponorogo using logistic regression. The data collection was conducted through Badan Pusat Statistik (BPS) in the book of Ponorogo Regency in Figure on 2012 to 2016. The existing data shows that in sixty months have been twenty-six times landslides occurrence in Ponorogo districts. The data statistically analyzed in simultaneous proves that contribution of rainy days and rainfall to landslide were included adequate correlation (Nagelkerke R Square = 25.4 % and Cox & Snell R Square = 36.9 %) and in partial test proves that rainy days have significant impact (sig. = 0.024) and rainfall does not significant impact (sig. = 0.291) (α = 0.05) to landslide occurrence in Ponorogo regency. The rainy days per month were abled applied to predict for possible landslide elsewhere.

Keywords: rainy days, rainfall, landslide, Ponorogo, logistic regression

1. Introduction

Ponorogo Regency is an area in East Java Province who are in a position 200 km northwest province capital, and 800 km to the capital city of Indonesia.The area of 1.371.78 km² is divided in 21 district that consists of 307 villages. Ponorogo Regency topography varies from lowlands to mountains. Based on existing data, a large district that is 79% Ponorogo situated at an altitude of fewer than 500 m above sea level, 14.4% are between 500 and 700 m above sea level and the remaining 5.9%is at the height of the above 700 m(Badan Perencanaan Pembangunan Daerah Ponorogo, 2013).

Based on the location of topography, climate, and rainfall, Ponorogo regency including areas that are often categorized landslides, especially in the hills and mountains (Yuniarta, Saido, & Purwana, 2015)of which there are in five districts that is Ngrayun, Slahung, Pudhak, Pulung, and Ngebel. According to data from BNPB and Kemenkes RI from

2012 to 2016, Ngrayun district was ranked the highest with nine times the landslide, Slahung with five times, Ngebel four times and the last place is Pudhak and Pulung with twice the landslide occurred. Almost all landslide events always start with rainfall in high intensity or rain for more than a day(Badan Nasional Penanggulangan Bencana, 2018; Kementerian Kesehatan Republik Indonesia, 2018).

Factors of rainy days and rainfall are also the ultimate set to be the cause of the occurrence of the landslide by Ubeku and Okeke (Ubechu & Okeke, 2017), as well as Paimin a Geoscientist who gave a statement that 25% of landslide factors are caused by rainy days for three days. (Paimin, Sukresno, & Pramono, 2009), the Department of Public Works also makes one of the fourteen factors that cause landslides is rainfall (Departemen Pekerjaan Umum, 2007), the book of Disaster Risks Indonesia also makes one of the four factors that cause landslides is rainfall (Amri et al., 2016). This study examines the extent to which the simultaneous correlation of rainy days every month and monthly rainfall on the occurrence of landslides in Ponorogo and the impact of rainy days every month and monthly rainfall on landslides in Ponorogo using logistic regression.

These are some studies related to landslides and logistic regression i.e, the first is the results perform that landslides terrace dramatically from 1946 to 2012 inthe capture area. The nearness and overlapping of human development with landslides terraced. However,the logistic regression results prove that variation in sensitivity to landslides was due to natural causes,with the exclusion of historical deforestation and recentlyestablished road systems. Accordingly, well-recoveredhistoricalwoodland sites might presently be landslide-prone areas(Y. C. Chuang & Shiu, 2018), second in Ambon Indonesia, eight landslide causative factors were respected in the landslide sensitivityevaluation. The causative factors were height, slope angle, slope aspect, closeness to stream network, lithology, the solidity of geological boundaries, closenessto faults, and closenessto the road network. The output sensitivitymaps were reclassified into five categorize ranging from very low to very high sensitivity using Jenks natural breaksmethod. Twenty percent of allmapped landslideswere used as the legalization of the sensitivitymodels. The legalization and the accuracy of each modelwere examined by calculating areas under recipient operating characteristic curves (ROCs),andthe areasnether the curve (AUC) for the success rate curves of FR, LR, and ANN were 0.688, 0.687, and 0.734, severally. The AUC for the prediction rate curve of FR, LR, and ANN were 0.668, 0.667, and 0.717, respectively (Aditian, Kubota, & Shinohara, 2018), the third is twelve landslide causative factors (namely, slope, slope aspect, highness, curvature, profile arch, plan arch, slope length, topographic dampness index, gap to river, gap

to road, gap to fault and yearly maximum 24- and 48-h rainfalls) were used in this landslide sensitivity analysis. These models were applied to the Kaoping River basin in Southwestern Taiwan to rate its show. Landslide inventory maps from 2008 to 2011 were congregated. The results prove that the RBF-SVM model makes better the logistic regression in the study area (Lin, Chang, Huang, & Ho, 2017).

2. The Methods

Logistic regression is a method of statistical analysis to describe the relationship between response variables (dependent variable) which has two or more categories with one or more explanatory variables (independent variable) scale or interval (Hosmer & Lemeshow, 2005). Logistic regression is a nonlinear regression, used to explain the relationship between X and Y which is nonlinear, Y-dislocated abnormality, diversity of non-constant response unexplained by ordinary linear regression model (Agresti, 1996).
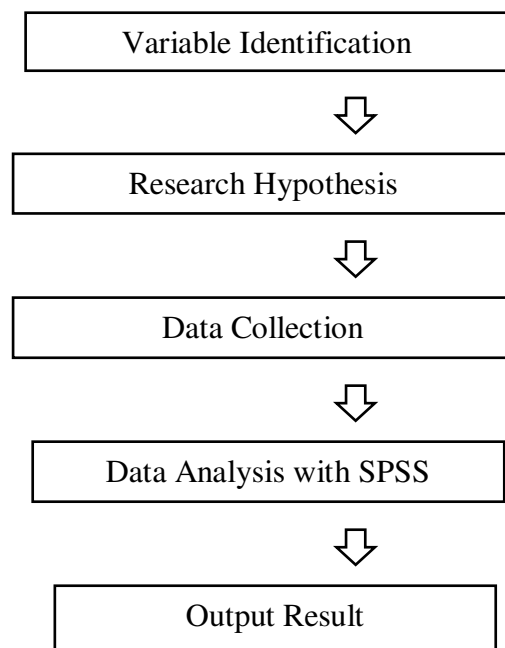

Figure 1. Research method with regression logistic

2.1 Variable Identification

Independent variable is landslide occurrence and the dependent variable is rainy days every month and rainfall every month

2.2 Research Hypothesis

1. Any simulant correlation between rainy days every month and rainfall every month to landslide occurrence

2. Impact between rainy days every month to landslide occurrence

3. Impact between rainfall every month to landslide occurrence

2.3 Data collection

This data was taken on five years from Ponorogo in Figure since 2012 to 2016 by Ponorogo Regional Development Planning Agencyand Central Bureau of Statistics (Badan Perencanaan Pembangunan Daerah Ponorogo, 2013, 2014, Badan Pusat Statistik Kabupaten Ponorogo, 2015a, 2015b, 2016), Ministry of Health of the Republic of Indonesia (Kementerian Kesehatan Republik Indonesia, 2018) and BPBD (Badan Nasional Penanggulangan Bencana, 2018).

Table 1. Data of rainy days every month, rainfall every month and landslide occurred status in 2012-2016 at Ponorogo regency

No.	Years	Month	Rainy Days Every Month	Rainfall Every Month	Landslide Status	No.	Years	Month	Rainy Days Every Month	Rainfall Every Month	Landslide Status
1	2012	1	21	15	No	31	2014	7	2	7	No
2	2012	2	13	16	Yes	32	2014	8	0	8	No
3	2012	3	16	12	No	33	2014	9	0	0	No
4	2012	4	12	17	No	34	2014	10	0	12	No
5	2012	5	7	15	No	35	2014	11	13	16	No
6	2012	6	2	7	No	36	2014	12	18	19	Yes
7	2012	7	1	3	No	37	2015	1	14	8	No
8	2012	8	0	0	No	38	2015	2	16	13	Yes
9	2012	9	1	4	No	39	2015	3	20	12	Yes
10	2012	10	3	11	No	40	2015	4	15	11	Yes
11	2012	11	11	19	No	41	2015	5	4	1	No
12	2012	12	23	17	No	42	2015	6	1	11	No
13	2013	1	23	17	No	43	2015	7	0	0	No
14	2013	2	18	19	No	44	2015	8	0	0	No
15	2013	3	15	14	No	45	2015	9	0	0	No
16	2013	4	13	19	Yes	46	2015	10	0	2	No
17	2013	5	15	12	Yes	47	2015	11	8	4	No
18	2013	6	12	13	Yes	48	2015	12	16	10	No
19	2013	7	8	11	No	49	2016	1	16	17	No
20	2013	8	0	0	No	50	2016	2	20	19	Yes
21	2013	9	0	0	No	51	2016	3	16	21	Yes
22	2013	10	3	17	No	52	2016	4	17	19	Yes

23	2013	11	13	18	No	53	2016	5	12	14	Yes
24	2013	12	19	19	No	54	2016	6	11	14	No
25	2014	1	18	17	Yes	55	2016	7	6	9	No
26	2014	2	14	12	No	56	2016	8	5	9	No
27	2014	3	13	17	No	57	2016	9	13	21	Yes
28	2014	4	11	15	No	58	2016	10	15	17	No
29	2014	5	6	14	No	59	2016	11	23	22	Yes
30	2014	6	4	22	No	60	2016	12	16	12	Yes

2.4 Data Analysis with SPSS

Data Validation Test
⇩
Variable Category Naming
⇩
Properness Model Test
⇩
Significance Test
⇩
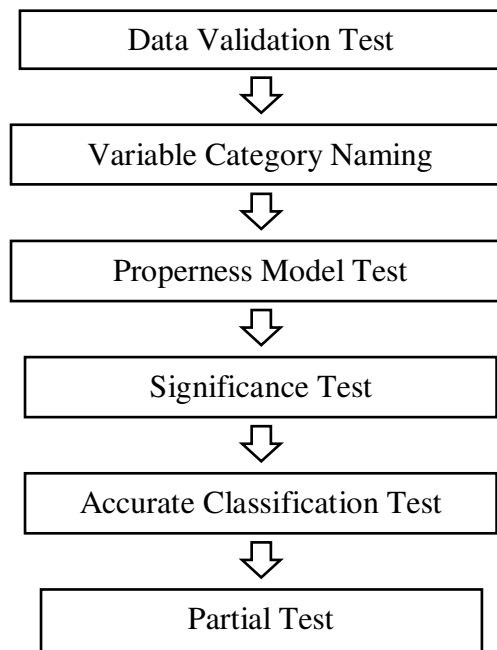Accurate Classification Test
⇩
Partial Test

Figure 2. Data Analysis method with regression logistic using SPSS

Figure 2 describes the logistic regression stage i.e. data validation test is used to know that data is valid or not, variable category naming is used to know the code of the landslide occurs or not, the properness model test is used for sufficiently explained the data (Goodness of fit), the significant test is used to know the correlation value of dependent variables to independent variable, the accuracy classification test is used for measure precision prediction in this study, and the last partial test is used to prove the significance value of dependent variables to independent variable. The statistic of logistic regression was produced the chi-square value that is used to check the correlation of rainy days and rainfall toward landslide occurrence. The accepted criteria of rainy days and rainfall to landslide occurrence can be seen if the chi-square value is lower than chi-square table.

2.5 Output Result

The output details referred to Figure 2 were shown in the result and discussion section.

3. Results and Discussion

Here are the results of calculations using logistic regression using SPSS tools (Reed & Wu, 2013)

3.1 Data Validation Test

Table 2. Case Processing Summary

Case Processing Summary			
Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	60	100.0
	Missing Cases	0	.0
	Total	60	100.0
Unselected Cases		0	.0
Total		60	100.0

Referred to table 2, the sixty data is valid and no missing cases.

3.2 Variable Category Naming

Table 3. Dependent Variable Encoding

Dependent Variable Encoding	
Original Value	Internal Value
No landslide occurred	0
Landslide occurred	1

Referred to table 3. Azero value indicates that is no landslides are occurring, and one value indicates that landslide occurred.

3.4 Properness Model Test

Table 4. Omnibus Tests of Model Coefficients

Omnibus Tests of Model Coefficients				
		Chi-square	df	Sig.
Step 1	Step	17.552	2	.000
	Block	17.552	2	.000
	Model	17.552	2	.000

The correlation between X1 and X2 to Y, Chi-Square technique obtained Chi-Square value 17.552 with Sig value 0.000 <0.05, it means together rainy days (X1) and rainfall (X2) associated with a landslide (Y).

Omnibus Tests:

H0 = variables x1 and x2 do not significantly affect Y

H1 = at least one between x1 and x2 significantly affect to y

Test criteria: H0 is rejected if sig value <0.05, or chi-square value > chi-square table (5.99)

Referred to Table 4. The output can be seen that significant = 0.000, it means less than 0.05 and chi-square value is 17.552, it means chi-square value is higher than chi-square table.

Decision: H0 rejected

Conclusion: the value of chi-square = 17.552 with significant value or p-value = 0.000 means with 95% confidence level, there is at least one free variable (x1, x2) that influence on the dependent variable, so the model can be used for further analysis

Table 5. Hosmer and Lemeshow Test

Hosmer and Lemeshow Test			
Step	Chi-square	df	Sig.
1	6.023	8	.645

The probability of 0.645> 0.05, meaning that the binary regression model is suitable for further analysis since there is no significant difference between the predicted classification and the observed classification.

Hosmer and Lemeshow Test:

H0 = The model has sufficiently explained the data (Goodness of fit)

H1 = Model is not enough to explain data

Test criteria:

H0 accepted If the p-value or significance > 0.05

Referred to Table 5. has a significance value of 0.645 > 0.05, it means that the significance value is greater than 0.05.

Decision: H0 accepted

Conclusion: the model has sufficiently explained the data (goodness of fit).

3.5 Significance Test

Table 6. Model Summary

Model Summary			
Step	-2 Log likelihoods	Cox & Snell R Square	Nagelkerke R Square
1	52.038a	.254	.369

Referred to table 7 indicates that the coefficient determinant of logistic regression that is 0,369 so it can be concluded that the contribution of variable X1 and X2 to Y is equal to 37%

4. Accurate Classification Test (Percentage Correct)

Table 7. Classification

Classification Table[a]	
Observed	Predicted

			Landslide occurred		Percentage Correct
			No	Yes	
Step 1	Landslide occurred	No	39	5	88.6
		Yes	9	7	43.8
	Overall Percentage				76.7

Referred to table 7 indicates thatprecision prediction in this study is 76.7%.

3.6 Partial Test

Table 8. Variable in the Equation

		Variables in the Equation					
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**
Step 1a	**Rainy Days**	.169	.075	5.121	1	.024	1.184
	Rainfall	.096	.091	1.114	1	.291	1.101
	Constant	-4.498	1.440	9.762	1	.002	.011

The partial test shows that only X2 is significant because of the value of Sig 0.024 <0.05, while X2 significant 0.291> 0.05 means that alone X1 has no significant effect on Y. A partial test of rainy days every month:

H0 = 0, the variable of rainy days does not significantly affect the landslide

H1 ≠ 0, the variable of rainy day significantly affects the landslide

Test criteria:

H0 is rejected if significance value < 0.05

Decision: H0 is rejected because significance value = 0.24

Conclusion: rainy days have significantly affected the landslide

A partial test of rainfall every month:

H0 = 0, the variable of rainfall does not significantly affect to landslide occurred

H1 ≠ 0, the variable of rainfall significantly affects the landslide

Test criteria:

H0 is rejected if significance value < 0.05

Decision: H0 is accepted because significance value = 0.291

Conclusion: rainfall does not have significantly affect to landslide occurred

Referred to table 2 to table 8, prove that in Ponorogo regency the factor of rainy days every month have significantly affectedto the landslide occurred and rainfall every month does not have significantly affected to the landslide occurred. So, for the next step we can

discuss that the numberof rainy days every month can be used to predict landslide occurs in Ponorogoand to analysis of landslide can use the others methods besides logistic regression for example artificial neural network (Logar, Turk, Marsden, & Ambrožič, 2017), support vector machine (Y. Chuang & Shiu, 2018), artificial neural networks (ANNs), boosted regression tree (BRT), classification and regression trees (CART), generalized linear model (GLM), generalized additive model (GAM), multivariate adaptive regression splines (MARS), naïve Bayes (NB),quadratic discriminant analysis (QDA), random forest (RF), and support vector machines (SVM) (Pourghasemi & Rahmati, 2018).

4. Conclusion

The data statistically from Ponorogo regency that analyzed using logistic regression method in simultaneous proves that contribution of rainy days and rainfall to landslide were included adequate correlation (Nagelkerke R Square = 25.4 % and Cox & Snell R Square = 36.9 %) and in partial test proves that rainy days have significant impact (sig. = 0.024) and rainfall does not significant impact (sig. = 0.291) (α = 0.05) to landslide occurrencein Ponorogo regency. For the next research, the rainy days per month were abled applied to predict for possible landslide elsewhere and landslide analized can used others algoriltm beside linear regression such as artificial neural network, support vector machine, boosted regression tree, generalized linear regression, etc.

References

Aditian, A., Kubota, T., & Shinohara, Y. (2018). Geomorphology Comparison of GIS-based landslide susceptibility models using frequency ratio , logistic regression , and arti fi cial neural network in a tertiary region of Ambon , Indonesia. *Geomorphology Journal*, *318*, 101–111. https://doi.org/10.1016/j.geomorph.2018.06.006

Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. Wiley. https://doi.org/10.1002/0470114754

Amri, M. R., Yulianti, G., Yunus, R., Wiguna, S., Adi, A. W., Ichwana, A. N., … Septian, R. T. (2016). *Risiko Bencana Indonesia*. Jakarta: Badan Nasional Penanggulangan

Bencana.

Badan Nasional Penanggulangan Bencana. (2018). Data Pantauan Bencana. Retrieved June 21, 2018, from http://geospasial.bnpb.go.id/pantauanbencana/data/index.php

Badan Perencanaan Pembangunan Daerah Ponorogo. (2013). *Pembangunan Ponorogo Dalam Angka 2013*. Ponorogo. Retrieved from https://ponorogokab.bps.go.id/publication/

Badan Perencanaan Pembangunan Daerah Ponorogo. (2014). *Pembangunan Ponorogo Dalam Angka 2014*. Ponorogo. Retrieved from https://ponorogokab.bps.go.id/publication

Badan Pusat Statistik Kabupaten Ponorogo. (2015a). *Ponorogo Dalam angka 2015*. Ponorogo. Retrieved from https://ponorogokab.bps.go.id/publication

Badan Pusat Statistik Kabupaten Ponorogo. (2015b). *Ponorogo Dalam angka 2017*. Ponorogo. Retrieved from https://ponorogokab.bps.go.id/publication

Badan Pusat Statistik Kabupaten Ponorogo. (2016). *Ponorogo Dalam angka 2016*. Ponorogo. Retrieved from https://ponorogokab.bps.go.id/publication

Chuang, Y. C., & Shiu, Y. S. (2018). Relationship between landslides and mountain development—Integrating geospatial statistics and a new long-term database. *Science of the Total Environment Journal*, *622–623*, 1265–1276. https://doi.org/10.1016/j.scitotenv.2017.12.039

Chuang, Y., & Shiu, Y. (2018). Science of the Total Environment Relationship between landslides and mountain development — Integrating geospatial statistics and a new long-term database. *Science of the Total Environment Journal*, *622–623*, 1265–1276. https://doi.org/10.1016/j.scitotenv.2017.12.039

Departemen Pekerjaan Umum. Pedoman Penataan Ruang Kawasan Rawan Bencana Longsor, Pub. L. No. 22 /PRT/M/2007, 148 (2007). Indonesia: Menteri Pekerjaan Umum Republik Indonesia. Retrieved from landspatial.bappenas.go.id/komponen/peraturan/the_file/permen22_2007.pdf%0A

Hosmer, D. W., & Lemeshow, S. (2005). Multiple Logistic Regression. In *Applied Logistic Regression* (pp. 31–46). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471722146.ch2

Kementerian Kesehatan Republik Indonesia. (2018). Pusat Krisis Kesehatan Kementerian Kesehatan Republik Indonesia. Retrieved June 11, 2018, from http://pusatkrisis.kemkes.go.id/

Lin, G., Chang, M., Huang, Y., & Ho, J. (2017). Assessment of susceptibility to rainfall-

induced landslides using improved self-organizing linear output map , support vector machine , and logistic regression. *Engineering Geology Journal*, *224*(May), 62–74. https://doi.org/10.1016/j.enggeo.2017.05.009

Logar, J., Turk, G., Marsden, P., & Ambrožič, T. (2017). Prediction of rainfall induced landslide movements by artificial neural networks. *Journal of Natural Hazards and Earth System Sciences Discussions*, (July), 1–18. https://doi.org/10.5194/nhess-2017-253

Paimin, Sukresno, & Pramono, I. B. (2009). *Teknik Mitigasi Banjir dan Tanah Longsor*. (A. N. Ginting, Ed.). Balikpapan: Tropenbos International Indonesia Programme. Retrieved from www.tropenbos.org

Pourghasemi, H. R., & Rahmati, O. (2018). Prediction of the landslide susceptibility: Which algorithm, which precision? *Catena Journal*, *162*(November), 177–192. https://doi.org/10.1016/j.catena.2017.11.022

Reed, P., & Wu, Y. (2013). Journal of Fluency Disorders Logistic regression for risk factor modelling in stuttering research ☽. *Journal of Fluency Disorders*, *38*(2), 88–101. https://doi.org/10.1016/j.jfludis.2012.09.003

Ubechu, B. O., & Okeke, O. . (2017). Landslide: Causes, Effects and Control. *International Journal of Current Multidisciplinary Studies*, *3*(03), 647–663.

Yuniarta, H., Saido, A. P., & Purwana, Y. M. (2015). Kerawanan Bencana Tanah Longsor Kabupaten Ponorogo. *Jurnal Matriks Teknik Sipil*, *3*(1), 194–201.