Journal of Fundamental and Applied Sciences

ISSN 1112-9867

Available online at http://www.jfas.info

Research Article

Special Issue

ESTIMATED PROBABILITY OF THE NUMBER OF BUILDINGS DAMAGED BY THE FLOODS AND THE COST OF REPAIRS

Sukono[1], M. Suyudi[1], S. Supian[1], Y. Hidayat[2,*], R. Budiono[3], M. K. A. Kamarudin[4], A. R. Hassan[4], F. Lananan[4] and R. Umar[4]

[1]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran

[2]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran

[3]Department of Biology, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran

Jl. Raya Bandung-Sumedang Km 21, Jatinangor 45363 Jawa Barat, Indonesia

[4]East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Gong Badak Campus, 23000 Kuala Nerus, Terengganu, Malaysia

Published online: 15 January 2018

ABSTRACT

Flood disasters often cause buildings damaged, for repairs required considerable cost. This paper analyzes estimates of the probability and the number of buildings damaged by the Citarum River flood in the framework of cost recovery planning. The probability estimation of building damage was carried out using a logistic regression model with a genetic algorithm approach. The number of buildings damaged and risk of losses are estimated using the principle of expected value. The analysis shows that the probability estimator of the building damage significantly follows the logistic regression model. Meanwhile, the estimated the building damage is 349, with a loss of IDR67.80 billion.

Author Correspondence, e-mail: yuyun.hidayat@unpad.ac.id

doi: http://dx.doi.org/10.4314/jfas.v10i1s.8

So the related parties (government, financial institutions and homeowners) should reserve the cost of repair at least worth the risk of loss, to face future flood disasters.

Keywords: Citarum River; logistic regression; genetic algorithm; losses risk; expected value.

1. INTRODUCTION

The escalation and intensity of Citarum river flood in Kabupaten Bandung are more increasingly and happened repeatedly in every year. Therefore, it gives impact to the residence that is resided near the Citarum watershed and its neighborhoods. One of the direct faced impacts is the building damage [12]. The building damage which appeared by flood destruction includes: hydrostatic pressure, buoyancy, and the effect towards the building caused by debris [1, 12]. The building damage, besides it is caused by the flood damage ability, also caused by the water pathology also the contaminant soluble in the liquid against the building material. [10]. According to the report from Balai Besar Wilayah Sungai (BBWS) Citarum, the causing effect of flood which hit Bale Endah area has a type of building resides by the residence mostly are the ordinary public building [12]. So by the increasing of escalation and intensity of floods in the area, thus type of building has a quite big probability of damage to be occurred [6, 13].

The occurrence probability of buildings damage by floods, suspected can be estimated using logistic regression models. However, to our knowledge there has been no previous research on the probability estimate building damage caused by flooding is done by using a logistic regression model. Therefore, the reference review of the estimated probability of damage to the building, based on its similarity to the failure of loan payments to a financial institution. It is based on several studies that have done by previous researchers. For example, in [16, 7] conducts research on bank failure (bankruptcy) with logistic regression prediction. The purpose of that study is to determine the bankruptcy indicators probability of a bank, in case of economic and financial crisis. To perform the estimation of logistic regression parameter model, one of which can be done by using a genetic algorithm approach [4]. This also demonstrated by some previous researches. For example, in [14] estimate the probability of non-performing loans from customers. A similar study has been conducted involving the use of a logistic regression model and genetic algorithms or other soft computing [5].

Based on the previous discussion, in this paper, the analysis of probability estimation and the number of building damage caused by flood disaster is done. The analysis of the problem is done using logistic regression model and genetics algorithm. This analysis aims to estimate the magnitude of the probability of damage to a building and also to estimate the cost of building damage and its risk of loss. So, it can be used for the planning of joint repair costs when the flood is occurred.

2. METHODHOLOGY

In this section, the intention is to conduct a discussion about materials and methodology used in this paper.

2.1. Materials

In this section we intend to explain the material used in the analysis. The data used in this analysis is obtained from interviews with community representatives in the Citarum river flood disaster in South Bandung area in 2016. The data consists of 100 samples in which n_1 = 17 is included in category 1 that the building damage occurred and n_0 = 83 in which included in category 0 that the building damage does not occurred. There are 8 factors that assumed to contributes to the causing of buildings damage by flooding includes F_1 is the distance between house building to the river flow, F_2 is the type of house building, F_3 is the strength level of water flow, F_4 is the strength level of attractions of other material brought by water flow, F_5 is the height level of floodwaters, F_6 is the average time of floodwater, F_7 is the average frequency of floods disaster occurred each year and F_8 is the sedimentation level of flood water content.

To test the normality, data observation need to be done because the dependent variable is the data of which its value is fluctuate between high and low. Striking differences in values can lead to bias in the data analysis, so it does not replace the real situation. Data normality test was performed using statistical software Minitab 16.

2.2. Methodology

In this section, the intention is to conduct discussions concerning the methodology used in the analysis. The discussions included logistic regression model, genetic algorithm and model parameter significance test.

2.2.1. Logistic Regression Model

This section deals with logistic regression models and the estimation methods. Binary logistic regression model is used to estimate the effect of several independent variable F_i ($i = 1,...,M$, M is the number of independent variable) respect to the dependent variable Z, which is a variable with binary value or dichotomous [3, 9]. The term binary or dichotomous is because the variable is possibly has values 0 and 1. The binary logistic regression model used in this analysis has the following formula [14]:

$$\pi(F_i) = \frac{e^{\sum_{j=0}^{J}\theta_k F_{ij}}}{1 + e^{\sum_{j=0}^{J}\theta_k F_{ij}}}, \quad i = 1,...,M \tag{1}$$

If the logarithmic natural is given to both of left and right hand side of Equation (1), then the following new equation can be formed [2], [14]:

$$g(F_i) = \ln\left(\frac{\pi(F_i)}{1-\pi(F_i)}\right) = \sum_{j=0}^{J} \theta_j F_{ij}, \quad i = 1,...,M \tag{2}$$

The parameter estimation of Equation (2) can be done using the likelihood method as discussed below.

2.2.2. Estimator of Likelihood Regression Logistic Model

This section aims to discuss likelihood estimators for logistic regression model. Suppose that $\boldsymbol{\theta}' = (\theta_0, \theta_1, ..., \theta_J)$ is a vector parameter. According to [14], the purpose is to estimate the parameter θ_j ($j = 0,1,...,J$) which contributes to Equation (2). Suppose that there are J independent variables $F_1,...,F_J$, the conditional density function of the independent variable Z and vector parameter $\boldsymbol{\theta}$ following Bernoulli distribution:

$$f(Z \mid \boldsymbol{\theta}) = \prod_{i=1}^{M} \pi_i^{Z_i}(1-\pi_i)^{1-Z_i}\ ; \quad Z_i = 0 \text{ or } 1 \tag{3}$$

Maximum Likelihood Estimator (MLE) is that the values of the vector parameter $\boldsymbol{\theta}$ can maximize the likelihood function of Equation (3).

To variable Z_i, the value 0 or 1 is given for each pair of (F_i, Z_i). If Z_i = 1 is given, then the contribution to likelihood function is $\pi(F_i)$ and if the value Z_i = 0 is given, then the contribution to likelihood function is $1-\pi(F_i)$. Therefore, the contribution to the likelihood function of the pair (F_i, Z_i) can be written as follows [8, 14]:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{M} \pi_i^{Z_i} (1-\pi_i)^{1-Z_i} \; ; \; Z_i = 0 \text{ or } 1 \tag{4}$$

By substituting Equation (1) to Equation (4), the following equation is obtained:

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{M} \left(e^{\sum_{j=0}^{J} \theta_j F_{ij}} \right)^{Z_i} \left(1 + e^{\sum_{j=0}^{J} \theta_j F_{ij}} \right)^{-1} \tag{5}$$

If the logarithmic natural is given to both of left and right hand side of Equation (5), then the following equation can be obtained [14]:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{M} \left\{ Z_i \sum_{j=0}^{J} \theta_j F_{ij} - \ln\left(1 + e^{\sum_{j=0}^{J} \theta_j F_{ij}} \right) \right\} \tag{6}$$

In this paper, the value of the elements of the vector parameter $\boldsymbol{\theta}$ in Equation (6) is estimated by using a genetic algorithm as discussed in the following.

2.2.3. Estimation by Using Genetic Algorithm

In this section we intend to discuss the genetic algorithm to maximize the log likelihood function in order to estimate parameters of the logistic regression model.

Genetic algorithm is a search method based on the mechanism of scientific genetics and natural selection [4]. Scientific genetic mechanisms reflect the individual's ability to mate and produce the offspring with characteristics similar to their parents. While natural selection describes that the creatures is able to sustain their life, if they are able to adapt to its surrounding environment [11]. Thus, his descendant is expected to have the best combination of characteristics from their parents and able to sustain future generations [15].

The general structure of a genetic algorithm is as following steps [14]:

a) Generating an initial population, the initial population is generated randomly to obtain an initial solution;

b) The population is made up of a number of chromosomes that represent the achieved solution;

c) Forming a new generation, the new generation's involving three operators namely reproduction/selection, crossover and mutation.

d) Evaluation of the solution, each process of population was evaluated by calculating the value of each chromosome fitness and the evaluation was conducted until stopping criteria are met. If the stopping criteria have not been met, a new generation will be set up again by repeating steps b).

Based on the general structure of the genetic algorithm, to maximize the log likelihood function of Equation (6), the following general genetic algorithm can be arranged [11, 15]:

1) Determining the initial population. The initial population determined is J , which are randomly generated. This number of initial random population is then converted into the form of decimal values θ_j where $j=1,...,J$.

2) Evaluation of chromosomes. The fitness value of a chromosome is the value of the log likelihood function which is given as Equation (6). Based on the fitness values, the biggest value is chosen for maximization program.

3) The calculation of population convergent percentage. Percent of population convergence p_c is the percentage of the number of individuals who produce the same and biggest fitness value. This p_c value is counted using following equation:

$$p_c = \frac{n}{pop} \times 100\% \tag{9}$$

where n is the number of individuals who produce the same and biggest fitness value and pop is the number of population.

4) Evaluation of stopping condition. The process of genetic algorithm will stop when the counter generation has reached the defined number of generations c_g that is c_g =1000 or percent of population conventions p_c to reach the defined threshold limit that is $\tau = 90\%$.

5) Chromosome selection. The selection process is based on roulette wheel selection. Since this is the maximization program, then the fitness evaluation value $eval(v_i)$, $i = 1,...,N$ is done based on Equation (6) by formula:

$$eval(v_i) = f(\boldsymbol{\theta}) \tag{10}$$

where $f(\boldsymbol{\theta})$ is the fitness value which refer to Equation (6) and $\boldsymbol{\theta}' = (\theta_0, \theta_1, ..., \theta_J)$.

6) Crossbreeding. The new populations from selection results are crossbreed by using Single-Point Crossover (SPX) method.
7) Mutation. Mutations of each generation is done by calculate $m \times pop\_size \times p_m$ where m is the number of mutation, pop_size is the size of population and p_m is a probability of mutation (the value is defined randomly).
8) Decoding. Decoding is the genes encoding process within a chromosome such that the value back to normal, which is by changing the coding into decimal values.

Furthermore, the genetic algorithm is used to analyze the observation data of building damage by Citarum river flood.

2.2.4. Test of Verification and Validation of Parameter for Logistic Regression Model

In this section, the purpose is to explain about the verification and validation of logistic regression model parameter.

2.2.4.1. Verification Test

In the verification of parameter estimation test of logistic regression model of this paper, the Wald statistical test is done. According to [14-15], the verification of parameter estimator θ_j ($j = 0,1,...,J$) is performed by Wald individually. The Wald statistical test t_{Stat} is performed where statistics t_{Stat} follows standard Normal distribution as follows:

$$\hat{t}_{Stat} = \frac{\hat{\theta}_j}{SE(\hat{\theta}_j)}; \quad j = 0,1,...,J \tag{11}$$

where $\hat{\theta}_j$ is the estimator of θ_j ($j = 0,1,...,J$) and $SE(\bar{\theta}_j)$ is the standard deviation of $\hat{\theta}_j$. The hypothesis of Wald test is $H_0 : \hat{\theta}_j = 0$ with alternative $H_1 :\ \hat{\theta}_j \neq 0$ ($j = 0,1,...,J$). The test criterion is reject H_0 if $\hat{t}_{Stat} < t_{\frac{1}{2}(1-\alpha)}$ or $\hat{t}_{Stat} > t_{\frac{1}{2}(\alpha)}$, otherwise accept H_0 whenever $t_{\frac{1}{2}(1-\alpha)} \le \hat{t}_{Stat} \le t_{\frac{1}{2}(\alpha)}$ where $t_{\frac{1}{2}(\alpha)}$ is the percentile of Normal standard distribution with level of significance $(1-\alpha)\%$.

2.2.4.2. Validity Test

In this paper the validation test performed using Likelihood Ratio Test, Hosmer and Lameshow Test and R-Square Test.

Likelihood Ratio Test. To test the validity of model logistic regression model parameter estimator can be done by using statistical likelihood ratio test $\hat{G}$ as follows [8, 14-15]:

$$\hat{G} = 2\left\{\sum_{i=1}^{M} Z_i \ln \hat{\pi}_i + \sum_{i=1}^{M}(1-Z_i)\ln(1-\hat{\pi}_i) - n_1 \ln n_1 - n_0 \ln n_0 + M \ln M\right\} \quad (12)$$

The hypothesis for the likelihood ratio test is H_0: $\hat{\theta}_0 = \hat{\theta}_1 = ... = \theta_J = 0$ ($j = 0,1,...,J$) with alternative H_1 : $\exists \hat{\theta}_0 \ne \hat{\theta}_1 \ne ... \ne \theta_J \ne 0$ ($j = 0,1,...,J$). Therefore, statistics $\hat{G}$ is asymptotically Chi-Square $\chi^2_{(\alpha,df)}$ distributed, then the test criterion used is rejecting H_0 if $\hat{G} > \chi^2_{(\alpha,df)}$, otherwise accept H_0 if $\hat{G} \le \chi^2_{(\alpha,df)}$ where α the defined significance is level and $df = J-1$ with J is the number of parameter of a model.

Hosmer and Lemeshow Test. The validation test of a model parameter estimator can be done by Hosmer and Lemeshow statistical test $\hat{\eta}$ as follows [8, 14-15]:

$$\hat{\eta} = \sum_{j=1}^{g} \frac{(\xi_j - n_j \overline{\pi}_j)}{n_j \overline{\pi}_j (1-\overline{\pi}_j)} \quad \text{or} \quad P_{Value} = \Pr(\hat{\eta}) \quad (13)$$

where $\hat{\xi}_j = \sum_{j=1}^{n_j} Z_j$ and $\overline{\pi}_j = \sum_{j=1}^{n_j}(m_j \overline{\pi}_j / n_j)$. The hypothesis for Hosmer and Lemeshow test is H_0: There is no difference between the observation and the model used with alternative H_0: There is a difference between the observation and the model used. Hosmer and Lemeshow statistics are following the Chi-Square $\chi^2_{(\alpha,df)}$, then the criterion used is reject H_0 if $\hat{\eta} > \chi^2_{(\alpha,df)}$ or $P_{Value} < (1-\alpha)$, otherwise accept H_0 if $\hat{\eta} \le \chi^2_{(\alpha,df)}$ or

$P_{Value} \geq (1-\alpha)$ where α is the defined level of significance and $df = g-2$ with usual value of $g = 10$.

R-Square. According to [14], Hosmer and Lemeshow using deterministic value R^2 within logistic regression model to perform the strong relation between independent and dependent variable. The statistics of R^2 can be determined by following equations [8, 14-15]:

$$R^2 = 1 - \exp\left\{ -\left(\frac{\hat{\ell}(\boldsymbol{\theta})}{N} \right) \right\} \tag{14}$$

where $\hat{\ell}(\boldsymbol{\theta})$ the estimator of maximum is log likelihood of Equation (6) and N is the number of observation data. If the value of $R^2 \to 1$, then the relation between independent and dependent variable is strong. Otherwise, if $R^2 \to 0$, then the relation between independent and dependent variable is weak.

3. RESULTS AND DISCUSSION

In this section, the intention is to conduct the discussions about the analyzed data, results and discussion based on the data processing.

As previously described in section 2.1 in multivariate analysis, the normality test should be done before the analyzed data is used. The purpose of this normality test is to ensure that the data is normally distributed [16]. Data normality test can be done by Minitab 16 software. If the data is normal distributed, then this data is used to estimate the parameter of a binary logistic regression model.

3.1. Results

In this section, the discussion about parameter estimation results of binary logistic regression model is done. The results of the normality test of analyzed data indicates that the data be normal distribution. Therefore, this data can then be used for parameter estimation which expressed as the vector parameter $\boldsymbol{\theta}' = (\theta_0, \theta_1, ..., \theta_8)$. The parameter estimator to be obtained is a parameter that maximizes likelihood log function of Equation (6). In this paper, the parameter estimation is done using genetic algorithm. The estimation process using genetic algorithm is done by using MATLAB 7.0 software. The estimation result of this parameter is given in Table 1.

Table 1. Parameters estimator and standard error

Factors of (F_j)	**Parameter Estimators ($\hat{\theta}_j$)**	**Standard Error** $SE(\hat{\theta}_j)$	**Ratio** $\hat{t}_j = \frac{\hat{\theta}_j}{SE(\hat{\theta}_j)}$	**Significances**
F_1	2.09867895	0.68455	3.065778906	Significance
F_2	1.21426465	0.39372	3.084081708	Significance
F_3	3.09974576	1.05834	2.928875182	Significance
F_4	2.97525321	1.03263	2.881238401	Significance
F_5	2.03458493	0.49261	4.130214429	Significance
F_6	1.09325381	0.35821	3.051991318	Significance
F_7	1.00316351	0.34790	2.883482380	Significance
F_8	1.12642730	0.36934	3.049838360	Significance
Constant	-2.30413812	1.03184	-2.233038184	Significance

Log likelihood statistic $\hat{G}$ = 27.835282450

For the parameter estimators given in Table 1, verification and validation test is required. Verification and validation test is done to ensure that the parameter estimators are significant, either partially or wholly toward to the estimated logistic regression model.

Verification test for logistic regression model parameter estimator is done using Wald statistic, referring to Equation (11). The hypothesis for the Wald test is H_0: $\hat{\theta}_j = 0$ with alternative H_1: $\hat{\theta}_j \neq 0$ ($j = 0,1,...,8$). If the defined significance is α = 0.95, then from normal standard distribution table, the obtained percentile values $t_{\frac{1}{2}(1-0.95)}$ = -0.27 and $t_{\frac{1}{2}(0.95)}$ = 0.27. Observing the results given in Table 1, for the value of $\hat{t}_j$ ($j = 0,1,...,8$), all are $\hat{t}_j < t_{\frac{1}{2}(1-0.95)}$ or $\hat{t}_j > t_{\frac{1}{2}(0.95)}$, then the hypothesis H_0 is rejected. It means that the whole logistic regression model parameter estimator is significant.

Furthermore, the validation of the test for logistic regression model parameter estimators is intended to test that the estimators of logistic regression model simultaneously is significantly affecting $\pi(F)$. Validation test in this paper was performed using Likelihood statistical Ratio by referring to the Equation (12). The hypothesis in this test is H_0: $\hat{\theta}_0 = \hat{\theta}_1 = ... = \theta_8 = 0$, with alternative H_1: $\exists\hat{\theta}_0 \neq \hat{\theta}_1 \neq ... \neq \theta_8 \neq 0$. If the defined level of significance $\alpha = 0.95$ and $df = 8$, then from Chi-Square distribution table the obtained percentile value is $\chi^2_{(1-0.95,8)} =$ 2.7326. While statistics $\hat{G}$ = 27.835282450, clearly that $\hat{G} > \chi^2_{(1-0.95,8)}$. Therefore, hypothesis H_0 is rejected. It means that the estimators of logistic regression model simultaneously is significantly affecting $\pi(F)$.

The next validation test is performed using Hosmer and Lemeshow statistical test with refer to Equation (13). The hypothesis for Hosmer and Lemeshow test is as follows, H_0: There is a difference between the observation data to its estimator model with alternative H_1: There is no difference between the observation data to its estimator model. Hosmer and Lemeshow Statistical test can also be done using statistics P_{Value} refers to Equation (13). In this test, the obtained probability is P_{Value}= 0.256395. Whenever the defined level of significance is α= 0.95, it clearly that $P_{Value} > \alpha$. Therefore, the hypothesis H_0 is rejected, means that no difference between two data observed to its estimator model.

Further, to measure the strength of relationship between independent and dependent variables, it can be performed using the coefficient of determination R^2 refers to Equation (14). Based on the data processing results, the determination value is R^2= 0.79534972. This indicates that there is exist a strong relationship between independent variable $F_1,...,F_8$ to the dependent variable $\pi(F)$.

3.2. Discussion

Based on the estimator of the logistic regression model parameters given in Table 1, whenever the parameter estimators are substituted into Equation (1), a logistic regression model can be used to estimate the damage probability caused by Citarum river flood.

As an example, an event is predicted to have following category: distance between the house to the river is 500 meters or category F_1 = 1; the type of building is a simple house or in a category F_2 = 1; the attractions level of the water flow is very strong or in category F_3 = 1; the attractions level of the other material carried by water is big or in the category F_4 = 1; the height level of inundation is above 1 meter or in category F_5 = 1; the average time of flooding occurred is less than 7 days or in category F_6 = 0; the frequency average of flooding occurred twice each yearly or in category F_7 = 1; and the level of sedimentation content by flood is very low or in the category F_8 = 0. If these category factor values and estimator of the logistic regression model parameters given in Table 1 substituted to Equation (1), the probability of building damage can be calculated as follows:

$$\pi(F) = \frac{e^{-2.304+2.099(1)+1.214(1)+3.100(1)+2.975(1)+2.035(1)+1.093(0)+1.003(1)+1.126(0)}}{1+e^{-2.304+2.099(1)+1.214(1)+3.100(1)+2.975(1)+2.035(1)+1.093(0)+1.003(1)+1.126(0)}}$$

$$= 0.99996$$

Thus, for the house building which is included in simple category with quite near distance with the flood characteristics described previously its probability of damage occurred is big that is 0.99996. These examples along with the other nine buildings, according to the characteristics of each building predicted the probability of the damage given in Table 2.

Table 2. Predicted probability of building damage based on characteristics

Buildings	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	**Probability**
1	1	1	1	1	1	0	1	0	0.99996
2	1	1	1	1	0	0	1	0	0.99969
3	0	1	0	1	0	1	0	1	0.98376
4	0	1	0	0	1	0	1	0	0.87523
5	1	0	1	0	1	0	1	0	0.99736
6	1	0	1	1	0	1	0	1	0.99969
7	1	1	1	0	1	0	0	1	0.99930
8	1	0	1	0	0	1	1	0	0.99325
9	0	0	0	1	0	0	0	0	0.66173
10	0	0	0	0	0	0	0	0	0.09079

Furthermore, based on the observation results, it is known that from 100 sample, the information obtained is that n_1 = 17 includes in 1 means that the damage buildings is occurred and n_0 = 83 includes category 0 means that the building damage does not occurred. The average of loss value caused by building damage is reaching IDR194,269,341.00 for each house. According to the information, it can be determined that the probability of building damage occurred is 0.17. If in the flooding season about 2053 number of houses predicted to be affected by Citarum River flood, then it is predicted that building damage occurred for about 349 numbers of houses. The total loss is predicted to reach IDR 67.80 billion. Therefore, the parties need to create a budget plan to repair buildings damage minimal equal to the total loss IDR67.80 billion.

4. CONCLUSION

In this paper, the probability estimation and the number of building damage caused by Citarum river flood are analyzed, which related to the cost recovery planning. According to the analysis, it can be concluded that the probability estimation of building damage occurred can be done using logistic regression model. The genetic algorithm [17] is performed to estimate the parameter of logistic regression model. The estimation results of the parameter of logistic regression model shows that 8 factor $F_1,...,F_8$ significantly affecting the probability function $\pi(F)$ in Equation (1). By estimator of Equation (1), for the simple house building with just 500 meters far from the river with characteristic of great flood, the estimated probability of damage occurred is 0.99996. In the next year flood season, if there are predicted 2053 homes affected by flooding [18-20], then there will be 349 homes were damaged with total losses reached IDR67.80 billion. Prediction of damage number and total loss can be taken into consideration for planning the repair cost budget.

5. ACKNOWLEDGEMENTS

Further thanks to the Rector, Director of DRPMI and Dean of FMIPA, Universitas Padjadjaran which has a grant program of the Academic Leadership Grant (ALG) under the coordination of Prof. Dr. Sudradjat and a grant program of Competence Research of Lecturer of Unpad (*Riset Kompetensi Dosen Unpad*/RKDU) under the coordination of Dr. Sukono which is a means to increase research activities and publications to researchers at the Universitas Padjadjaran.

6. REFERENCES

[1] Brody S D, Zahran S, Maghelal P, Grover H, Highfield W E. Examining the impact of planning and development decisions on property damage in Florida. Journal of the American Planning Association, 2007, 73(3):330-345

[2] Bayaga A. Multinomial logistic regression: Usage and application in risk analysis. Journal of Applied Quantitative Methods, 2010, 5(2):288-297

[3] Cramer J S. The origins of logistic regression. Tinbergen Institute discussion paper, T1.2002-119/4, Amsterdam: Tinbergen Institute, 2002, pp. 1-15

[4] Demir E, Akkus O. An introductory study on how the genetic algorithm works in the parameter estimation of Binary Logit Model? International Journal of Sciences: Basic and Applied Research, 2015, 19(2):162-180

[5] Eshima N, Tabata M, Zhi G. Path analysis with logistic regression models: Effect analysis

of fully recursive causal systems of categorical variables. Journal of the Japan Statistical Society, 2001, 31(1):1-4

[6] Francisco J P S. Property damage recovery and coping behavior of households affected by an extreme flood event in Marikina City, Metro Manila, Philippines. Philippine Institute for Development Studies Discussion Paper Series No. 2015-40, pp. 1-52

[7] Hauser R P, Booth D. Predicting bankruptcy with robust logistic regression. Journal of Data Science, 2011, 9(4):565-584

[8] Lu M, Yang W. Multivariate logistic regression analysis of complex survey data with application to BRFSS data. Journal of Data Science, 2012, 10(2):157-173

[9] Peng C Y, Lee K L, Ingersoll G M. An introduction to logistic regression analysis and reporting. Journal of Educational Research, 2002, 96(1):3-14

[10] Rehman K, Cho Y S. Building damage assessment using scenario based tsunami numerical analysis and fragility curves. Water, 2016, 8(3):1-17

[11] Roupec J. Advanced genetic algorithms for engineering design problems. Engineering Mechanics, 2011, 17(5-6):407-417

[12] Setiadi H A. Identifikasi kerusakan bangunan dan fungsi infrastruktur akibat banjir Citarum di Wilayah Kabupaten Bandung. Jurnal Sosial Ekonomi Pekerjaan Umum, 2013, 5(1):51-64

[13] Siriwardane E. The probability of rare disasters: Estimation and implications. Harvard Business School Finance working paper no. 16-061, 2015, pp. 1-80

[14] Sukono A S, Mamat M, Prafidya K. Credit scoring for cooperative of financial services using logistic regression estimated by genetic algorithm. Applied Mathematical Sciences, 2014, 8(1):45-57

[15] Hosmer D. W., Lameshow S. Apllied logistic regression. New York: John Wiley and Sons, 1989

[16] Yan X, Luo W, Li W, Chen W, Zhang C, Liu H. An improved genetic algorithm and its application in classification. International Journal of Computer Science Issues, 2013, 10(1):337-346

[17] Masrom S, Abidin S Z, Omar N, Rahman A S, Rizman Z I. Dynamic parameterizations of

particle swarm optimization and genetic algorithm for facility layout problem. ARPN Journal of Engineering and Applied Sciences, 2017, 12(10):3195-201

[18] Zaghdoudi T. Bank failure prediction with logistic regression. International Journal of Economics and Financial Issues, 2013, 3(2):537-543

[19] Saudi A S, Ridzuan I S, Balakrishnan A, Azid A, Shukor D M, Rizman Z I. New flood risk index in tropical area generated by using SPC technique. Journal of Fundamental and Applied Sciences, 2017, 9(4S):828-850

[20] Saudi A S, Kamarudin M K A, Ridzuan I S, Ishak R, Azid A, Rizman Z I. Flood risk index pattern assessment: Case study in Langat river basin. Journal of Fundamental and Applied Sciences, 2017, 9(2S):12-27

How to cite this article:
Sukono, Suyudi M, Supian S, Hidayat Y, Budiono R, Kamarudin MKA, Hassan AR, Lananan F, Umar R. Estimated probability of the number of buildings damaged by the floods and the cost of repairs. J. Fundam. Appl. Sci., 2018, 10*(1S), 112-127.*