

Modeling Risk Factors for Total Paralysis of Stroke Patients in RSUD Dr. R. Sosodoro Djatikoesoemo Bojonegoro Using Binary Logistic Regression

Puji Hidayatus Sholikhah¹, Sri Pingit Wulandari¹

Abstract—One of the causes of death in Bojonegoro Regency is stroke. Stroke is a nervous system function disorder that occurs suddenly and is caused by circulatory disorders of the brain. From 2015 to March 2018, stroke has always been ranked first as the type of disease that most patients suffer in RSUD Dr. R. Sosodoro Djatikoesoemo Bojonegoro. Stroke can cause sufferers to experience limited ability to carry out daily activities (eating, dressing, defecating, bathing, and moving) because there are limbs that are paralyzed so it is difficult to move. Therefore, this study aims to describe the characteristics of stroke patients and to analyze the risk factors that cause stroke patients get totally paralyzed using binary logistic regression. The data used in the form of secondary data obtained from the results of medical records of stroke patients who were hospitalized at RSUD Dr. R. Sosodoro Djatikoesoemo Bojonegoro from January to March 2018. Based on the results of the analysis showed that the majority of stroke patients are over 55 years old and males also the risk factors that cause stroke patients get totally paralyzed are the type of stroke and previous history of stroke.

Keywords—Binary Logistic Regression, Paralysis, Stroke

I. INTRODUCTION

Stroke is a disease of the brain in the form of impaired local and or global nerve function that appears suddenly, progressively, and quickly. Impaired nerve function in stroke is caused by non-traumatic brain blood circulation disorders. Nerve disorders can cause symptoms such as paralysis of the face or limbs, speech is not fluent, changes in consciousness, impaired vision, and so forth [1]. Paralysis is the most common defect experienced by stroke sufferers and can occur in various parts of the body. As a result of the paralysis, stroke patients have difficulty doing activities or doing work that could previously be done alone [2].

Several previous studies on stroke cases showed that the variables that influence the incidence of ischemic stroke include gender, age, and hypertension [3]. According to [4], explained that the variables that influence the incidence of stroke where the risk of stroke is found when affected by a stroke are age, hypertension, diabetes mellitus, and hyperlipidemia or high cholesterol. According to [5], variables that influence the death of stroke patients in Dr. Wahiddin Sudirohusodo Makassar includes history of previous stroke, type of stroke, gender, and age. Meanwhile, according to [6], the basic causes of stroke death in Padang Pariaman Regency, West Sumatra Province are age and gender.

One of the 10 diseases that cause death in Bojonegoro Regency is a stroke. Based on data obtained from Dr. R. Sosodoro Djatikoesoemo Bojonegoro in 2015, stroke was ranked first as the most common type of illness for 347 patients [7]. In addition, in 2016 to October 2018 stroke was still ranked first as the most common type of disease suffered by patients. In 2016 and 2017 there were 720 patients respectively while in 2018 that is until October there were 668 patients. Stroke patients have limited ability to perform daily activities because there are limbs that experience paralysis. According to medical records, stroke patients are said to experience total paralysis if in doing all daily activities (eating, dressing, defecating, bathing, and moving) must be totally assisted whereas if one of the daily activities is not totally assisted then considered not experiencing total paralysis.

Based on the problem of stroke that occurred in Bojonegoro Regency, will be doing such a description of the characteristics of stroke patients and analysis of risk factors that cause stroke patients to experience total paralysis in RSUD Dr. R. Sosodoro Djatikoesoemo Bojonegoro used the binary logistic regression method. RSUD Dr. R. Sosodoro Djatikoesoemo Bojonegoro was chosen because the hospital is a regional general hospital owned by the Bojonegoro Regency Government which is located in the city center and is the only type B hospital in Bojonegoro. The hospital has facilities to conduct medical check-ups (one of which is for early detection of strokes using CT-Scan 64 Slices).

¹Puji Hidayatus Sholikhah Department of Business Statistics, Institut Teknologi Sepuluh Nopember (ITS), Surabaya, Indonesia. E-mail: psholikhah@gmail.com, sri_pingit@statistika.its.ac.id.

Binary logistic regression is a data analysis method used to determine the effect of predictor variables (X) which are categorical or continuous on the response variable (Y) which is binary (dichotomous) [8]. The response variable used is the condition of stroke patients with category 0 for stroke patients who are not totally paralyzed and category 1 for stroke patients who are totally paralyzed. Whereas the predictor variables used were length of stay, type of stroke, genetic risk of stroke including age, gender, family history of stroke, and history of previous stroke and risk of lifestyle stroke including hypertension, diabetes mellitus, hyperlipidemia or high cholesterol, and body mass index of stroke patients.

II. METHOD

A. Data Source

The data used in this study are secondary data obtained from the medical records of stroke patients who are hospitalized in Dr. R. Sosodoro Djatikoesoemo Bojonegoro in January to March 2018 with 111 patients.

B. Variable of Observation

The variables used in this study are shown in Table 1.

TABLE 1.
VARIABLE OF OBSERVATION

Variable	Explanation	Category	Data Scale
Y	Condition of Stroke Patients	0 = Not Totally Paralyzed 1 = Totally Paralyzed	Nominal
X_1	Length of Stay	0 = < 8 Days 1 = \geq 8 Days	Nominal
X_2	Type of Stroke	0 = Non Hemorrhagic 1 = Hemorrhagic	Nominal
X_3	Age	0 = \leq 55 Years Old 1 = > 55 Years Old	Nominal
X_4	Gender	0 = Female 1 = Male	Nominal
X_5	Family History of Stroke	0 = No 1 = Yes	Nominal
X_6	Previous History of Stroke	0 = No 1 = Yes	Nominal
X_7	Severe Hypertension	0 = No 1 = Yes	Nominal
X_8	Diabetes Mellitus	0 = No 1 = Yes	Nominal
X_9	Hyperlipidemia (High Cholesterol)	0 = No 1 = Yes	Nominal
X_{10}	BMI	0 = Not Overweight 1 = Overweight	Nominal

The operational definitions of response variables and predictor variables are noted when stroke patients are admitted to the RSUD Dr. R. Sosodoro Djatikoesoemo Bojonegoro in January to March 2018.

C. Steps of Analysis

The steps of analysis in this study are as follows:

1. Describe stroke patients using pie chart.
2. Do an independence test between the response variable (Y) with each predictor variables (X) using the Chi-Square test.
3. Perform model specifications.
4. Do an analysis using binary logistic regression with the response variable (Y) is the condition of stroke patients including category 0 for stroke patients who are not totally paralyzed and category 1 for stroke patients who are totally paralyzed and predictor variables (X), which are significant variables in the test independence. Binary logistic regression analysis performed includes:
 - a. Estimating binary logistic regression parameters.
 - b. Testing the significance of the parameters simultaneously.
 - c. Testing the significance of the parameters partially.
 - d. Testing the goodness of fit from the binary logistic regression model.
 - e. Perform interpretation using odds ratio.
 - f. Checking the classification accuracy.
5. Make conclusions.

D. Pie Chart

Pie chart is a chart in the shape of a circle that shows the value of data as a percentage of the whole. Each value of the data series will represents a slice of a pie. The larger value of the data is the larger pie slice [9]. The example for a pie chart can be shown in Figure 1.

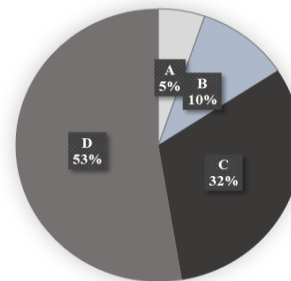


Figure 1. Pie Chart

E. Independence Test

Independence test is a test used to determine the relationship between two variables. Independence test has to do with a contingency table that is a table that contains data on the number or frequency or categories that describe two or more variables simultaneously. If the table contains data from two variables, then it is called a two-dimensional contingency table, if from three variables it is called a three-dimensional contingency table, and so on [10]. A two-dimensional contingency table with rows as many as r and columns as many as c is called $r \times c$ contingency table as shown in the Table 2.

TABLE 2.
TWO-DIMENSIONAL CONTINGENCY TABLE

X	Y					Total
	1	2	3	...	c	
1	n_{11}	n_{12}	n_{13}	...	n_{1c}	$n_{1.}$
2	n_{21}	n_{22}	n_{23}	...	n_{2c}	$n_{2.}$
3	n_{31}	n_{32}	n_{33}	...	n_{3c}	$n_{3.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
r	n_{r1}	n_{r2}	n_{r3}	...	n_{rc}	$n_{r.}$
Total	$n_{.1}$	$n_{.2}$	$n_{.3}$...	$n_{.c}$	$n_{..}$

Explanation

$l = 1, 2, 3, \dots, r$ and $m = 1, 2, 3, \dots, c$

n_{lm} = Number of observations in l^{th} row and m^{th} column

$n_{l.}$ = Total of observations in l^{th} row

$n_{.m}$ = Total of observations in m^{th} column

$n_{..}$ = Total all of observations

The hypothesis for independent testing is

$H_0 : P_{lm} = P_{l.} \times P_{.m}$ (There is no a relationship between observed variables)

$H_1 : P_{lm} \neq P_{l.} \times P_{.m}$ (There is a relationship between observed variables)

The test statistics used for the independence test is shown in Equation 1.

$$\chi^2 = \sum_{l=1}^r \sum_{m=1}^c \frac{(n_{lm} - \hat{e}_{lm})^2}{\hat{e}_{lm}} \quad (1)$$

with

$$\hat{e}_{lm} = \frac{n_{l.} \times n_{.m}}{n_{..}} \quad (2)$$

Explanation

P_{lm} = Probability of observations in l^{th} row and m^{th} column

$P_{l.}$ = Total probability of observations in l^{th} row

$P_{.m}$ = Total probability of observations in m^{th} column

n_{lm} = Number of observations in l^{th} row and m^{th} column

\hat{e}_{lm} = Expectation value in l^{th} row and m^{th} column

$n_{l.}$ = Total of observations in l^{th} row

$n_{.m}$ = Total of observations in m^{th} column

$n_{..}$ = Total all of observations

By using a significant level α then H_0 will be rejected if the value of $\chi^2 > \chi_{\alpha((r-1) \times (c-1))}^2$ or if P-value $< \alpha$.

F. Binary Logistic Regression

Binary logistic regression is a data analysis method used to determine the effect of categorical or continuous predictor variables (X) on binary (dichotomous) response variable [8]. The results of the response variable (Y) consist of two categories denoted by number 1 if "successful" and number 0 if "failed" so that the response variable (Y) follows the Bernoulli distribution for every single observation. The

probability function for each observation is shown in Equation 3.

$$f(y_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (3)$$

where $y_i = 0, 1$ and $i = 1, 2, 3, \dots, n$

Based on Equation 3 if $y_i = 0$ so that $f(y_i) = 1 - \pi(x_i)$ and if $y_i = 1$ so that $f(y_i) = \pi(x_i)$. Binary logistic regression model with p predictor variables can be shown in Equation 4.

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)}} \quad (4)$$

Equation 4 can be described by using the logit transformation of $\pi(x)$ to simplify the estimation of the regression parameters so that the logit model obtained is a linear function of the parameters. The logit model can be shown in Equation 5.

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (5)$$

1) Parameter Estimate

According to [8], to estimate the parameters in logistic regression, the Maximum Likelihood Estimation (MLE) method is used. The MLE method estimates parameters by maximizing the likelihood function and has the condition that a data must follow a certain distribution. In binary logistic regression, each observation follows the Bernoulli distribution so that its likelihood function can be determined.

If x_i and y_i are pairs of predictor variable and response variable for the i^{th} observation ($i = 1, 2, 3, \dots, n$) and it is assumed that each pair is independent, then the probability function for each pair can be shown in Equation 6.

$$f(x_i) = \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \quad (6)$$

where $y_i = 0, 1$ and

$$\pi(x_i) = \frac{e^{\left(\sum_{j=0}^p \beta_j x_j \right)}}{1 + e^{\left(\sum_{j=0}^p \beta_j x_j \right)}} \quad (7)$$

The likelihood function obtained from the merging of the distribution function of each pair that is shown in Equation 8.

$$l(\beta) = \prod_{i=1}^n f(x_i) = \prod_{i=1}^n (1 - \pi(x_i)) e^{y_i \ln \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right)} \quad (8)$$

The likelihood function in Equation 8 will be easier if it is maximized in the form $\ln l(\beta)$ and expressed as $L(\beta)$ that is shown in Equation 9.

$$L(\beta) = \ln l(\beta) = \sum_{j=0}^p \left[\sum_{i=1}^n y_i x_{ij} \right] \beta_j - \sum_{i=1}^n \ln \left[1 + e^{\sum_{j=0}^p \beta_j x_{ij}} \right] \quad (9)$$

The maximum value of β is obtained by derivating $L(\beta)$ and the results obtained are equal to zero. The first derivative of $L(\beta)$ towards β is shown in Equation 10.

$$\frac{\partial L(\beta)}{\partial \beta_j} = \sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \left(\frac{\sum_{j=0}^p \beta_j x_{ij}}{1 + e^{\sum_{j=0}^p \beta_j x_{ij}}} \right) \quad (10)$$

so that

$$\sum_{i=1}^n y_i x_{ij} - \sum_{i=1}^n x_{ij} \pi(x_i) = 0 \quad (11)$$

where $j = 0, 1, 2, \dots, p$

The second derivative of $L(\beta)$ towards β is shown in Equation 12.

$$\frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b} = - \sum_{i=1}^n x_{ia} x_{ib} \pi(x_i) (1 - \pi(x_i)) \quad (12)$$

with $a, b = 0, 1, 2, \dots, p$

To get the estimate value of β , the iteration process is carried out by using Newton Raphson's iteration method. This method uses the first and second derivatives of $L(\beta)$. The steps for Newton Raphson's iteration to get the estimate value of β are as follows:

- a. Determine the initial value of $\beta^{(0)}$ then use Equation 13 to get $\pi(x_i)^{(0)}$.

$$\pi(x_i)^{(0)} = \frac{\exp\left(\sum_{j=0}^p \beta_j^{(0)} x_{ij}\right)}{1 + \exp\left(\sum_{j=0}^p \beta_j^{(0)} x_{ij}\right)} \quad (13)$$

- b. From $\pi(x_i)^{(0)}$ in step (a) then the vector $\mathbf{q}^{(0)}$ and Hessian matrix $\mathbf{H}^{(0)}$ are obtained. That are can be shown in Equation 14 and Equation 15.

$$\mathbf{q} = \begin{bmatrix} \frac{\partial L(\beta)}{\partial \beta_0} \\ \frac{\partial L(\beta)}{\partial \beta_1} \\ \vdots \\ \frac{\partial L(\beta)}{\partial \beta_p} \end{bmatrix} \quad (14)$$

$$\mathbf{H} = \begin{bmatrix} h_{00} & h_{01} & \cdots & h_{0p} \\ h_{10} & h_{11} & \cdots & h_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ h_{p0} & h_{p1} & \cdots & h_{pp} \end{bmatrix} \quad (15)$$

$$\text{with } h_{ab} = \frac{\partial^2 L(\beta)}{\partial \beta_a \partial \beta_b}; a, b = 0, 1, 2, \dots, p \quad (16)$$

- c. To get estimate iterations of β then Newton Raphson's formula is used that is shown in Equation 17.

$$\boldsymbol{\beta}^{(t+1)} = \boldsymbol{\beta}^{(t)} - (\mathbf{H}^{(t)})^{-1} \mathbf{q}^{(t)} \quad (17)$$

with $t = 0, 1, 2, \dots$ until convergent.

Iteration will stop if obtained $\|\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}\| \leq \varepsilon$ where ε is a very small number.

2) Parameter Significance Test

Significance test on the β coefficients of the model that has been obtained is done by simultaneously and partially. Each explanation is as follows:

- i) Test the Significance of Parameters Simultaneously

Test the significance of the parameters simultaneously used to determine the predictor variables (β coefficients) simultaneously have an effect on the response variable or not [8]. The hypothesis used in the parameter significance test simultaneously is as follows.

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{There is at least one } \beta_j \neq 0, \text{ where } j = 1, 2, \dots, p$$

The test statistics used for the parameter significance test simultaneously is shown in Equation 18.

$$G = -2 \ln \frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\prod_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{1-y_i}} \quad (18)$$

$$\text{with } n_1 = \sum_{i=1}^n y_i \text{ and } n_0 = \sum_{i=1}^n (1 - y_i), n = n_1 + n_0 \quad (19)$$

Explanation

n_1 = Total of observations that are success ($y = 1$)

n_0 = Total of observations that are not success ($y = 0$)

n = Total all of observations

$\hat{\pi}_i$ = The i^{th} estimate probability that is obtained from

Equation 4

G test statistics is Likelihood Ratio Test that follows the Chi-Squared distribution so that by using a significant level of α so that H_0 will be rejected if the value of $G > \chi_{\alpha(p)}^2$

where p is the number of parameters in the model without β_0 (the number of predictor variables) or H_0 will be rejected if the P-value $< \alpha$.

- ii) Test the Significance of Parameters Partially

Test the significance of the parameters partially is used to determine the predictor variables (β coefficients) partially (individually) have an effect on the response variable or not [8]. The hypothesis used in the partial significance test is as follows.

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, \text{ where } j = 1, 2$$

The test statistics used for the parameter significance test partially is shown in Equation 20.

$$W = \frac{\hat{\beta}_j^2}{(SE(\hat{\beta}_j))^2} \tag{20}$$

Explanation

W = Wald test statistics

$\hat{\beta}_j$ = Parameter estimate value of the j^{th} predictor variable

$SE(\hat{\beta}_j)$ = The value of standard error from parameter estimate of the j^{th} predictor variable

Wald test statistics is following the Chi-Squared distribution so that by using a significant level of α then H_0 will be rejected if the value of $W > \chi_{\alpha(p)}^2$ where p is the number of predictor variables or H_0 will be rejected if the P-value $< \alpha$.

3) Goodness of Fit Test

Goodness of fit test is used to determine the model obtained based on simultaneously logistic regression is feasible or not, in other words is between the results of observations and the possibility of the prediction results of the model there are differences or not [8]. Goodness of fit test is carried out by using the Hosmer and Lemeshow test with the hypothesis

H_0 : Model is fit (the results of observations and the possibility of prediction results of the model is not differ significantly)

H_1 : Model is not fit (the results of observations and the possibility of prediction results of the model is differ significantly)

The test statistics used for the goodness of fit test is shown in Equation 21.

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - n'_k \bar{\pi}_k)^2}{n'_k \bar{\pi}_k (1 - \bar{\pi}_k)} \tag{21}$$

Explanation

\hat{C} = Hosmer and Lemeshow test statistics

g = Total of groups

O_k = Observations in the k^{th} group

n'_k = Number of observations in the k^{th} group

$\bar{\pi}_k$ = Average value of estimate probability in the k^{th} group

\hat{C} test statistics is following the Chi-Squared distribution so that by using a significant level of α then H_0 will be rejected if the value of $\hat{C} > \chi_{\alpha(g-2)}^2$ or H_0 will be rejected if P-value $< \alpha$.

4) Odds Ratio

Odds Ratio (OR) is used to interpret the parameter coefficients in logistic regression. OR is a measure that shows the comparison value for the probability of the occurrence of an event with the probability of the un-occurrence of an event. The formula for calculate OR can be shown in Equation 22.

$$OR = \frac{\pi(1) / [1 - \pi(1)]}{\pi(0) / [1 - \pi(0)]} = e^{\beta_j} \tag{22}$$

with $\pi(1) = \frac{e^{\beta_0 + \beta_j}}{1 + e^{\beta_0 + \beta_j}}$ and $\pi(0) = \frac{e^{\beta_0}}{1 + e^{\beta_0}}$

where $j = 1, 2, 3, \dots, p$

If the value of $OR < 1$, it means that there is a negative relationship between response variable and predictor variable in every time the change of the value of the predictor variable. If the value of $OR = 1$, it means that there is no a relationship between response variable and predictor variable. Then if the value of $OR > 1$, it means that there is a positive relationship between response variable and predictor variable in every time the change of the value of the predictor variable [8].

5) Classification Accuracy

Whether or not the classification of data can be known through the accuracy of the model classification. An evaluation that looks at the chance of classifying errors by a classification function is called an evaluation classification procedure. The measure used in this procedure is Apparent Error Rate (APER), that is a value that shows the proportion of samples that was wrongly predicted by the classification function [11]. Table 3 is shown the classification accuracy.

TABLE 3.

CLASSIFICATION ACCURACY

The Result of Observations	Prediction	
	y = 0	y = 1
y = 0	A	B
y = 1	C	D

A is the number of observations with category $y = 0$ that is predicted as category $y = 0$ by the model. B is the number of observations with category $y = 0$ that is predicted as category $y = 1$ by the model. C is the number of observations with category $y = 1$ that is predicted as category $y = 0$ by the model. D is the number of observations with category $y = 1$ that is predicted as category $y = 1$ by the model. So that, the formula for calculate APER can be shown in Equation 23.

$$APER = \frac{B + C}{A + B + C + D} \times 100\% \tag{23}$$

Then, the classification accuracy can be calculated with the formula in Equation 24.

$$\text{Classification Accuracy} = 100\% - APER \tag{24}$$

G. Definition of Stroke

Stroke is defined as a disorder of the nervous system that occurs suddenly and is caused by brain blood circulation disorders. Disorders of the brain's blood circulation can be in the form of blocked blood vessels of the brain or rupture of blood vessels in the brain. The brain that is supposed to get a supply of oxygen and nutrients becomes disturbed. Lack of oxygen supply to the brain will cause the death of nerve cells

(neurons). Impaired brain function will cause stroke disorders [12].

1) Classification of Stroke

Based on anatomic pathology and its causes, stroke is classified into two types, ischemic and hemorrhagic.

i) Ischemic Stroke

Ischemic stroke is a blockage of blood vessels that causes blood flow to the brain partially or completely stopped. Ischemia stroke is generally caused by atherothrombosis of cerebral arteries, both large and small. Blockages that occur in ischemic stroke can be along the pathways of arteries leading to the brain [13].

ii) Hemorrhagic Stroke

Hemorrhagic stroke is caused by bleeding into brain tissue (intracerebrum hemorrhagia or intracerebrum hematoma) or bleeding into the subarachnoid space, that is the narrow space between the surface of the brain and the layer of tissue that covers the brain (subarachnoid hemorrhagia). Hemorrhagic stroke can occur when intracerebrum vascular lesions rupture so that is causing bleeding into the subarachnoid space or directly into brain tissue [14]. Hemorrhagic stroke are rare but if it occurs its usually very severe and deadly [15].

2) Previous Researches

Someone suffering from stroke because they have a risk of stroke. The risk of stroke is something that magnifies a person to suffer a stroke. There are two main groups for risk of stroke. The first group is genetically determined or associated with normal bodily functions so that it can not be modified, such as age, gender, race, family history of stroke, and Transient Ischemic Attack or previous stroke. Then the second group is from someone's lifestyle and can be modified, such as hypertension, diabetes mellitus, hyperlipidemia or high cholesterol, alcohol consumption, and smoking [16]. According to [12], genetic stroke risk includes age, gender, race, family history of stroke, and previous history of stroke. Then the risk of lifestyle stroke includes hypertension, diabetes mellitus, dyslipidemia or cholesterol, smoking, and obesity. According to [17], genetic stroke risks are age, gender, and race. Then the risk of lifestyle stroke are hypertension, heart disease, diabetes mellitus, hyperlipidemia or high cholesterol, smoking, and an unhealthy lifestyle. Based on the risk of stroke, the predictor variables used for genetic stroke risk include age, gender, family history of stroke, and previous history of stroke. Then the risk of lifestyle stroke include hypertension, diabetes mellitus, hyperlipidemia or high cholesterol, and body mass index.

i) Age

Stroke can occurs at any age. The risk of having a stroke increases with the age, it means that how the older a person is, the risk for having a stroke will be easier. The risk will increases after the age of 55 years old [17].

ii) Gender

Stroke attacks men 19% more than women, it means that men have an easier risk for having a stroke than women [17].

iii) Family History of Stroke

The risk of stroke will increase in someone with a family history of stroke. Someone with a family history of stroke is more possible to suffering diabetes mellitus and hypertension. This supports the hypothesis that the increase in the incidence of stroke in families with strokes is a result of the reduced risk of stroke [12].

iv) Previous History of Stroke

Someone who had a stroke will be possible to having a stroke again. Someone who has experienced Transient Ischemic Attack (TIA), nine times more at risk of stroke than those who have never experienced TIA [17].

v) Hypertension

Hypertension increases the risk of stroke 2 to 4 times without depending on the other risks of stroke [18]. A person is said have severe hypertension if the systolic blood pressure is more than 180 mmHg or the diastolic blood pressure is more than 120 mmHg [19].

vi) Diabetes Mellitus

Diabetes mellitus increases the risk of stroke 2 times. An increase in blood sugar levels is directly related to the risk of stroke. A person is said have diabetes mellitus when the blood sugar level (blood sugar level after eating or glucose load) is more than equal to 200 mg/dl [12].

vii) Hyperlipidemia (High Cholesterol)

High blood cholesterol can increases the risk of stroke. A person's fat profile can be determined by Low Density Lipoprotein or LDL cholesterol (bad cholesterol that carries cholesterol from the liver into cells), High Density Lipoprotein or HDL cholesterol (good cholesterol that carries cholesterol from cells into the liver), and triglycerides. Someone said have hyperlipidemia (high cholesterol) if the total cholesterol is more than 200 mg/dl. Total cholesterol is obtained from the sum of LDL cholesterol, HDL cholesterol, and one-fifth of triglycerides [12].

viii) Body Mass Index (BMI)

Someone who is overweight has a high risk for suffering a stroke. Someone who is overweight has a risk of stroke 2,46 times compared to those who are not overweight [20]. BMI is an anthropometric measurement to assess whether a body component complies with normal or ideal standards. BMI is obtained by dividing body weight (kg) by height squared (m^2). Someone said to be overweight if BMI is more than equal to 23 [21].

III. RESULTS AND DISCUSSION

This chapter will explain the results of the analysis and discussion of the modeling of risk factors for total paralysis in stroke patients who are hospitalized at Dr. R. Sosodoro Djatikoesoemo Bojonegoro in 2018 used binary logistic

regression. The explanation of the results of the analysis is as follows.

A. Characteristics of Stroke Patients

Characteristics of stroke patients are explained as follows.

1) Based on Length of Stay

Characteristics of stroke patients based on the length of stay can be shown in Figure 2.

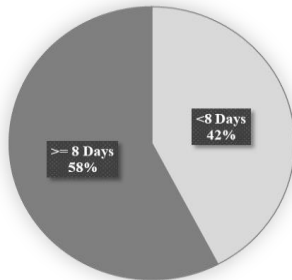


Figure 2. Length of Stay

Figure 2 shows that the majority of stroke patients were hospitalized for more than equal to 8 days. This is shown by percentage of 58%. The remaining for 42% is shown that stroke patients were hospitalized for less than 8 days.

2) Based on Type of Stroke

Characteristics of stroke patients based on the type of stroke can be shown in Figure 3.

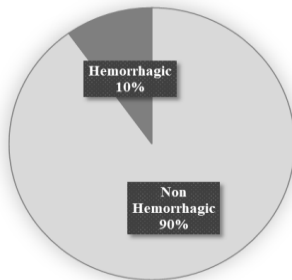


Figure 3. Type of Stroke

Figure 3 shows that the majority of stroke patients have non hemorrhagic type of stroke. This is shown by percentage of 90%. The remaining for 10% is shown that stroke patients have hemorrhagic type of stroke.

3) Based on The Age of Stroke Patients

Characteristics of stroke patients based on the age can be shown in Figure 4.

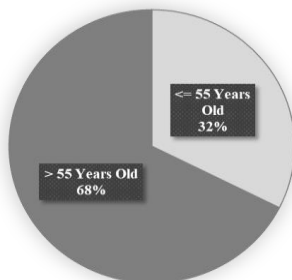


Figure 4. The Age of Stroke Patients

Figure 4 shows that the majority of stroke patients are in more than 55 years old. This is shown by percentage of 68%.

The remaining for 32% is shown that stroke patients are in less than equal to 55 years old.

4) Based on The Gender of Stroke Patients

Characteristics of stroke patients based on the gender can be shown in Figure 5.

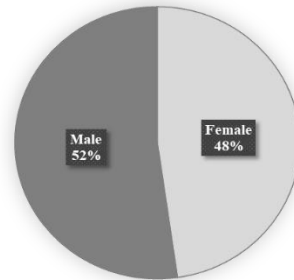


Figure 5. The Gender of Stroke Patients

Figure 5 shows that the majority of stroke patients are males. This is shown by percentage of 52%. The remaining for 48% is shown that stroke patients are females.

5) Based on Family History of Stroke

Characteristics of stroke patients based on the family history of stroke can be shown in Figure 6.

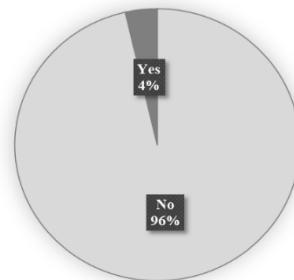


Figure 6. Family History of Stroke

Figure 6 shows that the majority of stroke patients had not the family history of stroke. This is shown by percentage of 96%. The remaining for 4% is shown that stroke patients had the family history of stroke.

6) Based on Previous History of Stroke

Characteristics of stroke patients based on the previous history of stroke can be shown in Figure 7.

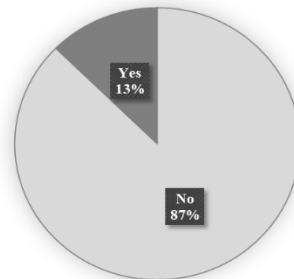


Figure 7. Previous History of Stroke

Figure 7 shows that the majority of stroke patients had not the previous history of stroke. This is shown by percentage of 87%. The remaining for 13% is shown that stroke patients had the previous history of stroke.

7) Based on Severe Hypertension

Characteristics of stroke patients based on the severe hypertension can be shown in Figure 8.

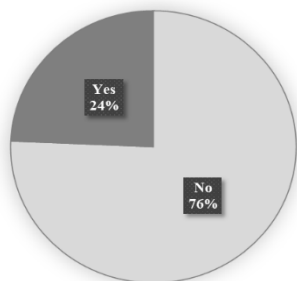


Figure 8. Severe Hypertension

Figure 8 shows that the majority of stroke patients had not the severe hypertension. This is shown by percentage of 76%. The remaining for 24% is shown that stroke patients had the severe hypertension.

8) Based on Diabetes Mellitus

Characteristics of stroke patients based on the diabetes mellitus can be shown in Figure 9.

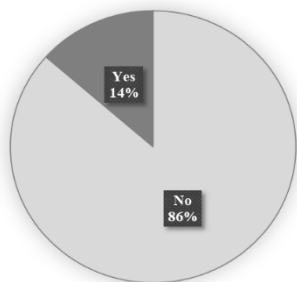


Figure 9. Diabetes Mellitus

Figure 9 shows that the majority of stroke patients had not the diabetes mellitus. This is shown by percentage of 86%. The remaining for 14% is shown that stroke patients had the diabetes mellitus.

9) Based on Hyperlipidemia (High Cholesterol)

Characteristics of stroke patients based on the hyperlipidemia (high cholesterol) can be shown in Figure 10.

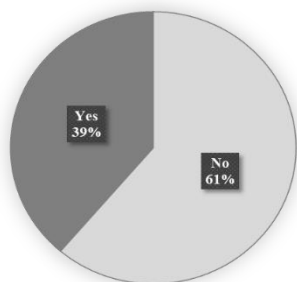


Figure 10. Hyperlipidemia (High Cholesterol)

Figure 10 shows that the majority of stroke patients had not the hyperlipidemia (high cholesterol). This is shown by percentage of 61%. The remaining for 39% is shown that stroke patients had the hyperlipidemia (high cholesterol).

10) Based on Body Mass Index (BMI)

Characteristics of stroke patients based on the BMI can be shown in Figure 11.

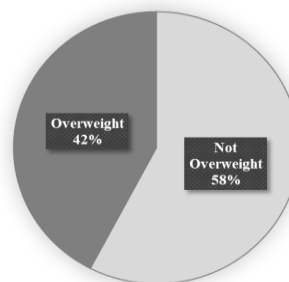


Figure 11. Body Mass Index (BMI)

Figure shows that the majority of stroke patients have the BMI in not overweight category. This is shown by percentage of 58%. The remaining for 42% is shown that stroke patients have the BMI in overweight category.

B. Independence Test for Risk Factors for Total Paralysis in Stroke Patients

Independence test is used to determine that there is or there is no a relationship between response variables, that is the condition of stroke patients with predictor variables such as length of stay, type of stroke, age, gender, family history of stroke, previous history of stroke, severe hypertension, diabetes mellitus, hyperlipidemia (high cholesterol), and BMI.

By using a significant level α of 0,05 then H_0 will be rejected if the value of χ^2 is greater than $\chi^2_{0,05(1)}$ that is 3,841 or if P-value is less than α . The results of the analysis for the independence test are shown in Table 4.

TABLE 4.
INDEPENDENCE TEST

Variable	χ^2	P-value	Decision
Length of Stay (X_1)	3,261	0,071	Failed to reject H_0
Type of Stroke (X_2)	8,268	0,004	Reject H_0
Age (X_3)	0,005	0,944	Failed to reject H_0
Gender (X_4)	1,930	0,165	Failed to reject H_0
Family History of Stroke (X_5)	0,503	0,478	Failed to reject H_0
Previous History of Stroke (X_6)	5,241	0,022	Reject H_0
Severe Hypertension (X_7)	0,593	0,441	Failed to reject H_0
Diabetes Mellitus (X_8)	0,114	0,735	Failed to reject H_0
Hyperlipidemia (High Cholesterol) (X_9)	1,070	0,301	Failed to reject H_0
BMI (X_{10})	3,261	0,071	Failed to reject H_0

Table 4 shows that the decision reject H_0 are in variables of type of stroke (X_2) and previous history of stroke (X_6) that proved by value of χ^2 from each variables that are greater than $\chi^2_{0,05(1)}$ that is 3,841 and strengthened by P-value from each variables that are less than α that is 0,05 so the conclusion is the type of stroke and previous stroke history have a relationship with the condition of stroke patients.

C. Binary Logistic Regression Analysis for Total Paralyzed Risk Factors in Stroke Patients

Binary logistic regression analysis is used to determine the predictor variables that influence the response variable, that is the condition of stroke patients. The predictor variable used were significant variables from the independence test, that are type of stroke and previous stroke history. Explanation of the results of the binary logistic regression analysis is as follows.

6) Estimation of Binary Logistic Regression Parameters

Estimation of binary logistic regression parameters was used to determine the model formed for the total risk factors for total paralysis in stroke patients. This analysis also tested the significance of parameters simultaneously and partially. The explanation of this analysis is as follows.

iii) Test the Significance of Parameters Simultaneously

Test the significance of the parameters simultaneously used to determine the predictor variables (β coefficients) simultaneously have an effect on the response variable (the condition of stroke patients) or not.

By using a significant level α of 0,05 then H_0 will be rejected if the value of G is greater than $\chi^2_{0,05(2)}$ or if P-value is less than α . The results of the analysis for the parameter significance test simultaneously is shown in Table 5.

TABLE 5.
SIMULTANEOUSLY TEST

G	$\chi^2_{0,05(2)}$	P-value
10,827	5,991	0,004

Table 5 shows that based on the parameter significance test simultaneously obtained G value of 10,827 that is greater than the value of $\chi^2_{0,05(2)}$ that is 5,991 and strengthened by P-value of 0,004 that is less than α of 0,05 so the decision is reject H_0 that means there is at least one of predictor variable that has significant effect to the condition of stroke patients.

iv) Test the Significance of Parameters Partially

Test the significance of the parameters partially is used to determine the predictor variables (β coefficients) partially (individually) have an effect on the response variable (the condition of stroke patients) or not.

By using a significant level α of 0,05 then H_0 will be rejected if the value of W is greater than $\chi^2_{0,05(1)}$ or if P-value is less than α . The results of the analysis for the parameter significance test partially is shown in Table 6.

Table 6 shows that based on the parameter significance test partially, the decision to reject H_0 is found in all predictor variables as evidenced by the W value of each variable that are greater than the value of $\chi^2_{0,05(1)}$ and strengthened by the P-value of each variable that are less than α 0,05 so that the conclusion is the type of stroke and previous stroke history significantly have an influence to the condition of stroke patients.

TABLE 6.
PARTIALLY TEST

Variable	$\hat{\beta}$	W	$\chi^2_{0,05(1)}$	P-value	Decision
Type of Stroke (X_2)	2,129	7,495		0,006	Reject H_0
Previous History of Stroke (X_6)	1,761	5,452	3,841	0,020	Reject H_0
Constant	-2,854	37,848		0,000	Reject H_0

v) Binary Logistic Regression Model

Based on the results of the parameter significance test simultaneously and partially it can be seen that the predictor variables that significantly influence to the condition of stroke patients are stroke type and previous stroke history, so the logit model obtained is as follows.

$$g(x) = -2,854 + 2,129X_{2(1)} + 1,761X_{6(1)}$$

7) Goodness of Fit Test for Binary Logistic Regression Model

Goodness of fit test is used to determine the model formed for the risk factors for total paralysis in stroke patients is fit or not.

By using a significant level α of 0,05 then H_0 will be rejected if the value of \hat{C} is greater than $\chi^2_{0,05(1)}$ or if P-value is less than α . The results of the goodness of fit test is shown in Table 7.

TABLE 7.
GOODNESS OF FIT TEST

\hat{C}	$\chi^2_{0,05(1)}$	P-value
0,043	3,841	0,835

Table 7 shows that based on the goodness of fit test, obtained \hat{C} value of 0,043 that is less than the value of $\chi^2_{0,05(1)}$ that is 3,841 and strengthened by P-value of 0,835 that is greater than α of 0,05 so the decision is failed to reject H_0 that means model is fit (the results of observations and the possibility of prediction results of the model is not differ significantly).

8) Interpretation of the Binary Logistic Regression Model

Interpretation of binary logistic regression models for risk factors for total paralysis in stroke patients by using the odds ratio. The odds ratio results are shown in Table 8.

TABLE 8.
THE RESULTS OF ODDS RATIO

Variable	Odds Ratio
Type of Stroke (X_2)	8,406
Previous History of Stroke (X_6)	5,816

Table 8 shows the odds ratio that can be explained as follows:

- a. Stroke patients with hemorrhagic stroke types tend to experience total paralysis by 8,406 times greater than stroke patients with non-hemorrhagic stroke types.

b. Stroke patients who have previous history of stroke tend to experience total paralysis by 5,816 times greater than stroke patients who do not have previous history of stroke.

9) Checking the Accuracy of the Binary Logistic Regression Classification Model

The classification accuracy check is used to find out whether or not the classification of observations made. The classification accuracy check for total paralysis risk factors in stroke patients based on significant predictor variables in the partially test that are the type of stroke (X_2) and previous history of stroke (X_6) can be shown in the following Table 9.

TABLE 9.

THE CLASSIFICATION ACCURACY CHECK		
The Results of Observations	Prediction	
	Not Totally Paralyzed	Totally Paralyzed
Not Totally Paralyzed	99	0
Totally Paralyzed	11	1

Table 9 shows the results of the classification accuracy check that is equal to 99 stroke patients with total paralysis condition precisely classified as total paralysis by the model, 11 stroke patients with total paralysis condition are classified as total paralysis by the model, and 1 stroke patient with total paralysis conditions precisely classified as total paralysis by the model. Based on these results the classification accuracy can be calculated as follows:

$$APER = \frac{0+11}{99+0+11+1} \times 100\% = 9,91\%$$

so that,

$$Classification Accuracy = 100\% - 9,91\% = 90,09\%$$

Based on the calculations, the accuracy of the model classification is 90,09%, it means that the condition of stroke patients who are appropriately classified by the model is 90,09%.

IV. CONCLUSION

The conclusions from the results of the analysis and discussion are as follows:

- 1) Most stroke patients are hospitalized for more than the same as 8 days, have a type of non-hemorrhagic stroke, over 55 years old, male, do not have a family history of stroke and previous history of stroke, do not have severe hypertension, diabetes mellitus, hyperlipidemia (high cholesterol), and being overweight.
- 2) Risk factors that cause stroke patients to experience total paralysis are stroke type and previous stroke history where patients with hemorrhagic stroke type tend to be at risk of total paralysis 8,406 times greater than patients with non-hemorrhagic stroke types and patients with previous stroke history tend to be at risk total paralysis 5,816 times greater than patients who have no previous stroke history.

REFERENCES

[1] Riskesdas, "Riset Kesehatan Dasar Riskesdas 2013", Accessed on 11th September 2018 from www.depkes.go.id, 2013.

[2] L. Lingga, "All About Stroke Hidup Sebelum dan Pasca Stroke" Jakarta: PT Elex Media Komputindo, 2013.

[3] Z. A. Aulia, "Pemodelan Kasus Stroke Berdasarkan Jenisnya Menggunakan Analisis Regresi Logistik Biner di Rumah Sakit Umum Haji Surabaya", Accessed on 20th September 2018 from repository.its.ac.id, 2017.

[4] S. P. Kristiyawati, "Analisis Faktor Risiko yang Berhubungan dengan Kejadian Stroke di Rumah Sakit Panti Wilasa Citarum Semarang", Accessed on 22nd September 2018 from http://lib.ui.ac.id, 2008.

[5] Amran, "Analisis Faktor Risiko Kematian Penderita Stroke", Accessed on 7th November 2018 from https://amranepid.files.wordpress.com, 2012.

[6] L. Mihardja, et al, "Angka Kematian dan Faktor Risiko Stroke Sebagai Penyebab Dasar Kematian di Kabupaten Padang Pariaman Provinsi Sumatera Barat", Accessed on 6th December 2018 from https://media.neliti.com, 2016.

[7] Pemkab Bojonegoro, "10 Penyakit Pembunuh di Bojonegoro", Accessed on 11th September 2018 from bojonegorokab.go.id, 2015.

[8] D. W. Hosmer, and S. Lemeshow, "Applied Logistic Regression Second Edition", Canada: John Wiley & Sons Inc, 2000.

[9] J. J. Parsons, D. Oja, R. Ageloff, and P. Carey, "New Perspectives on Microsoft[®] Excel[®] 2010 Introductory", Boston: Course Technology Cengage Learning, 2011.

[10] A. Agresti, "An Introduction to Categorical Data Analysis Second Edition", Hoboken New Jersey: John Wiley & Sons Inc, 2007.

[11] R. A. Johnson and D. W. Wichern, "Applied Multivariate Statistical Analysis Sixth Edition", New Jersey: Pearson Education Inc, 2007.

[12] R. Pinzon and L. Asanti, "Awat Stroke! Pengertian, Gejala, Tindakan, Perawatan, & Pencegahan", Yogyakarta: ANDI, 2010.

[13] American Stroke Association (ASA), "Types of Stroke", Accessed from http://www.strokeassociation.org, 2011.

[14] V. Felgin, "Stroke", Jakarta: PT Bhuana Ilmu Populer, 2006.

[15] A. Krishna, "Mengenal Keluhan Anda Info Kesehatan Umum untuk Masyarakat", Accessed on 7 November 2018 from www.informasimedika.com, 2013.

[16] Y. Yueniawati, "Pencitraan pada Stroke". Malang: UB Press, 2016.

[17] W. Sari, et al, "Care Your Self Stroke Cegah dan Obati Sendiri", Jakarta: Penebar Plus (Penebar Swadaya Grup), 2016.

[18] Perreu and Bogusslavsky, In: "Awat Stroke! Pengertian, Gejala, Tindakan, Perawatan, & Pencegahan", Yogyakarta: ANDI, 2003.

[19] American Heart Association (AHA), "Understanding Blood Pressure Reading", Accessed on 24th April 2019 from https://www.heart.org, 2019.

[20] Oki, et al, In: "Awat Stroke! Pengertian, Gejala, Tindakan, Perawatan, & Pencegahan", Yogyakarta: ANDI, 2006.

[21] T. Sudargo, et al, "Pola Makan dan Obesitas", Yogyakarta: Gadjah Mada University Press, 2014.