Cloud-Based Multinomial Logistic Regression

by Suryono Suryono

Cloud-Based Multinomial Logistic Regression for Analyzing Maternal Mortality Data in Postpartum Period

Radite Purwahana
Master of Information System
Diponegoro University
Semarang, Indonesia
raditepurwahana16@gmail.com

Suryono Suryono
Physics Departement of Science and Mathematics
Diponegoro University
Semarang, Indonesia
suryono@fisika.undip.ac.id

Jatmiko Endro Suseno
Physics Departement of Science and Mathematics
Diponegoro University
Semarang, Indonesia
jatmikoendro@gmail.com

***Abstract*— The analysis used in dealing with maternal mortality factors in the postpartum period can be used as a reference in preventing maternal death in the postpartum period. Appropriate analysis is needed to reduce maternal mortality rates in the postpartum period. This study uses multinomial logistic regression to analyze the data of mothers dying in the postpartum period based on the main variables causing maternal death. Multinomial logistic regression process is carried out by looking at data records of variables that influence maternal mortality. In the first experiment using data from midwife visits for seven days, the results of the multinomial logistic regression process with the highest maternal mortality occurred on the fourth day with anogenital variables reaching a percentage of 32.4% of the causes of maternal death. Multinomial logistic regression processes are combined with cloud computing technology so that data can be processed more quickly and can be used together.**

Keywords—Cloud Computing, Multinomial Logistic Regression, Postpartum, Maternal Mortality

I. Introduction

Postpartum is the most vulnerable time when the recovery process of the mother's womb back to normal so that it needs regular monitoring by medical personnel, the puerperium usually lasts six weeks or forty days [1].

Maternal Mortality Rates (MMR) is a scale used to determine the number of maternal deaths. As per 2013 statistics, maternal mortality rates amounted to 289,000 people worldwide each year. This statistic has shown an increase of 2000 cases more than in 2010 which is 287,000 worldwide. Sub-Saharan Africa contributes 62% (179,000) of deaths worldwide in 2013. This statistic shows a substantial increase of 17,000 cases in the region compared to 2010 statistics [2]. Therefore, a system that can monitor the development of the maternal mortality index in the postpartum period is needed, so that it can more quickly show the results of any policies carried out by the government to succeed or not to suppress that number [1].

At present, the technology used to reduce the risk of maternal death in the postpartum period uses microcontroller technology, combined with the Least Mean Square (LMS-based algorithm) to detect the patient's heart rate in real-time. This system has a disadvantage because it sees maternal mortality based on one variable, namely heart rate [3].

The application of logistic regression methods is very relevant in measuring the relationship between categorical dependent variables and is usually a continuous independent variable by changing the dependent variable into a probability value. Multinomial logistic regression refers to cases where the results can have three or more possible types. Multinomial logistic regression method is used if the response variable is categorical (nominal or ordinal) with predictive variables that are continuous or categorical [4].

In this study proposing cloud computing technology as a new technology in business networks and models, this technology has succeeded in attracting many domestic and foreign enthusiasts including various fields of computer science and technology. Cloud-based systems can provide a better model and provide convenience in resource sharing, development, and speed in processing data [5].

Seeing the utilization of technology that has not been maximized in managing maternal data died during the postpartum period, in this study a multinomial logistic regression cloud computing system will be built to analyze data on mothers dying during postpartum.

II. Proposed Method

A. *Tools and Materials*

The material used in this study is the postpartum maternal examination visit data conducted by the midwife for seven consecutive days then the data is stored in a database. After the data is stored in the database then the data is grouped based on which data needs to be analyzed to do which factor classification influences maternal death using data analysis method of the multinomial logistic regression category.

Data were obtained from Bhakti Bunda clinic health service data spread across several cities in Central Java as a sample of the health service population of all cities in Central Java Province for the period of 2017-2018. Monitoring focus using data contained in the examination of vital signs in the form of blood pressure, respiration rate, pulse, temperature, eyes, abdomen, extremities, and anogenitalia.

B. *Research Stages*

To achieve the research objectives, there are several steps that have been carried out. Beginning with the business process carried out in this study as follows:

- Data of maternal examination visits in the postpartum period were used as inputs to be

978-1-5386-7440-6/18/$31.00 ©2018 IEEE

analyzed using multinomial logistic regression methods.

- Determine the initial model, likelihood ratio test, the final model, and the estimated opportunity,

For business processes in the maternal death system during postpartum. We design in the form of the information system framework shown in Figure 1:

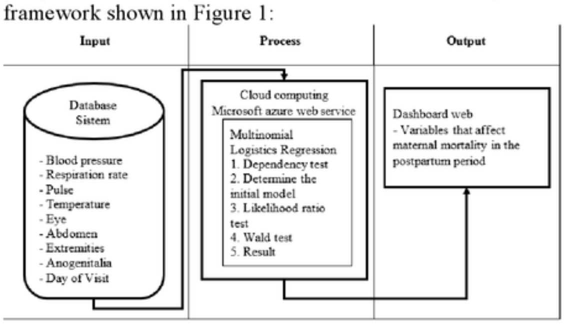

Fig. 1. System Framework Maternal Mortality

1) Logistic Regression Method

Multinomial logistic regression models are used to analyze the relationship between one response variable in the form of a death index and several independent variables. For example "x" the independent variable size "(p + 1)" and the response variable "Y (r category)" has the categories "j = 0, 1, 2, ..., r-1" with the probability of response "π1, π2, ..., πr-1 "and "$\sum_{j=1}^{r-1} \pi_j = 1$" [6]. Conditional probability:

$$P(y = j|x) = \pi j(x), j = 1, 2, \ldots, r - 1 \quad (1)$$

The logistic regression model nominal r category has r-1 logit function if taken "y = 0" as the base category, then the logit function is owned:

$$g_1\,(x) = \ln\left[\frac{P(y=1|x)}{P(y=0|x)}\right] = \ln\left[\frac{\pi_1(x)}{\pi_0\,(x)}\right] \quad (2)$$

....

$$g_{r-1}(x)\,(x) = \ln\left[\frac{P(y=r-1|x)}{P(y=0|x)}\right] = \ln\left[\frac{\pi_{r-1}(x)}{\pi_0\,(x)}\right]$$

From conditional probability "P(y=j|x)=π_j(x), j=1, 2, ..., r-1" can be written:

$$\pi_0 = \frac{1}{1 + e^{g_1(x)} + e^{g_2(x)} + \ldots + e^{g_{r-1}(x)}} \quad (3)$$

$$\pi_1 = \frac{e^{g_1(x)}}{1 + e^{g_1(x)} + e^{g_2(x)} + \ldots + e^{g_{r-1}(x)}}$$

$$\pi_2 = \frac{e^{g_2(x)}}{1 + e^{g_1(x)} + e^{g_2(x)} + \ldots + e^{g_{r-1}(x)}}$$

....

$$\pi_{r-1} = \frac{e^{g_{r-1}(x)}}{1 + e^{g_1(x)} + e^{g_2(x)} + \ldots + e^{g_{r-1}(x)}}$$

With:

$$g_1(x) = \beta_{10} + \beta_{11}x_1 + \ldots + \beta_{1p}x_p \quad (4)$$

$$g_1(x) = \beta_{10} + \beta_{11}x_1 + \ldots + \beta_{1p}x_p$$

....

$$g_{r-1}(x) = \beta_{(r-1)0} + \beta_{(r-1)1}x_1 + \ldots + \beta_{(r-1)p}x_p$$

Parameter estimation in multinomial logistic regression analysis is used to determine the initial model, parameter estimation is used maximum likelihood method with Newton raphson iteration which requires the first partial derivative and the second partial derivative of the likelihood function.

The likelihood function for a sample of "n" observations independently is:

$$l(\beta) = \prod_{i=1}^{n} \pi_o(x_1)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}} \quad (5)$$

Then the first partial derivative is carried out so that the general shape is obtained:

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial} = \sum_{i=1}^{n} x_{ki}(y_{ji} - \pi_{ji}) \quad (6)$$

Followed by the second partial derivative.

$$\frac{\partial^2 L(\beta)}{\partial \beta_{jk} \partial \beta_{j'k'}} = \sum_{i=1}^{n} x_{k'i} x_{ki} \pi_{ji} \pi_{j'i} \quad (7)$$

With Newton Raphson's iteration method, the initial estimation is selected for "β" such as "$\hat{\beta} = 0$" then calculated "M (Y-π(x))" and "$\hat{I}(\hat{\beta})$", then calculated inverse "$\hat{I}(\hat{\beta})$" where each "i+1" is calculated as a new estimate:

$$\hat{\beta}^{(g+1)} = \hat{\beta}^{(g)} + \hat{I}(\hat{\beta})^{-1}\{M(Y - \pi(x))\} \quad (8)$$

The iteration ends when it gets:

$$\hat{\beta}^{(g+1)} \cong \hat{\beta}^{(g)} \quad (9)$$

It is a test by comparing a model containing "$\hat{\beta}^{(g+1)}$" as an independent variable and "$\hat{\beta}^{(g)}$" as a model that does not contain independent variables. This test is also used to analyze whether the significance model or model does not match the hypothesis.

H_0: $\beta_{1k} = \beta_{2k} = \ldots = \beta_{(r-1)p} = 0$
H_1: one of $\beta_{jk} \neq 0$, with j=1, 2, ..., r-1, k= 1, 2, ..., p

Likelihood statistical test.

$$G = -2\ln\left[\frac{\text{likelihood without independent variables}}{\text{likelihood with independent variables}}\right] \quad (10)$$

Test criteria: reject "H_0" when "G > $X^2_{(\alpha,v)}$" with "$v = (c-1)\sum_{k=1}^{p}(J_k - 1)$".

Wald test is used to find out which independent variables have a stronger relationship with the response variable.

Hypothesis: H_0: $\beta_{jk}= 0$
H_1: $\beta_{jk} \neq 0$ with j= 1, 2, ..., r-1, k=1, 2, ..., p

The significance level is "α".

$$\text{Test statistics: } Wjk = \left\{\frac{\hat{\beta}_{jk}}{se(\hat{\beta}_{jk})}\right\}^2 \quad (11)$$

Test criteria: reject H_0 when $W_{jk} >> X^2_{(\alpha,1)}$

The odds ratio is defined as the ratio of a successful variable value to a variable that has failed. In other words, the odds ratio explains how much influence the success variable is compared to the failed variable on an experiment or observation. In the case of research with multinomial logistic regression, this value can be seen from the value of Exp (B) on the results of data analysis contained in SPSS 16 output. These results will show the effect of each independent variable on the dependent variable.
The odds ratio is defined as:

$$\theta = \frac{\Omega}{\Omega} = \frac{\pi_{1|1} - \pi_1}{\pi_{2|2} - \pi_2} = \frac{\pi_1}{\pi_2} \times \frac{1 - \pi_2}{1 - \pi_1} \quad (12)$$

The odds ratio can be interpreted as a relative risk when:

$$\frac{1 - \pi_2}{1 - \pi_1} \approx 1 \quad (13)$$

2) Cloud Computing Maternal Mortality

Cloud computing is an Internet-based technology and tries to disguise its complexity for clients. Cloud providers use virtualization technology combined with self-service capabilities to calculate resources through the network infrastructure, especially the internet. Cloud has several VMs (VMs) hosted as infrastructure. In the cloud, customers only need to pay for what they use. Cloud computing refers to two applications provided as services, namely through the internet and system hardware and software in the data center that provides these services [7].

The cloud computing architecture in this study begins by including maternal mortality data in the postpartum period into the cloud database to be analyzed using multinomial logistic regression methods. The cloud computing system scheme in this study can be seen in Figure 2.

III. Result and Discussion

The results of the data calculation of the mother died in the postpartum period using the data obtained at the midwife's visit for seven consecutive days, can be seen in Table I.

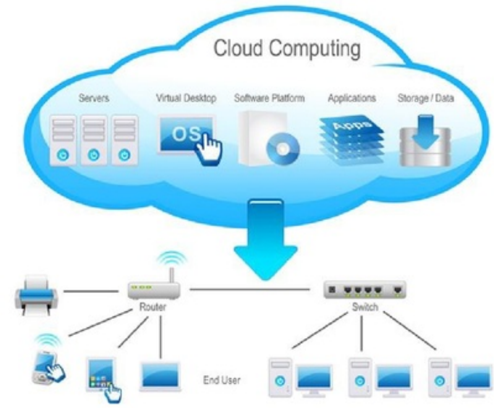

Fig. 2. Architecture Cloud Computing

TABLE I. VISIT DATA

Condition	Vital Signs	H1	H2	H3	H4	H5	H6	H7
Normal	blood pressure	4	5	3	2	5	3	5
	respiration rate	3	5	4	5	2	5	3
	pulse	2	2	3	5	2	5	2
	body temperature	3	1	2	2	5	2	5
	eye	2	3	5	1	4	1	4
	abdomen	5	2	2	3	3	3	3
	ekstemitas	2	5	5	2	3	5	3
	anogenitalia	1	3	2	5	2	2	2
Abnormal	blood pressure	3	2	1	3	5	1	5
	respiration rate	2	1	3	2	2	3	2
	pulse	5	3	2	5	2	5	1
	body temperature	4	1	5	2	5	2	3
	eye	3	3	3	1	4	1	5
	abdomen	2	5	2	3	3	3	2
	ekstemitas	2	2	2	2	3	5	1
	anogenitalia	1	1	1	5	2	2	3

A. *Opportunity Estimation Analysis*

The results of the seven-day midwife visit data were used as input for processing using multinomial logistic regression methods. Obtained opportunity estimates and exponential estimates of opportunities. The estimated opportunities can be seen in Table II.

TABLE II. ESTIMATION OF OPPORTUNITIES

Variable	P1	P2	P3	P4	P5	P6
	-0.815	-0.292	-0.565	0.762	-0.121	-0.179
VS1	0.568	-0.139	-0.41	-1.381	0.232	-0.69
VS2	0.937	0.393	0.837	-0.343	0.021	0.702
VS3	1.754	0.74	1.02	0.504	0.501	1.423
VS4	0.773	-1.157	0.382	-1.393	0.436	-0.474
VS5	0.321	-0.178	0.397	-2.202	0.098	-1.284
VS6	1.249	0.562	0.29	-0.513	0.402	0.404
VS7	0.928	0.776	1.065	-0.685	0.64	1.144

Furthermore, an exponential value of opportunity estimates is needed to be used in the analysis of opportunity estimates. Exponential estimation of opportunities can be seen in Table III.

TABLE III. EXPONENTIAL ESTIMATION OF OPPORTUNITIES

Variable	Exp g1(x)	Exp g2(x)	Exp g3(x)	Exp g4(x)	Exp g5(x)	Exp g6(x)
blood pressure	0.781	0.650	0.377	0.538	1.117	0.419
respiration rate	1.130	1.106	1.313	1.520	0.905	1.687
pulse	2.557	1.565	1.576	3.547	1.462	3.469
body temperature	0.959	0.235	0.833	0.532	1.370	0.520
eye	0.610	0.625	0.845	0.237	0.977	0.232
abdomen	1.543	1.310	0.760	1.283	1.324	1.252
ekstemitas	1.120	1.623	1.649	1.080	1.680	2.625
anogenital	0.443	0.747	0.568	2.143	0.886	0.836

B. *Analysis of Maternal Mortality Factors in the Postpartum Period*

The main factors that influence maternal death can be seen from the eight main factors in the form of blood pressure, respiration rate, pulse, temperature, eyes, abdomen, extremities, and anogenitalia. Using the multinomial logistic regression method can be seen the highest factor of maternal death in the postpartum period for seven days.

The experiment used data from midwife visits during the seven days obtained results, namely the first day of the highest percentage of death in abdomen variables with a percentage of 18,2%, for the second day the highest mortality was in the variable of abdomen with a percentage of 15,5% and ekstemitas with a percentage of 15,1%, than the third day are eye with a percentage of 18,7%, for the fourth day the highest mortality was shown in the variable anogenital with a percentage of 32,4%, for the fifth day the highest death was shown in the variable body temperature with a percentage of 25,1%, for the sixth day the highest death was shown in the variable ekstemitas with a percentage of 24,4% and the seventh day of death was the highest in the eye variable with a percentage of 22,1%. The percentage can be seen in Tables IV.

From the results of the study, in analyzing the data of mothers died during postpartum using cloud computing technology. Cloud computing is able to produce software systems that are able to process data more quickly in analyzing the influence of variables that appear in the data analysis of mothers dying during postpartum using multinomial logistic regression methods.

IV. Conclusion

From the results of the study in analyzing the data of mothers dying in the postpartum period, the results showed that the patient's condition had no effect on the death of the mother during postpartum because the mother who died in the postpartum period was not necessarily in an abnormal condition but the effect was eight variables of vital signs. namely the first day of the highest percentage of death in abdomen variables with a percentage of 18,2%, for the second day the highest mortality was in the variable of abdomen with a percentage of 15,5% and ekstemitas with a percentage of 15,1%, than the third day are eye with a percentage of 18,7%, for the fourth day the highest mortality was shown in the variable anogenital with a percentage of 32,4%, for the fifth day the highest death was shown in the variable body temperature with a percentage of 25,1%, for the sixth day the highest death was shown in the variable ekstemitas with a percentage of 24,4% and the seventh day of death was the highest in the eye variable with a percentage of 22,1%.

TABLE IV. MOTHER'S DEATH ESTIMATION IN POSTPARTUM PERIOD

Vital Signs	D1	D2	D3	D4	D5	D6	D7
blood pressure	0.160	0.133	0.077	0.110	0.229	0.086	0.205
respiration rate	0.130	0.128	0.152	0.176	0.104	0.195	0.115
pulse	0.169	0.103	0.104	0.234	0.096	0.229	0.066
body temperature	0.176	0.043	0.153	0.098	0.251	0.096	0.184
eye	0.135	0.138	0.187	0.052	0.216	0.051	0.221
abdomen	0.182	0.155	0.090	0.151	0.156	0.148	0.118
ekstemitas	0.104	0.151	0.153	0.100	0.156	0.244	0.093
anogenitalia	0.067	0.113	0.086	0.324	0.134	0.126	0.151

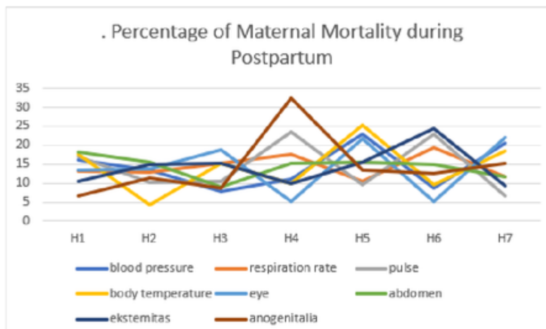

Fig. 3. Percentage of Maternal Mortality during Postpartum

References

[1] N. Sivadas, A. Radhakrishnan, B. S. Aswathy, V. M. Kumar, and K. K. Gulia, "Dynamic changes in sleep pattern during post-partum in normal pregnancy in rat model," Behav. Brain Res., vol. 320, pp. 264–274, 2017.

[2] V. Mukami, R. Millham, and T. Puckree, "Comparison of frameworks and models for analyzing determinants of maternal mortality and morbidity," 2016 IST-Africa Conf. IST-Africa 2016, no. Maternal Mortality, 2016.

[3] R. Arias-Ortega, M. J. Gaitán-González, and O. Yáñez-Suarez, "Implementation of a real-time algorithm for maternal and fetal heart rate monitoring in a digital signal controller platform.," Conf. Proc. IEEE Eng. Med. Biol. Soc., vol. 2010, pp. 2354–7, 2010.

[4] E. Scholar, "Comparative Analysis using Multinomial Logistic Regression," pp. 119–124, 2014.

[5] L. Wei, Z. Qianyuan, Y. Xiudi, and Z. Yayun, "The Analysis and Implementation of Index Management System Based on Cloud Computing," 2012 Fourth Int. Conf. Comput. Inf. Sci., pp. 392–395, 2012.

[6] D. W. Hosmer and S. Lemeshow, "Applied Logistic Regression," Wiley Series in Probability and Sattistics, no. 1. p. 373, 2000.

[7] F. Sabahi, "Virtualization-level security in cloud computing," 2011 IEEE 3rd Int. Conf. Commun. Softw. Networks, pp. 250–254, 2011.

Cloud-Based Multinomial Logistic Regression

ORIGINALITY REPORT

4%	2%	2%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	icicos.org Internet Source	1%
2	Submitted to Cardiff University Student Paper	1%
3	Submitted to Manipal University Student Paper	1%
4	Ralph Blumenhagen. "Orientifolds of K3 and Calabi-Yau Manifolds with Intersecting D-branes", Journal of High Energy Physics, 07/08/2002 Publication	<1%
5	www.nyu.edu Internet Source	<1%
6	Submitted to University of Lancaster Student Paper	<1%
7	Jeppe Olsen. "Optimization and Characterization of a Multiconfigurational Self-Consistent Field (MCSCF) State", Advances in Chemical Physics, 01/01/1983	<1%

Publication

8	Submitted to University of Technology, Sydney Student Paper	<1%
9	F. Ponti, J.C. Piani, R. Suglia. "Residual Gas Model for On-Line Estimation for Inlet and Exhaust Continuous VVT Engine Configuration", IFAC Proceedings Volumes, 2004 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		

Cloud-Based Multinomial Logistic Regression

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4