VOL. 15, NO. 20, OCTOBER 2020 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

THE RISK MODELING OF DIABETES BASED ON PARAMETRIC AND NONPARAMETRIC BINARY LOGISTIC REGRESSION

Suliyanto and Marisa Rifada
Department of Mathematics, Airlangga University, Surabaya, Indonesia
E-Mail: marisa.rifada@fst.unair.ac.id

ABSTRACT

The parametric binary logistic regression assumes that the logit function is known to be expressed as a linear function in the parameter, while the nonparametric binary logistic regression assumes that the logit function is unknown and can be approximated by the Generalized Additive Model (GAM) or Local Likelihood Logit Estimation (LLLE) method. The GAM method assumes that the logit function is the sum of the nonparametric regression functions of each predictor variable with the known link function. The LLLE method assumes that the logit function is a linear function in the parameter, where the parameters depend on arbitrary fixed points and the likelihood logit function depends on the multivariate kernel weighting. In this study we compared the risk prediction of diabetes based on three approaches, i.e parametric binary logistic regression, nonparametric binary logistic regression using the GAM method, and nonparametric binary logistic regression with the LLLE method. The results of classification accuracy in risk prediction of diabetes using the parametric binary logistic regression approach of 80.2%, the GAM method of 88.89%, and the LLLE method of 100%. So, the best approach model is obtained by nonparametric binary logistic regression with the LLLE method.

Keywords: diabetes, GAM, local likelihood logit estimation, nonparametric binary logistic regression, parametric binary logistic regression.

1. INTRODUCTION

Diabetes is one of the most serious chronic diseases. Diabetes can occur when the pancreas doesn't produce enough insulin (a hormone used to regulate blood sugar), or when the body can not effectively use the insulin produced by the body itself (Basar, *et al*, 2015) and (Prayitno *et al*, 2016). High blood glucose causes diabetes is not easily controlled so it will cause some of the consequences that can occur that include causing serious damage to the heart, blood vessels, eyes, kidneys, and neurological disorders. Diabetes is a major cause of some diseases that attack the body, and can cause death (George, *et al*, 2013), (Goyal and Gafazzo, 2013), and (Ojugo, *et al*, 2015).

The global incidence of type II diabetes mellitus is increasing rapidly. The World Health Organization predicts that the number of people with type II diabetes mellitus will double to at least 350 million worldwide by 2030 unless appropriate action is taken (WHO, 2003). The model of the relationship between the risk of diabetes incidence in patients and the factors thought to influence it will be more realistic if formulated mathematically to see how much influence these factors have on the chances of patients affected by Diabetes. The statistical method that is suitable for analyzing the relationship between categorical scale response variables and categorical and continuous scale predictor variables is logistic regression. Binary logistic regression is a regression with dichotomous response variables consisting of two categories (Agresti, 2002). There are two ways to approach the logistic regression model, i.e. the parametric approach and the nonparameteric approach. The parametric approach assumes that the regression model for each individual observation has the same parameters, whereas the nonparametric approach assumes that not all individuals have the same parameters.

Diabetes Mellitus research in patients treated in hospitals using logistic regression was conducted by (Rahman, *et al*, 2013) to determine risk factors and early detection of diabetes events and concluded that risk factors associated with the incidence of diabetes were general obesity, central obesity, vegetable consumption and fruit, smoking and stress. Trisnawati and Setyogoro (Trisnawati, *et al*, 2013) identified risk factors for the incidence of type II Diabetes Mellitus in the Cengkareng District Health Center in West Jakarta using a binary logistic regression model and the results showed that the risk factors affecting type II diabetes mellitus were age, family history, physical activity, body mass index blood pressure, stress and cholesterol levels. Research conducted by (Rahman, *et al*, 2013) and (Trisnawati, *et al*, 2013) in the case of Diabetes Mellitus using binary logistic regression with a parametric approach. In general, the physical condition of each patient is different from one another, this allows the characteristics of each patient to be different so that the parametric binary logistic regression approach is less suitable to be applied. In this regard, the more suitable approach is nonparametric binary logistic regression. In the nonparametric binary logistic regression model known as the nonparametric regression model with binary response variables to estimate the model used the Generalized Additive Models (GAM) method (Hastie and Tibshirani, 1990) or local likelihood logit estimation (Frölich, 2006). The GAM method assumes that the regression function is the sum of unknown nonparametric regression functions of each component of the predictor variable with the known link function. The local likelihood logit estimation method assumes that the logit function is a linear function in the parameter, where the parameters depend on arbitrary fixed points and the likelihood logit function depends on the multivariate kernel weighting.

Suliyanto *et al.* (2018) have discussed the estimation of nonparametric binary logistic regression models based on kernel estimators using the GAM method. Modeling the case of Diabetes Mellitus patients using a nonparametric binary logistic regression approach using two methods, namely Generalized Additive Model (GAM) and Local Likelihood Logit Estimation (LLLE) has been discussed by (Suliyanto and Rifada, 2019). Rifada, *et al* (2018) conducted a logistic regression analysis based on the local scoring algorithm with case study of Type II Diabetes Mellitus in Surabaya Indonesia and has results of classification accuracy of 88.89%. In this study we discusses the modeling of cases of Type II Diabetes Mellitus patients in Surabaya Haji Hospital using three approaches, i.e. parametric binary logistic regression, nonparametric binary logistic regression with the GAM method, and nonparametric binary logistic regression with the local likelihood logit estimation method.

2. MATERIALS AND METHODS

2.1 Parametric Binary Logistic Regression

Binary logistic regression model is a regression model with categorical Y response variables consisting of two categories, namely $Y = \{0, 1\}$ and the success probability of experimental results is based on logistical distribution, while the predictor variables are continuous and categorical (Hosmer, *et. al*, 2013). The binary logit model is a binary logistic regression model with the logit link function, i.e.

$$G(X_i) = \mathrm{X_i}\beta\,; i = 1, 2, \ldots, n \tag{1}$$

where $G(\boldsymbol{X_i}) = \ln\left[\frac{\pi(X_i)}{1-\pi(X_i)}\right]$ is a logit link function, $\boldsymbol{X_i} = (1, X_{1i}, X_{2i}, \ldots, X_{pi})$ is a vector from the predictor variable of observation to i, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ is a parameter vector that corresponds to the predictor variable..

2.2 Nonparametric Binary Logistic Regression with GAM Method

The nonparametric regression model estimation using the kernel function is called the Nadaraya-Watson estimator which was introduced by (Staniswalis, 1989).The Generalized Additive Models (GAM) method is one method of the nonparametric regression approach introduced by (Hastie and Tibshirani, 1990). The GAM method relating to nonparametric regression assumes that the nonparametric regression function is expressed as the sum of nonparametric regression functions of each component x whose form is unknown with the link G function known, i.e.:

$$\mathrm{G}\big(\mathrm{E}(\mathrm{Y}|\mathrm{X} = \mathrm{x})\big) = \textstyle\sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{f_j}(\mathrm{x_j}) \tag{2}$$

where $f_j(x_j)$ is a nonparametric regression function of the predictor variable x_j.

2.3 Nonparametric Binary Logistic Regression with LLLE Method

Independent samples $(Y_i, \boldsymbol{X_i})\,; i = 1, 2, \ldots, n$, with $Y_i = \{0, 1\}$ and $\boldsymbol{X_i} = (1, X_{1i}, X_{2i}, \ldots, X_{pi})$ assumed to meet the nonparametric binary logistic regression model as follows:

$$\mathrm{G}(\mathrm{X_i}, \mathrm{x_0}) = \mathrm{X_i}\ \beta(\mathrm{x_0})\,; \mathrm{i} = 1, 2, \ldots, \mathrm{n} \tag{3}$$

where $G(\boldsymbol{X_i}, \mathbf{x_0}) = \ln\left[\frac{\pi(X_i\,,\mathbf{x_0})}{1-\pi(X_i\,,\mathbf{x_0})}\right]$ is the logit link function, $\boldsymbol{\beta}(\mathbf{x_0})$ is a parameter vector that corresponds to the predictor variable. According to (Frölich, 2006) to estimate $\boldsymbol{\beta}(\mathbf{x_0})$ in equation (5) uses the local likelihood logit estimation method.

2.4 Local Likelihood Estimation

According to (Frölich, 2006), in general the method of local likelihood estimation is to obtain $\widehat{\boldsymbol{\beta}}$ so that

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\arg\max} \textstyle\sum_{i=1}^{n} \ln f(Y_i\,, g(\boldsymbol{X_i}\,,\boldsymbol{\beta}))\ K_h(\boldsymbol{X_i} - \mathbf{x}) \tag{4}$$

where $\ln f(Y_i\,, g(X_i, \beta))$ is log-pdf from observation $(Y_i, \boldsymbol{X_i})$ and $g(\boldsymbol{X_i}, \boldsymbol{\beta}) = \boldsymbol{E}(Y|\boldsymbol{X_i})$. Local likelihood estimation was introduced by (Tibshirani and Hastie, 1987) and its properties were analyzed by (Fan, *et al*, 1995), (Fan and Gijbels, 1996), (Fan, *et al*, 1998), (Eguchi, *et al*, 2003). The lone local estimation has been used to estimate density and hazard function functions, but is rarely used and applied to estimate regression functions with binary response variables.

2.5 Local Likelihood Logit Estimation

Given paired data $(Y_i, \boldsymbol{X_i})\,; i = 1, 2, \ldots, n$ with Y_i's response variable has two categories and predictor variables $\boldsymbol{X_i} = (1, X_{1i}, X_{2i}, \ldots, X_{pi})$ has a continuous type. It is assumed that the paired data meets the nonparametric binary logistic regression model. Frölich (2006) proposes that to estimate the nonparametric binary logistic regression method, the local likelihood logit estimation method obtained from (3) by replacing the log-probability density function from (Y_i, X_i), so we get the local likelihood logit estimation as follows:

$$\begin{aligned}\ell = \textstyle\sum_{i=1}^{n} [Y_i \ln \pi(\boldsymbol{X_i}, \mathbf{x_0}) +\\ +(1 - Y_i)\ln(1 - \pi(\boldsymbol{X_i}, \mathbf{x_0}))] K_h(\boldsymbol{X_i} - \mathbf{x_0})\end{aligned} \tag{5}$$

where $K_h(\boldsymbol{X_i} - \mathbf{x_0})$ is the multiplication of kernel functions from continuous type predictor variables. X_{ji} and x_j each is the value of the observation to i of the predictor variable X_j and fixed point arbitrary to j. K is an univariate symmetry kernel function, and h_j is the bandwidth of the predictor variable X_j.

2.6 Selection of Optimal Bandwidth

The selection of bandwidth (h) is very important in obtaining a regression function estimator based on a nonparametric approach. Bandwidth is a balance controller between the smoothness of the function of the data, if h is very small then the estimated function obtained will be very rough and go to the data while if h is very large, the estimated function obtained will be very smooth and towards the average of the response variable. Therefore, in choosing h it is expected that the value is optimal. There are two methods to obtain optimal h, namely the Cross Validation (CV) method and the Generalized Cross Validation (GCV) method (Eubank, 1988). If a criterion for h is limited to a linear estimator class, then for each h there is a matrix $A(h)$ of size of $n \times n$ so that

$$\widehat{m}(x) = A(h)\boldsymbol{Y} \tag{6}$$

The optimal h value is obtained by minimizing GCV as follows:

$$GCV(h) = \frac{MSE(h)}{(\,n^{-1}\, tr[\mathbf{I} - A(h)]\,)^2} \tag{7}$$

where $MSE(h) = n^{-1}\sum_{i=1}^{n}[\,y_i - \widehat{m}(x_i)]^2$. For the local likelihood estimation method, Staniswalis (1989) suggests using criteria $CV_{ML}(h)$ that correspond to the likelihood function as follows:

$$CV_{ML}(h) = \sum_{i=1}^{n} Y_i \ln g\big(X_i\,, \hat{\beta}(x_0)_{-X_i|h}\big) + \\ +(1 - Y_i)\ln\big[1 - g\big(X_i\,, \hat{\beta}(x_0)_{-X_i|h}\big)\big] \tag{8}$$

where $\hat{\boldsymbol{\beta}}(\boldsymbol{x_0})_{-\boldsymbol{X_i}|h}$ is the leave-one-out coefficient estimate to estimate $E[Y|\boldsymbol{X} = \boldsymbol{X_i}]$ obtained from sample data without observing i.

3. RESULTS AND DISCUSSIONS

3.1 Data Source

The data used in risk modeling of Type II Diabetes Mellitus is a sample of 81 patients in the Internal Medicine Poly Surabaya Hajj General Hospital in 2018 diagnosed by doctors suffering from Type II Diabetes Mellitus or not Type II Diabetes Mellitus. Response variable (Y) is the condition of outpatients in the poly internal medicine at the Haji General Hospital Surabaya diagnosed by doctors suffering from Type II Diabetes Mellitus or not, while the predictor variables include age (X_1), body mass index (BMI) (X_2), waist circumference (X_3), and pressure systolic blood (X_4).

3.2 Parametric Binary Logistic Regression

The risk modeling of the incidence of Type II Diabetes Mellitus using the parametric binary logistic regression approach shows that the results of parameter testing are significantly significant for α = 5% with a P-value of 0.000 as presented in table 1 as follows:

Table-1. Results of parameter testing together.

		Chi-square	df	P-value
Step 1	Step	32.334	4	0.000
	Block	32.334	4	0.000
	Model	32.334	4	0.000

This means that the variables of age, BMI, waist circumference and systolic blood pressure jointly influence the risk of Type II Diabetes Mellitus. The results of individual parameter testing along with the Odd Ratio (OR) values are presented in Table-2 as follows:

Table-2. Results of individual parameter testing.

Variable	Coefficient	SE	Wald	df	P-value	OR
X_1	0.046	0.025	3.499	1	0.061	1.047
X_2	0.223	0.096	5.364	1	0.021	1.249
X_3	0.017	0.034	0.263	1	0.608	1.017
X_4	0.037	0.015	5.923	1	0.015	1.037
Constant	-14.325	3.684	15.123	1	0.000	0.000

Based on Table-2 it can be concluded that for $\alpha = 5\%$, the variable BMI and systolic blood pressure have a significant effect on the risk of the incidence of Type II Diabetes Mellitus with P-value values of 0.021 and 0.015 respectively. The OR value of the variable BMI is 1.249 meaning that if the BMI rises by 1 unit, then the risk of patients affected by Type II Diabetes Mellitus increases by 24.9%. Furthermore, the OR value of the variable systolic blood pressure is 1.037 meaning that if the systolic blood pressure rises by 1 unit, then the risk of patients affected by Type II Diabetes Mellitus increases by 3.7%. The suitability test results of parametric binary logistic regression model using the Hosmer and Lemeshow test for $\alpha = 5\%$ to show that the model matches the P-value of 0.799 as presented in Table-3 as follows:

Table-3. Results of the suitability test for parametric binary logistic regression model.

Step	Chi-square	df	P-value
1	4.608	8	0.799

VOL. 15, NO. 20, OCTOBER 2020 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

The modeling results using the parametric binary logistic regression approach obtained the classification accuracy value of 80.2% with a cut value of 0.5.

3.3 Nonparametric Binary Logistic Regression with the GAM Method

The first step to obtain a nonparametric binary logistic regression model using the GAM method is to determine the optimal bandwidth for each predictor variable using the Nadaraya Watson estimator, namely bandwidth which has a minimum GCV value. The optimal bandwidth results for each predictor variable are presented in Table-4 as following:

Table-4. Optimal bandwidth for each predictor variable.

Variable	Optimal bandwidth	Minimum GCV
X_1	6.09	0.2195633419576
X_2	5.85	0.2368853651003
X_3	2.06	0.2014526551331
X_4	2.77	0.1550234791469

Furthermore, the optimal bandwidth is used to determine the initial value of the smoothing function $\hat{f}_j(X_j)$ for each predictor variable. The next step is iterating using the local scoring algorithm so the estimation of nonparametric binary logistic regression model using the GAM method based on the Kernel estimator is obtained. The plot of estimating the probability of someone affected by Type II Diabetes Mellitus based on each predictor variable is as follows:

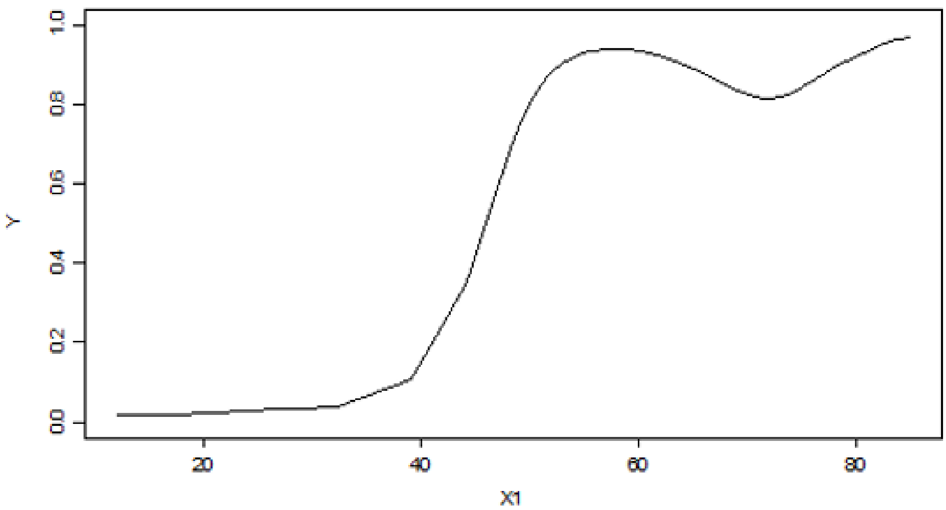

Figure-1. Plot of estimated probability of the incidence of Type II Diabetes Mellitus against age.

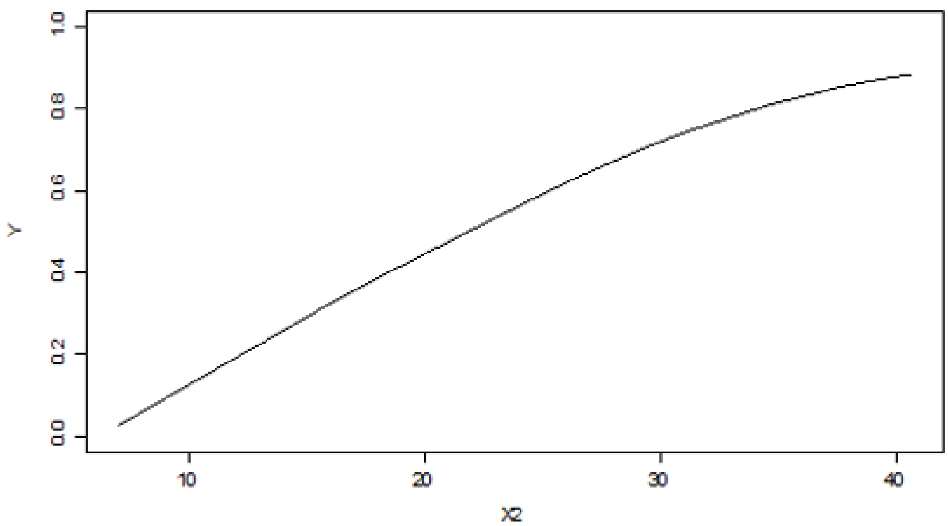

Figure-2. Plot estimation of the probability of Type II Diabetes Mellitus against BMI.

ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

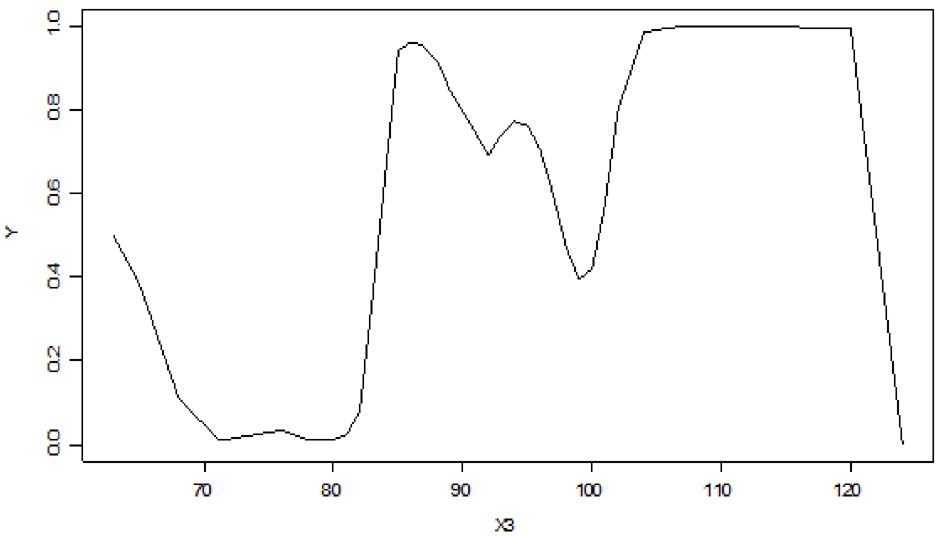

Figure-3. Plot estimation of the probability of Type II Diabetes Mellitus against circumference waist.

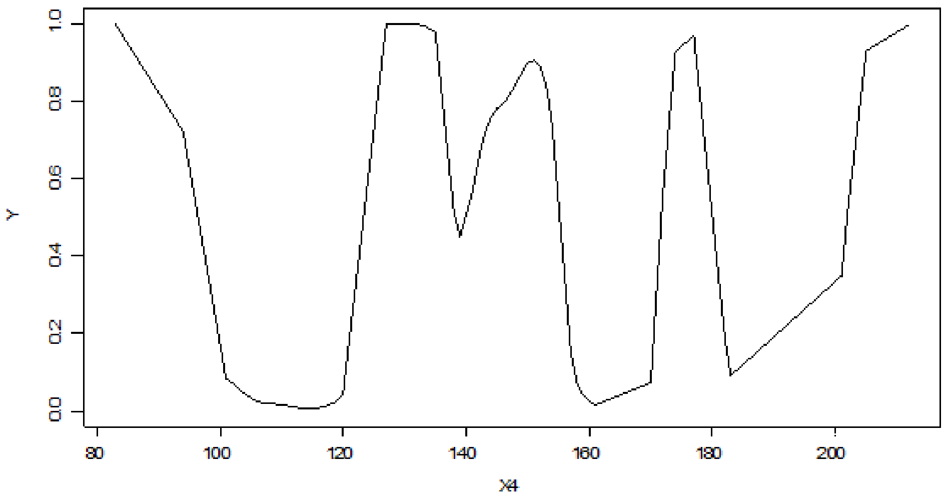

Figure-4. Plot estimation of the probability of Type II Diabetes Mellitus against pressure systolic blood.

Based on Figure-1 it can be seen that the probability of a person suffering from Type II Diabetes Mellitus increases with age up to around 60 years, after which it tends to decrease. This is in line with the results of a study by (Sue Kirkman, *et al,* 2013) which states that the incidence of diabetes increases with age until around age 65, after which both incidence and prevalence appear to be decreasing. Diabetes is often found in older adults because at that age physiological function decreases and there is a decrease in insulin secretion or resistance so that the ability of the body to function at high blood glucose control is less optimal. Figure-2 shows that the relationship between BMI and the prevalence of Type II Diabetes Mellitus forms a linear trend which shows that the more an individual's BMI increases, there is a significant increase in the risk probability of Type II Diabetes Mellitus. This is in line with the research by (Michael, *et al*, 2014) which states that the more an individual's BMI increases, there is a significant increase in the prevalence of Type II Diabetes Mellitus. The relationship between waist circumference and the prevalence of Type II Diabetes Mellitus is shown by Figure-3. The result forms fluctuate trend. This is because in this research we not distinguish by gender. The normal waist circumference limits for both men and women are different. The values of 102 cm for men and 88 cm for women, recommended as cut off points by National Heart Lung and Blood Institute (NHLBI) of waist circumference for healthy people (Flegal, 2007). This is in line with a study by (Dagan, *et al*, 2013) which concluded that differences between the gender were statistically significant for waist circumference. So, in the future research we emphasize the need to investigate men and women separately. Figure-4 shows that the relationship between Systolic Blood Pressure and the prevalence of Type II Diabetes Mellitus forms fluctuate trend. This is

ARPN Journal of Engineering and Applied Sciences

because blood pressure at the time of measurement depends on the patient's condition. No matter which device is used to measure blood pressure, it must be recognised that blood pressure is a variable haemodynamic phenomenon, which is influenced by many factors, not least being the circumstances of measurement itself. These influences on blood pressure can be significant, often accounting for rises in systolic blood pressure greater than 20 mm Hg, and if they are ignored, or unrecognised, hypertension will be diagnosed erroneously and inappropriate management instituted. These factors have to be carefully considered in all circumstances of blood pressure measurement-self measurement by patients, conventional measurement, measurement with automated devices whether in a doctor's surgery, an ambulance, a pharmacy, or in hospital using sophisticated technology (Beevers, *et al*, 2001). Furthermore, we calculate the accuracy classification of nonparametric approach using the GAM method in the incidence of Type II Diabetes Mellitus has a validity of 88.89% with a cut value of 0.5.

3.4 Nonparametric Binary Logistic Regression with the LLLE Method

The first step to obtain a nonparametric binary logistic regression model with the LLLE method is to determine the optimal bandwidth for each predictor variable, namely bandwidth which has a maximum $CV_{ML}(h)$ value. Some bandwidth values for each predictor variable along with the $CV_{ML}(h)$ value is presented in Table-5 as follows:

Table-5. Bandwidth value along with $CV_{ML}(h)$ value.

h_1	h_2	h_3	h_4	$CV_{ML}(h)$
24.34333	11.21	0.01	0.01	-42.8098
24.34333	11.21	0.01	43.01	-112.101
24.34333	11.21	0.01	86.01	-112.101
24.34333	11.21	20.34333	0.01	-36.6991
24.34333	11.21	20.34333	43.01	-50.3274

Based on Table-5 is obtained the optimal bandwidth value for each predictor variables X_1, X_2, X_3, and X_4 is 24.34333, 11.21, 20.34333, and 0.01 respectively. Next, estimate the parameters for each observation (arbitrary fixed point) using optimal bandwidth and some of the results presented in table 6 as follows:

Table-6. Parameter estimate for each observation.

Observation	X_1	X_2	X_3	X_4	$\hat{\beta}_0(x_0)$	$\hat{\beta}_1(x_0)$	$\hat{\beta}_2(x_0)$	$\hat{\beta}_3(x_0)$	$\hat{\beta}_4(x_0)$
1	65	18.52	71	101	-9.6154	0.0383	-0.0002	0.0556	0.0083
2	12	17.86	64	103	-9.6153	0.0420	-0.0001	0.0562	0.0084
3	48	31.50	88	110	-9.6153	0.0271	0.0042	0.0590	0.0080
4	39	20.52	72	154	-9.6155	0.0973	-0.0003	0.0804	-0.0160
5	39	19.25	65	102	-9.6153	0.0405	-0.0002	0.0560	0.0087

Table-7. Estimated success probability and OR for each predictor variable.

Observation	Pr(success)	OR(X_1)	OR(X_2)	OR(X_3)	OR(X_4)
1	0.087568	1.03904	0.99980	1.05717	1.00833
2	0.009461	1.04289	0.99990	1.05781	1.00844
3	0.108109	1.02747	1.00421	1.06078	1.00803
4	0.075707	1.10219	0.99970	1.08372	0.98413
5	0.028966	1.04133	0.99980	1.05760	1.00874

Based on Table-6, estimate the probability of patients affected by type II diabetes mellitus and Odd Ratio (OR) for each predictor variable are obtained as presented in Table-7. Based on Table-7, it can be interpreted that the probability of a patient 1 affected by type II diabetes mellitus when having 65 years of age, a BMI of 18.52, a waist circumference of 71 and a systolic blood pressure of 101 is 0.087568. OR value of the age variable for patient 1 is 1.03904, this mean that if the patient's age rises 1 year then the risk of the patient affected by type II diabetes mellitus rises by 3.9%.

The value of the classification accuracy of the nonparametric binary logistic regression model with the LLLE method of 100%. In summary the value of the accuracy of the classification of the three method, i.e parametric and nonparametric binary logistic regression models is presented in table 8 as follows:

Table-8. The classification accuracy value of the three binary logistic regression models.

Model	Classification Accuracy
Parametric binary logistic regression	80.2%
Nonparametric binary logistic regression using the GAM method	88.89%
Nonparametric binary logistic regression using the LLLE method	100%

4. CONCLUSIONS

Based on the result of applying logistic regression model with parametric and nonparametric approach to the modeling of risk prediction of diabetes, it is concluded that the parametric binary logistic regression has a validity of 80.2%, the nonparametric binary logistic regression with the GAM method has a validity of 88.89%, and nonparametric binary logistic regression with the local likelihood logit estimation method has a validity of 100%. Based on the value of the accuracy of the classification, the best approach model is obtained by nonparametric binary logistic regression with the local likelihood logit estimation method.

ACKNOWLEDGEMENTS

The authors thanks to the Ministry of Research, Technology and Higher Education of the Republic of Indonesia for financial support of this research through The Fundamental Research University Grant 2019. The authors also thank to the anonymous referees for their valuable suggestions which let to the improvement on the manuscript.

REFERENCES

Agresti A. 2002. Categorical data Análisis. 2nd Edition. New York: John Wiley and Sons.

Basar., Md Abul., Alvi., Hassan Nomani., Bokul., Gazi Nowrin., Khan M., Shahriar., Anowar., Farzana., Huda., Mohammad Nurul., Al Mamun and Khondaker Abdullah. 2015. A Review on Diabetes Patient Lifestyle Management Using Mobile Application. 18th International Conference on Computer and Information Technology. 379-385.

Beevers G., Lip G.Y.H and O'Brien E. 2001. Blood pressure measurement. BMJ. 322(7292): 981-985.

Dagan S.S., Segev S., Novikov I and Dankner R. 2013. Waist circumference vs body mass index in association with cardiorespiratory fitness in healthy men and women: a cross sectional analysis of 403 subjects. Nutrition Journal. 12: 12.

Eguchi S., Kim T and Park B. 2003. Local Likelihood Method: a bridge over parametric and nonparametric regression. Journal of Nonparametric Statistics. 15: 665-683.

Eubank R.L. 1988. Spline Smoothing and Nonparametric Regression. New York: Marcel Deker, and Basel.

Fan J., Heckman N.E and Wand M.P. 1995. Local Polinomial Kernel Regression for Generalized Linear Models and Quasi-Likelihood Functions. Journal of the American Statistical Association. 90: 141-150.

Fan J and Gijbels I. 1996. Local Polinomial Modelling and its Applications. London: Chapman and Hall.

Fan J., Farmen M and Gijbels I. 1998. Local Maximum Likelihood Estimation and Inference. Journal of the Royal Statistical Society, Series B. 60: 591-608.

Flegal K.M. 2007. Waist circumference of healthy men and women in the United States. International Journal of Obesity. 31: 1134-1139.

Frölich M. 2006. Non-parametric regression for binary dependent variables. Econometrics Journal. 9: 511-540.

George., Eleni I., Protopappas., Vasilios C., Mougiakakou and Stavroula G. 2013. Short-term vs. Long-term Analysis of Diabetes Data: Application of Machine Learning and Data Mining Techniques.

Goyal S and Cafazzo J.A. 2013. Mobile phone health apps for diabetes management: Current evidence and future developments. Qjm. 106(12): 1067-1069.

Hastie T.J and Tibshirani R.J. 1990. Generalized Additive Models, London: Chapman & Hall.

Hosmer D.W., Lemeshow S and Sturdivant R.X. 2013. Applied Logistic Regression. Third Edition. New York: John Wiley and Sons.

Michael L Ganz., Neil Wintfeld., Qian Li., Veronica Alas., Jakob Langer and Mette Hammer. 2014. The association of body mass index with the risk of type 2 diabetes: a case control study nested in an electronic health records system in the United States. Diabetology & Metabolic Syndrome. 6: 50.

Ojugo A.A, Eboka A.O, Yoro R.E, Yerokun M.O and Efozia F.N. 2015. Hybrid Model for Early Diabetes Diagnosis. Second International Conference on Mathematics and Computers in Science and in Industry. 55-64.

Prayitno., Agus., Wibawa., Andi Dharma., Purnomo., Mauridhi and Hery. 2016. Early Detection Study of Kidney Organ Complication Caused by Diabetes Mellitus using Iris Image Color Constancy. International Conference of Information, Communication Technology and System (ICTS). : 146-149.

Rahman A.W., Amiruddin R and Noor B.N. 2013. Faktor Risiko dan Deteksi Dini Kejadian Diabetes Mellitus Tipe 2 di Kecamatan Tempe Kabupaten Wajo. Jurnal Media Kesehatan Masyarakat Indonesia. 2(1).

Rifada M., Suliyanto., Eko Tjahjono and Kesumawati A. 2018. The Logistic Regression Analysis with Nonparametric Approach based on Local Scoring Algorithm (Case Study: Diabetes Mellitus Type II Cases in Surabaya of Indonesia). International Journal of Advances in Soft Computing and its Application. 10(3): 167-178.

Staniswalis J. 1989. The Kernel Estimate of a Regression Function in Likelihood-Based Models, Journal of the American Statistical Association. 84: 276-283.

Sue Kirkman M, *et al*. 2012. Diabetes in Older Adults. Diabetes Care. 35(12): 2650-2664.

Suliyanto., Rifada M., Eko Tjahjono and Sediono. 2018. The Estimation of Binary Nonparametric Regression Model based on The Kernel Estimator by Generalized Additive Models Method, International Conference on Research, Implementation and Education of Mathematics and Science 5 Regular Conference Proceeding. Yogyakarta. 65-72.

Suliyanto and Rifada M. 2019. Modeling of Cases of Diabetes Mellitus Patients Based on Non-parametric Binary Logistic Regression Approach with the Generalized Additive Model and Local Likelihood Logit Estimation Methods, Proceeding of Global Conference on Engi-neering and Applied Science, Sapporo, Hokkaido, Japan. 7(1618): 94-103.

Tibshirani R and Hastie T. 1987. Local Likelihood Estimation. Journal of the American Statistical Association. 82: 559-567.

Trisnawati S.K and Setyogoro S. 2013. Faktor Risiko Kejadian Diabetes Melitus Tipe II Di Puskesmas Kecamatan Cengkareng Jakarta Barat Tahun 2012. Jurnal Ilmiah Kesehatan. 5(1).

WHO. 2003. Screening for Type 2 Diabetes: Report of a World Health Organization and International Diabetes Federation meeting. Accessed September 9, 2019 via http://www.who.int/diabetes/publications/en/screening_mnc03.pdf