International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 5, May - 2022, Pages: 201-206

Predicting of Diabetes Mellitus Type 2 Risk Using Nonparametric Ordinal Logistic Regression Based on Smoothing Spline Estimator

Sakinah Priandi[1], Nur Chamidah[2], Sediono[3], Toha Saifudin[4], M. Fariz Fadillah Mardianto[5]

[1,2,3,4,5]Statistics Study Program, Universitas Airlangga, Surabaya, Indonesia
[1]sakinah.priandi-2018@fst.unair.ac.id
[2]Correspondence author: nur-c@fst.unair.ac.id
[3]sediono101@gmail.com
[4]tohasaifudin@fst.unair.ac.id
[5]m.fariz.fadillah.m@fst.unair.ac.id

Abstract: *Diabetes mellitus is one of the deadliest diseases in the world. Indonesia ranks seventh for the highest diabetes sufferers in the world. The prevalence of diabetics in Indonesia reaches 6.2%. According to these results, most diabetics suffer from diabetes mellitus type 2. Dyslipidemia is one of the factors that affect diabetes mellitus type 2. Dyslipidemia is a disorder of fat metabolism which is characterized by an increase in plasma fat levels. The main abnormality of fat levels is an increase in LDL cholesterol and triglycerides. To predict the risk of diabetes mellitus type 2, we need to build a model. In statistical analysis, there are two approaches to estimating the model, namely parametric and nonparametric. In this study, we predicted the risk of diabetes mellitus type 2 based on LDL cholesterol and triglyceride levels using nonparametric ordinal logistic regression (GAM) based on a smoothing spline estimator and compared it with the parametric ordinal logistic regression (GLM) approach. Based on stability test, namely the press's Q value based on nonparametric ordinal logistic regression (GAM) model based on the smoothing spline estimator is stable or consistent with the accuracy value and the sensitivity value for the category diabetes mellitus type 2 are 85% and 0.93 respectively. Meanwhile, the Press's Q value based on parametric ordinal logistic regression (GLM) model is instable or inconsistent with accuracy value and the sensitivity value for the category diabetes mellitus type 2 are 37.5% and 0.47 respectively. This means ordinal logistic regression using a nonparametric approach (GAM) based on a smoothing spline estimator is better than a parametric approach (GLM) to predict the risk of diabetes mellitus type 2.*

Keywords—Diabetes Mellitus Type 2; LDL Cholesterol; Triglycerides; Ordinal Logistic Regression; Generalized Additive Model; Smoothing Spline Estimator; Generalized Linear Model

1. INTRODUCTION

Diabetes Mellitus is one of the deadliest diseases in the world. This disease is better known as the silent killer because it can attack all important organs of the human body and cause various kinds of complications in the human body. Types of diabetes mellitus can be classified into type 1 diabetes mellitus and diabetes mellitus type 2. Diabetes mellitus type 2 accounts for between 90% and 95% of diabetes, with the highest proportion in low- and middle-income countries [1].

Indonesia ranks seventh for the highest diabetics. The prevalence of diabetics in Indonesia reaches 6.2%, which means that there are more than 10.8 million people with diabetes. From these results, most sufferers suffer from diabetes mellitus type 2 [2]. On the other hand, this is also due to the condition of the Covid-19 pandemic. Covid-19 infection also affects people with diabetes mellitus. The covid-19 infection has resulted in significant changes in the metabolism of patients with significant increases in blood glucose which can lead to increased insulin resistance and associated hyperglycemia [3]. Based on data from the Indonesian Ministry of Health as of October 13, 2020, it shows that of the 1488 Covid-19 patients who died, around 11.6% suffered from diabetes mellitus [2].

One of the risk factors that influence diabetes mellitus is dyslipidemia. Dyslipidemia is a fat metabolism disorder characterized by an increase or decrease in plasma fat levels. The main abnormality in the fat content is an increase in LDL fat content and an increase in triglyceride levels [4]. The risk of diabetes mellitus increases with the increase in triglyceride levels. A person who has diabetes mellitus type 2 is characterized by hypertriglyceridemia. Reduced insulin secretion can cause triglyceride levels to increase affecting the state of diabetes mellitus type 2 [5]. In addition, high LDL cholesterol will have a relationship with diabetes mellitus [6].

Based on this research, the researcher used the ordinal logistic regression method. Where, diabetes mellitus type 2 is divided into three levels, namely not suffering from diabetes mellitus type 2/normal (Y=0), pre-diabetes mellitus type 2 (Y=1), and diabetes mellitus type 2 (Y=2). Ordinal logistic regression is a method that can be used to find the relationship between response variables that are polychotomous (have more than two categories) and have levels with more than one predictor variable [7]. There are 2 approaches to ordinal logistic regression, namely the parametric approach with Generalized Linear Models (GLM) and the nonparametric approach with Generalized Additive Models (GAM). One of the nonparametric estimators is the smoothing spline. In one of the studies, a smoothing spline was used to estimate the internal capacity of Lithium-Ion batteries and showed that the proposed method produces good results with low RMSE [8]. The results of this study support that the smoothing spline is an estimator that has the best flexibility in estimating nonparametric regression functions compared to others [9].

Based on the facts and previous research, to get better results, researchers will predict risk factors for diabetes mellitus type 2 using nonparametric ordinal logistic regression (GAM) approaches with smoothing spline estimator and compare it with the parametric ordinal logistic regression (GLM). With this study, it is hoped that the prediction results obtained can be used to predict the possible risk of a person developing diabetes mellitus type 2 to prevent the occurrence of diabetes mellitus in Indonesia.

2. LITERATURE REVIEW

2.1 Diabetes Mellitus Type 2

Diabetes Mellitus Type 2 is a metabolic disease caused by an increase in blood sugar in the body due to decreased insulin secretion by pancreatic beta cells or impaired insulin function (insulin resistance). Insulin resistance means that diabetes mellitus type 2 is caused by the failure of insulin target cells or the inability to respond to insulin normally. Many of these causes of insulin resistance result from obesity, lack of exercise, and aging [10]. One way to detect diabetes is to take the Fasting Blood Sugar (GDP) test.

2.2 LDL Cholesterol

Cholesterol is a type of fatty substance produced by cells in the body, and about a quarter of the amount produced is produced in the liver [11]. Cholesterol is classified into two, namely LDL and HDL. LDL cholesterol (Low-Density Lipoprotein) is also called bad cholesterol. This LDL plays a role in carrying cholesterol throughout the body that is needed through the arterial wall network. If the patient eats too much LDL, it will accumulate cholesterol in the arteries, thus it can cause plaques and result in the narrowing of the arteries.

2.3 Triglycerides

Triglycerides are one of the main types of fat that are transported in the blood and stored in body fat tissue. The function of triglycerides is to store calories and provide energy for the body. The two main sources of plasma triglycerides are exogenous carried by chylomicrons (eg carbohydrates and fats from food) and endogenous (derived from the liver) and endogenous carried by Very Low-Density Lipoprotein (VLDL). Individuals with diabetes mellitus type 2 are characterized by hypertriglyceridemia (a condition in which triglycerides are elevated with or without other lipoprotein disorders) [5].

2.4 *Generalized Additive Model* (GAM)

Generalized Additive Models (GAM) is a model that can be used to estimate nonparametric models by replacing the linear form of Generalized Linear Models (GLM) with an additive form that uses certain smoothing [12]. The GAM model in ordinal logistic regression can be expressed in the following equation:

$$g(\gamma_i) = \theta_q + \sum_{j=1}^{p} f_j(x_{ij}), \quad i = 1,2,\dots,n \qquad (1)$$

In the ordinal logistics model, the g function is a link function for the cumulative logit model, i.e $ln\left(\frac{\gamma_i}{1-\gamma_i}\right)$, thus γ_i can be obtained as follows:

$$\gamma_i = \frac{\exp\left(\theta_q + \sum_{j=1}^{p} f_j(x_{ij})\right)}{\left[1+\exp\left(\theta_q + \sum_{j=1}^{p} f_j(x_{ij})\right)\right]}, q = 1,2,\dots,K-1 \qquad (2)$$

The regression function in GAM is generally estimated using a local scoring algorithm. This is because the local scoring algorithm can accommodate an additive nonparametric regression model whose distribution of the response variable belongs to the exponential family [12]. The steps in the local scoring algorithm for $(m = 0,1,2,\dots)$ with following steps:

1. Defining $\hat{f}_j^{(0)}(x_{ij}) = 0, \hat{\theta}^{(0)} = logit(\sum_i n_i g_{ik} / \sum_i n_j)$
2. $logit(\hat{\boldsymbol{\gamma}}_i^{(m)}) = \hat{\theta} + 1\sum_{j=1}^{p} f_j^{(m)}(x_{ij})$ (3)
3. $\pi_i^{(m)} = L^{-}\hat{\boldsymbol{\gamma}}_i^{(m)}$ (4)
4. $\boldsymbol{C}_i^{(m)} = diag[\gamma_k(1-\gamma_k)], k = 1,2,..K$ (5)
5. $\boldsymbol{A}_i^{(m)} = A\left(\hat{\boldsymbol{\gamma}}_i^{(m)}{}_i, n_i\right)$ (6)
6. $\boldsymbol{W}_i^{(m)} = \left(\boldsymbol{C}_i^{(m)}\right)^t \boldsymbol{A}_i^{(m)}\left(\boldsymbol{C}_i^{(m)}\right)$ (7)
7. $\boldsymbol{z}_i^{(m)} = \hat{\theta} + \left(\boldsymbol{C}_i^{(m)}\right)^{-1}\left(\boldsymbol{g}_i - \hat{\boldsymbol{\gamma}}_i^{(m)}\right)$ (8)
8. $\theta^{(m+1)} = \left(\sum_{i=1}^{n} \boldsymbol{W}_i^{(m)}\right)^{-1} \sum_{i=1}^{n} \boldsymbol{W}_i^{(m)} \boldsymbol{z}_i$ (9)
9. Obtaining the adjusted dependent value
$$z_i = \eta^{(m)}(\boldsymbol{x}_i) + \frac{\boldsymbol{1}^T \boldsymbol{W}_i^{(m)}\left[\left(\boldsymbol{C}_i^{(m)}\right)^{-1}\left(\boldsymbol{g}_i - \hat{\boldsymbol{\gamma}}_i^{(m)}\right)\right]}{\boldsymbol{1}^T \boldsymbol{W}_i^{(m)} \boldsymbol{1}} \qquad (10)$$
10. Defining $\hat{\boldsymbol{\alpha}}^{(m)} = \bar{z}^{(m)}$
11. $R_{ij}^{(m)} = z_i - \hat{\alpha}^{(m)} - 1\left(\sum_{s=1}^{j-1} f_s^{(m)}(x_{is}) - \sum_{s=j+1}^{p} f_s^{(m)}(x_{is})\right)$ (11)
12. Calculating smoothing function
$$f_j^{(m+1)}(x_{ij}) = A(\lambda_{ij}) R_{ij}^{(m+1)} \qquad (12)$$

13. Iterating steps (10) to (12) to find the value of the backfitting convergence criterion $\left|\frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(f_j^{(m-1)}(x_{ij})-f_j^{(m)}(x_{ij})\right)^2}{1+\sum_{i=1}^{n}\sum_{j=1}^{p}\left(f_j^{(m-1)}(x_{ij})\right)^2}\right| \leq \varepsilon$ (13)
14. Calculating the value of $\gamma_i^{(m+1)}$, $\pi_i^{(m+1)}$ and weighting matrix $\boldsymbol{W}_i^{(m+1)}$
15. Calculating adjusted dependent variable $z_i^{(m)}$
16. Calculating deviance value (D)

Iterating steps (1) to (16) to obtain a convergent value $\left|D^{(m+1)} - D^{(m)}\right| < \varepsilon$

2.5 Smoothing Spline Estimator

Smoothing is a process that can eliminate coarse data by following the shape of the data pattern. The smoothing $\hat{\boldsymbol{f}}(\boldsymbol{x})$ can be obtained based on the observed data, namely pairs of predictor variables and response variables. The smoothing function estimator is a function estimator that can map the data well and has a small error variance. Therefore, by using $\boldsymbol{n}$ observational data, $\boldsymbol{f}(\boldsymbol{x})$ can be obtained by minimizing the penalized least square (PLS) function, which is:

$$\sum_{i=1}^{n}\{y_i - f(x_i)\}^2 + \lambda\int_a^b\{f''(t)\}^2\,dt \qquad (14)$$

with λ is constant and $a \leq x_1 \leq \cdots \leq x_n \leq b$ [13]

2.6 Generalized Linear Model (GLM)

The Generalized Linear Model (GLM) is an extension of the linear regression model with the assumption that the predictor has a linear effect but does not assume a certain distribution of the response variable and is used when the response variable is a member of the exponential family [14]. The GLM model in ordinal logistic regression can be expressed in the following equation:

$$g(\gamma_i) = \theta_q + \sum_{j=1}^{p}\beta_j x_{ij}, \quad i = 1,2,\ldots,n \qquad (15)$$

In the ordinal logistics model, the g function is a link function for the cumulative logit model, i.e $ln\left(\frac{\gamma_i}{1-\gamma_i}\right)$, thus γ_i can be obtained as follows:

$$\gamma_i = \frac{\exp\left(\theta_q+\sum_{j=1}^{p}\beta_j x_{ij}\right)}{\left[1+\exp\left(\theta_q+\sum_{j=1}^{p}\beta_j x_{ij}\right)\right]}, q = 1,2,\ldots,K-1 \qquad (16)$$

By assuming as many categories as K, the probability of the response variables for the parametric logistic regression model in each category is obtained as follows.

$$\pi_0(x) = \frac{\exp\left(\theta_0+\sum_{j=1}^{p}\beta_j x_{ij})\right)}{1+\exp\left(\theta_0+\sum_{j=1}^{p}\beta_j x_{ij})\right)} \qquad (17)$$

$$\pi_1(x) = \frac{\exp\left(\theta_1+\sum_{j=1}^{p}\beta_j x_{ij}\right)}{1+\exp\left(\theta_1+\sum_{j=1}^{p}\beta_j x_{ij}\right)} - \frac{\exp\left(\theta_0+\sum_{j=1}^{p}\beta_j x_{ij}\right)}{1+\exp\left(\theta_0+\sum_{j=1}^{p}\beta_j x_{ij}\right)} \qquad (18)$$

$$\pi_2(x) = 1 - \frac{\exp\left(\theta_1+\sum_{j=1}^{p}\beta_j x_{ij}\right)}{1+\exp\left(\theta_1+\sum_{j=1}^{p}\beta_j x_{ij}\right)} \qquad (19)$$

2.7 Press's Q

Press's Q is a measure to determine the stability in classification. Press's Q test statistics show the extent to which these groups can be separated from the existing variables.

Press's Q hypothesis test can be written as follows:
H_0 = Unstable/inconsistent model classification results
H_1 = Stable/consistent model classification results

Press's Q test can be written as follows:

$$Press'sQ = \frac{(N-nK)^2}{N(K-1)} \sim X^2{}_{(1)} \qquad (20)$$

The results of the *Press's Q* calculation are compared with the Chi-square value with the critical area being H_0 is rejected if *Press's Q*>$X^2{}_{(df,\alpha)}$[15].

2.8 Deviance

To test the fitness of the model by comparing the actual model to the alleged model, we need a test statistic called deviance

The fitness test of the ordinal logit model using the deviance test is carried out with the following hypotheses:
H_0: ordinal logistics model is fit
H_1 : ordinal logistics model is not fit

The calculation of deviance values for ordinal logistic regression models is as follows :

$$D = 2\sum_{i=1}^{n}\left[y_{i1}\ln\left(\frac{y_{i1}}{\hat{\pi}_1(\boldsymbol{x}_i)}\right) + y_{i2}\ln\left(\frac{y_{i2}}{\hat{\pi}_2(\boldsymbol{x}_i)}\right) + \cdots + y_{iq}\ln\left(\frac{y_{iq}}{\hat{\pi}_q(\boldsymbol{x}_i)}\right)\right] \qquad (21)$$

The null hypothesis is rejected if the value of $D > \chi^2{}_{((q-1)(J-p-1),\alpha)}$ with J is the number of levels of different predictor variables [16].

2.9 Confusion Matrix

A confusion Matrix (also known as a contingency table or error matrix) is one of the techniques that can be used to access the results of the performance of an algorithm. The Confusion Matrix contains information that compares the predicted results from the model generated by the classification with the actual results [17]. Confusion Matrix tables can be used in classification evaluation matrices such as accuracy, sensitivity, and others.

Table 1: *Confusion Matrix 3x3*

		Actual		
		A	B	C
Prediction	A	AA	BA	CA
	B	AB	BB	CB

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 6 Issue 5, May - 2022, Pages: 201-206

C	AC	BC	CC

The level of accuracy is a comparison between the correct clarified data and the whole data [18].

$$Accuracy = \frac{AA+BB+CC}{n} X\ 100\% \quad (22)$$

Sensitivity is the percentage, or proportion, of the number of positive categories that are classified correctly with all samples having positive categories [19]

$$Sensitivity_A = \frac{AA}{AA+AB+AC} \quad (23)$$

3. MATERIAL AND METHOD

This section will explain the research data, sampling techniques, and research variables.

3.1 RESEARCH DATA

The data used in this study is a combination of primary data and secondary data obtained through filling out questionnaires, interviews, and viewing the medical records of diabetes mellitus outpatients at the Haji General Hospital Surabaya since October 2020. In this study, we use 40 respondent.

3.2 Research Variables

In this study, there were 2 predictor variables analyzed with the risk of diabetes mellitus type 2. Research variables, operational definitions, and data scales are presented in Table 2.

Table 2: Research Variables

Research Variable	Operational Definition	Data Scale
Diabetes Occurrence	Measured by Fasting Blood Sugar (GDP) which is categorized into 3 categories	Ordinal: 0. Not suffering from diabetes/normal mellitus type 2 1. Pre-diabetes mellitus type 2 2. Diabetes mellitus type 2
LDL Cholesterol	The number of LDL cholesterol levels in the patient's blood examination results	Ratio
Triglycerides	The number of triglyceride levels in the patient's blood	Ratio

4. RESULT AND DISCUSSION

The data used in this study amounted to 40 data consisting of 10 patients not suffering from diabetes mellitus type 2, 15 patients suffering from pre-diabetes mellitus type 2, and 15 patients suffering from diabetes mellitus type 2. The data used in predicting models with predictor variables including LDL cholesterol and triglyceride levels.

4.1 Predicting of Diabetes Mellitus Type 2 Risk Using Ordinal Nonparametric Logistic Regression Based on Smoothing Spline Estimator

The predictive result of the estimation model of diabetes mellitus type 2 risk using nonparametric ordinal logistic regression based on smoothing spline estimator can bee seen in the value of the classification accuracy trought a confussion matrix (Table 3).

Table 3: Classification Accuracy Using Smoothing Spline

		Actual		
		Normal	Pre-diabetes mellitus type 2	Diabetes mellitus type 2
Prediction	Normal	6	1	1
	Pre-diabetes mellitus type 2	1	14	0
	Diabetes Mellitus Type 2	3	0	14

$$Accuracy = \frac{6+14+14}{40} X\ 100\% = 85\%$$

Based on the above calculation, the accuracy of the nonparametric ordinal logistic regression model based on smoothing spline estimator is 85%, so it can be seen that the nonparametric ordinal logistic regression model based on the smoothing spline estimator can explain 85% of the occurrence of diabetes mellitus type 2 for the sample data.

$$Sensitivity_{(diabetesmellitustype\ 2)} = \frac{14}{1+0+14} = 0.93$$

The sensitivity value for the category of diabetes mellitus type 2 is 0.93. It explains that the probability of a person being declared as a diabetes mellitus type 2 patient and suffering from diabetes mellitus type 2 is 0.93.

To determine the stability of the classification or the accuracy of the model, a Press's Q test is needed. The value of Press's Q in the nonparametric ordinal logistic regression

(GAM) model based on smoothing spline estimator is as follows:

$$Press'sQ = \frac{(40-(34x3))^2}{40(3-1)} = 48.05$$

Because the value of Press's Q is greater than the value of $x^2_{(0.05,1)} = 3.84$ it can be concluded that the nonparametric ordinal logistic regression (GAM) model based on the smoothing spline estimator is stable or consistent.

After getting the predicted values from model, then the deviance value is calculated as the model fitness criteria. Deviance values for nonparametric ordinal logistic regression model can be summarized in the following table:

Table 4: Deviance value for nonparametric ordinal logistic regression model

Model	Deviance Value
Nonparametric ordinal logistic regression based on smoothing spline estimator	58.50

Because the value of deviance is smaller than the value of $x^2_{(0.05,74)} = 95.08$ it can be concluded that the nonparametric ordinal logistic regression (GAM) model based on the smoothing spline estimator is fit.

4.2 Comparation of Prediction of Risk Diabetes Mellitus Type 2 Using Parametric Ordinal Logistic Regression (GLM)

If we estimate the risk of diabetes mellitus type 2 using parametric ordinal logistic regression, we get the estimation equation:

$$\hat{\pi}_0(x_i) = \frac{\exp(-0.604+0.007X_1-0.002X_2)}{1+\exp(-0.604+0.007X_1-0.002X_2)} \quad (24)$$

$$\hat{\pi}_1(x_i) = \frac{\exp(1.030+0.007X_1-0.002X_2)}{1+\exp(1.030+0.007X_1-0.002X_2)} - \frac{\exp(-0.604+0.007X_1-0.002X_2)}{1+\exp(-0.604+0.007X_1-0.002X_2)} \quad (25)$$

$$\hat{\pi}_2(x_i) = 1 - \frac{\exp(1.030+0.007X_1-0.002X_2)}{1+\exp(1.030+0.007X_1-0.002X_2)} \quad (26)$$

The following is the estimated value of $\hat{\pi}_k(x_i)$which is the largest and will be classified in the k-th category

The value of the classification accuracy of the ordinal logistic regression model using a parametric approach (GLM) through a confusion matrix (Table 5)

Table 5: Classification Accuracy Using Parametric

		Actual		
		Normal	Pre-diabetes mellitus type 2	Diabetes mellitus type 2
Prediction	Normal	0	0	0
	Pre-diabetes mellitus type 2	4	8	8
	Diabetes Mellitus Type 2	6	7	7

$$Accuracy = \frac{8+7}{40} X\ 100\% = 37.5\%$$

Based on the above calculation, the accuracy of the parametric ordinal logistic regression model is 37.5%, so it can be seen that the parametric ordinal logistic regression model can explain 37.5% of the occurrence of diabetes mellitus type 2 for the sample data.

$$Sensitivity_{(diabetes\ mellitustype\ 2)} = \frac{7}{0+8+7} = 0.47$$

The sensitivity value for the category of diabetes mellitus type 2 is 0.47. it explains that the probability of a person being declared as a diabetes mellitus type 2 patient and suffering from diabetes mellitus type 2 is 0.47.

The value of Press's Q in the parametric ordinal logistic regression (GLM) model is as follows:

$$Press'sQ = \frac{(40-(15x3))^2}{40(3-1)} = 0.31$$

Because the value of Press's Q is smaller than the value of $x^2_{(0.05,1)} = 3.84$, thus it can be concluded that the resulting parametric ordinal logistic regression (GLM) model is unstable or inconsistent.

Deviance values for parametric ordinal logistic regression model can be summarized in the following table:

Table 6: Deviance value for parametric ordinal logistic regression model

Model	Deviance Value
Parametric ordinal logistic regression	85.82

Because the value of deviance is smaller than the value of $x^2_{(0.05,74)} = 95.08$ it can be concluded that the parametric ordinal logistic regression (GLM) model is fit.

5. CONCLUSIONS

Prediction of diabetes mellitus type 2 risk is effected by LDL cholesterol and triglycerides by using the proposed model approach, that is, additive nonparametric ordinal logistic regression model (GAM) based on smoothing spline estimator is better than parametric ordinal logistic linear regression model (GLM). Based on the validity test that is Press's Q value statistics, the proposed model ordinal logistic nonparametric (GAM) approach is valid to predict diabetes mellitus type 2 risk by cholesterol LDL and triglycerides, but the logistic parametric regression model (GLM) approach is not valid

6. ACKNOWLEDGMENTS

The authors thank patient Haji Hospital Surabaya as respondents for helping in this research such that this research could run smoothly. The authors also thank Airlangga University for supporting and facilitating this research.

7. REFERENCES

[1] World Health Organization. (2019). Classification of Diabetes Mellitus 2019. Retrieved 20 September 2021, from https://www.who.int/publications/i/item/classification-of-diabetes-mellitus

[2] Kompas.(2021).Pasien diabetes meningkat selama pandemic, Indonesia peringkat 7 tertinggi di dunia. Retrieved 19 September 2021, from https://www.kompas.com/sains/read/2021/04/07/110100223/pasien-diabetes-meningkat-selama-pandemi-indonesia-peringkat-7-tertinggi?page=all

[3] Ceriello, A.(2020). Hyperglycemia and COVID-19: What was known and what is really new?,diabetes research and clinical practice, 167, 108383

[4] Arifin, A.Y., Ernawati, F., and Prihatini, M. (2019). Hubungan Kadar Glukosa Darah Terhadap Peningkatan Kadar Lemak Darah Pada Populasi Studi Kohor Kecamatan Bogor Tengah 2018. Jurnal Biotek Medisiana Indonesia, **8**(2), 87-93

[5] American Diabetes Association.(2012). Diagnosis and Classification of Diabetes Mellitus.Diabetes Care, S64-71

[6] Nainggolan, O., Kristanto, A.Y., and Edison, H.(2013). Determinan Diabetes Mellitus Analisis Baseline Data Studi Kohort Penyakit Tidak Menular Bogor 2011.Buletin Penelitian Sistem Kesehatan, **16**(3), 331-339

[7] Hosmer, D.W., Lemeshow, S., and Studivant, R.X. (2013).Applied Logistic Regression (Vol. 398). John Wiley &Sons : New York

[8] Lin, P.,Cabrera, J., Denis Y. W., Yang, F. and Tsui, K. L. (2020). SOH Estimation and SOC Recalibration of Lithium-Ion Battery with Incremental Capacity Analysis & Cubic Smoothing Spline.Journal of The Electrochemical Society, 167(9)

[9] Lestari, B., Fatmawati, Budiantara,I. N. and Chamidah, N. (2020). Estimation of Regression Function in Multi-Response Nonparametric Regression Model Using Smoothing Spline and Kernel Estimators. Journal of Physics: Conference Series, 1097

[10] Fatimah, R. N.(2015). Diabetes Melitus tipe 2.Jurnal Majority, 4(5), 93-101

[11] Ministry of Health of Republic of Indonesia. (2018). Apa Itu Kolestrol. Kemenkes RI: Jakarta. Retrieved 13 October 2021, fromhttps://www.p2ptm.kemkes.go.id/infographic-p2ptm/hipertensi-penyakit-jantung-dan-pembuluh-darah/apa-itu-kolestrol

[12] Hastie, T., and Tibshirani, R. (1987). Non-parametric Logistic and Proportional Odds Regression.Journal of the Royal Statistikal Society : Series C (Applied Statistiks), 36(3), 260-276

[13] Aydın, D.(2007). A Comparison of the Nonparametric Regression Models using Smoothing Spline and Kernel Regression.World Academy of Science, Engineering and Technology, 36, 253-257

[14] Nelder, J.A., and Wedderburn. (1972). Generalized Linier Model. Journal of the Royal Statistical Society. Series A (General), 135(3)

[15] Meyers, L. S., Gamst, G. and Guarino, A. J.(2016). Applied Multivariate Research. SAGE Publication, London

[16] Agresti, A.(2002).Categorical Data Analysis 2 nd edition. John Wiley & Sons, New York

[17] Ortega, J. H.J.C., Lagman, A. C., Natividad, L. R. Q., Bantug, E. T., Resureccion, M. R., & Manalo, L. O.(2020). Analysis of Performance of Classification Algorithm in Mushroom Poisonous Detection using Confusion Matrix Analysis. International Journal, 9, 1.3

[18] Wardani, L. K. (2020). Perbandingan Seleksi Fitur Term Frequency & Tri-Gram Character Menggunakan Algoritma Naïve Bayes Classifier (Nbc) pada Tweet Hashtag #2019gantipresiden, Journal Kajian Ilmu dan Teknologi, 9(1), 103-114

[19] Han, J., Kamber, M., and Pei, J.(2012). Data Mining: Concepts and Techniques, 3 rd edition.Morgan Kaufmann Publishers Inc, United States of America