Data and Metadata. 2025; 4:912
doi: 10.56294/dm2025912

ORIGINAL

Nonparametric Bi-Response Ordinal Logistic Regression Model for Diabetes Mellitus and Hypertension Risks Based on Multivariate Adaptive Regression Spline

Modelo de regresión logística ordinal birrespuesta no paramétrica para los riesgos de diabetes mellitus e hipertensión basado en spline de regresión adaptativa multivariante

Maylita Hasyim[1,2], Nur Chamidah[3,4], Toha Saifudin[3,4], Budi Lestari[5]

[1]Airlangga University, Doctoral Study Program of Mathematics and Natural Sciences, Faculty of Science and Technology. Surabaya 60115, Indonesia.
[2]Universitas Bhinneka PGRI, Study Program of Mathematics Education, Faculty of Social and Humanities. Tulungagung 66221, Indonesia.
[3]Airlangga University, Department of Mathematics, Faculty of Science and Technology. Surabaya 60115, Indonesia.
[4]Airlangga University, Research Group of Statistical Modeling in Life Science, Faculty of Science and Technology. Surabaya 60115, Indonesia.
[5]University of Jember, Department of Mathematics, Faculty of Mathematics and Natural Sciences. Jember 68121, Indonesia.

Cite as: Hasyim M, Chamidah N, Saifudin T, Lestari B. Nonparametric Bi-Response Ordinal Logistic Regression Model for Diabetes Mellitus and Hypertension Risks Based on Multivariate Adaptive Regression Spline. Data and Metadata. 2025; 4:912. https://doi.org/10.56294/dm2025912

Submitted: 10-07-2024 **Revised:** 20-11-2024 **Accepted:** 27-04-2025 **Published:** 28-04-2025

Editor: Dr. Adrián Alejandro Vitón Castillo

Corresponding Author: Nur Chamidah

ABSTRACT

This study discusses the application of nonparametric regression for bi-response ordinal logistic modeling based on the Multivariate Adaptive Regression Spline (MARS) estimator in assessing the risk of diabetes mellitus and hypertension. The MARS estimator provides greater flexibility by allowing for nonlinearity and interactions among predictors, making it well-suited for modeling health-related risk factors. Parameter estimation in this study is conducted using the Maximum Likelihood Estimation (MLE) method. However, due to the non-linearity of the first derivative of the log-likelihood function, the Berndt-Hall-Hall-Hausman (BHHH) numerical iteration method is applied to obtain parameter estimates. The complexity of the likelihood function poses challenges in constructing the Hessian matrix, necessitating an approximation of the second derivative using the first derivative in the BHHH method. The analysis identifies Age, Body Mass Index (BMI), and Total Cholesterol as significant predictor variables influencing the risk of diabetes mellitus and hypertension. Model evaluation is carried out using accuracy, the Area Under the Curve (AUC), and the Apparent Error Rate (APER). The results demonstrate an accuracy of 82,44 %, indicating strong classification performance. Additionally, the AUC value of 73,42 % suggests the model falls within the good category, while the APER value of 17,56 % confirms the model's stability and reliability. The findings suggest that the MARS-based bi-response ordinal logistic regression model effectively captures the relationship between significant risk factors of diabetes mellitus and hypertension.

Keywords: Nonparametric Bi-Response Ordinal Logistic Regression; Diabetes Mellitus; Mars; Bmi; High Blood Pressure.

RESUMEN

Este estudio analiza la aplicación de la regresión no paramétrica para el modelado logístico ordinal de birespuesta basado en el estimador Spline de Regresión Adaptativa Multivariante (SRAM) para evaluar el riesgo de diabetes mellitus e hipertensión. El estimador SRAM proporciona mayor flexibilidad al permitir la no linealidad y las interacciones entre predictores, lo que lo hace adecuado para modelar factores de riesgo relacionados con la salud. La estimación de parámetros en este estudio se realiza utilizando el método de Estimación de Máxima Verosimilitud (EMV). Sin embargo, debido a la no linealidad de la primera derivada de la función de log-verosimilitud, se aplica el método de iteración numérica Berndt-Hall-Hall-Hausman (BHHH) para obtener estimaciones de parámetros. La complejidad de la función de verosimilitud plantea desafíos en la construcción de la matriz Hessiana, lo que requiere una aproximación de la segunda derivada utilizando la primera derivada en el método BHHH. El análisis identifica la edad, el Índice de Masa Corporal (IMC) y el colesterol total como variables predictoras significativas que influyen en el riesgo de diabetes mellitus e hipertensión. La evaluación del modelo se realiza mediante la precisión, el Área Bajo la Curva (ABC) y la Tasa de Error Aparente (TEA). Los resultados demuestran una precisión del 82,44 %, lo que indica un excelente rendimiento de clasificación. Además, el valor del ABC del 73,42 % sugiere que el modelo se encuentra en la categoría de bueno, mientras que el valor de TEA del 17,56 % confirma su estabilidad y fiabilidad. Los hallazgos sugieren que el modelo de regresión logística ordinal de birespuesta basado en SRAM captura eficazmente la relación entre los factores de riesgo significativos de diabetes mellitus e hipertensión.

Palabras clave: Regresión Logística Ordinal de Birespuesta no Paramétrica; Diabetes Mellitus; SRAM; IMC; Presión Arterial Alta.

INTRODUCTION

In regression analysis, a nonparametric regression approach is used when the shape of the relationship between the response variable and predictor variables is not assumed to be a specific pattern.(1,2) This approach it does not rely on the assumption of a specific curve shape, thus providing greater flexibility.(3,4) The model estimation of the relationship pattern is derived from the observed pattern in the data.(5,6) The ability of nonparametric regression to find the shape of the regression curve pattern is supported by the existence of parameters in each type of nonparametric regression approach which makes the estimation of the regression curve pattern more flexible.(7) The nonparametric regression methods that provide flexibility in parameter estimation have been developed so far, including a spline approach consisting of a truncated spline(8,9,10,11,12,13) and Multivariate Adaptive Regression Spline (MARS).(14,15,16) Spline methods is mostly developed because it has excellent flexibility and interpretation among other nonparametric regression methods or approaches. According to (17), spline is a polynomial function that has segmented properties. With this segmented nature, splines are able to provide more flexibility than ordinary polynomials.(18) Therefore, splines have statistical properties that are useful for analyzing relationships in regression.(19,20,21) Spline in nonparametric regression continues to evolve until the adaptive model, where this model has the ability to adjust better in following the shape of the data pattern. The adaptive computation approach in the development of nonparametric regression has been much in demand and applied, one of which is Multivariate Adaptive Regression Spline (MARS).(14)

The MARS method is an adaptive approach that combines spline and Recursive Partitioning Regression (RPR).(15) When there are several predictors involved, the spline approach is limited in its ability to determine the location and quantity of knots employed. The knot selection process on the truncated spline will produce so many combinations regarding the number of predictors, knot positions, and also the number of knots. Since the MARS determines knots through an adaptive process rather than seeking them individually from the combination, it can overcome the shortcoming of the truncated spline in this instance. The adaptive process in the MARS is carried out using a stepwise algorithm, consisting of forward and backward steps. In the forward stepwise process, the MARS method constructs a model by incorporating truncated spline basis functions (knots and interactions) to achieve the maximum number of base functions. The backward stepwise process then refines the model by selecting the most influential basis functions from the forward stepwise stage, aiming to create a more parsimonious model. This selection is based on minimizing the Generalized Cross Validation (GCV) value to improve the estimation of the response variable.(16) The MARS also has the benefit of being appropriate for high-dimensional data instances since it can handle interactions between predictor variables represented by basis function. The MARS model is able to cover the weakness of the RPR model which is not continuous at the knot, because the basis function in the selected MARS model is polynomial with a continuous derivative at each knot point.(7)

The MARS modeling has been developed depending on the type of response can be divided into continuous and categorical response regression models. According to (17) mentioned that the MARS is also a modern statistical classification method that has utilized the flexibility of the model and estimated a distribution within each class that ultimately provides a clustering rule. Thus, the MARS is also suitable for the case of calculating the accuracy of data classification that requires the response variable to be categorical. The MARS method with categorical responses (binary and ordinal) can serve as a modern statistical classification method, where classification in the MARS is based on the logistic regression approach. Logistic regression is an analysis used to see the relationship between categorical response variables and categorical and continuous predictor variables.[22] The logistic regression equation is obtained from the estimated form of the probability function of a success event or a certain event occurring, which then on this probability function is carried out logit transformation so that a logit link function is formed. This logit link function is the MARS model, or referred to as the logit MARS model.

Previous researchers who examined the MARS method with a single response in the form of categories include Kishartini et al.[23] who applied the MARS method to classify work status, Annur et al.[24] who applied the MARS method to determine the factors that influence student study period, Binadari et al.[25] who compared logistic regression modeling with MARS applied to the response of major interest, and Serrano et al.[26] who applied the MARS method to identify gender differences. Meanwhile, previous researchers who applied the MARS approach with a single continuous response include Nisai and Budiantara[27] who modeled Dengue Fever (DHF) cases using survival analysis with the MARS approach, Otok et al.[16] modeled the lecturer performance index using survival analysis with the MARS approach, and Wang et al.[28] analyzed the probabilistic stability of earth dam slopes using MARS.

The studies mentioned above developed or applied the MARS method in the case of a single response only. In the real cases, we often find cases where we must employ the MARS approach with more than one response, such as bi-response MARS and multi-response MARS models. The bi-response or multi-response regression model consists of several equations with the assumption that there is a correlation between responses. In this case, we can accommodate this correlation by using a covariance matrix that is used as a weight matrix when estimating model parameters.[29] Several studies that discuss and apply the MARS method to the case of multi-response regression models are modeling welfare indicators in Java using bi-response MARS by Ampulembang et al.[29], developing the MARS model in the form of a multivariate response and making its application by Milborrow[30], development completion of continuous bi-response nonparametric regression models using the MARS method by Ampulembang[31], and modeling bi-response MARS using earth package for regression problems by Eyduran et al.[32]. Therefore, the novelty of this research lies in the theoretical development of parameter estimation in the bi-response ordinal logistic nonparametric regression model using the MARS estimator. Furthermore, it will be developed from the application aspect, namely designing algorithms and programs to apply the parameter estimation theory of the ordinal logistic nonparametric regression model based on the MARS estimator in modeling the risk of non-communicable diseases such as diabetes mellitus and hypertension.

Hypertension and diabetes mellitus are non-communicable diseases that are the main burden that the Indonesian government must solve, considering that the prevalence and causes of death due to these diseases are increasing every year.[33] Diabetes mellitus is a serious chronic disease that occurs when the pancreas does not produce enough insulin or when the body cannot effectively use the insulin it produces.[34] The global report on diabetes explains that the number of cases and prevalence of diabetes mellitus has continued to increase over the past few decades.[35] This data is supported by the Riskesdas 2018 which shows that diabetes mellitus is ranked fourth in the group of non-communicable diseases in Indonesia.[33] Based on the gender category, people with diabetes mellitus in Indonesia are more female (1,8 %) than male (1,2 %). The data above show that age and gender are factors that are thought to affect the risk of diabetes mellitus. In addition, physical conditions such as body mass index (BMI) determined by weight and height, cholesterol and uric acid are also thought to affect the risk of diabetes mellitus. In patients with type-2 diabetes mellitus, an increase in blood sugar levels often occurs along with an increase in blood pressure.[36] According to Waeber et al.[37], hypertension is a major risk factor for diabetes mellitus. Diabetes mellitus and hypertension cannot be cured but can be controlled, and there is a significant relationship between them. Thus, theoretically and scientifically, diabetes mellitus and hypertension have a correlation (relationship).[36]

These studies mentioned above generally model hypertension and diabetes mellitus risk data separately as a uni-response regression model, there is only one study by Hardine et al.[38] which modeled diabetes mellitus and hypertension as a bi-response nonparametric regression model. The research about nonparametric regression modeling conducted by Hardine et al.[38], has not accommodated interactions between predictor variables and the number of predictor variables is only one. Meanwhile this research uses the MARS approach in estimating the parameters of the nonparametric logistic regression model which is able to accommodate interactions between predictor variables through the basis function and also the number of predictor variables is more variety. Based on this fact, it is important to model diabetes mellitus and hypertension as a bi-response case, because there is a correlation between these two variables. Factors that are thought to affect the risk of diabetes mellitus and hypertension and the interaction between factors will be accommodated in the nonparametric bi-response ordinal logistic regression model based on the MARS estimator. On the other hand, the MARS method is also able to analyze the classification accuracy of the resulting nonparametric regression model. This modeling is expected to be useful for the government and stakeholders in the health sector in determining preventive effort policies to minimize the incidence of diabetes mellitus and hypertension in Indonesia, while for the community it is expected to add insight in managing lifestyles to avoid diabetes mellitus and hypertension.

METHOD

Data Set and Research Variables

The data used in this study are secondary data obtained from the website: https://www.kaggle.com/datasets/tourdeglobe/fatty-liver-disease. The data is a dataset collected in a patient program undergoing medical examination, specifically for patient data who are declared to have diabetes mellitus and/or hypertension.

Table 1. Dataset of DM Cases

Patient Number	Diabetes Mellitus Status (Y_1)	Hypertension Status (Y_2)	Age (X_1)	Gender (X_2)	Body Mass Index (X_3)	Total Cholesterol (X_4)
1	1	2	53	2	34,95	103
2	1	2	33	2	31,02	102
3	1	2	23	2	25,91	144
⋮	⋮	⋮	⋮	⋮	⋮	⋮
664	3	2	32	2	48.23	200

Details about the dataset utilized in this study are given in table 1. The response variables in this dataset include the presence of hypertension and diabetes mellitus. Next, we present information on factors that are believed to affect diabetes mellitus and hypertension, including age, gender, body mass index, and total cholesterol users. There are 664 patients in this data set. Here, category 1 denotes normal, category 2 denotes stage-1 diabetes mellitus, and category 3 denotes stage-2 diabetes mellitus. The diabetes mellitus status is as the first response variable (Y_1) on an ordinal scale.[39] The second response variable, namely the status of hypertension (Y_2) is on an ordinal scale, where 1 represents normal, 2 represents stage-1 hypertension, and 3 represents stage-2 hypertension.[33] There is an interval scale for the age variable (X_1). Next, category 1 denotes male and category 2 denotes female, and the gender variable (X_2) has a nominal scale. Furthermore, the body mass index variable (X_3) and total cholesterol X_4 have a scale of ratio.

Bi-response Ordinal Logistic Regression

When the response variables are polychotomous and have an ordinal scale, ordinal logistic regression is a regression analysis used to examine the relationship between the predictor and response variables.[40] The cumulative logit model can be used for ordinal logistic regression. In this model, the ordinal response variable Y is expressed in cumulative probability. The cumulative probability Y is expressed as follows:[22]

$$P(Y \leq r|X_j) = \pi(x_j) = \frac{exp\left(\theta_r + \sum_{j=1}^{p} \alpha_j x_{ji}\right)}{1 + exp\left(\theta_r + \sum_{j=1}^{p} \alpha_j x_{ji}\right)} \quad (1)$$

Where x_j=(x_1i,x_2i,…,x_{pi}) : the i-th (i=1,2,…,n)observation predictor variable for each p predictor variable, while r=1,2,…,r is the response variable category. Equation (1) is a proportional odds model, where each cumulative logit model has a different intercept θ but the same effect α_j.

Estimating ordinal logistic regression parameters involves decomposing them using the logit transformation $P(Y\leq r|X_j)$ in equation (1), which is described by the following equation:[22]

$$g_r(x) = \text{logit}\, P(Y \leq r|X_j) = ln\left(\frac{P(Y \leq r|X_j)}{1 - P(Y \leq r|X_j)}\right) = \theta_r + \sum_{j=1}^{p} \alpha_j x_{ji} \quad (2)$$

For example, the cumulative probability of the r -th category response is explained by the following equations, if there are three response categories, namely r=1,2,3.

$$P(Y \leq 1|x_j) = \frac{exp\left(\theta_1+\sum_{j=1}^{p} \alpha_j x_{ji}\right)}{1+exp\left(\theta_1+\sum_{j=1}^{p} \alpha_j x_{ji}\right)} \text{ and } P(Y \leq 2|x_j) = \frac{exp\left(\theta_2+\sum_{j=1}^{p} \alpha_j x_{ji}\right)}{1+exp\left(\theta_2+\sum_{j=1}^{p} \alpha_j x_{ji}\right)} \quad (3)$$

If equation (2) is applied to three response categories, namely r=1,2,3, then the cumulative logit model for each response category can be described as follows:

$$\hat{g}_1(x)=\ln\left(\frac{P(Y\le 1|x)}{1-P(Y\le 1|x)}\right)=\ln\left(\frac{P(Y\le 1|x)}{P(Y>1|x)}\right)=\theta_1+\left(\alpha_1X_1+\alpha_2X_2+\ldots+\alpha_pX_p\right)$$
$$\hat{g}_2(x)=\ln\left(\frac{P(Y\le 2|x)}{1-P(Y\le 2|x)}\right)=\ln\left(\frac{P(Y\le 2|x)}{P(Y>2|x)}\right)=\theta_2+\left(\alpha_1X_1+\alpha_2X_2+\ldots+\alpha_pX_p\right) \quad (4)$$

The bi-response ordinal logistic regression model develops the ordinal logistic regression model, where two ordinal scale response variables are correlated. For example, the first response variable is denoted by   and has as many as   categories, while denotes the second response variable and has as many as   categories, then the bi-response ordinal logistic regression model is expressed as follows:[41]

$$\tilde{g}_1(x)=\text{logit}\big(P(Y_1\le a|x)\big)=\text{logit}\big(F_a(x)\big)=ln\left(\frac{F_a(x)}{1-F_a(x)}\right)=\theta_{1a}+\alpha_1^Tx$$
$$\tilde{g}_2(x)=\text{logit}\big(P(Y_2\le b|x)\big)=\text{logit}\big(F_b(x)\big)=ln\left(\frac{F_b(x)}{1-F_b(x)}\right)=\theta_{2b}+\alpha_2^Tx \quad (5)$$

Where a=1,2,...,A-1 and b=1,2,...,B-1; {θ_1a,θ_2b,Δ_{ab} } is intercept parameters that meet the requirement $\theta_{11}\le\theta_{12}\le\ldots\le\theta_1$a and $\theta_{21}\le\theta_{22}\le\ldots\le\theta_2$b;x=[(x$_1$ x$_2$...x$_k$)]^Tis a vector of predictor variables; a_1=[(a_{11} a_{12}...a_1k)]^T and a_2=[(a_{21} a_{22}...a_2k)]^T are vetors of parameters; $F_{a\bullet}$ (x)=P(Y$_1$≤a|x) is the marginal cumulative probability of variable Y$_1$ being less than or equal to category-a with respect to x; and $F_{\bullet b}$ (x)=P(Y$_2$≤b|x) is the marginal cumulative probability of variable Y$_2$ being less than or equal to category-b with respect to x.

Thus, the marginal cumulative probabilities $F_{a\bullet}$ (x) and $F_{\bullet b}$ (x) are obtained as follows:[41]

$$F_{a\bullet}(\boldsymbol{x})=\frac{exp(\theta_{1a}+\alpha_1^T\boldsymbol{x})}{1+exp(\theta_{1a}+\alpha_1^T\boldsymbol{x})} \text{ and } F_{\bullet b}(\boldsymbol{x})=\frac{exp(\theta_{2b}+\alpha_2^T\boldsymbol{x})}{1+exp(\theta_{2b}+\alpha_2^T\boldsymbol{x})} \quad (6)$$

The Maximum Likelihood Estimation (MLE) method can be used to estimate parameters for the bi-response ordinal logistic regression model.[42] For example, suppose that (x$_1$i,x$_2$i,...,x$_{pi}$,y$_1$i,y$_2$i)is paired data from n independent random samples, (x$_1$i,x$_2$i,...,x$_{pi}$) is data from   predictor variables, and (y$_1$i,y$_2$i) is data from two categorical response variables on an ordinal scale. Next, Y$_1$ has as many as A categories, and Y$_2$ has as many as B categories. Then, there are Y$_{abi}$ random variables with a multinomial distribution for each probability of π_{abi}. Thus, the joint probability density function between variables Y$_1$ and Y$_2$ is given by the following equation:

$$P(Y_{11i}=y_{11i},\ldots,Y_{ABi}=y_{ABi})=\prod_{a=1}^{A}\prod_{b=1}^{B}\pi_{abi}^{y_{abi}} \quad (7)$$

Hence, the likelihood function is obtained as follows:

$$L(\boldsymbol{\mu})=\prod_{i=1}^{n}\prod_{a=1}^{A}\prod_{b=1}^{B}\pi_{abi}^{y_{abi}} \quad (8)$$

Where, $\mu=[\theta_1^{(1)}\ \theta_2^{(1)}\ \theta_1^{(2)}\ \theta_2^{(2)}\ \alpha_0^{(1)}\ \alpha_1^{(1)}\ \alpha_2^{(1)}\ldots\alpha_p^{(1)}\ \alpha_0^{(2)}\ \alpha_1^{(2)}\ \alpha_2^{(2)}\ldots\alpha_p^{(2)}]^T$

The principle of the MLE method is to estimate the parameters of the bi-response ordinal logistic regression model by maximizing the likelihood function. To simplify the calculation, a logarithm transformation is performed on the likelihood function as follows:

$$\begin{aligned}\boldsymbol{\ell}(\boldsymbol{\mu})=ln\,L\,(\boldsymbol{\mu})&=\sum_{i=1}^{n}\sum_{a=1}^{A}\sum_{b=1}^{B}y_{abi}\,ln(\pi_{abi})\\&=\sum_{i=1}^{n}[y_{11i}\,ln\,\pi_{11i}+y_{12i}\,ln\,\pi_{12i}+\ldots+y_{ABi}\,ln\,\pi_{ABi}]\end{aligned} \quad (9)$$

Based on equation (9), the next step is to perform the first partial derivative of the ln-likelihood function with respect to the parameters, and then set it to zero. The results of the first partial derivative obtained are nonlinear functions on the parameters to be estimated, so a numerical method is needed to obtain parameter estimates, namely using the Berndt-Hall-Hall-Hausman (BHHH) iteration method.

Multivariate Adaptive Regression Spline (MARS)

The MARS is a nonparametric regression method with an adaptive approach that combines truncated spline regression and RPR. The MARS method can overcome the weaknesses of truncated splines because the determination of knots in the MARS is done through an adaptive process. The adaptive process in the MARS was carried out using a stepwise algorithm, which included forward and backwards. The MARS also overcomes the weaknesses of the RPR method, which is not continuous at the knot point. The advantage of the MARS model is that it can model high-dimensional data and accommodate interactions between predictor variables. The MARS model is obtained from the forward and backward stepwise algorithm as follows:[19]

$$+\sum_{m=1}^{M}\alpha_m\prod_{k=1}^{K_M}\left[S_{km}\left(x_{v(k,m)}-t_{km}\right)\right] \quad (10)$$

Where α_0 is main of basis function, α_m is coefficient of basis function m, M is maximum of basis function (non-constant basis function), K_m is degree of interaction, $x_{v(k,m)}$ is predictor variable, t_{km} is knot point, and:

$$S_{km}=\begin{cases}1 & \text{if the data is on the right hand side of the knot point}\\ -1 & \text{if the data is on the left hand side of the knot point}\end{cases}$$

The MARS is also a modern statistical classification method that utilize model flexibility and estimates a distribution within each class, ultimately providing a grouping rule.[43] Classification in the MARS is based on the logistic regression approach. Thus, the logit link function of the MARS model is as follows:[19]

$$f(x)=logit\ P(Y\le r|X_j)=\alpha_0+\sum_{m=1}^{M}\alpha_m\prod_{k=1}^{K_M}\left[S_{km}\left(x_{v(k,m)}-t_{km}\right)\right] \quad (11)$$

In the MARS, selecting the optimum (the best) model is based on the Generalized Cross Validation (GCV) value of the model has the most minimum value among the other models.[44] The GCV function is given as follows:

$$GCV(M)=\frac{ASR}{\left[1-\frac{C(\widehat{M})}{n}\right]^2}=\frac{\frac{1}{n}\sum_{i=1}^{n}\left[f(x_i)-\hat{f}_M(x_i)\right]^2}{\left[1-\frac{C(\widehat{M})}{n}\right]^2} \quad (12)$$

Where f(x_i) is response variable, f_M (x_i) is the estimated value of response variable on M basis function, n is the number of observation, K_m is degree of interaction, C(M)=C(M)+dM, d is value when each basis function reaches the optimization that are d=2 (for additive model) and d=3 (for interaction model).

Evaluation of Classification Procedures

Table 2. The (9×9)-Confusion Matrix

Actual	**Prediction**								
	y_{11}	y_{12}	y_{13}	y_{21}	y_{22}	y_{23}	y_{31}	y_{32}	y_{33}
y_{11}	n_{1111}	$n_{11,12}$	$n_{11,13}$	$n_{11,21}$	$n_{11,22}$	$n_{11,23}$	$n_{11,31}$	$n_{11,32}$	$n_{11,33}$
y_{12}	$n_{12,11}$	$n_{12,12}$	$n_{12,13}$	$n_{12,21}$	$n_{12,22}$	$n_{12,23}$	$n_{12,31}$	$n_{12,32}$	$n_{12,33}$
y_{13}	$n_{13,11}$	$n_{13,12}$	$n_{13,13}$	$n_{13,21}$	$n_{13,22}$	$n_{13,23}$	$n_{13,31}$	$n_{13,32}$	$n_{13,33}$
y_{21}	$n_{21,11}$	$n_{21,12}$	$n_{21,13}$	$n_{21,21}$	$n_{21,22}$	$n_{21,23}$	$n_{21,31}$	$n_{21,32}$	$n_{21,33}$
y_{22}	$n_{22,11}$	$n_{22,12}$	$n_{22,13}$	$n_{22,21}$	$n_{22,22}$	$n_{22,23}$	$n_{22,31}$	$n_{22,32}$	$n_{22,33}$
y_{23}	$n_{23,11}$	$n_{23,12}$	$n_{23,13}$	$n_{23,21}$	$n_{23,22}$	$n_{23,23}$	$n_{23,31}$	$n_{23,32}$	$n_{23,33}$
y_{31}	$n_{31,11}$	$n_{31,12}$	$n_{31,13}$	$n_{31,21}$	$n_{31,22}$	$n_{31,23}$	$n_{31,31}$	$n_{31,32}$	$n_{31,33}$
y_{32}	$n_{32,11}$	$n_{32,12}$	$n_{32,13}$	$n_{32,21}$	$n_{32,22}$	$n_{32,23}$	$n_{32,31}$	$n_{32,32}$	$n_{32,33}$
y_{33}	$n_{33,11}$	$n_{33,12}$	$n_{33,13}$	$n_{33,21}$	$n_{33,22}$	$n_{33,23}$	$n_{33,31}$	$n_{33,32}$	$n_{33,33}$

Source: Fahmy[46]

Classification procedure evaluation is an evaluation that looks at the chances of classification errors made by a classification function.(45) A confusion matrix is formed to evaluate the strength of the model obtained in the classification procedure. The confusion matrix is a table that summarizes the performance of the classification model.(46) The (9×9)-confusion matrix is presented in table 2.

The elements of the confusion matrix are used to find several values of the model's strength, Area Under Cover (AUC), and APER.(47) The Area Under Cover (AUC) measure can evaluate classification with unbalanced data cases. If the data has three or more categories, the average AUC calculation can be used as follows:(48)

$$\left.\begin{aligned} AUC_{Total} &= \frac{2}{c(c-1)}\sum_{a<b} AUC(a,b) \\ AUC(a,b) &= \frac{AUC(a|b)+AUC(b|a)}{2} \end{aligned}\right\} \quad (13)$$

Where c is the number of class:

$$AUC(a|b) = \frac{n_{ab,mn}}{n_{ab,mn}+n_{ab,m(n+1)}+\cdots+n_{ab,mq}}$$
$$AUC(b|a) = \frac{n_{ab,mn}}{n_{ab,mn}+n_{a(b+1)b,mn}+\cdots+n_{aq,mn}}$$

The AUC value criteria are presented in table 3.(49)

Table 3. AUC Value Criteria

AUC Value	Criteria
0,9<AUC≤1,0	Excellent
0,8<AUC≤0,9	Very Good
0,7<AUC≤0,8	Good
0,6<AUC≤0,7	Sufficient
0,5<AUC≤0,6	Bad
AUC<0,5	Test is not useful

Source: Šimundić(49)

Furthermore, the test statistics determine the extent to which the classified groups can be separated using the existing variables that have stability in classification accuracy, APER (Apparent Error Rate) is used. The APER value states the proportion of samples that are incorrectly classified by the classification function formulated as follows:(50)

$$APER(\%) = \frac{Total\ nmber\ of\ misclassified\ samples}{Total\ number\ of\ samples} \quad (14)$$

Analysis Method

This research follows these stages for data analysis:

a. Conduct data exploration from response and predictor variables to determine each research variable's descriptive statistics.

b. Conducting dependency testing between response variables using the Mantel-Haenszel test.

c. Determining the MARS bi-response ordinal regression model.

d. Forming basis functions for each response in the MARS bi-response ordinal regression model.

e. Dividing the data into two parts, namely in-sample data (90 % of the total data) and out-sample data (10 % of the total data).

f. Determining the nonparametric bi-response ordinal logistic regression model with the MARS estimator based on the results of step (d) using in-sample data.

g. Estimating parameters from the bi-response ordinal logistic regression model with the MARS estimator for each response through Berndt-Hall-Hall-Hausman (BHHH) iteration.

h. Evaluating the classification procedure on in-sample data through AUC (according to equation (13)), and APER values (according to equation (14)).

i. Evaluating the model on out-sample data by determining the 9×9 confusion matrix to obtain accuracy, AUC, and APER values.

RESULTS AND DISCUSSION

This section discusses the estimation of bi-response nonparametric logistic regression models with the MARS estimator and implements the estimation theory on diabetes mellitus and hypertension risk data.

Estimation of Bi-response Ordinal Logistic Nonparametric Regression Model Based on MARS Estimator

Given paired data (x_1i,x_2i,...,x_{pi},y_1i) and (x_1i,x_2i,...,x_{pi},y_2i) with i=1,2,...,n, x_1,x_2,...,x_p is predictor variable and y_1 "," y_2 is response variable, n indicates the number of observations. Suppose the relationship between the predictor and response variables is expressed in a regression function f, whose form is unknown and can be approached using a bi-response ordinal logistic regression model. In that case, the following model is obtained:

$$\begin{aligned} y_{1i} &= f_1(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_{1i} \\ y_{2i} &= f_2(x_{1i}, x_{2i}, \dots, x_{pi}) + \varepsilon_{2i} \end{aligned} \qquad (15)$$

The regression functions f_1 and f_2 in equation (15) are nonparametric regression functions whose forms are assumed to be unknown because these functions are approximated by the MARS regression function as follows.

The MARS regression function for response 1 (f_1) is written as follows in equation (16):

$$f^{(1)}(x_{1i}, x_{2i}, \dots, x_{pi}) = \alpha_0^{(1)} + \sum_{m_1=1}^{M_1} \alpha_{m_1}^{(1)} \prod_{k_1=1}^{K_{M_1}} \left[s_{k_1m_1} . \left(x_{v(k_1,m_1)_i} - t_{k_1m_1} \right) \right] \qquad (16)$$

Where:

$$x_{v(k_1,m_1)} \in \{x_j\}_{j=1}^{p}, t_{k_1m_1} \in \left\{ x_{v(k_1,m_1)_i} \right\}_{i=1}^{n}, m_1 = 1,2,\dots,M_1$$

If s_{k1m1}=+1, then:

$$+\left(x_{v(k_1,m_1)i} - t_{k_1m_1}\right)_+ = \begin{cases} x_{v(k_1,m_1)_i} - t_{k_1m_1}, \text{ if } x_{v(k_1,m_1)_i} > t_{k_1m_1} \\ 0, \text{otherwise} \end{cases}$$

If s_{k1m1}=-1, then:

$$-\left(x_{v(k_1,m_1)i} - t_{k_1m_1}\right)_+ = \begin{cases} t_{k_1m_1} - x_{v(k_1,m_1)_i}, \text{ if } t_{k_1m_1} > x_{v(k_1,m_1)_i} \\ 0, \text{otherwise} \end{cases}$$

The MARS regression function for response 2 is written as follows in equation (17):

$$f^{(2)}(x_{1i}, x_{2i}, \dots, x_{pi}) = \alpha_0^{(2)} + \sum_{m_2=1}^{M_2} \alpha_{m_1}^{(2)} \prod_{k_2=1}^{K_{M_2}} \left[s_{k_2m_2} . \left(x_{v(k_2,m_2)} - t_{k_2m_2} \right) \right] \qquad (17)$$

Where:

$$x_{v(k_2,m_2)} \in \{x_j\}_{j=1}^{p}, t_{k_2m_2} \in \left\{ x_{v(k_2,m_2)_i} \right\}_{i=1}^{n}, m_2 = 1,2,\dots,M_2$$

If s_{k1m1}=+1, then:

$$+\left(x_{v(k_2,m_2)i} - t_{k_2m_2}\right)_+ = \begin{cases} x_{v(k_2,m_2)_i} - t_{k_2m_2}, \text{ jika } x_{v(k_2,m_2)_i} > t_{k_2m_2} \\ 0, \text{sebaliknya} \end{cases}$$

If s_{k1m1}=-1, then:

$$-\left(x_{v(k_2,m_2)i} - t_{k_2m_2}\right)_+ = \begin{cases} t_{k_2m_2} - x_{v(k_2,m_2)_i}, \text{ jika } t_{k_2m_2} > x_{v(k_2,m_2)_i} \\ \qquad 0, \text{sebaliknya} \end{cases}$$

Equation (16) and equation (17) then obtain the nonparametric bi-response regression function with the MARS estimator as follows:

$$\begin{aligned} f^{(1)}\left(x_{1i}, x_{2i}, \ldots, x_{pi}\right) &= \alpha_0^{(1)} + \sum_{m_1=1}^{M_1} \alpha_{m_1}^{(1)} B_{m_{1}i}(\underset{\sim}{x}, \underset{\sim}{t}) \\ f^{(2)}\left(x_{1i}, x_{2i}, \ldots, x_{pi}\right) &= \alpha_0^{(2)} + \sum_{m_2=1}^{M_2} \alpha_{m_2}^{(2)} B_{m_{2}i}(\underset{\sim}{x}, \underset{\sim}{t}) \end{aligned} \tag{18}$$

Next, the bi-response ordinal logistic regression model will be defined with the MARS regression function estimator, which has been described in equation (18).

Ordinal logistic regression model with MARS estimator for response variable 1 in equation (19):

$$\begin{aligned} \hat{g}_1^{(1)}(x) &= \ln\frac{P(Y\le 1)}{\left(1-P(Y\le 1)\right)} = \ln\frac{P(Y\le 1)}{P(Y>1)} \\ &= \ln\frac{P(Y=1)}{P(Y=2)+P(Y=3)} = \theta_1^{(1)} + \left[\alpha_0^{(1)} + \sum_{m_1=1}^{M_1}\alpha_{m_1}^{(1)}\prod_{k_1=1}^{K_{M_1}}\left[s_{k_1m_1}.\left(x_{v(k_1,m_1)} - t_{k_1m_1}\right)\right]\right] \\ \hat{g}_2^{(1)}(x) &= \ln\frac{P(Y\le 2)}{\left(1-P(Y\le 2)\right)} = \ln\frac{P(Y\le 2)}{P(Y>2)} \\ &= \ln\frac{P(Y=1)+P(Y=2)}{P(Y=3)} = \theta_2^{(1)} + \left[\alpha_0^{(1)} + \sum_{m_1=1}^{M_1}\alpha_{m_1}^{(1)}\prod_{k_1=1}^{K_{M_1}}\left[s_{k_1m_1}.\left(x_{v(k_1,m_1)} - t_{k_1m_1}\right)\right]\right] \end{aligned} \tag{19}$$

For π_1=P(Y=1);π_2=P(Y=2);π_3=P(Y=3) to response 1, it can be written as follows:

$$\begin{aligned} ln\frac{\pi_1}{\pi_2+\pi_3} &= \theta_1^{(1)} + \left[\alpha_0^{(1)} + \sum_{m_1=1}^{M_1}\alpha_{m_1}^{(1)}\prod_{k_1=1}^{K_{M_1}}\left[s_{k_1m_1}.\left(x_{v(k_1,m_1)} - t_{k_1m_1}\right)\right]\right] \\ ln\frac{\pi_1+\pi_2}{\pi_3} &= \theta_2^{(1)} + \left[\alpha_0^{(1)} + \sum_{m_1=1}^{M_1}\alpha_{m_1}^{(1)}\prod_{k_1=1}^{K_{M_1}}\left[s_{k_1m_1}.\left(x_{v(k_1,m_1)} - t_{k_1m_1}\right)\right]\right] \end{aligned} \tag{20}$$

Meanwhile, the ordinal logistic regression model with the MARS estimator for response variable 2 is written in equation (21):

$$\begin{aligned} \hat{g}_1^{(2)}(x) &= \ln\frac{P(Y\le 1)}{\left(1-P(Y\le 1)\right)} = \ln\frac{P(Y\le 1)}{P(Y>1)} \\ &= \ln\frac{P(Y=1)}{P(Y=2)+P(Y=3)} = \theta_1^{(2)} + \left[\alpha_0^{(2)} + \sum_{m_2=1}^{M_2}\alpha_{m_1}^{(2)}\prod_{k_2=1}^{K_{M2}}\left[s_{k_2m_2}.\left(x_{v(k_2,m_2)} - t_{k_2m_2}\right)\right]\right] \\ \hat{g}_2^{(2)}(x) &= \ln\frac{P(Y\le 2)}{\left(1-P(Y\le 2)\right)} = \ln\frac{P(Y\le 2)}{P(Y>2)} \\ &= \ln\frac{P(Y=1)+P(Y=2)}{P(Y=3)} = \theta_2^{(2)} + \left[\alpha_0^{(2)} + \sum_{m_2=1}^{M_2}\alpha_{m_1}^{(2)}\prod_{k_2=1}^{K_{M_2}}\left[s_{k_2m_2}.\left(x_{v(k_2,m_2)} - t_{k_2m_2}\right)\right]\right] \end{aligned} \tag{21}$$

For π_1=P(Y=1);π_2=P(Y=2);π_3=P(Y=3) to response 2, it can be written as follows:

$$
\begin{aligned}
ln\frac{\pi_1}{\pi_2+\pi_3} &= \theta_1^{(2)} + \left[\alpha_0^{(2)} + \sum_{m_2=1}^{M_2} \alpha_{m_1}^{(2)} \prod_{k_2=1}^{K_{M_2}} \left[s_{k_2m_2}.\left(x_{v(k_2,m_2)} - t_{k_2m_2}\right)\right]\right] \\
ln\frac{\pi_1+\pi_2}{\pi_3} &= \theta_2^{(2)} + \left[\alpha_0^{(2)} + \sum_{m_2=1}^{M_2} \alpha_{m_1}^{(2)} \prod_{k_2=1}^{K_{M_2}} \left[s_{k_2m_2}.\left(x_{v(k_2,m_2)} - t_{k_2m_2}\right)\right]\right]
\end{aligned}
\quad (22)
$$

Next, to determine the likelihood function of the ordinal bi-response random variable, nine random variables are formed, including (y_{11}i,y_{12}i,y_{13}i,y_{21}i,y_{22}i,y_{23}i,y_{31}i,y_{32}i,y_{33}i), which follows a multinomial distribution with each probability (π_{11}i,π_{12}i,π_{13}i,π_{21}i,π_{22}i,π_{23}i,π_{31}i,π_{32}i,π_{33}i). The joint probability density function of variables Y1 and Y2 is:

$$
\begin{aligned}
&P(Y_{11i=}y_{11i}, Y_{12i=}y_{12i}, Y_{13i=}y_{13i}, Y_{21i=}y_{21i}, Y_{22i=}y_{22i}, Y_{23i=}y_{23i}, Y_{31i=}y_{31i}, Y_{32i=}y_{32i}, Y_{33i=}y_{33i}) \\
&= \pi_{11i}^{y_{11i}} \pi_{12i}^{y_{12i}} \pi_{13i}^{y_{13i}} \pi_{21i}^{y_{12i}} \pi_{22i}^{y_{22i}} \pi_{23i}^{y_{23i}} \pi_{31i}^{y_{31i}} \pi_{32i}^{y_{32i}} \pi_{33i}^{y_{33i}}
\end{aligned}
\quad (23)
$$

For estimating the parameters of the bi-response ordinal logistic regression model with the MARS estimator using the Maximum Likelihood Estimation (MLE) method. The principle of the MLE method is to estimate the model parameters, namely:

$$\boldsymbol{\mu} = \left[\theta_1^{(1)} \quad \theta_2^{(1)} \quad \theta_1^{(2)} \quad \theta_2^{(2)} \quad \alpha_0^{(1)} \quad \alpha_1^{(1)} \quad \alpha_2^{(1)} \quad \ldots \quad \alpha_p^{(1)} \quad \alpha_0^{(2)} \quad \alpha_1^{(2)} \quad \alpha_2^{(2)} \quad \ldots \quad \alpha_p^{(2)}\right]^T \quad (24)$$

By maximizing the likelihood function. Here, the likelihood function of the μ parameter can be written:

$$L(\boldsymbol{\mu}) = \prod_{i=1}^{n} \prod_{m=1}^{3} \prod_{n=1}^{3} \pi_{mni}^{y_{mni}} = \prod_{i=1}^{n} \pi_{11i}^{y_{11i}} \pi_{12i}^{y_{12i}} \pi_{13i}^{y_{13i}} \pi_{21i}^{y_{12i}} \pi_{22i}^{y_{22i}} \pi_{23i}^{y_{23i}} \pi_{31i}^{y_{31i}} \pi_{32i}^{y_{32i}} \pi_{33i}^{y_{33i}} \quad (25)$$

To simplify the calculation, an ln transformation is performed on the likelihood function to form the following ln-likelihood function:

$$
\begin{aligned}
\ell\left(\grave{\mathbf{i}}\right) = \ln L\left(\grave{\mathbf{i}}\right) &= \sum_{i=1}^{n} \begin{bmatrix} y_{11i}\ln\pi_{11i} + y_{12i}\ln\pi_{12i} + y_{13i}\ln\pi_{13i} + y_{21i}\ln\pi_{21i} + y_{22i}\ln\pi_{22i} \\ +y_{23i}\ln\pi_{23i} + y_{31i}\ln\pi_{31i} + y_{32i}\ln\pi_{32i} + y_{33i}\ln\pi_{33i} \end{bmatrix} \\
&= \sum_{i=1}^{n}\sum_{a=1}^{3}\sum_{b=1}^{3} y_{abi}\ln\pi_{abi}
\end{aligned}
\quad (26)
$$

The next step, the ln-likelihood function in equation (27), is performed as the first partial derivative on the parameters to be estimated and then equated to zero. The results of the log-likelihood function derivative on the parameters are described as follows:

a. The first derivative is obtained from the estimated results of the $\theta_1^{(1)}$ parameter as follows:

$$\frac{\partial\boldsymbol{\ell}(\boldsymbol{\mu})}{\partial\theta_1^{(1)}} = \sum_{i=1}^{n} \left[\begin{matrix} \left(\frac{y_{11i}}{\pi_{11i}} - \frac{y_{12i}}{\pi_{12i}} - \frac{y_{21i}}{\pi_{21i}} + \frac{y_{22i}}{\pi_{22i}}\right)\rho_{11i} + \left(\frac{y_{12i}}{\pi_{12i}} - \frac{y_{13i}}{\pi_{13i}} - \frac{y_{22i}}{\pi_{22i}} + \frac{y_{23i}}{\pi_{23i}}\right)\rho_{12i} + \\ \left(\frac{y_{13i}}{\pi_{13i}} - \frac{y_{23i}}{\pi_{23i}}\right) \end{matrix}\right]\omega_{1\bullet i} \quad (27)$$

Where, $\omega_{1\cdot}$=$F_{1\cdot}$ (1-$F_{1\cdot}$).

b. The first derivative is obtained from the estimated results of the $\theta_2^{(1)}$ parameter as follows:

$$\frac{\partial \ell(\boldsymbol{\mu})}{\partial \theta_2^{(1)}} = \sum_{i=1}^{n} \left[\left(\frac{y_{21i}}{\pi_{21i}} - \frac{y_{22i}}{\pi_{22i}} - \frac{y_{31i}}{\pi_{31i}} + \frac{y_{32i}}{\pi_{32i}} \right) \rho_{21i} + \left(\frac{y_{22i}}{\pi_{22i}} - \frac{y_{32i}}{\pi_{32i}} - \frac{y_{32i}}{\pi_{32i}} + \frac{y_{33i}}{\pi_{33i}} \right) \rho_{22i} + \left(\frac{y_{23i}}{\pi_{23i}} - \frac{y_{33i}}{\pi_{33i}} \right) \right] \omega_{2\square i} \quad (28)$$

Where, $\omega_{2\cdot}$=$F_{2\cdot}$ (1-$F_{2\cdot}$).

c. The first derivative is obtained from the estimated results of the $\theta_1^{(2)}$ parameter as follows:

$$\frac{\partial \ell(\boldsymbol{\mu})}{\partial \theta_1^{(2)}} = \sum_{i=1}^{n} \left[\left(\frac{y_{11i}}{\pi_{11i}} - \frac{y_{12i}}{\pi_{12i}} - \frac{y_{21i}}{\pi_{21i}} + \frac{y_{22i}}{\pi_{22i}} \right) \tau_{11i} + \left(\frac{y_{21i}}{\pi_{21i}} - \frac{y_{22i}}{\pi_{22i}} - \frac{y_{31i}}{\pi_{31i}} + \frac{y_{32i}}{\pi_{32i}} \right) \tau_{21i} + \left(\frac{y_{31i}}{\pi_{31i}} - \frac{y_{32i}}{\pi_{32i}} \right) \right] \delta_{\square 1i} \quad (29)$$

Where δ_1=F_1 (1-F_1).

d. The first derivative is obtained from the estimated results of the $\theta_{2(2)}$ parameter as follows:

$$\frac{\partial \ell(\boldsymbol{\mu})}{\partial \theta_2^{(2)}} = \sum_{i=1}^{n} \left[\left(\frac{y_{12i}}{\pi_{12i}} - \frac{y_{13i}}{\pi_{13i}} - \frac{y_{22i}}{\pi_{22i}} + \frac{y_{23i}}{\pi_{23i}} \right) \tau_{12i} + \left(\frac{y_{22i}}{\pi_{22i}} - \frac{y_{23i}}{\pi_{23i}} - \frac{y_{32i}}{\pi_{32i}} + \frac{y_{33i}}{\pi_{33i}} \right) \tau_{22i} + \left(\frac{y_{32i}}{\pi_{32i}} - \frac{y_{33i}}{\pi_{33i}} \right) \right] \delta_{\square 2i} \quad (30)$$

Where, δ_2=F_2 (1-F_2).

The results of the first derivative of the ln-likelihood function for each estimated parameter are non-linear, so numerical iteration is needed to obtain the estimated value of each parameter. This study uses the Berndt-Hall-Hall-Hausman (BHHH) iteration.

Implementation of Bi-response Ordinal Logistic Nonparametric Regression Model Based on MARS Estimator on Diabetes Mellitus and Hypertension Risk Data

To describe the distribution of data on the response variable, a contingency table is presented as follows in table 4.

Table 4. Descriptive of Response Variables

Diabetes Mellitus (DM)	Hypertension (HT)			Total
	1 (normal)	2 (stage-1 HT)	3 (stage-2 HT)	
1 (normal)	74	179	14	267
2 (stage-1 DM)	31	87	47	165
3 (stage-2 DM)	22	119	91	232
Total	127	385	152	664

Table 4 above shows that out of 664 respondents who were the research sample, 74 people (11,14 %) were in normal condition in both responses (diabetes mellitus and hypertension). Meanwhile, respondents who had normal category of stage 2 diabetes mellitus but had stage 1 hypertension were 179 people (26,96 %) and had stage 2 hypertension were 14 people (2,11 %). Respondents with the stage 1 diabetes mellitus category but normal hypertension were 31 people (4,67 %), those in the stage 1 hypertension category were 87 people (13,10 %), and those suffering from stage 2 hypertension were 47 people (7,08 %). Respondents who had stage 2 diabetes mellitus status but normal hypertension condition were 22 people (3,31 %), stage 1 hypertension was 119 people (17,92 %) and respondents who suffered from stage 2 diabetes mellitus and stage 2 hypertension were 91 people (13,70 %).

Next, a descriptive description of the predictor variables is presented, which have a continuous scale (interval or ratio) as follows in table 5.

Table 5. Descriptive of Continuous Scale Predictor Variables

Variable	Minimum	Maximum	Mean	Standard Deviation
Age (X_1)	18	71	46,77	10,38
Body Mass Index (X_3)	18,12	56	32,11	4,95
Total Cholesterol (X_4)	31,2	291	137,86	36,71

Table 5 shows that the average age of respondents is around 46 years, with the lowest age being 18 years and the highest age being 71 years. In this study, respondents had an average Body Mass Index (BMI) of 32,11 kg/m2, with the largest BMI being 56 kg/m2 and the smallest BMI being 18,12 kg/m2. In contrast, the average LDL cholesterol level of respondents was 137,86 mg/dL with the highest LDL cholesterol level being 291 mg/dL and the lowest being 31,2 mg/dL. The description of the predictor variables with a categorical measurement scale, namely gender (X_2), is presented in table 6 below.

Table 6. Descriptive of Categorical Scale Predictor Variable

Gender (X_2)	Diabetes Mellitus (DM)			Total	Hypertension (HT)			Total
	1 (normal)	2 (stage 1 DM)	3 (stage 2 DM)		1 (normal)	2 (stage 1 HT)	3 (stage 2 HT)	
Male	98	80	131	309	46	176	87	309
Female	169	85	101	355	81	209	65	355
Total	267	165	232	664	127	385	152	664

In table 6, it can be seen that in the male gender, the largest number has a diabetes mellitus response category, namely 131 people (19,73 %), while in the hypertension response, the most significant number of men is in the pre-hypertension category, namely 176 people (26,51 %). Respondents based the female gender in the largest diabetes mellitus response were in normal conditions, namely, 169 people (25,45 %), while in the hypertension response, the largest number was in the pre-hypertension category, namely 209 people (31,48 %).

Modelling the risk of diabetes mellitus (Y_1) and the incidence of hypertension (Y_2) using a bi-response logistic regression model based on the MARS estimator begins with a dependency test between response variables. This test aims to determine whether there is a dependency between the two response variables so that it is feasible to analyze them bi-response (bivariate). The test was carried out using the Mantel-Haenszel test statistic. The dependency test results concluded that there was dependency between the response variables, thus fulfilling the assumptions for conducting a bi-response analysis.

The results of the estimation using OSS-R obtained an ordinal bi-response MARS model formed with a minimum GCV value of 0,4601877 for response model 1 (f ̂((1)) (x)) and a GCV of 0,3751423 for response model 2 (f ̂((2)) (x)); both models can be seen in equation (31) and equation (34) as follows:

MARS Ordinal Response Model 1 (Y_1):

$$\begin{aligned}\hat{f}^{(1)}(x) = {} & 2{,}18938124 - 0{,}04985149BF_1 + 0{,}09586090BF_2 - 0{,}01189331BF_3 + \\ & 0{,}04402806BF_4 + 0{,}08493989BF_5 - 0{,}04407316BF_6 - 0{,}13735926BF_7 + \\ & 0{,}11902546BF_8 + 0{,}25710867BF_9 + 0{,}00221592BF_{10} + 0{,}01346406BF_{11}\end{aligned} \tag{31}$$

Where:

BF_1= h(168-X_4); BF_2= h(X_1-36)h(X_3-27,99); BF_3= h(42-X_1)h(168-X_4); BF_4=h(60-X_1)h(X_3-27,99); BF_5=h(X_1-48)h(X_3-27,99); BF_6=h(56-X_1)h(168-X_4); BF_7=h(X_3-27,99)h(91-X_4); BF_8=h(X_3-27,99)h(104-X_4); BF_9=h(X_3-27,99)h(99-X_4); BF_{10}=h(36-X_3)h(168-X_4); BF_{11}=h(X_3-22)h(168-X_4)

After logit transformation, the MARS ordinal cumulative logit model for response variable 1 (Y_1) is presented as follows in equation (32).

$$\begin{aligned}\hat{g}_1^{(1)}(x) &= 4{,}515 + \begin{bmatrix} 2{,}18938124 - 0{,}04985149BF_1 + 0{,}09586090BF_2 - 0{,}01189331BF_3 + \\ 0{,}04402806BF_4 + 0{,}08493989BF_5 - 0{,}04407316BF_6 - 0{,}13735926BF_7 + \\ 0{,}11902546BF_8 + 0{,}25710867BF_9 + 0{,}00221592BF_{10} + 0{,}01346406BF_{11} \end{bmatrix} \\ &= 6{,}70438124 - 0{,}04985149BF_1 + 0{,}09586090BF_2 - 0{,}01189331BF_3 + \\ &\quad 0{,}04402806BF_4 + 0{,}08493989BF_5 - 0{,}04407316BF_6 - 0{,}13735926BF_7 + \\ &\quad 0{,}11902546BF_8 + 0{,}25710867BF_9 + 0{,}00221592BF_{10} + 0{,}01346406BF_{11} \\ \hat{g}_2^{(1)}(x) &= 5{,}966 + \begin{bmatrix} 2{,}18938124 - 0{,}04985149BF_1 + 0{,}09586090BF_2 - 0{,}01189331BF_3 + \\ 0{,}04402806BF_4 + 0{,}08493989BF_5 - 0{,}04407316BF_6 - 0{,}13735926BF_7 + \\ 0{,}11902546BF_8 + 0{,}25710867BF_9 + 0{,}00221592BF_{10} + 0{,}01346406BF_{11} \end{bmatrix} \\ &= 8{,}15538124 - 0{,}04985149BF_1 + 0{,}09586090BF_2 - 0{,}01189331BF_3 + \\ &\quad 0{,}04402806BF_4 + 0{,}08493989BF_5 - 0{,}04407316BF_6 - 0{,}13735926BF_7 + \\ &\quad 0{,}11902546BF_8 + 0{,}25710867BF_9 + 0{,}00221592BF_{10} + 0{,}01346406BF_{11}\end{aligned} \tag{32}$$

MARS Ordinal Response Model 2 (Y_2):

$f^{(2)}$ (x)=1,74830762+0,02391818BF_1+ 0,06008681BF_2 (33)

Where:

BF_1= h(X_1-36);
BF_2= h(X_3-22,08)

After logit transformation, the MARS ordinal cumulative logit model for response variable 2 (Y_2) is presented as follows in equation (34).

$$\begin{aligned}\hat{g}_1^{(2)}(x) &= -1,8444+\left[1,74830762+0,02391818BF_1+\ 0,06008681BF_2\ \right]\\ &= -0,09609238+0,02391818BF_1+\ 0,06008681BF_2\\ \hat{g}_2^{(2)}(x) &= 1,1231+\left[1,74830762+0,02391818BF_1+\ 0,06008681BF_2\right]\\ &= 2,87140762+0,02391818BF_1+\ 0,06008681BF_2\end{aligned} \quad (34)$$

Based on equation (31), it can be seen that the MARS ordinal model of response variable 1 (Y_1) contains the interaction of two predictor variables. There are three variables included in the model, namely variables X_1 (age), X_3 (body mass index) and X_4 (total cholesterol). In equation (35), it can be seen that the MARS ordinal model of response variable 2 (Y_2) does not contain interactions between predictor variables, and there are two variables included in the model, namely variables X_1 (age) and X_3 (body mass index). The factors that influence diabetes mellitus contained in equation (31) and hypertension contained in equation (33) are supported by several previous relevant studies.

The next step is to form a nonparametric ordinal bi-response logistic model based on the obtained MARS estimator. The model is formed using 90 % of the 664 in-sample data, namely 598 data. The bi-response ordinal logistic regression model based on the MARS estimator for response variable 1 (Y_1) is as follows in equation (35) and equation (36).

$$\begin{aligned}\hat{g}_1^{(1)}(x) &= 4,871861+\begin{bmatrix}2,18938124-0,04985149BF_1+\ 0,09586090BF_2-\\ 0,01189331BF_3+\ 0,04402806BF_4+\ 0,08493989BF_5-\\ 0,04407316BF_6-0,13735926BF_7+0,11902546BF_8+\\ 0,25710867BF_9+0,00221592BF_{10}+0,01346406BF_{11}\end{bmatrix}\\ &= 7,06124224-0,04985149BF_1+\ 0,09586090BF_2-\ 0,01189331BF_3+\\ &\quad 0,04402806BF_4+\ 0,08493989BF_5-0,04407316BF_6-0,13735926BF_7+\\ &\quad 0,11902546BF_8+0,25710867BF_9+0,00221592BF_{10}+0,01346406BF_{11}\end{aligned} \quad (35)$$

$$\begin{aligned}\hat{g}_2^{(1)}(x) &= 6,342155+\begin{bmatrix}2,18938124-0,04985149BF_1+\ 0,09586090BF_2-\\ 0,01189331BF_3+\ 0,04402806BF_4+\ 0,08493989BF_5-\\ 0,04407316BF_6-0,13735926BF_7+0,11902546BF_8+\\ 0,25710867BF_9+0,00221592BF_{10}+0,01346406BF_{11}\end{bmatrix}\\ &= 8,53153624-0,04985149BF_1+\ 0,09586090BF_2-\ 0,01189331BF_3+\\ &\quad 0,04402806BF_4+\ 0,08493989BF_5-0,04407316BF_6-0,13735926BF_7+\\ &\quad 0,11902546BF_8+0,25710867BF_9+0,00221592BF_{10}+0,01346406BF_{11}\end{aligned} \quad (36)$$

The bi-response ordinal logistic regression model based on the MARS estimator for response variable 2 (Y_2) is as follows equation (37) and equation (38).

$$\begin{aligned}\hat{g}_1^{(2)}(x) &= -1,863083 + \left[1,74830762 + 0,02391818BF_1 + 0,06008681BF_2\right] \\ &= -0,11477538 + 0,02391818BF_1 + 0.06008681BF_2\end{aligned} \quad (37)$$

$$\begin{aligned}\hat{g}_2^{(2)}(x) &= 1,155045 + \left[1,74830762 + 0,02391818BF_1 + 0,06008681BF_2\right] \\ &= 2,90335262 + 0,02391818BF_1 + 0.06008681BF_2\end{aligned} \quad (38)$$

Next, the classification procedure will be evaluated by determining the accuracy, AUC, and APER values. These three values are obtained by first forming a confusion matrix. Here is the confusion matrix of in-sample data in table 7.

Table 7. Confusion Matrix of In-sample Data

Actual	Prediction								
	y_{11}	y_{12}	y_{13}	y_{21}	y_{22}	y_{23}	y_{31}	y_{32}	y_{33}
y_{11}	19	40	0	0	1	0	1	7	0
y_{12}	59	85	0	0	0	0	1	17	0
y_{13}	7	4	0	0	0	0	0	1	0
y_{21}	10	15	0	0	0	0	2	2	0
y_{22}	43	28	0	0	0	0	3	8	0
y_{23}	26	9	0	0	0	0	2	2	0
y_{31}	7	10	0	0	0	0	1	0	0
y_{32}	78	26	0	0	0	0	2	0	0
y_{33}	67	10	0	0	0	0	3	2	0

Referring to equation (13), the in-sample data accuracy value obtained was “Accuration = “ 82,44147 %. This accuracy value is close to 100 %, so based on the overall calculation of the classification accuracy of the accuracy value obtained, it can be concluded that the model formed is good to classify patients with diabetes mellitus and hypertension. Furthermore, referring to equation (14), the AUC value obtained is AUC=73,4224 %. This AUC value is in the “good” criteria range. The APER value based on equation (15) is APER=17,55853 %. Based on the APER value, which is smaller than 30 %, it can be seen that the bi-response ordinal logistic regression model based on the MARS estimator has stability and consistency in statistical classification.

In the classification model, the classification procedure for out-sample data, including the Accuracy, AUC, and APER values, must be evaluated. Table 8 shows the confusion matrix of out-sample data.

Table 8. Confusion Matrix of Out-sample Data

Actual	Prediction								
	y_{11}	y_{12}	y_{13}	y_{21}	y_{22}	y_{23}	y_{31}	y_{32}	y_{33}
y_{11}	5	1	0	0	0	0	0	0	0
y_{12}	16	1	0	0	0	0	0	0	0
y_{13}	2	0	0	0	0	0	0	0	0
y_{21}	2	0	0	0	0	0	0	0	0
y_{22}	3	2	0	0	0	0	0	0	0
y_{23}	6	2	0	0	0	0	0	0	0
y_{31}	4	0	0	0	0	0	0	0	0
y_{32}	10	3	0	0	0	0	0	0	0
y_{33}	7	2	0	0	0	0	0	0	0

Furthermore, referring to equation (14), the AUC value obtained is AUC=72,04534 %. This AUC value is in the “good” criteria range. The APER value based on equation (15) isAPER=9,090909 %. Based on the APER value, which is smaller than 30 %, it can be seen that the bi-response ordinal logistic regression model based on the MARS estimator has stability and consistency in statistical classification.

CONCLUSIONS

Based on the results of the analysis and discussion, several conclusions can be drawn. The parameter estimation of the bi-response ordinal logistic regression model with the MARS estimator is obtained using the Maximum Likelihood Estimation (MLE) method. However, due to the non-linearity of the first derivative of the log-likelihood function, the BHHH numerical iteration method is employed to estimate the parameters. The complexity of the likelihood function results in difficulties in forming the Hessian matrix, leading to an approximation of the second derivative using the first derivative in the BHHH method. In modeling the risk of diabetes mellitus and hypertension using the nonparametric ordinal bi-response logistic regression method with the MARS estimator, it was found that the significant predictor variables influencing these risks are Age, BMI, and Cholesterol. The AUC value of 73,42 % falls within the "good" category, while the APER value of 17,56 % is below 30 %, confirming the model's stability and consistency. These findings suggest that the developed model effectively captures the relationship between risk factors and the likelihood of diabetes mellitus and hypertension.

BIBLIOGRAPHIC REFERENCES

1. Chamidah N, Lestari B, Susilo H, Alsagaff MY, Budiantara IN, Aydin D. Spline Estimator in Nonparametric Ordinal Logistic Regression Model for Predicting Heart Attack Risk. Symmetry (Basel) [Internet]. 2024 Oct 30;16(11):1440. Available from: https://doi.org/10.3390/sym16111440

2. Lestari B, Chamidah N, Aydin D, Yilmaz E. Reproducing Kernel Hilbert Space Approach to Multiresponse Smoothing Spline Regression Function. Symmetry 2022, 14(11): 2227. https://doi.org/10.3390/sym14112227.

3. Eubank RL. Spline smoothing and nonparametric regression. (No Title). 1988.

4. Chamidah N, Lestari B, Budiantara IN, Aydin D. Estimation of Multiresponse Multipredictor Nonparametric Regression Model Using Mixed Estimator. Symmetry 2024, 16(4): 386. https://doi.org/10.3390/sym16040386.

5. Fernandes AAR, Budiantara IN, Otok BW, Suhartono. Spline estimator for bi-responses nonparametric regression model for longitudinal data. Appl Math Sci [Internet]. 2014;8:5653–65. Available from: http://www.m-hikari.com/ams/ams-2014/ams-113-116-2014/47566.html.

6. Aydin D, Yilmaz E, Chamidah N, Lestari B, Budiantara IN. Right-censored nonparametric regression with measurement error. Metrika 2024, 87(3). https://doi.org/10.1007/s00184-024-00953-5.

7. Ampulembang AP, Otok BW, Rumiati AT, Budiasih. Bi-responses nonparametric regression model using MARS and its properties. Appl Math Sci [Internet]. 2015;9:1417–27. Available from: http://www.m-hikari.com/ams/ams-2015/ams-29-32-2015/5127.html

8. Sriliana I, Budiantara IN, Ratnasari V. The performance of mixed truncated spline-local linear nonparametric regression model for longitudinal data. MethodsX. 2024;12:102652.

9. Sriliana I, Budiantara IN, Ratnasari V. A truncated spline and local linear mixed estimator in nonparametric regression for longitudinal data and its application. Symmetry (Basel). 2022;14(12):2687.

10. Juniar MA, Fania A, Ulya D, Ramadhan R, Chamidah N. Modelling Crime Rates in Indonesia Using Truncated Spline Estimator. BAREKENG J Ilmu Mat dan Terap. 2024;18(2):1201–16.

11. Hidayati L, Chamidah N, Budiantara IN. Spline truncated estimator in multiresponse semiparametric regression model for computer based national exam in West Nusa Tenggara. In: IOP Conference Series: Materials Science and Engineering. IOP Publishing; 2019. p. 52029.

12. Chamidah N, Lestari B, Saifudin T. Modeling of blood pressures based on stress score using least square spline estimator in bi-response nonparametric regression. Int J Innov Creat Change. 2019;5(3):1200–16.

13. Chamidah N, Lestari B, Budiantara IN, Saifudin T, Rulaningtyas R, Aryati A, Wardani P, Aydin D. Consistency and Asymptotic Normality of Estimator for Parameters in Multiresponse Multipredictor Semiparametric Regression Model. Symmetry 2022, 14(2): 336. https://doi.org/10.3390/sym14020336.

14. Gackowski M, Szewczyk-Golec K, Pluskota R, Koba M, Mądra-Gackowska K, Woźniak A. Application

of Multivariate Adaptive Regression Splines (MARSplines) for Predicting Antitumor Activity of Anthrapyrazole Derivatives. Int J Mol Sci [Internet]. 2022 May 4;23(9):5132. Available from: https://www.mdpi.com/1422-0067/23/9/5132.

15. Hasyim M, Chamidah N, Saifudin T. Estimation of uniresponse ordinal logistic nonparametric regression model based on multivariate adaptive regression spline. In 2024. p. 020010. Available from: https://pubs.aip.org/aip/acp/article-lookup/doi/10.1063/5.0242167.

16. Hasyim M, Prastyo DD. Modelling lecturer performance index of private university in Tulungagung by using survival analysis with multivariate adaptive regression spline. J Phys Conf Ser [Internet]. 2018 Mar;974:012065. Available from: https://iopscience.iop.org/article/10.1088/1742-6596/974/1/012065.

17. Pramudita DT, Budiantara IN, Ratnasari V. Comparison of selection optimal knot using Cross Validation and Generalized Cross Validation for nonparametric regression truncated spline longitudinal data. In 2024. p. 020011. Available from: https://pubs.aip.org/aip/acp/article-lookup/doi/10.1063/5.0211356.

14. Roy SS, Roy R, Balas VE. Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. Renew Sustain Energy Rev. 2018;82:4256–68.

15. Friedman JH, Roosen CB. An introduction to multivariate adaptive regression splines. Stat Methods Med Res. 1995.

16. Otok BW, Rumiati AT, Ampulembang AP, Azies H Al. ANOVA Decomposition and Importance Variable Process in Multivariate Adaptive Regression Spline Model. Int J Adv Sci Eng Inf Technol [Internet]. 2023 Jun 5;13(3):928–34. Available from: https://ijaseit.insightsociety.org/index.php/ijaseit/article/view/17674.

17. Raykov T, Marcoulides GA. An introduction to applied multivariate analysis. Routledge; 2008.

18 Aydin D, Chamidah N, Lestari B, Mohammad S, Yilmaz E. Local Polynomial Estimation for multi-response semiparametric regression models with right censored data. Communications in Statistics-Simulation and Computation, 2025, 54(3). https://doi.org/10.1080/03610918.2025.2476595.

19. Utami TW, Chamidah N, Saifudin T, Lestari B, Aydin D. Estimation of Biresponse Semiparametric Regression Model for Longitudinal Data Using Local Polynomial Kernel Estimator. Symmetry 2025, 17(3): 392. https://doi.org/10.3390/sym17030392.

20. Lestari B, Chamidah N, Budiantara IN, Aydin D. Determining confidence interval and asymptotic distribution for parameters of multiresponse semiparametric regression model using smoothing spline estimator. Journal of King Saud University – Science 2023, 35: 102664. https://doi.org/10.1016/j.jksus.2023.102664.

21. Aydin D, Yilmaz E, Chamidah N, Lestari B. Right-censored partially linear regression model with error in variables: application with carotid endarterectomy dataset. International Journal of Biostatistics, 2023, 20(1), pp. 1–34. https://doi.org/10.1515/ijb-2022-0044.

22. Agresti A. An Introduction to Categorical Data Analysis: Second Edition. An Introduction to Categorical Data Analysis: Second Edition. 2006. 1–356 p.

23. Kishartini K, Safitri D, Ispriyanti D. Multivariate Adaptive Regression Splines (MARS) Untuk Klasifikasi Status Kerja Di Kabupaten Demak. J Gaussian. 2014;3(4):711–8.

24. Annur M, Dahlan JA, Agustina F. Penerapan Metode Multivariate Adaptive Regression Spline (MARS) untuk Menentukan Faktor yang Mempengaruhi Masa Studi Mahasiswa FPMIPA UPI. J EurekaMatika. 2015;3(1):135–55.

25. Binadari R, Wilandari Y, Suparti S. Perbandingan Metode Regresi Logistik Biner Dan Multivariate Adaptive Regression Spline (Mars) Pada Peminatan Jurusan SMA (Studi Kasus SMA Negeri 2 Semarang). J Gaussian. 2015;4(4):987–96.

26. Serrano NB, Sánchez AS, Lasheras FS, Iglesias-Rodríguez FJ, Valverde GF. Identification of gender

differences in the factors influencing shoulders, neck and upper limb MSD by means of multivariate adaptive regression splines (MARS). Appl Ergon. 2020;82:102981.

27. Nisai SF, Budiantara IN. Analisis Survival dengan Pendekatan Multivariate Adaptive Regression Splines pada Kasus Demam Berdarah Dengue (DBD). J Sains dan Seni ITS. 2012;1(1):15838.

28. Wang L, Wu C, Gu X, Liu H, Mei G, Zhang W. Probabilistic stability analysis of earth dam slope under transient seepage using multivariate adaptive regression splines. Bull Eng Geol Environ. 2020;79:2763–75.

29. Ampulembang AP, Otok BW, Rumiati AT, Budiasih. Modeling of Welfare Indicators in Java Island Using Biresponses MARS. Int J Appl Math Stat. 2016;54(2):66–75.

30. Milborrow S. Derived from mda: Mars by T. Hastie and R. Tibshirani. earth: Multivariate adaptive regression splines. R package version 4.4. 3. Retrived from https://cran.r-project.org/web/packages/earth/index.html;2015.

31. Ampulembang AP. Pengembangan Model Regresi Nonparametrik Birespon Kontinu Menggunakan Metode MARS. Institut Teknologi Sepuluh Nopember; 2017.

32. Eyduran E, Çanga D, Sevgenler H, Çelik AE. Use of Bootstrap Aggregating Bagging MARS to Improve Predictive Accuracy for Regression Type Problems. presented at the 11. Uluslararası İstatistik Kongresi. 2019;

33. Kemenkes. Laporan Riskesdas 2018 Nasional.pdf. Lembaga Penerbit Balitbangkes. 2018. p. hal 156.

34. Schutta MH. Diabetes and hypertension: epidemiology of the relationship and pathophysiology of factors associated with these comorbid conditions. J Cardiometab Syndr. 2007;2(2):124–30.

35. Saeedi P, Petersohn I, Salpea P, Malanda B, Karuranga S, Unwin N, et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas. Diabetes Res Clin Pract. 2019;157:107843.

36. Yildiz M, Esenboğa K, Oktay AA. Hypertension and diabetes mellitus: highlights of a complex relationship. Curr Opin Cardiol. 2020;35(4):397–404.

37. Waeber B, Feihl F, Ruilope L. Diabetes and hypertension. Blood Press. 2001;10(5–6):311–21.

38. Hardine DR, Abdullah A, Ikbal M, Chamidah N. Pemodelan Kadar Gula Darah dan Tekanan Darah Pada Remaja Penderita Diabetes Miletus Tipe II dengan pendekatan Regresi Nonparametrik Birespon Berdasarkan Estimator Spline. In: Seminar Nasional Matematika dan Aplikasiny. 2017. p. 308–12.

39. Committee PP, Classification A. Standards of medical care in diabetes-2010. Diabetes Care. 2010;33(SUPPL. 1).

40. Mayawi M, Nurhayati N, Talib T, Bustan AW, Laamena NS. Ordinal Logistic Regression Analysis of Factors that Affecting the Blood Sugar Levels Diabetes Mellitus Patients. Pattimura Int J Math. 2023;2(1):33–42.

41. Molenberghs G, Lesaffre E. Marginal Modeling of Correlated Ordinal Data Using a Multivariate Plackett Distribution. J Am Stat Assoc. 1994;89(426):633.

42. Society IB. Global Cross-Ratio Models for Bivariate , Discrete , Ordered Responses Author (s): Jocelyn R . Dale Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2530704. 2009;42(4):909–17.

43. Bhatia AS, Mukherjee BN. Comparison of Some Multinomial Classification Rules: A Case Study. Behaviormetrika. 1993 Jan;20:91–106.

44. Friedman JH. Estimating Functions of Mixed Ordinal and Categorical Variables Using Adaptive Splines. 1993;(108):73-113.

45. Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis. 2007. p. 671–757.

46. Fahmy AM. Confusion Matrix in Three-class Classification Problems: A Step-by-Step Tutorial. J Eng Res. 2023;7(1):0-0.

47. Rey deCastro B. Cumulative ROC curves for discriminating three or more ordinal outcomes with cutpoints on a shared continuous measurement scale. PLoS One. 2019;14(8):1–16.

48. Hand DJ, Till RJ. A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. Mach Learn. 2001;45:171–86.

49. Šimundić AM. Measures of Diagnostic Accuracy: Basic Definitions. EJIFCC. 2009 Jan;19:203–11.

50. Härdle W, Simar L. Applied Multivariate Statistical Analysis. Appl Multivar Stat Anal. 2003;(April).

ACKNOWLEDGMENT

The authors thank the Center for Education Financial Services and the Indonesia Endowment Funds for Education for funding this research.

FINANCING

This research was funded by The Center for Education Financial Services and The Indonesia Endowment Funds for Education, grant number: 1146/J5.2.3./BPI.06/10/2021.

CONFLICT OF INTEREST

The authors declare that no conflict of interest is associated with this research. The research process was conducted objectively and independently, without influence from any stakeholder that could benefit personally or institutionally.

AUTHORSHIP CONTRIBUTION

Conceptualization: Maylita Hasyim, Nur Chamidah, Toha Saifudin.
Data curation: Maylita Hasyim, Nur Chamidah, Toha Saifudin.
Formal analysis: Maylita Hasyim, Nur Chamidah, Budi Lestari.
Research: Maylita Hasyim, Nur Chamidah, Toha Saifudin.
Methodology: Maylita Hasyim, Nur Chamidah, Budi Lestari.
Project management: Maylita Hasyim.
Resources: Maylita Hasyim, Nur Chamidah.
Software: Maylita Hasyim, Toha Saifudin.
Supervision: Nur Chamidah, Toha Saifudin, Budi Lestari.
Validation: Nur Chamidah, Toha Saifudin, Budi Lestari.
Display: Maylita Hasyim.
Drafting - original draft: Maylita Hasyim.
Writing - proofreading and editing: Nur Chamidah, Budi Lestari.