BAREKENG: Journal of Mathematics and Its Applications
March 2026 Volume 20 Issue 1 Page 0255-0270
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20no1pp0255-0270

COMPARISON OF BINARY PROBIT REGRESSION AND FOURIER SERIES NONPARAMETRIC LOGISTIC REGRESSION IN MODELING DIABETES STATUS

Bambang Widjanarko Otok [1], Muhammad Zulfadhli [2*], Riwi Dyah Pangesti [3], Muhammad Idham Kurniawan [4], Albertus Eka Putra Haryanto [5], Darwis [6], Iwan Kurniawan [7]

[1,2,4] *Department of Statistics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember*
Jln. Teknik Kimia, Keputih, Sukolilo, Surabaya, Jawa Timur, 60111, Indonesia

[3] *Department of Statistics, Faculty of Mathematics and Natural Sciences, Bengkulu University*
Jln. WR. Supratman, Kandang Limun, Muara Bangka Hulu, Sumatera, Bengkulu 38371, Indonesia

[5] *Regional Economic Development Institute (REDI)*
Jln. Arief Rahman Hakim No.152, Keputih, Sukolilo, Surabaya, Jawa Timur, 60111, Indonesia

[6] *Islamic Religious Education, Faculty of Education and Teaching, STAIN Majene*
Jln. Balai Latihan Kerja No.7, Totoli, Banggae, Majene, Sulawesi Barat, 91415, Indonesia

[7] *Public Sector Business Administration, Politeknik STIA LAN Bandung*
Jln. Hayam Wuruk No.34-38, Citarum, Bandung Wetan, Bandung, Jawa Barat, 40115, Indonesia

*Corresponding author's e-mail: * **muhammadzulfadhli23@gmail.com***

Article Info

Article History:

Received: 3rd March 2025
Revised: 28th April 2025
Accepted: 25th July 2025
Available online: 24th November 2025

Keywords:

Categorical data;
Bernoulli distribution;
Binary Probit Regression (BPR);
Binary Logistic Regression (BLR) ;
Fourier Series Nonparametric Binary Logistic Regression (FSNBLR).

ABSTRACT

Diabetes mellitus is a chronic disease with a rising global prevalence, including in Indonesia. Early detection and accurate modeling are crucial for effective prevention and management. Binary Logistic Regression (BLR) is commonly used for binary outcome modeling; however, in practice, the relationship between binary outcomes and continuous predictors is often nonlinear, making BLR less suitable. To address these limitations, alternative methods such as Binary Probit Regression (BPR) and Flexible Semiparametric Nonlinear Binary Logistic Regression (FSNBLR) have been developed. This study aims to compare the performance of BLR, BPR, and FSNBLR models in classifying diabetes mellitus cases at Hajj General Hospital Surabaya. All three models were estimated using the Maximum Likelihood Estimation (MLE) method. Since the resulting estimators do not have closed-form solutions, numerical iteration using the Newton-Raphson method was applied. Model performance was assessed using Area Under the Curve (AUC), accuracy, sensitivity, and specificity. The FSNBLR model outperformed both the BLR and BPR models. It achieved the highest AUC value of 81.86%, while BLR (66.30%) and BPR (66.30%). That is indicated FSNBLR superior discriminative ability. In addition, the FSNBLR model recorded higher accuracy, sensitivity, and specificity compared to the other two models. The FSNBLR model demonstrated better predictive performance in identifying diabetes mellitus cases, especially in scenarios involving nonlinear relationships between predictors and the outcome variable. These findings suggest that flexible semiparametric approaches offer greater effectiveness in medical classification tasks, particularly for chronic conditions like diabetes mellitus.

How to cite this article:
B. W. Otok, M. Zulfadhli, R. D. Pangesti, M. I. Kurniawan, A. E. P. Haryanto, Darwis and I. Kurniawan., "COMPARISON OF BINARY PROBIT REGRESSION AND FOURIER SERIES NONPARAMETRIC LOGISTIC REGRESSION IN MODELING DIABETES STATUS AT HAJJ GENERAL HOSPITAL SURABAYA," *BAREKENG: J. Math. & App.*, vol. 20, iss. 1, pp. 0255-0270, Mar, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*
Research Article · Open Access

1. INTRODUCTION

Binary Logistic Regression (BLR) is a statistical technique used to model the relationship between a binary outcome variable and one or more predictors. The model assumes that the logit (log-odds) of the event probability is linearly related to the independent variables. However, this approach can be inadequate for data that follow a probit link, since the underlying distribution and link function differ, which may lead to bias and a poor model fit when the logit link is applied incorrectly. To address this issue, Binary Probit Regression (BPR) provides an alternative that accommodates such characteristics. Nonetheless, in certain situations, the relationship between the dependent and independent variables is more complex and may require a nonparametric framework. One of the advancing methods that employs the logit function is Nonparametric Binary Logistic Regression (NBLR). [1].

Nonparametric Binary Logistic Regression (NBLR) is an effective tool for exploring the relationship between response and predictor variables when the regression curve's form is not predetermined. The NBLR curve is assumed to be smooth since it is confined to a specific function space, and the data themselves are expected to define the estimation form, free from the researcher's subjective influence. This characteristic makes NBLR highly adaptable, as it relies on smoothing techniques applied to the observed data. There are various smoothing methods available, including Spline estimators [2], Fourier Series estimator [3], Wavelet estimator [4], Kernel estimator [5], Local Polynomial estimator [6], and Multivariate Adaptive Regression Splines [7], [8].

Spline estimators are utilized when the data exhibit changing patterns defined by knot points [2] making them suitable for cases where the relationship is smooth yet varies in curvature across different regions. The local polynomial estimator helps reduce bias and asymptotic variance in nonparametric regression, especially when dealing with multiple response variables [9], and is ideal for modeling local trends in complex, multidimensional datasets. Wavelet estimators are particularly useful for handling signals with Gaussian additive noise, particularly when the data include localized structures or abrupt changes that must be captured at multiple scales [10]. Kernel estimators are commonly chosen for smoothing noisy data, where basic nonparametric local averaging is sufficient [5]. The Fourier Series estimator, on the other hand, is ideal for data with repetitive patterns [3]. Among these, the Fourier Series method was chosen due to its specialization and widespread use in situations where both response and predictor variables show recurring patterns that follow a specific trend [11]. This method provides an optimal balance between precision and computational complexity in additive nonparametric regression models [12]. applicable for both univariate and multivariate predictor variables [3], [10], [13], [14].

The Fourier Series method was first introduced by [3], and then [10] explored its application in nonparametric regression., [15] subsequently applied the Fourier Series in semiparametric regression, while. [16] expanded it to bivariate semiparametric regression using Fourier Series. This was later developed into Fourier Series nonparametric mixture regression models by [17], [18], [19], and Fourier Series semiparametric mixture models by [20]. However, prior studies have mainly focused on quantitative data, as seen in [21], [22], [23], [24]. In practice, however, the relationships between responses and predictors frequently involve categorical response variables.

Several researchers have created nonparametric regression estimators for categorical data, including Local Likelihood Logit Estimation [25], Decision Tree approaches [26], and B-Spline function [27] More recently, nonparametric regression estimators for categorical data, such as Fourier Series Nonparametric Logistic Regression (FSNBLR), have been developed [28]. Additionally, Fourier Series has been studied as a smoothing technique in nonparametric regression because it can represent complex, periodic, or oscillating relationships by combining sine and cosine functions. Originally developed for continuous signals, the Fourier basis can be adapted to model nonlinear relationships in categorical response data by transforming predictor variables and capturing latent cyclical or wave-like structures. This method is particularly effective when the relationship between predictors and categorical responses is nonlinear and non-monotonic, which is often observed in medical and behavioral data.

Earlier research has predominantly focused on comparing traditional methods such as Binary Logistic Regression (BLR) and Binary Probit Regression (BPR), without considering more recent techniques like FSNBLR, which were specifically designed to address nonlinear patterns in categorical data. Traditional methods like BLR and BPR assume a linear relationship between the predictors and the transformed response variable, which limits their ability to capture the complex or recurring patterns often present in real-world data, such as those seen in medical conditions like diabetes mellitus. However, no study to date has compared

BLR, BPR, and FSNBLR estimators. Hence, this study seeks to compare the performance of BLR, BPR, and FSNBLR methods in the context of diagnosing diabetes mellitus at Hajj General Hospital Surabaya.

To date, no study has compared the BLR, BPR, and FSNBLR estimators. Therefore, this research aims to evaluate and compare the performance of BLR, BPR, and FSNBLR methods in diagnosing diabetes mellitus at Hajj General Hospital Surabaya, with the goal of identifying the most effective method for handling categorical response data that exhibit nonparametric patterns.

2. RESEARCH METHODS

To derive the BPR, BLR, and FSNBLR estimators for categorical data, several steps are required: building the BPR, BLR, and FSNBLR models, defining the Log-Likelihood function, and calculating its derivatives with respect to each parameter in the model. Finally, the process is completed through numerical iterations using the Newton–Raphson method..

2.1 Probability Distribution

Let $x_1, x_2, \ldots, x_p$ denote the p predictor variables. The response variable Y is assumed to follow a Bernoulli distribution [1] with its probability distribution expressed as:

$$Y \sim B\big(1, \pi(x)\big), \pi(x) = \pi\big(x_1, x_2, \ldots, x_p\big)$$

where the probability of success is given by:

$$P(Y_i = 1) = \pi(x_i)$$

and the probability of failure is given by:

$$P(Y_i = 0) = 1 - \pi(x_i)$$

$\pi(x_i)$ is defined in the probability distribution function $P(Y_i = y_i)$, where i represents the observation index $(i = 1,2,\ldots,n)$ as follows.

$$P(Y_i = y_i) = \pi(x_i)^{y_i}\big(1 - \pi(x_i)\big)^{1-y_i} = \left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right)^{y_i}\big(1 - \pi(x_i)\big) \tag{1}$$

2.2 BLR Estimator

2.2.1 Logit Function (Link Function)

Eq. (1) can be rewritten as the natural logarithmic function

$$\ln P(Y_i = y_i) = y_i \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) + \ln\big(1 - \pi(x_i)\big). \tag{2}$$

When expressed in exponential form, Eq. (2) takes the shape of an exponential family distribution function as shown below.

$$\exp(\ln P(Y_i = y_i)) = \exp\left(y_i \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) + \ln\big(1 - \pi(x_i)\big)\right), \tag{3}$$

Eq. (3) defines the exponential family distribution function as follows.

$$f(y_i, \theta) = \exp\left(\frac{y_i\, \theta - b(\theta)}{a(\emptyset)} + c(\theta, \emptyset)\right). \tag{4}$$

As a result, its probability distribution function is part of the exponential family of distributions..

$$P(Y_i = y_i) = \exp\left(\frac{y_i \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right)}{1} + \ln\big(1 - \pi(x_i)\big)\right), \tag{5}$$

where

$$\theta = \ln\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) \qquad a(\emptyset) = 1$$

$$b(\theta) = \ln(1-\pi(\boldsymbol{x_i})) \qquad c(\theta,\emptyset) = 0.$$

The variable θ in Eq. (5) represents the logit function, and the logit function for the resulting regression is:

$$\theta = \ln\left(\frac{\pi(\boldsymbol{x_i})}{1-\pi(\boldsymbol{x_i})}\right). \tag{6}$$

The logit function, serving as the link function, simplifies the regression model and facilitates parameter estimation. This is achieved by applying a logit transformation.

2.2.2 Logit Transformation Model

The logit transformation model is expressed as follows:

$$\ln\left(\frac{\pi(\boldsymbol{x_i})}{1-\pi(\boldsymbol{x_i})}\right) = f(x_{1i}, \ \dots \ , x_{pi}), \tag{7}$$

where $f(x_{1i}, \ \dots \ , x_{pi})$ in Eq. (7) as follows.

$$\ln\left(\frac{\pi(\boldsymbol{x_i})}{1-\pi(\boldsymbol{x_i})}\right) = \sum_{j=1}^{p}(\beta_0 + \beta_j x_{ji})\,;\ i = 1,2,\dots n. \tag{8}$$

By using Eq. (8), BLR model is obtained as follows.

$$\pi(\boldsymbol{x_i}) = \frac{\exp\sum_{j=1}^{p}(\beta_0 + \beta_j x_{ji})}{1+\exp\sum_{j=1}^{p}(\beta_0 + \beta_j x_{ji})}\ ;\ i = 1,2,\dots,n, \tag{9}$$

where β_0 and β_j , $j = 1,2,\dots,p$ are the model parameters of the logit function.

2.2.3 Likelihood Function $l(\boldsymbol{\beta})$

The form of the likelihood function $l(\beta)$

where

$$\beta = (\beta_0 \quad \beta_1 \quad \cdots \quad \beta_p),$$

It is derived using the Maximum Likelihood Estimation (MLE) method.

$$l(\beta) = \prod_{i=1}^{n} P(Y_i = y_i) = \pi(x_i)^{\sum_{i=1}^{n} y_i}\big(1-\pi(x_i)\big)^{n-\sum_{i=1}^{n} y_i}. \tag{10}$$

In logistic regression, the model parameters are typically estimated using the Maximum Likelihood Estimation (MLE) approach. This procedure involves obtaining the parameter values that maximize the log-likelihood function, which is achieved by setting its first-order derivatives to zero and solving the resulting system of equations. Through this maximization process, the MLE method identifies the parameter estimates that best explain the observed data.

2.2.4 Log-Likelihood Function $L(\boldsymbol{\beta})$

The likelihood Eq. (10) can be easily maximized as shown by $\ln l(\beta)$.

$$\begin{aligned} L(\beta) = \ln[l(\beta)] &= \sum_{i=1}^{n} y_i \ln[\pi(x_i)] + \sum_{i=1}^{n} (1-y_i)\ln[1-\pi(x_i)] \\ &= \sum_{i=1}^{n} \left\{y_i\left(f(x_{1i}, \ \dots \ , x_{pi})\right) - \ln[1+\exp(f(x_{1i}, \ \dots \ , x_{pi}))]\right\}. \end{aligned} \tag{11}$$

2.2.5 Newton-Raphson Iteration

$$\beta^{(t+1)} = \beta^{(t)} - \left(H(\beta)^{(t)}\right)^{-1} g(\beta)^{(t)}, \tag{12}$$

where $\beta^{(t)}$ is the β of the t-th iteration, $t = 1,2,\ldots$, is convergent.

$$\beta^{(t)} = \begin{pmatrix} {\beta_1}^{(t)} & {\beta_2}^{(t)} & \cdots & {\beta_p}^{(t)} \end{pmatrix}.$$

$\boldsymbol{g}(\boldsymbol{\beta})$ is the gradient vector of β

$$\boldsymbol{g}(\boldsymbol{\beta}) = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1}, \frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2}, \cdots, \frac{\partial L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_p}\right)^{\boldsymbol{T}}, \tag{13}$$

$\boldsymbol{H}(\boldsymbol{\beta})$ represents the Hessian matrix of β in Eq. (12), as given by the following equation.

$$\boldsymbol{H}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial^2 L(\boldsymbol{\beta})}{\partial {\boldsymbol{\beta}_1}^2} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_2} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1 \partial \boldsymbol{\beta}_p} \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_1} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial {\boldsymbol{\beta}_2}^2} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2 \partial \boldsymbol{\beta}_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_p \partial \boldsymbol{\beta}_1} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_p \partial \boldsymbol{\beta}_2} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial {\boldsymbol{\beta}_p}^2} \end{bmatrix}. \tag{14}$$

2.2.6 Estimator $\hat{\boldsymbol{\beta}}$

The estimate, $\hat{\beta}$ can be obtained from the Newton-Raphson iteration equation when

$$\left|\beta^{(t+1)} - \beta^{(t)}\right| < \varepsilon,\ \varepsilon = 0.000001. \tag{15}$$

Therefore, the estimator $\hat{\beta}$ is expressed as:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 & \hat{\beta}_2 & \cdots & \hat{\beta}_p \end{pmatrix}.$$

Using the result from the estimator $\hat{\beta}$, the BLR model in Eq. (16) can be written as::

$$\hat{\pi}(\boldsymbol{x_i}) = \frac{\exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_p x_{pi}\right)}{1 + \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \cdots + \hat{\beta}_p x_{pi}\right)}, \tag{16}$$

where, $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$ are the estimated coefficients of the logit function, and p represents the number of predictor variables..

2.3 BPR Estimator

2.3.1 Probit Function

Eq. (1) can be written as a probit function (Φ)

$$\Phi(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]. \tag{17}$$

2.3.2 Probit Transformation Model

The probit transformation model is expressed as follows:

$$\Phi^{-1}(x_i) = x_i^T \gamma, i = 1, 2, \ldots, n; \tag{18}$$

$$z = \Phi^{-1}\left(\pi(x_i)\right) = x_i^T \gamma; \tag{19}$$

$$\pi(x_i) = \Phi\left(x_i^T \gamma\right); \tag{20}$$

$$z = \Phi^{-1}\left(\pi(x_i)\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p. \tag{21}$$

2.3.3 Likelihood Function $\boldsymbol{l}(\boldsymbol{\gamma})$

The likelihood function $l(\gamma)$

where

$$\gamma = (\gamma_0 \quad \gamma_1 \quad \cdots \quad \gamma_p), \tag{22}$$

It is derived using the Maximum Likelihood Estimation (MLE) method.

$$l(\gamma) = \prod_{i=1}^{n} \mathrm{P}\,(y_i = 1|x_i)^{y_i}[1 - P(y_i = 1|x_i)]^{1-y_i}, \tag{23}$$

$$l(\gamma) = \prod_{i=1}^{n} \Phi\,(x_i^T\gamma)^{y_i}[1 - \Phi(x_i^T\gamma)]^{1-y_i}. \tag{24}$$

2.3.4 Log-Likelihood Function $L(\boldsymbol{\gamma})$

The likelihood function Eq. (24) can be easily maximized as $\ln l(\gamma)$.

$$L(\gamma) = \ln l(\gamma) = \sum_{i=1}^{n}[y_i \ln(\Phi) + (1 - y_i)\ln(1 - \Phi)]\,. \tag{25}$$

2.3.5 Newton-Raphson Iteration

$$\gamma^{(t+1)} = \gamma^{(t)} - \left(H(\gamma)^{(t)}\right)^{-1} g(\gamma)^{(t)}, \tag{26}$$

where $\gamma^{(t)}$ is the γ of the t-th iteration, $t = 1,2,\ldots$, is convergent.

$$\gamma^{(t)} = \left(\gamma_1{}^{(t)} \quad \gamma_2{}^{(t)} \quad \cdots \quad \gamma_p{}^{(t)}\right).$$

$\boldsymbol{g}(\boldsymbol{\gamma})$ is the gradient vector of γ

$$\boldsymbol{g}(\boldsymbol{\gamma}) = \left(\frac{\partial \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}_1}, \frac{\partial \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}_2}, \cdots, \frac{\partial \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\beta}_p}\right)^{\boldsymbol{T}}, \tag{27}$$

and $\boldsymbol{H}(\boldsymbol{\gamma})$ is the Hessian matrix of γ in Eq. (28), with the following equation.

$$\boldsymbol{H}(\boldsymbol{\gamma}) = \begin{bmatrix} \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1{}^2} & \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1 \partial \boldsymbol{\gamma}_2} & \cdots & \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_1 \partial \boldsymbol{\gamma}_p} \\ \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_2 \partial \boldsymbol{\gamma}_1} & \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_2{}^2} & \cdots & \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_2 \partial \boldsymbol{\gamma}_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_p \partial \boldsymbol{\gamma}_1} & \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_p \partial \boldsymbol{\gamma}_2} & \cdots & \frac{\partial^2 \boldsymbol{L}(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}_p{}^2} \end{bmatrix}. \tag{28}$$

2.3.6 Estimator $\hat{\boldsymbol{\gamma}}$

The estimate, $\hat{\gamma}$ can be obtained from the Newton-Raphson iteration equation when:

$$\left|\gamma^{(t+1)} - \gamma^{(t)}\right| < \varepsilon,\ \varepsilon = 0.0000001. \tag{29}$$

Thus, the estimator $\hat{\gamma}$ is expressed as:

$$\hat{\gamma} = (\hat{\gamma}_1 \quad \hat{\gamma}_2 \quad \cdots \quad \hat{\gamma}_p).$$

Using the result from the estimator $\hat{\gamma}$, the BPR model in Eq. (30) can be written as:

$$P(y_i = 1) = \Phi\big(\hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \cdots + \hat{\gamma}_p x_p\big), i = 1, 2, \ldots, n;$$

$$P(y_i = 0) = 1 - \Phi\big(\hat{\gamma}_0 + \hat{\gamma}_1 x_1 + \hat{\gamma}_2 x_2 + \cdots + \hat{\gamma}_p x_p\big), i = 1, 2, \ldots, n; \tag{30}$$

where, $\hat{\gamma}_0, \hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_p$ are the estimated coefficients of the probit function, and p represents the number of predictor variables..

2.4 FSNBLR Estimator

2.4.1 Logit Transformation Model

The logit transformation model is expressed as follows:

$$\ln\left(\frac{\pi(\boldsymbol{x_i})}{1-\pi(\boldsymbol{x_i})}\right) = f(x_{1i}, \ \dots \ , x_{pi}), \tag{31}$$

where f represents a regression equation or function (regression curve) that follows an additive model. Since The function $f(x_{1i}, \ \dots \ , x_{pi})$ in Eq. (31) can be approximated by a multivariate Fourier Series function as follows..

$$\ln\left(\frac{\pi(\boldsymbol{x_i})}{1-\pi(\boldsymbol{x_i})}\right) = \sum_{j=1}^{p}\left(b_j x_{ji} + \frac{1}{2}a_{0j} + \sum_{k=1}^{K} a_{kj}\cos kx_{ji}\right); \ i = 1,2,3, \dots n. \tag{32}$$

By using Eq. (32), FSNBLR model is obtained as follows.

$$\pi(x_i) = \frac{\exp\sum_{j=1}^{p}\left(b_j x_{ji} + \frac{1}{2}a_{0j} + \sum_{k=1}^{K} a_{kj}\cos kx_{ji}\right)}{1+\exp\sum_{j=1}^{p}\left(b_j x_{ji} + \frac{1}{2}a_{0j} + \sum_{k=1}^{K} a_{kj}\cos kx_{ji}\right)}; \ i = 1,2,3, \dots, n; \tag{33}$$

where, b_j, a_{0j} and a_{kj} , $j = 1,2,3, \dots, p$, $k = 1,2,3, \dots, K$ are the model parameters of the Fourier Series function.

2.4.2 Likelihood Function $l(\boldsymbol{\beta})$

The likelihood function $l(\beta)$ is expressed as:

where

$$\beta = (b_1 \quad a_{01} \quad a_{11} \quad \dots \quad a_{K1} \ \vdots \ \cdots \ \vdots \ b_p \quad a_{0p} \quad a_{1p} \quad \dots \quad a_{Kp}),$$

It is derived using the Maximum Likelihood Estimation (MLE) method.

$$l(\beta) = \prod_{i=1}^{n} P(Y_i = y_i) = \pi(x_i)^{\sum_{i=1}^{n} y_i}\left(1-\pi(x_i)\right)^{n-\sum_{i=1}^{n} y_i}. \tag{34}$$

Parameter estimation in logistic regression can be performed using the MLE method by maximizing the first derivative of the log-likelihood function. The likelihood function in Eq. (34) can be easily maximized using the expression $\ln l(\beta)$.

2.4.3 Log-Likelihood Function $L(\boldsymbol{\beta})$

$$\begin{aligned} L(\beta) = \ln[l(\beta)] &= \sum_{i=1}^{n} y_i \ln[\pi(x_i)] + \sum_{i=1}^{n} (1-y_i)\ln[1-\pi(x_i)] \\ &= \sum_{i=1}^{n}\left\{y_i\left(f(x_{1i}, \ \dots \ , x_{pi})\right) - \ln\left[1+\exp(f(x_{1i}, \ \dots \ , x_{pi}))\right]\right\}. \end{aligned} \tag{35}$$

The estimator $\hat{\beta}$ is obtained by taking the partial derivatives of the function in Eq. (35) with respect to b_j, a_{0j}, a_{kj} and then setting them equal to 0

$$\frac{\partial L(\beta)}{\partial b_j} = 0 \ ; \ j = 1,2,3, \dots, p;$$

$$\frac{\partial L(\beta)}{\partial a_{0j}} = 0 \ ; \ j = 1,2,3, \dots, p;$$

$$\frac{\partial L(\beta)}{\partial a_{kj}} = 0 \ ; \ k = 1,2,3, \dots, K \ ; \ j = 1,2,3, \dots, p.$$

The estimator $\hat{b}$ is derived using Eq. (36).

$$\sum_{i=1}^{n}\{(y_i - \pi(x_i))x_{ji}\} = 0. \tag{36}$$

The estimator $\hat{a}_0$ is derived using Eq. (37).

$$\sum_{i=1}^{n}\left\{\frac{1}{2}(y_i - \pi(x_i))\right\} = 0. \tag{37}$$

The estimator $\hat{a}_k$ will be obtained using Eq. (38).

$$\sum_{i=1}^{n}\left\{\sum_{k=1}^{K}\cos kx_{ji}\,(y_i - \pi(x_i))\right\} = 0. \tag{38}$$

2.4.4 Newton-Raphson Iteration

The derivative of $L(\beta)$ Eq. (35) against b_j, a_{0j}, a_{kj} as formulated in the implicit equation, does not yield a closed-form solution. Therefore, numerical iteration using the Newton–Raphson method is required to proceed..

$$\beta^{(t+1)} = \beta^{(t)} - \left(H(\beta)^{(t)}\right)^{-1} g(\beta)^{(t)}, \tag{39}$$

where $\beta^{(t)}$ is the β of the t-th iteration, $t = 1,2,\ldots,$ is convergent.

$$\beta^{(t)} = \left(b_1{}^{(t)} \quad a_{01}{}^{(t)} \quad a_{11}{}^{(t)} \quad \ldots \quad a_{K1}{}^{(t)} \quad \vdots \quad \cdots \quad \vdots \quad b_p{}^{(t)} \quad a_{0p}{}^{(t)} \quad a_{1p}{}^{(t)} \quad \ldots \quad a_{Kp}{}^{(t)}\right).$$

$\boldsymbol{g}(\boldsymbol{\beta})$ is the gradient vector of β

$$\boldsymbol{g}(\boldsymbol{\beta}) = \left(\frac{\partial L(\boldsymbol{\beta})}{\partial b_1}, \frac{\partial L(\boldsymbol{\beta})}{\partial a_{01}}, \frac{\partial L(\boldsymbol{\beta})}{\partial a_{11}}, \ldots, \frac{\partial L(\boldsymbol{\beta})}{\partial a_{K1}}, \ldots, \frac{\partial L(\boldsymbol{\beta})}{\partial b_p}, \frac{\partial L(\boldsymbol{\beta})}{\partial a_{0p}}, \frac{\partial L(\boldsymbol{\beta})}{\partial a_{1p}}, \ldots, \frac{\partial L(\boldsymbol{\beta})}{\partial a_{Kp}}\right)^{T}, \tag{40}$$

and $\boldsymbol{H}(\boldsymbol{\beta})$ represents the Hessian matrix of β in Eq. (40), defined as follows.

$$\boldsymbol{H}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial^2 L(\boldsymbol{\beta})}{\partial b_1{}^2} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial b_1 \partial a_{01}} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial b_1 \partial a_{Kp}} \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial a_{01} \partial b_1} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial a_{01}{}^2} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial a_{01} \partial a_{Kp}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L(\boldsymbol{\beta})}{\partial a_{Kp} \partial b_1} & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial a_{Kp} \partial a_{01}} & \cdots & \frac{\partial^2 L(\boldsymbol{\beta})}{\partial a_{Kp}{}^2} \end{bmatrix}. \tag{41}$$

The components of the vector $\boldsymbol{g}(\boldsymbol{\beta})$ Eq. (40) are derived from the first-order partial derivatives of $L(\beta)$ with respect to b_j, a_{0j}, a_{kj}, Likewise, the entries of the matrix $\boldsymbol{H}(\boldsymbol{\beta})$ Eq. (41) are obtained from the second-order partial derivatives of $L(\beta)$ with respect to b_u, a_{0u}, a_{ku}.

Second derivative of $L(\beta)$ function with respect to b_u

$$\frac{\partial^2 L(\beta)}{\partial b_u \partial b_j} = -\sum_{i=1}^{n} x_{ji} x_{ui}\, \pi(x_i)(1 - \pi(x_i)). \tag{42}$$

In the same manner as Eq. (42), the second derivative of the parameter combination is derived as follows..

$$\frac{\partial^2 L(\beta)}{\partial a_{ku} \partial b_j} = -\sum_{i=1}^{n}\sum_{k=1}^{K} \pi(x_i)(1 - \pi(x_i))x_{ji}\cos kx_{ui}\,. \tag{43}$$

Second derivative of $L(\beta)$ function with respect to a_{0u}

$$\frac{\partial^2 L(\beta)}{\partial a_{0u} \partial a_{0j}} = -\frac{1}{4}\sum_{i=1}^{n} \pi(x_i)\,(1 - \pi(x_i))\,. \tag{44}$$

In the same manner as Eq. (44), the second derivative of the parameter combination is obtained as follows.

$$\frac{\partial^2 L(\beta)}{\partial a_{ku}\partial a_{0j}} = -\frac{1}{2}\sum\nolimits_{i=1}^{n}\sum_{k=1}^{K}\pi(x_i)\big(1-\pi(x_i)\big)\cos kx_{ui}\,. \tag{45}$$

Second derivative of $L(\beta)$ function with respect to a_{ku}

$$\frac{\partial^2 L(\beta)}{\partial a_{ku}\partial a_{kj}} = -\sum\nolimits_{i=1}^{n}\sum_{k=1}^{K}\cos kx_{ji}\sum_{k=1}^{K}\cos kx_{ui}\,\pi(\boldsymbol{x_i})\big(1-\pi(\boldsymbol{x_i})\big)\,. \tag{46}$$

In the same manner as Eq. (46), the second derivative of the parameter combination is obtained as follows.

$$\frac{\partial^2 L(\boldsymbol{\beta})}{\partial a_{0u}\partial a_{kj}} = -\frac{1}{2}\sum\nolimits_{i=1}^{n}\sum_{k=1}^{K}\pi(\boldsymbol{x_i})\big(1-\pi(\boldsymbol{x_i})\big)\cos kx_{ji}\,. \tag{47}$$

where, b_j, a_{0j} and a_{kj} , $j, u = 1,2,3,\dots,p$, $j \neq u$, $k = 1,2,3,\dots,K$ are the model parameters of the Fourier Series function.

2.4.5 Estimator $\hat{\boldsymbol{\beta}}$

The estimate, $\hat{\beta}$ can be obtained from the Newton-Raphson iteration equation when:

$$\left|\beta^{(t+1)} - \beta^{(t)}\right| < \varepsilon,\ \varepsilon = 0.000001 \tag{48}$$

Thus, the estimator $\hat{\beta}$ is given by

$$\hat{\beta} = \begin{pmatrix}\hat{b}_1 & \hat{a}_{0_1} & \hat{a}_{1_1} & \dots & \hat{a}_{K_1} & \vdots & \cdots & \vdots & \hat{b}_p & \hat{a}_{0_p} & \hat{a}_{1_p} & \dots & \hat{a}_{K_p}\end{pmatrix}.$$

Based on the result of the estimator $\hat{\beta}$, FSNBLR model Eq. (49) can be written:

$$\hat{\pi}(x_i) = \frac{\exp\left(\hat{b}_1 x_{1i} + \frac{1}{2}\hat{a}_{01} + \hat{a}_{11}\cos x_{1i} + \cdots + \hat{a}_{K1}\cos Kx_{1i} + \cdots + \hat{b}_p x_{pi} + \frac{1}{2}\hat{a}_{0p} + \hat{a}_{1p}\cos x_{pi} + \cdots + \hat{a}_{Kp}\cos Kx_{pi}\right)}{1 + \exp\left(\hat{b}_1 x_{1i} + \frac{1}{2}\hat{a}_{01} + \hat{a}_{11}\cos x_{1i} + \cdots + \hat{a}_{K1}\cos Kx_{1i} + \cdots + \hat{b}_p x_{pi} + \frac{1}{2}\hat{a}_{0p} + \hat{a}_{1p}\cos x_{pi} + \cdots + \hat{a}_{Kp}\cos Kx_{pi}\right)}, \tag{49}$$

where, $\hat{b}_1, \hat{a}_{01}$ and $\hat{a}_{k1}$ are the estimator model of the Fourier Series function for predictor variable x_1, while $\hat{b}_p, \hat{a}_{0p}$ and $\hat{a}_{kp}$ are for predictor variable x_p, K represents the number of oscillation parameters, and p denotes the number of predictor variables.

2.5 Hypothesis Test for Parameter Model

Hypothesis test for parameter model consists of simultaneous and partial tests. Hypothesis test for simultaneous uses the Likelihood Ratio Test (LRT) and hypothesis test for partial uses the Wald test.

2.5.1 Simultaneous

The simultaneous test is conducted to determine the significance of parameter θ as a whole or simultaneously, where θ is parameters for the model.

Hypothesis:

H_0 : $\theta_1 = \theta_2 = \theta_3 = \cdots = \theta_j = \cdots = \theta_k = 0$,
H_1 : There is at least one $\theta_j \neq 0$.

Statistics test for simultaneous test Eq. (50):

$$G^2 = -2\sum\nolimits_{i=1}^{n}\left[y_i \ln\left(\frac{\hat{\pi}(x_i)}{y_i}\right) + (1-y_i)\ln\left(\frac{1-\hat{\pi}(x_i)}{1-y_i}\right)\right]. \tag{50}$$

Decision:
Reject H_0 when $G^2 > \chi^2_{(v,a)}$ or $p - value < a$.

2.5.2 Partial

Hypothesis:

$H_0: \theta_j = 0, j = 1,2, \dots, k;$
$H_1: \theta_j \neq 0.$

Statistics test for partial test Eq. (51):

$$W = \frac{\hat{\theta}_j}{\widehat{SE}(\hat{\theta}_j)}. \tag{51}$$

Decision:
Reject H_0 when $W > \chi^2_{(v,a)}$ or $p - value < a$.

3. RESULTS AND DISCUSSION

This study utilizes secondary data on diabetes mellitus status obtained from the Internal Medicine Clinic of Hajj General Hospital Surabaya, collected in August 2018. The dataset consists of one dependent variable (y) and three independent variables (x). as summarized in Table 1. These variables are employed as inputs in estimating the BLR, BPR, and FSNBLR models to assess the performance of each method in modeling diabetes status.

Table 1. Variable Description

Variable	Notation	Description	Unit	Scale
Response	y	Status of Type 2 Diabetes Mellitus	0 = Doesn't have Diabetes Mellitus 1 = Has Diabetes Mellitus	Nominal
Predictor	x_1	Age	Year	Ratio
	x_2	Body Mass Index	kg/m^2	Ratio
	x_3	Abdominal Circumference	cm	Ratio

Based on Table 1, these variables were chosen for their medical relevance and their availability in the patient records at Hajj General Hospital Surabaya. They are utilized to model the probability of having Type 2 Diabetes Mellitus using the BLR, BPR, and FSNBLR methods. The dataset includes 60 patients, comprising 39 diagnosed with diabetes mellitus and 21 without, as depicted in Fig. 1.

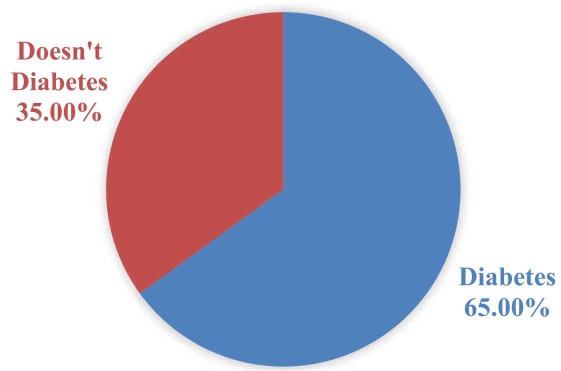

Figure 1. Status of Type 2 Diabetes Mellitus

3.1 Descriptive Analytics

Descriptive analysis is conducted to examine the characteristics of each variable in the dataset, as presented in Table 2.

Table 2. Descriptive Statistics of Research Variables

Category	Variable	Mean	StDev	Minimum	Maximum
Doesn't have Diabetes Mellitus	Age	47.0952	16.4405	17	71
	IMT	22.3681	4.47414	16.02	31.25
	Abdominal Mass	84.9524	13.1851	64	119
Has Diabetes Mellitus	Age	62.9487	8.72066	51	83
	IMT	25.5018	3.54516	18.49	33.78
	Abdominal Mass	93.6410	8.39952	82	115

Table 2 summarizes the characteristics of the variables related to diabetes mellitus status. The dataset contains no missing values, and diagnostic checks confirm the absence of multicollinearity among the predictors. The conceptual predictor structure used in this study is presented in Fig. 2.

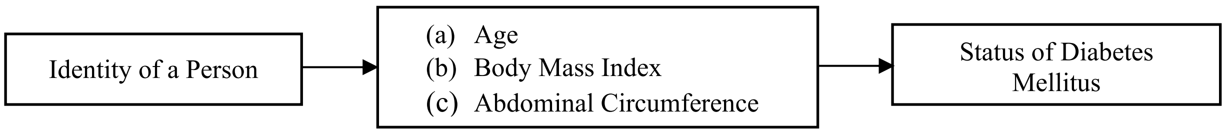

Figure 2. Conceptual Diagram of Variables

As shown in Fig. 2, diabetes mellitus is a multifaceted issue, one aspect of which is an individual's identity. According to the Central Bureau of Statistics, a person's identity includes factors such as age, body mass index, and abdominal circumference.

We created scatterplots for each predictor variable, categorized into several groups, to examine the relationship with the proportion of individuals diagnosed with diabetes mellitus $(y = 1)$ in each group. The plot shows the proportion of patients diagnosed with diabetes mellitus relative to the total number of patients in each group. The scatterplot is displayed in Fig. 3 below..

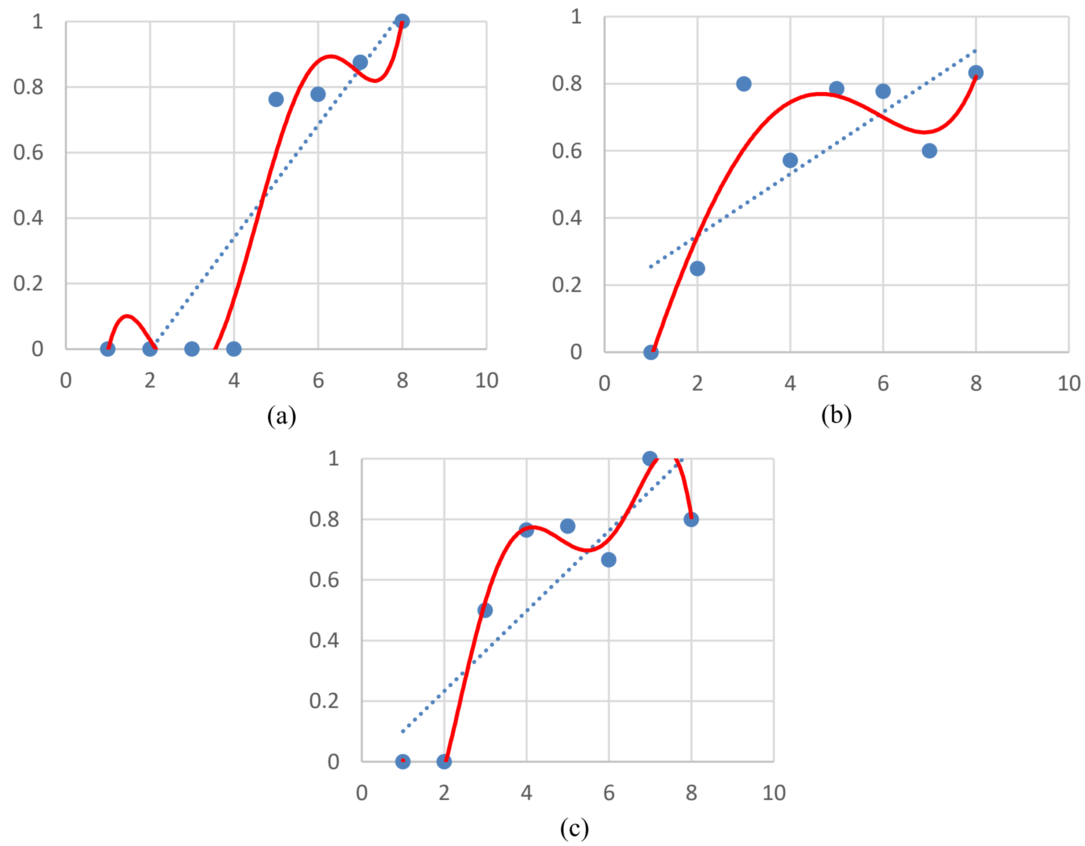

Figure 3. Scatterplots of Several Data Groups Versus the Number of Has Diabetes Mellitus in the Group (a) Age, (b) Body Mass Index, (c) Abdominal Circumference

As shown in Fig. 3, the probability of having diabetes mellitus ($y = 1$) for variable x_1, x_2, and x_3 exhibits a repeating pattern and follows an upward trend. Therefore, the logit function, which assumes a linear relationship, does not accurately capture the pattern observed in this case.

To model the status of diabetes mellitus, we use the BLR, BPR, and FSNBLR methods. The parameter estimates and significant parameters in the resulting model are presented below..

3.2 BLR Model

The BLR model follows Eq. (9) as follows.

$$\pi(x_i) = \frac{\exp\sum_{j=1}^{p}(\beta_0+\beta_j x_{ji})}{1+\exp\sum_{j=1}^{p}(\beta_0+\beta_j x_{ji})} \; ; \; i = 1,2,\dots,n$$

where, β_0 and β_j , $j = 1,2,\dots,p$ are the model parameters of the logit function.

3.2.1 Parameter Estimation in BLR Model Results

According to the BLR model in Eq. (16), the parameter estimation results for the diabetes mellitus data are as follows.

$$\hat{\pi}(\boldsymbol{x_i}) = \frac{\exp(-11.781 + 0.118x_{1i} + 0.132x_{2i} + 0.026x_{3i})}{1 + \exp(-11.781 + 0.118x_{1i} + 0.132x_{2i} + 0.026x_{3i})}$$

3.2.2 Significant Parameter in BLR Model Results

Based on the parameter estimation of the BLR model, the significant parameters for the diabetes mellitus data are presented in Table 3.

Table 3. Significant Parameter in BLR Model

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-11.781	4.40281	-2.676	0.00745
x_1	0.118	0.04147	2.847	0.00441
x_2	0.132	0.11749	1.124	0.26102
x_3	0.026	0.05101	0.52	0.60282

The results inTable 3 indicate that only the age variable is significant in the model. Hence, age has an impact on an individual's diabetes mellitus status..

3.3 BPR Model

The BPR model is as follows Eq. (18).

$$\Phi(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right]$$

3.3.1 Parameter Estimation in BPR Model Results

According to the BPR model in Eq. (30), the parameter estimation results for the diabetes mellitus data are as follows.

$$P(y_i = 1) = \Phi(6.997 + 0.071x_1 + 0.078x_2 + 0.015x_3)$$
$$P(y_i = 0) = 1 - \Phi(6.997 + 0.071x_1 + 0.078x_2 + 0.015x_3)$$

3.3.2 Significant Parameter in BPR Model Results

Based on parameter estimation BPR model, the results of significant parameter in the BPR model for data on diabetes mellitus can be seen in Table 4.

Table 4. Significant Parameter in BPR Model

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	6.997	2.42416	-2.886	0.0039
x_1	0.071	0.02263	3.13	0.00175
x_2	0.078	0.06947	1.133	0.25738
x_3	0.015	0.02947	0.516	0.60551

The results in Table 4 show that only the age variable is significant in the model. Therefore, age influences an individual's diabetes mellitus status..

3.4 FSNBLR Model

3.4.1 Selecting Optimal Oscillation Parameters

The selection of oscillation parameters in the FSNBLR model was guided by the minimum AIC value. To prevent unnecessary model complexity while preserving interpretability, the number of oscillation parameters was restricted. Using an R-based algorithm, the AIC values for each parameter combination are summarized in Table 5.

Table 5. Minimum AIC Results for Each Number of Oscillation Parameter

Number of Oscillation Parameter	Oscillation Parameter Combination			AIC (K)
	x_1	x_2	x_3	
$K = 1$	1	1	1	61.543
$K = 2$	1	2	1	57.837
$K = 3$	1	2	1	**57.837**

According to Table 5, the FSNBLR model with the optimal oscillation settings is obtained using the combination $x_1 = 1, x_2 = 2, x_3 = 1$ as it produces the lowest AIC value. This combination was selected by evaluating different parameter configurations and choosing the one that yielded the smallest AIC for each level of oscillation complexity, thereby achieving the best balance between model fit and complexity.

3.4.2 Parameter Estimation in FSNBLR Model Results

According to the FSNBLR model in Eq. (33), the parameter estimation results for the diabetes mellitus data in Eq. (49) are as follows.

$$\hat{\pi}(x_i) = \frac{\exp(-15.56 + 0.16x_{1i} + 0.81\cos x_{1i} + 0.22x_{2i} - 0.65\cos x_{2i} + 1.40\cos 2x_{2i} + 0.01x_{3i} - 1.40\cos x_{3i})}{1 + \exp(-15.56 + 0.16x_{1i} + 0.81\cos x_{1i} + 0.22x_{2i} - 0.65\cos x_{2i} + 1.40\cos 2x_{2i} + 0.01x_{3i} - 1.40\cos x_{3i})}$$

More details can be seen in Table 6.

Table 6. Parameter Estimation in FSNBLR

Parameters	Estimations
β_0	-15.564
b_1	0.165
$a_{1,1}$	0.813
b_2	0.229
$a_{1,2}$	-0.659
$a_{2,2}$	1.408
b_3	0.016
$a_{1,3}$	-1.401

3.4.3 Significant Parameter in FSNBLR Model Results

Based on the parameter estimation of the FSNBLR model, the significant parameters for the diabetes mellitus data are shown in Table 7.

Table 7. Significant Parameter in FSNBLR Model

	Estimate	Std. Error	z value	Pr(>|z|)
Intercept	-15.565	5.9572	-2.613	0.00898
x_1	0.16555	0.05548	2.984	0.00285
x_2	0.81312	0.70369	1.156	0.24788
x_3	0.2293	0.15428	1.486	0.13721
x_4	-0.6596	0.66471	-0.992	0.32105
x_5	1.40888	0.65126	2.163	0.03052
x_6	0.01656	0.05883	0.281	0.77836
x_7	-1.40055	0.66829	-2.096	0.03611

The results in Table 7 indicate that the variables age, body mass index, and abdominal circumference are significant in the model. Therefore, age influences an individual's diabetes mellitus status.

3.5 Comparison of BLR, GWBLR, FSNBLR

3.5.1 Getting the Best Model Based on Deviance Value

The regression model selected is the one with the smallest deviance value. The results from the deviance statistical test are shown in Table 8.

Table 8. Comparison of Deviance Values

Methods	Deviance Values
BLR	53.007
BPR	52.728
FSNBLR	**41.837**

According to Table 8, the FSNBLR model yields the lowest deviance value (41.837) compared to BLR (53.007) and BPR (52.728). This indicates that FSNBLR provides the best fit for predicting diabetes mellitus status, as a smaller deviance value reflects a superior model performance..

Getting the Best Classification Based on AUC & Press's Q Value

The selected FSNBLR model achieved the highest AUC and the lowest Press’s Q. The results of the classification test are presented in Table 9.

Table 9. Comparison of AUC and Press's Q

Methods	Accuracy	Sensitivity	Specificity	AUC	Press’s Q	Chi Square
BLR	73.33%	42.85%	89.74%	66.30%	13.067	51.829
BPR	73.33%	42.85%	89.74%	66.30%	13.067	51.623
FSNBLR	85%	71.42%	92.31%	81.86%	29.400	49.879

According to Table 9, Case 1 shows that the FSNBLR model attains a higher AUC (81.86%) compared with BLR (66.30%) and BPR (66.30%). Furthermore, the larger Press’s Q value for FSNBLR (29.400) demonstrates its superior classification ability and a stronger tendency to rejec H0 or Press's Q > Chi Square. These findings are derived from diabetes mellitus patient data from Hajj General Hospital Surabaya, with a visual comparison presented Fig. 4.

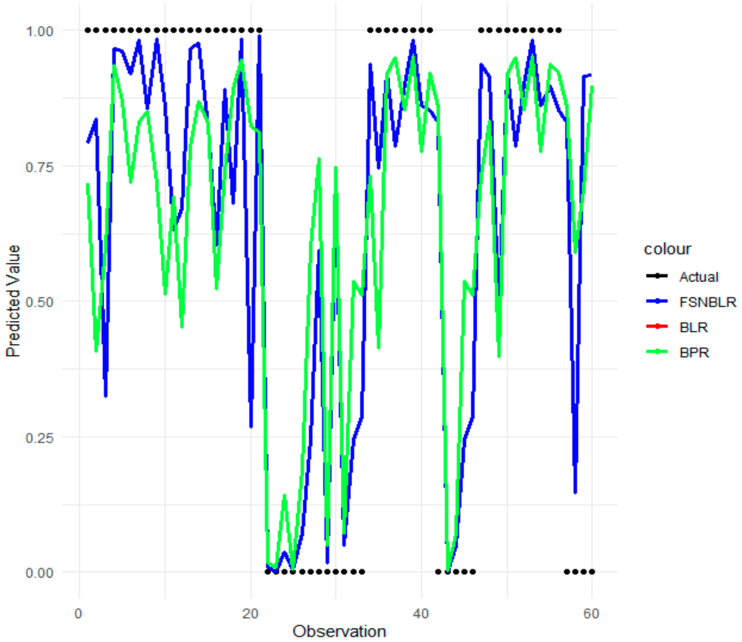

Figure 4. Comparison of Predicted Values in BLR, BPR, and FSNBLR

As shown in Fig. 4, the plots demonstrate that the predicted values from the three methods fluctuate, and no method consistently outperforms the others. However, for specific cases, such as the one discussed in this article, FSNBLR tends to perform better than both BLR and BPR. The FSNBLR model is superior as it provides odds estimates that closely align with the actual values in nearly all of the selected observations.

4. CONCLUSION

Based on the discussion that has been described, these findings align with the theoretical advantage of the FSNBLR model, which incorporates oscillatory components (e.g., cosine functions) to capture nonlinear and repeating patterns in the data. This makes FSNBLR especially effective for modeling categorical response variables influenced by predictors with non-monotonic or cyclical relationships—common in complex medical data such as diabetes mellitus risk factors.

Based on the data used in this study, the FSNBLR model is the best for predicting the status of diabetes mellitus, as shown below. The FSNBLR model for categorical data is:

$$\hat{\pi}(x_i) = \frac{\exp(-15.56 + 0.16x_{1i} + 0.81\cos x_{1i} + 0.22x_{2i} - 0.65\cos x_{2i} + 1.40\cos 2x_{2i} + 0.01x_{3i} - 1.40\cos x_{3i})}{1 + \exp(-15.56 + 0.16x_{1i} + 0.81\cos x_{1i} + 0.22x_{2i} - 0.65\cos x_{2i} + 1.40\cos 2x_{2i} + 0.01x_{3i} - 1.40\cos x_{3i})}$$

The deviance value for the FSNBLR model (41.837) is lower than that of the BLR (53.007) and BPR (52.728), further confirming that the FSNBLR model provides a better fit to the data. With an AUC value of 81.86%, the FSNBLR model outperforms both BLR and BPR in terms of estimation. Additionally, the FSNBLR model shows higher accuracy, sensitivity, and specificity compared to BLR and BPR, indicating superior performance.

Author Contributions

Bambang Widjanarko Otok: Formal Analysis, Funding Acquisition, Writing - Review and Editing. Muhammad Zulfadhli: Investigation, Resources, Validation, Writing - Review and Editing; Riwi Dyah Pangesti: Methodology, Visualization, Writing - Review and Editing. Muhammad Idham Kurniawan: Software, Data Curation, Writing - Original Draft. Albertus Eka Putra Haryanto: Investigation, Project Administration, Writing – Original Draft. Darwis: Investigation, Resources, Validation. Iwan Kurniawan: Resources, Writing - Original Draft. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors would like to thank Institut Teknologi Sepuluh Nopember, Bengkulu University, Regional Economic Development Institute (REDI), STAIN Majene, Politeknik STIA LAN Bandung for their technical support, laboratory facilities, and other non-authorial contributions that supported the completion of this research.

Declarations

The authors declare no conflicts of interest to report study.

REFERENCES

[1] F.E. Harrell Jr., *REGRESSION MODELING STRATEGIES: WITH APPLICATIONS TO LINEAR MODELS, LOGISTIC AND ORDINAL REGRESSION, AND SURVIVAL ANALYSIS*, 2nd ed., Springer, 2015. doi: https://doi.org/10.1198/tech.2003.s158.

[2] J.J. Faraway, *EXTENDING THE LINEAR MODEL WITH R: GENERALIZED LINEAR, MIXED EFFECTS AND NONPARAMETRIC REGRESSION MODELS*, 2nd ed., Chapman & Hall/CRC, 2016.doi: https://doi.org/10.1201/9781315382722

[3] M. M. Panja and B. N. Mandal, *WAVELET BASED APPROXIMATION SCHEMES FOR SINGULAR INTEGRAL EQUATIONS*. 2020. doi: https://doi.org/10.4324/9780429244070.

[4] Y. Farida, I. Purwanti, and N. Ulinnuha, "COMPARING GAUSSIAN AND EPANECHNIKOV KERNEL OF NONPARAMETRIC REGRESSION IN FORECASTING ISSI (INDONESIA SHARIA STOCK INDEX)," *BAREKENG: Journal of Mathematics and Applications*, vol. 16, no. 1, pp. 323-332, 2022. doi: https://doi.org/10.30598/barekengvol16iss1pp321-330.

[5] B. Pratama, A. Suryono, N. Auliyah, and N. Chamidah, "COMPARISON OF LOCAL POLYNOMIAL REGRESSION AND ARIMA IN PREDICTING THE NUMBER OF FOREIGN TOURIST VISITS TO INDONESIA," *BAREKENG: Journal of Mathematics and Applications*, vol. 18, no. 1, pp. 53-64, 2024.doi: https://doi.org/10.30598/barekengvol18iss1pp0043-0052

[6] M.D. Cattaneo, M. Jansson, and X. Ma, "LOCAL REGRESSION DISTRIBUTION ESTIMATORS," *Journal of Econometrics*, vol. 240, no. 2, p. 105074, 2024.doi: https://doi.org/10.1016/j.jeconom.2021.01.006

[7] S.D.P. Yasmirullah, B.W. Otok, J.D.T. Purnomo, and D.D. Prastyo, "PARAMETER ESTIMATION OF MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) WITH STEPWISE APPROACH TO MULTI DRUG-RESISTANT TUBERCULOSIS (MDR-TB) MODELING IN LAMONGAN REGENCY," *Journal of Physics: Conference Series*, vol. 1752, p. 012017, IOP Publishing, 2021.doi: https://doi.org/10.1088/1742-6596/1752/1/012017

[8] S.D.P. Yasmirullah, B.W. Otok, J.D.T. Purnomo, and D.D. Prastyo, "MODIFICATION OF MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS)," *Journal of Physics: Conference Series*, vol. 1863, p. 012017, IOP Publishing, 2021. doi: https://doi.org/10.1088/1742-6596/1863/1/012078.

[9] P.A. Morettin and R.F. Porto, "ESTIMATION OF NONPARAMETRIC REGRESSION MODELS BY WAVELETS," *São Paulo Journal of Mathematical Sciences*, vol. 16, pp. 539–565, 2022. doi: https://doi.org/10.1007/s40863-021-00240-5.

[10] M.D. Pasarella, Sifriyani, and F.D.T. Amijaya, "NONPARAMETRIC REGRESSION MODEL ESTIMATION WITH THE FOURIER SERIES APPROACH AND ITS APPLICATION TO THE ACCUMULATIVE COVID-19 DATA IN INDONESIA," *BAREKENG: Journal of Mathematics and Its Application*, vol. 16, no. 4, pp. 1167–1174, 2022. doi: https://doi.org/10.30598/barekengvol16iss4pp1167-1174.

[11] B.B. Jena, S.K. Paikray, and M. Mursaleen, "ON THE DEGREE OF APPROXIMATION OF FOURIER SERIES BASED ON A CERTAIN CLASS OF PRODUCT DEFERRED SUMMABILITY MEANS," *Journal of Inequalities and Applications*, vol. 2023, no. 18, 2023. doi: https://doi.org/10.1186/s13660-023-02927-z.

[12] D.S. Sheiso, "APPROXIMATION OF FUNCTIONS USING FOURIER SERIES AND ITS APPLICATION TO THE SOLUTION OF PARTIAL DIFFERENTIAL EQUATIONS," *Science Journal of Applied Mathematics and Statistics*, vol. 10, no. 4, pp. 57-84, 2022. doi: https://doi.org/10.2139/ssrn.4775015.

[13] W. Gazali, "THE ORIGIN OF THE BASIC FORMULA OF THE FOURIER SERIES," *Engineering, Mathematics and Computer Science Journal (EMACS)*, vol. 5, no. 1, 2023. doi: https://doi.org/10.21512/emacsjournal.v5i1.9398.

[14] M.L. Maslakov and V.V. Egorov, "FOR THE PROBLEM OF PHASE PROBABILITY DENSITY FUNCTION ESTIMATION," *Numerical Analysis and Applications*, vol. 15, no. 2, pp. 125–137, 2022. doi: https://doi.org/10.15372/SJNM20220205.

[15] Kuzairi, Miswanto, and M.F.F. Mardianto, "SEMIPARAMETRIC REGRESSION BASED ON FOURIER SERIES FOR LONGITUDINAL DATA WITH WEIGHTED LEAST SQUARE (WLS) OPTIMIZATION," *Journal of Physics: Conference Series*, vol. 1836, no. 1, 2021. doi: https://doi.org/10.1088/1742-6596/1836/1/012038.

[16] H. Husain, I. N. Budiantara, and I. Zain, "MIXED ESTIMATOR OF SPLINE TRUNCATED, FOURIER SERIES, AND

KERNEL IN BIRESPONSE SEMIPARAMETRIC REGRESSION MODEL," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 880, no. 1, 2021, doi: https://doi.org/10.1088/1755-1315/880/1/012046.

[17] I. Wayan Sudiarsa, I. Nyoman Budiantara, S. Suhartono, and S. W. Purnami, "COMBINED ESTIMATOR FOURIER SERIES AND SPLINE TRUNCATED IN MULTIVARIABLE NONPARAMETRIC REGRESSION," *Appl. Math. Sci.*, vol. 9, no. 97–100, pp. 4997–5010, 2015, doi: https://doi.org/10.12988/ams.2015.55394.

[18] N. P. A. M. Mariati, I. N. Budiantara, and V. Ratnasari, "COMBINATION ESTIMATION OF SMOOTHING SPLINE AND FOURIER SERIES IN NONPARAMETRIC REGRESSION," *J. Math.*, vol. 2020, 2020, doi: https://doi.org/10.1155/2020/4712531.

[19] I.N. Budiantara, V. Ratnasari, M. Ratna, W. Wibowo, N. Afifah, D.P. Rahmawati, and M.A.D. Octavanny, "MODELING PERCENTAGE OF POOR PEOPLE IN INDONESIA USING KERNEL AND FOURIER SERIES MIXED ESTIMATOR IN NONPARAMETRIC REGRESSION," *Investigacion Operacional*, vol. 40, pp. 538-551, 2019.

[20] K. Nisa, I.N. Budiantara, and A.T. Rumiati, "MULTIVARIABLE SEMIPARAMETRIC REGRESSION MODEL WITH COMBINED ESTIMATOR OF FOURIER SERIES AND KERNEL," in *IOP Conference Series: Earth and Environmental Science*, pp. 012028, IOP Publishing, 2017.doi: https://doi.org/10.1088/1755-1315/58/1/012028

[21] A.T. Ampa, I.N. Budiantara, and I. Zain, "MODELING THE LEVEL OF DRINKING WATER CLARITY IN SURABAYA CITY DRINKING WATER REGIONAL COMPANY USING COMBINED ESTIMATION OF MULTIVARIABLE FOURIER SERIES AND KERNEL," *Sustainability*, vol. 14, pp. 13663, 2022. doi: https://doi.org/10.3390/su142013663.

[22] M. Ramli, I.N. Budiantara, and V. Ratnasari, "A METHOD FOR PARAMETER HYPOTHESIS TESTING IN NONPARAMETRIC REGRESSION WITH FOURIER SERIES APPROACH," *MethodsX*, vol. 11, p. 102468, 2023. doi: https://doi.org/10.1016/j.mex.2023.102468.

[23] P.H. Jou and S.H. Mirhashemi, "FREQUENCY ANALYSIS OF EXTREME DAILY RAINFALL OVER AN ARID ZONE OF IRAN USING FOURIER SERIES METHOD," *Applied Water Science*, vol. 13, p. 16, 2023. doi: https://doi.org/10.1007/s13201-022-01823-z.

[24] L. Laome, I.N. Budiantara, and V. Ratnasari, "POVERTY MODELING WITH SPLINE TRUNCATED, FOURIER SERIES, AND MIXED ESTIMATOR GEOGRAPHICALLY WEIGHTED NONPARAMETRIC REGRESSION," in *AIP Conference Proceedings*, AIP Publishing, 2024. doi: https://doi.org/10.1063/5.0206173.

[25] S. Suliyanto, M. Rifada, and E. Tjahjono, "ESTIMATION OF NONPARAMETRIC BINARY LOGISTIC REGRESSION MODEL WITH LOCAL LIKELIHOOD LOGIT ESTIMATION METHOD (CASE STUDY OF DIABETES MELLITUS PATIENTS AT SURABAYA HAJJ GENERAL HOSPITAL)," in *Symposium on Biomathematics 2019*, pp. 1551-7616, AIP Conference Proceedings, Bali, 2020. doi: https://doi.org/10.1063/5.0025807.

[26] H. Hamie, A. Hoayek, B. El-Ghoul, and M. Khalifeh, "APPLICATION OF NON-PARAMETRIC STATISTICAL METHODS TO PREDICT PUMPABILITY OF GEOPOLYMERS FOR WELL CEMENTING," *Journal of Petroleum Science and Engineering*, vol. 212, p. 110333, 2022. doi: https://doi.org/10.1016/j.petrol.2022.110333.

[27] T. Wang, W. Tang, Y. Lin, and W. Su, "SEMI-SUPERVISED INFERENCE FOR NONPARAMETRIC LOGISTIC REGRESSION," *Statistics in Medicine*, vol. 42, pp. 2573–2589, 2023. doi: https://doi.org/10.1002/sim.9737.

[28] M. Zulfadhli, I.N. Budiantara, and V. Ratnasari, "NONPARAMETRIC REGRESSION ESTIMATOR OF MULTIVARIABLE FOURIER SERIES FOR CATEGORICAL DATA," *MethodsX*, p. 102983, 2024. doi: https://doi.org/10.1016/j.mex.2024.102983.

[29] V. Ratnasari, S. H. Utama, and A. T. R. Dani, "TOWARD SUSTAINABLE DEVELOPMENT GOALS (SDGS) WITH STATISTICAL MODELING: RECURSIVE BIVARIATE BINARY PROBIT," *IAENG Int. J. Appl. Math.*, vol. 54, no. 8, pp. 1515–1521, 2024.