The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 2026

Volume 40, Pages 264-272

ICBAST 2026: International Conference on Basic Sciences and Technology

A Truncated Spline Nonparametric Logistic Regression: Simultaneous Testing and Model Comparison

Afiqah Saffa Suriaslan
Sepuluh Nopember Institute of Technology

I Nyoman Budiantara
Sepuluh Nopember Institute of Technology

Vita Ratnasari
Sepuluh Nopember Institute of Technology

Dursun Aydin
Mugla Sıtkı Kocman University

Abstract: Regression is widely used in statistical analysis. For binary response variables, Logistic Regression has been widely applied due to its interpretability. However, conventional Binary Logistic Regression (BLR) generally assumes a linear relationship between predictors and the response, which may limit its ability to capture complex data patterns. To address this limitation, this study examines a Truncated Spline Nonparametric Logistic Regression (TSNLR) model that provides greater flexibility in modeling nonlinear relationships through the use of knot points. Statistical inference of the proposed model is examined using simultaneous hypothesis testing to assess the significance of predictor variables. The TSNLR model is applied to Human Development Index (HDI) data of cities/districts in Indonesia, which are classified into high and low HDI categories. Model performance is evaluated and compared with BLR using classification performance metrics. These findings suggest that TSNLR is an effective alternative for modeling binary response data with complex predictor relationships and can offer improved interpretability in applied statistical analysis.

Keywords: Truncated spline, Binary logistic, Simultaneous test, Human Development Index (HDI)

Introduction

Regression has become an analytical framework in statistical, enabling the prediction of outcomes and trends based on historical data (Yahya, 2022). Over the past few decades, regression analysis has significantly advanced from classical statistical approaches to more advanced methods (Joshi & Dhakal, 2021; Marasco et al., 2022). In the case of binary response variables that follow a Bernoulli distribution, binary logistic regression becomes one of the most widely used methods (Hosmer & Lemeshow, 2000). Through this approach, model parameter estimates can be obtained and used to explain the effect of predictor variables on the probability of an event occurring.

Various studies have examined binary logistic regression (Manoharan et al., 2020, Saifudin et al., 2024; Alfardus & Rawat, 2024) specially from the perspective of statistical inference, such as parameter hypothesis testing (Rusliyadi et al., 2021; Yuniarsih et al., 2024) and confidence interval formation (Gupta et al., 2020; Zhang et al., 2020). However, most of these studies in the fiel of binary logistic regression have only focused on the assumption of a linear relationship between predictor variables and logit functions, thereby limiting the flexibility of the model when faced with data with more complex patterns.

In practice, the relationship between predictor and response variables is often nonlinear (Eubank, 1999). This condition has encouraged the development of nonparametric regression approaches that do not depend on specific function forms, such as truncated splines (Budiantara, 2005). This approach is considered more effective because this method works by dividing the predictor variable domain into several subintervals through knot points, so that changes in relationship patterns in each subinterval can be modeled more accurate (Hazelton & Turlach, 2011), (Permatasari et al., 2021). Further developments have led to the use of nonparametric approaches in binary logistic regression, including local likelihood approaches to logistic models (Suliyanto et al., 2020), Fourier series functions (Zulfadhli et al., 2025), and truncated spline function integration (Suriaslan, et al., 2025a, 2025b).

Based on previous studies, this article aims to develop a binary logistic regression model using a truncated spline approach and examine its statistical inference through simultaneous parameter significance testing. The proposed modeling framework is applied to data on the Human Development Index (HDI) of districts/cities in Indonesia in 2023, which are categorized into high and low HDI levels. Furthermore, the performance of the proposed method is compared with traditional binary logistic regression to evaluate improvements in modeling performance.

Method

The research methodology is presented using a workflow diagram, as shown in Figure 1.

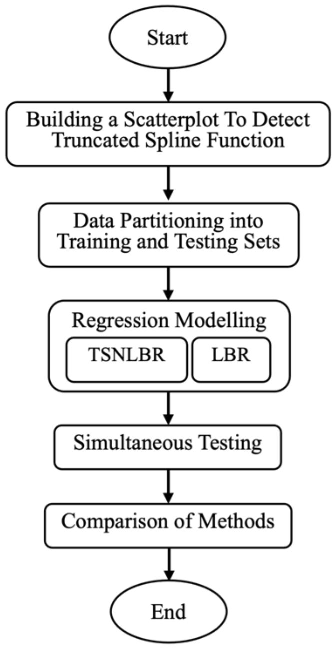

Figure. 1. Analysis workflow

Data Application

This study used secondary data obtained from the Indonesia Central Bureau of Statistics (BPS). The sample consists of 514 districts/cities in Indonesia for the year 2023, of which 305 are classified as having high HDI, while 209 are classified as having low HDI. The details of the variables are presented in Table I.

Table 1. Variable research

Variables	Description
y	Human Development Index (HDI) 0 = Low HDI 1 = High HDI
v_1	Household access to clean water (%)
v_2	Poverty rate
v_3	Rate of open unemployment
v_4	High school participation rate

HDI modeling using the TSNBLR method begins by exploring the relationship between the logit of the response variable and the predictor variable. This exploration is presented in a scatter diagram in Fig. 2.

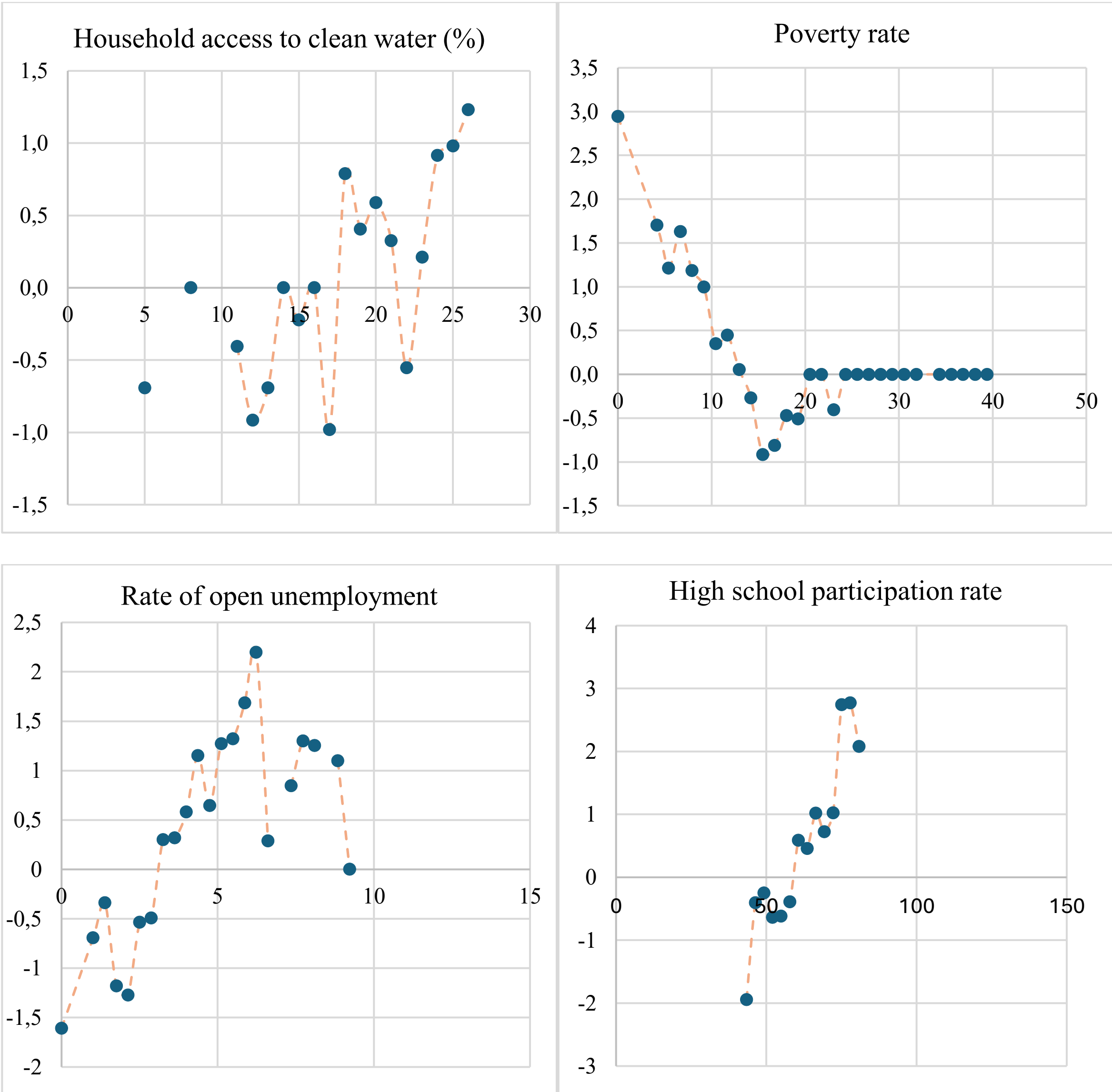

Figure 2. Scatterplots of the relationships between logit function of HDI and predictor variables

Based on the scatter plot in Fig. 2, it can be initially identified that there is a tendency for the relationship to fluctuate in certain sub-intervals. Therefore, in this study, it is appropriate to model the HDI using the Truncated Spline approach.

Binary Logistic Regression (BLR)

Agresti (2002) explains that the logistic regression model is expressed as $\pi(\boldsymbol{w_i})$, where $\pi(\boldsymbol{w_i})$ is defined as follows:

$$\pi(\boldsymbol{w_i}) = \frac{\exp(\beta_0 + \beta_1 w_{1i} + \beta_2 w_{2i} + \cdots + \beta_p w_{pi})}{1 + \exp(\beta_0 + \beta_1 w_{1i} + \beta_2 w_{2i} + \cdots + \beta_p w_{pi})} \quad (1)$$

The response variable in logistic regression is binary and will have a probability of $\pi(\boldsymbol{v_i})$ if it is 1, and probability of $1 - \pi(\boldsymbol{v_i})$ if it is 0.

Simultaneous tests can determine whether the regression model used in general has sufficient ability to explain the relationship between the response variable and the predictor variable. The hypotheses used in simultaneous tests are as follows:

$H_0: \beta_1 = \beta_2 = \cdots \beta_p = 0$
H_1: At least there is one $\beta_j \neq 0$, $j = 1,2, \dots, p$

The test statistics used based on the Likelihood Ratio Test (LRT) method are as follows (Hosmer & Lemeshow, 2000):

$$G^2 = 2\left[\ln l\left(\widehat{\mathbf{\Omega}}\right) - \ln l(\widehat{\omega})\right] \quad (2)$$

Where

ω : set of parameters under H_0
$\mathbf{\Omega}$: set consisting of parameters under the population

The null hypothesis H_0 is rejected if the test statistic G^2 exceeds the chi-square critical value $\chi^2_{(\propto,v)}$ or, equivalently, if the $p - value$ is less than a. Rejection of H_0 indicates that at least one model parameter is statistically significant.

Truncated Spline Nonparametric Logistic Regression (TSNLR*)*

Given response variable (Y_i) is bernoulli distributed, and $v_{1i}, v_{2i}, \dots, v_{pi}$; $i = 1, 2, \dots$, n, are as many as p predictor variables. The probability distribution formulated as (Hosmer & Lemeshow, 2000):

$$P(Y_i = y_i) = \pi(\boldsymbol{v_i})^{y_i}\left(1 - \pi(\boldsymbol{v_i})\right)^{1-y_i} \quad (3)$$

When (3) is formulated in exponential form, it belongs to the exponential family of distributions. The logit function is defined as:

$$\ln\left(\frac{\pi(v_i)}{1-\pi(v_i)}\right) = f(v_{1i}, v_{2i}, \dots, v_{3i}) \quad (4)$$

Then a logit transformation is performed to obtain a logistic regression model, is expressed as:

$$\ln\left(\exp f(v_{1i}, v_{2i}, \dots, v_{3i})\right) = \ln\left(\frac{\pi(v_i)}{1-\pi(v_i)}\right) \qquad \pi(\boldsymbol{v_i}) = \frac{\exp f(v_{1i}, v_{2i}, \dots, v_{3i})}{1+\exp f(v_{1i}, v_{2i}, \dots, v_{3i})} \quad (5)$$

The function $f(v_{1i}, v_{2i}, \dots, v_{3i})$ in (5) is approximated by a truncated spline function with knot points $K_{1j}, K_{2j}, \dots, K_{3j}$, where j is 1, 2, …, p. Consequently, the TSNBLR model can be expressed as (Suriaslan, et al., 2025a):

$$\pi(\boldsymbol{v_i}) = \frac{\exp\left(\sum_{j=1}^{p}\sum_{k=0}^{m}\gamma_{jk}v_{ji}^{k} + \sum_{j=1}^{p}\sum_{u=1}^{r}\gamma_{j(m+u)}\left(v_{ji} - K_{ju}\right)_+^m\right)}{1+\exp\left(\sum_{j=1}^{p}\sum_{k=0}^{m}\gamma_{jk}v_{ji}^{k} + \sum_{j=1}^{p}\sum_{u=1}^{r}\gamma_{j(m+u)}\left(v_{ji} - K_{ju}\right)_+^m\right)} \quad (6)$$

where, γ_0, γ_{jk}, and $\gamma_{j(m+u)}, j = 1,2, \dots, p, k = 1,2, \dots, m, u = 1,2, \dots, r$ are the model parameters in the truncated spline function. Truncated function for spline is represented by the following equation:

$$\left(v_{ji} - K_{ju}\right)_+^m = \begin{cases} \left(v_{ji} - K_{ju}\right) & , v_{ji} \geq K_{ju} \\ 0 & , v_{ji} < K_{ju} \end{cases} \quad (7)$$

The simultaneous test evaluates the significance of the full set of model parameters γ in the model, with hypothesis:

$H_0: \gamma_0 = \gamma_{11} = \cdots = \gamma_{p1} = \cdots = \gamma_{p(m+r)} = 0$
H_1: at least one $\gamma_{jk} \neq 0$ or $\gamma_{j(m+u)} \neq 0, j = 1, 2, \dots, p,\ k = 1, 2, \dots, (m+u),\ u = 1, 2, \dots, r$

The test statistic used is the Likelihood Ratio Test (LRT), is formulated as (Suriaslan, et al., 2025b):

$$G^2 = 2[\ln l(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) - \ln l(\boldsymbol{\gamma}_\omega)] - 2[\ln l(\hat{\boldsymbol{\gamma}}_\omega) - \ln l(\boldsymbol{\gamma}_\omega)] \tag{8}$$

Equation (8) is approximated by a second-order Taylor series expansion around the MLE under the full model $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}$

$$\ln l(\boldsymbol{\gamma}_\omega) \approx \ln l(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) + g(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})(\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) - \frac{1}{2}(\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})](\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) \tag{9}$$

Because $g(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})$ is zero in (9), it can be expressed as follows:

$$2[\ln l(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) - \ln l(\boldsymbol{\gamma}_\omega)] \approx (\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega)^T - [\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})](\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega) \tag{10}$$

The ln-likelihood function $\ln l(\boldsymbol{\gamma}_\omega)$ expanding around $\hat{\boldsymbol{\gamma}}_\omega$ gives:

$$\ln l(\boldsymbol{\gamma}_\omega) \approx \ln l(\hat{\boldsymbol{\gamma}}_\omega) + g(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})(\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_\omega) - \frac{1}{2}(\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})](\boldsymbol{\gamma}_\omega - \hat{\boldsymbol{\gamma}}_\omega) \tag{11}$$

As $g(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})$ is zero in (11), it may be defined as follows:

$$2[\ln l(\hat{\boldsymbol{\gamma}}_\omega) - \ln l(\boldsymbol{\gamma}_\omega)] \approx (\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})] - (\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega) \tag{12}$$

Based on (10) and (11), the LRT statistic becomes:

$$G^2 \approx (\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\boldsymbol{\gamma}_{\boldsymbol{\Omega}})](\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega) - (\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})](\hat{\boldsymbol{\gamma}}_\omega - \boldsymbol{\gamma}_\omega) \tag{13}$$

Based on (Pawitan, 2001), the maximum likelihood estimator $\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}$ satisfies the asymptotic distribution:

$$(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_{\boldsymbol{\Omega}}) \xrightarrow{d} N(\mathbf{0}, [\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})]^{-1})$$

where,

$$\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}}) = \begin{bmatrix} [\mathrm{I}_{11}] & [\mathrm{I}_{12}] \\ [\mathrm{I}_{21}] & [\mathrm{I}_{22}] \end{bmatrix}$$

The quadratic form $(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})](\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega)$ in (13) can be decomposed into:

$$(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega)^T[\mathrm{I}(\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}})](\hat{\boldsymbol{\gamma}}_{\boldsymbol{\Omega}} - \boldsymbol{\gamma}_\omega) = \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix}^T \begin{bmatrix} [\mathrm{I}_{11}] & [\mathrm{I}_{12}] \\ [\mathrm{I}_{21}] & [\mathrm{I}_{22}] \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\gamma}}_1 \\ \hat{\boldsymbol{\gamma}}_2 - \boldsymbol{\gamma}_{0\omega} \end{bmatrix} \tag{14}$$

thus obtained:

$$G^2 \approx \hat{\boldsymbol{\gamma}}_1{}^T([\mathrm{I}_{11}] - [\mathrm{I}_{12}][\mathrm{I}_{22}]^{-1}[\mathrm{I}_{21}])\hat{\boldsymbol{\gamma}}_1 \tag{15}$$

Equation (15) is simplified to

$$G^2 \approx \hat{\boldsymbol{\gamma}}_1{}^T[\mathrm{I}^{11}]^{-1}\boldsymbol{\gamma}_1 \tag{16}$$

where,

$\hat{\boldsymbol{\gamma}}_1 \xrightarrow{d} N\left(\mathbf{0}, [\mathrm{I}^{11}]_{(p(m+r)\times p(m+r))}\right)$, and $[\mathrm{I}^{11}]^{-\frac{1}{2}}\hat{\boldsymbol{\gamma}}_1 \xrightarrow{d} N\left(\mathbf{0}, \mathrm{I}_{(p(m+r)\times p(m+r))}\right)$

Assuming the null hypothesis H_0 is true, the LRT statistic converges asymptotically to a chi-square distribution with $p(m + r)$ degrees of freedom.

$$G^2 \approx \left[[\mathrm{I}^{11}]^{-\frac{1}{2}}\hat{\boldsymbol{\gamma}}_1\right]^T \left[[\mathrm{I}^{11}]^{-\frac{1}{2}}\boldsymbol{\gamma}_1\right] \approx Z^T Z \xrightarrow{d} \chi^2_{p(m+r)}$$

Thus it can be concluded $G^2 \xrightarrow{d} \chi^2_{p(m+r)}$

For simultaneous testing of model parameters, the rejection region of the null hypothesis H_0 at significance level α is defined according to the distribution of the test statistic, expressed as:

$$G^2 > \chi^2_{(\alpha,p(m+r))}$$

Results and Discussion

The dataset is divided into training and testing sets with an 80:20 ratio, where the training data are used for model estimation and parameter inference, while the testing data are employed to evaluate predictive performance. Modeling was performed considering all combinations of predictor variables and knot points. The best model selection was determined based on the minimum Akaike Information Criterion (AIC) value. The TSNBLR results showed that the best model was obtained when all four predictor variables were included in the model with the optimal knot point combination, are $K_{1u} = 1$, $K_{2u} = 1$, $K_{3u} = 1$, and $K_{4u} = 4$, as presented in Table 2.

Table 2. TSNLR Number Of Knots

No.	Number of Knot	AIC
1	1, 1, 1, 1	427.67
2	2, 1, 1, 1	429.60
3	3, 1, 1, 1	429.83
4	4, 1, 1, 1	431.30
	⋮	
193	**1, 1, 1, 4**	**423.67**
194	2, 1, 1, 4	425.56
	⋮	
255	3, 4, 4, 4	435.20
256	4, 4, 4, 4	436.45

Based on the Table 2, it can be seen that the optimal knot point model are K_{11} = 90.27, K_{21} = 9.62, K_{31} = 4.02, K_{41} = 58.96, K_{42} = 62.81, K_{43} = 64.76, and K_{44} = 67,64.

Table 3. Parameter Estimation Of TSNLR

Parameter	Estimation	Parameter	Estimation
$\hat{\gamma}_0$	-7.869	$\hat{\gamma}_{32}$	-0.830
$\hat{\gamma}_{11}$	0.040	$\hat{\gamma}_{41}$	0.066
$\hat{\gamma}_{12}$	-0.017	$\hat{\gamma}_{42}$	-0.082
$\hat{\gamma}_{21}$	-0.197	$\hat{\gamma}_{43}$	1.113
$\hat{\gamma}_{22}$	-0.058	$\hat{\gamma}_{44}$	-1.618
$\hat{\gamma}_{31}$	0.832	$\hat{\gamma}_{45}$	0.744

Based on the results of the TSNLR parameter obtained in Table, the model for HDI can be written in the following equation:

$$\text{Logit } \pi(\boldsymbol{v_i})) = -7.869 + 0.040v_{1i} - 0.017(v_{1i} - 90.27) - 0.197v_{2i} - 0.058(v_{2i} - 9.62)_+ + 0.832v_{3i} - 0.830(v_{3i} - 4.02)_+ + 0.066v_{4i} - 0.082(v_{4i} - 58.96)_+ + 1.113(v_{4i} - 62.81)_+ - 1.618(v_{4i} - 64.76)_+ + 0.744(v_{4i} - 67{,}64)_+$$

Table 4. Simultaneous Test Of TSNLR

Test Statistic	Value
G^2	292.494
χ^2	19.675
p-value	3.337e-56

Based on the Simultaneous Hypothesis Test Results, the test statistic value G^2 is 292.494 with a p-value of 3.337×10^{-56}. Meanwhile, from the chi-square distribution table, the critical value $\chi^2_{(0{,}005;11)}$ = 19.675 was obtained. Because the G^2 value is greater than the $\chi^2_{(0{,}005;11)}$ value, or equivalently p-value < 0.05, the null hypothesis is rejected. Thus, it can be concluded that there is at least one predictor variable that has a significant effect on the HDI.

As material for model evaluation, at this stage a conventional binary logistic regression model is constructed.

Table 5. Parameter Estimation Of BLR

Parameter	Estimation
$\hat{\beta}_0$	-7.959
$\hat{\beta}_1$	0.037
$\hat{\beta}_2$	-0.228
$\hat{\beta}_3$	0.278
$\hat{\beta}_4$	0.106

Based on the results of the BLR parameter obtained in Table, the model for HDI can be written in the following equation:

$$\text{logit}(\pi(\boldsymbol{x}_i)) = -7.959 + 0.0375x_{1i} - 0.228x_{2i} + 0.278x_{3i} + 0.106x_{4i}$$

Table 6. Simultaneous Test Of BLR

Test Statistic	Value
G^2	271.489
χ^2	9.488
p-value	1.523e-57

Based on the simultaneous hypothesis test, the test statistic G^2 is 271.489with a p-value of 1.523×10^{-57}. Since G^2 exceeds the critical value$\chi^2_{(0,005;11)}$ = 9.488, or equivalently the p-value is less than 0.05, the null hypothesis is rejected. This indicates that at least one predictor variable has a significant effect on HDI.

Table 7. Comparison of Models

Test Statistic	TSNLR	BLR
Accuracy	82.9%	81.1%
Sensitivity	88%	87.7%
Specificity	75.2%	73.3%
Precision	84.1%	83.1%
F1	86.03%	85.3%
AUC	89.5%	88.4%

Based on the results of testing on the test data, TSNLR showed better classification performance than binary logistic regression. TSNLR produced an accuracy of 82.9%, higher than the logistic model at 81.1%. The sensitivity of TSNLR models is 88%, which higher than BLR and shows ability in identifying areas with high HDI categories. However, the specificity of TSNLBR is higher, at 75.2% compared to 73.3% for BLR, so TSNLBR is better at recognizing areas with low HDI. In addition, the precision and F1-score values of TSNLBR are 84.1% and 86.03%, respectively, which are higher than those of the BLR model. However, the AUC value of TSNLR is slightly higher, at 89.5% compared to 88.4% for TSNBLR, indicating that TSNLR has slightly better discrimination capabilities at various classification thresholds. Overall, TSNLR showed better classification performance than BLR on most evaluation metrics.

Conclusion

This study shows that Truncated Spline Nonparametric Logistic Regression (TSNLR) is an effective methodological approach in modeling the Human Development Index (HDI), especially when the relationship between predictor and response variables is change in certain subinterval. The selection of the optimal knot point combination based on the AIC criterion produces a statistically significant model that is capable of improving classification performance compared to conventional Binary Logistic Regression (BLR), particularly in terms of accuracy, specificity, precision, and F1-score. Although the difference in AUC values between the two models is relatively small, the overall evaluation results confirm that the TSNLR approach provides better modeling flexibility without sacrificing discrimination ability, thus contributing as a competitive and relevant methodological alternative in the analysis of binary response data with complex relationship patterns.

Recommendations

Building on the current findings, future research may extend the TSNLR by considering higher order spline functions, such as quadratic or cubic forms. In addition, incorporating spatial data settings could be explored to better capture regional interdependencies. Further extensions may also involve semiparametric modeling to allow flexible relationships between the variables.

Scientific Ethics Declaration

* The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM journal belongs to the authors.

Conflict of Interest

* The authors declare that they have no conflicts of interest

Funding

* The authors would like to express their gratitude to the Ministry of Education, Culture, Research, and Technology of the Republic of Indonesia for funding this research through the "Master Program of Education Leading to a Doctoral Degree for Excellent Graduates (PMDSU) Batch VII", under Contract Numbers 017/C3/DT.05.00/PL/2025 and 1177/PKS/ITS/2025.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Basic Sciences and Technology (www.icbast.net) held in Konya/Türkiye on April 02-05, 2026.

References

Agresti, A. (2002). *Categorical data analysis*. John Wiley & Sons

Alfardus, A., & Rawat, D. B. (2024). Binary logistic regression based intrusion detection system for CAN bus security. In *Proceedings of the 2024 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI 2024* (pp. 141–147).

Budiantara, N. I. (2005). Model of truncated polynomial spline family in semiparametric regression. *Berkala MIPA, 15*(3), 55-61.

Eubank, L. R. (1999). *Nonparametric regression and spline smoothing* (2nd ed.). CRS Press.

Gupta, C., Podkopaev, A., & Ramdas, A. (2020). Distribution-free binary classification: Prediction sets, confidence intervals and calibration. *Advances in Neural Information Processing Systems, 33*, 3711–3723.

Hazelton, M. L., & Turlach, B. A. (2011). Semiparametric regression with shape-constrained penalized splines. *Computational Statistics and Data Analysis, 55*(10), 2871–2879.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York, Chichester.

Joshi, R. D., & Dhakal, C. K. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. *International Journal of Environmental Research and Public Health, 18*(14).

Manoharan, H., Teekaraman, Y., Kirpichnikova, I., Kuppusamy, R., Nikolovski, S., & Baghaee, H. R. (2020). Smart grid monitoring by wireless sensors using binary logistic regression. *Energies, 13*(15).

Marasco, S., Marano, G. C., & Cimellaro, G. P. (2022). Evolutionary polynomial regression algorithm combined with robust Bayesian regression. *Advances in Engineering Software, 167*.

Pawitan, Y. (2001). *In all likelihood: Statistical modelling and inference using likelihood*. Clarendon Press.

Permatasari, E. O., Nasuha, F., & Prawirosastro, C. L. (2021). Modeling the level of open unemployment in Central Kava with multivariate adaptive regression splines (MARS) approach. *Journal Asro,* 12(01), 66-74.

Rusliyadi, M., Ardil, Y. W. Y., & Winarno, K. (2021). Binary logistics regression model to analyze factors influencing technology adoption. In *Proceedings of the International Symposium Southeast Asia Vegetable 2021 (SEAVEG 2021)*, 460.

Saifudin, T., Panjaitan, L. S., Falasifah, S., & Yan Dwi. (2024). Application of geographically weighted logistic regression in modeling the human development index in East Java. *Jurnal Aplikasi Statistika & Komputasi Statistik, 16*(1), 43–57.

Suliyanto, Rifada, M., & Tjahjono, E. (2020). Estimation of nonparametric binary logistic regression model with local likelihood logit estimation method: Case study of diabetes mellitus patients at Surabaya Hajj General Hospital. *AIP Conference Proceedings, 2264*.

Suriaslan, A. S., Budiantara, I. N., & Ratnasari, V. (2025). Nonparametric regression estimation using multivariable truncated splines for binary response data. *MethodsX, 14*.

Suriaslan, A. S., Budiantara, I. N., Ratnasari, V., Zulfadhli, M., & Dhewy, R. C. (2025). An empirical study on unmet need: A statistical inference framework for truncated spline nonparametric binary logistic regression (TSNLBR). *Journal of Cultural Analysis and Social Change, 10*(2), 868–877.

Yahya, A. (2022). Systematic review of regression algorithms for predictive analytics. *Qubahan Techno Journal, 1*(4), 1–14.

Yuniarsih, E. T., Salam, M., Jamil, M. H., & Tenriawaru, A. N. (2024). Determinants determining the adoption of technological innovation of urban farming: Employing binary logistic regression model in examining Rogers' framework. *Journal of Open Innovation: Technology, Market, and Complexity, 10*(2).

Zhang, Z., Shi, X., Xiang, X., Wang, C., Xiao, S., & Su, X. (2020). Bootstrap confidence intervals for the optimal cutoff point to bisect estimated probabilities from logistic regression. *Statistical Methods in Medical Research, 29*(6), 1514–1526.

Zulfadhli, M., Budiantara, I. N., Ratnasari, V., & Suriaslan, A. S. (2025). Semiparametric regression estimator of Fourier series for categorical data. *International Journal of Applied Mathematics, 55*(7), 2157–2164.

Author(s) Information

Afiqah Saffa Suriaslan
Sepuluh Nopember Institute of Technology
Kampus ITS- Sukolilo, Surabaya 60111, Indonesia
ORCID iD: https://orcid.org/0009-0001-8176-1527

I Nyoman Budiantara
Sepuluh Nopember Institute of Technology
Kampus ITS- Sukolilo, Surabaya 60111, Indonesia
ORCID iD: https://orcid.org/0000-0001-6572-4083
Contact e-mail: i_nyoman_b@statistika.its.ac.id*

Vita Ratnasari
Sepuluh Nopember Institute of Technology
Kampus ITS- Sukolilo, Surabaya 60111, Indonesia
ORCID iD: https://orcid.org/0000-0002-7579-4752

Dursun Aydin
Mugla Sıtkı Kocman University
48000 Mentese/Mugla, Turkiye
ORCID iD: https://orcid.org/0000-0001-8393-1270

*Corresponding author(s) contact email

To cite this article:

Suriaslan, A.S., Budiantara, I. N., Ratnasari, V., & Aydin, D. (2026). A truncated spline nonparametric logistic regression: Simultaneous testing and model comparison. *The Eurasia Proceedings of Science, Technology, Engineering and Mathematics (EPSTEM), 40*, 264-272.