BAREKENG: Journal of Mathematics and Its Applications
June 2026 Volume 20 Issue 2 Page 1627–1642
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss2pp1627-1642

MODELLING THE NUMBER OF CRIMES IN EAST JAVA USING A TRUNCATED SPLINE SEMIPARAMETRIC REGRESSION APPROACH

Yahya Vigo Tri Saputra [1], **Moh. Hafiyusholeh** [2*], **Hani Khaulasari** [3], **Yuniar Farida** [4], **Putroue Keumala Intan** [5]

[1,2,3,4,5] *Department of Mathematics, Faculty of Science and Technology, Universitas Islam Negeri Sunan Ampel Surabaya*
Jln. Ahmad Yani No.117, Surabaya, 60237, Indonesia

*Corresponding author's e-mail: * **hafiyusholeh@uinsa.ac.id***

Article Info

Article History:
Received: 7th July 2025
Revised: 9th August 2025
Accepted: 18th September 2025
Available Online: 26th January 2026

Keywords:
Crime rate;
Generalized Cross Validation;
Ramsey's Reset Test;
Semiparametric Regression;
Spline Truncated.

ABSTRACT

High crime rates can lead to unrest and financial losses for the community. East Java is one of the provinces with high crime rates, with a total of 21,046 reported crimes in 2023. This study aims to identify the factors that influence crime rates in East Java and evaluate the goodness of the model through truncated spline semiparametric regression. Truncated spline semiparametric regression is a combination of parametric and nonparametric methods that can adjust changes in data patterns through the presence of knot points. The data used in this study were sourced from the Central Statistics Agency, including variables such as the number of people living in poverty, average years of schooling, gross regional domestic product, population, Gini ratio, per capita expenditure, and open unemployment rate. The results of the analysis indicate that the predictor variables have a significant influence on the number of crimes simultaneously. Partially, the variables that influence the number of crimes in East Java Province are average years of schooling, population, Gini ratio, per capita expenditure, and open unemployment rate. The best regression model is obtained using the combination knot point (4,2,4,3) with a minimum GCV value of 49636.60. The coefficient of determination obtained is 93.60%, indicating that the predictor variables can explain 93.60% of the variation in the crime rate, while the remaining 6.40% is attributed to variables outside the scope of the study.

How to cite this article:

Y. V. T. Saputra, M. Hafiyusholeh, H. Khaulasari, Y. Farida and P. K. Intan., "MODELLING THE NUMBER OF CRIMES IN EAST JAVA USING A TRUNCATED SPLINE SEMIPARAMETRIC REGRESSION APPROACH", *BAREKENG: J. Math. & App.,* vol. 20, no. 2, pp. 1627-1642, Jun, 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

Criminal acts or crimes are actions committed by individuals that violate the law, disturb the peace, and harm others [1]. These acts can be committed by anyone, regardless of gender or age. Crime is a social problem that remains a serious concern and a threat to society. Crime in Indonesia remains prevalent, as evidenced by the increase in crime rates over the past two years, particularly in East Java. East Java is the easternmost province of Java Island, comprising 38 administrative units, including 29 regencies and 9 cities. Uneven economic growth combined with high population density can create social disparities that may trigger criminal behavior. In 2023, East Java Province was still classified as a province with a high crime rate, with 21,046 reported criminal cases [2]. External and internal factors, such as economic, environmental, educational, and religious factors, as well as an individual's intentions, can contribute to the rise in criminal rates in East Java [3].

The high crime rate in East Java is evident in several ongoing cases, such as the bicycle theft incident in 2022 in an elite residential area of Surabaya [4]. The next crime case was a murder that occurred in Gresik in 2023 [5]. In 2024, in Kediri City, there was a break-in at an Automatic Teller Machine (ATM) belonging to Bank Jatim and an assault on a boarding school student by his senior [6].

Various factors, including environmental, individual, and family factors, significantly contribute to the occurrence of criminal acts that harm community life. The varied impacts of criminal acts necessitate preventive measures to combat the growth of crime rates in East Java, which involves identifying the triggering factors. Varied studies have been conducted to distinguish the variables that cause crime, such as the gini coefficient [7]; average educational attainment and per capita spending, [8], [9]; population size, human development index, number of individuals below the poverty line, open joblessness rate, and gross regional domestic output [10], [11].

Based on the findings regarding the factors influencing crime, it is important to further explore these aspects through a more rigorous methodological approach. Additionally, the determination of aspects related to crime in East Java requires analysis using several scientific methods. These methods include panel data regression [12] and nonparametric regression [13]. However, each approach still has shortcomings, including the fact that panel data regression and nonparametric regression only consider one pattern of relationship between variables, which is linear or nonlinear. Given these issues, a method is needed that can capture two different relationship patterns, linear and nonlinear, as well as produce a flexible model, enabling the proposed solution to utilize semiparametric regression, specifically truncated spline regression [14], [15].

Semiparametric regression is an algorithm that combines parametric and nonparametric components and is implemented when only part of the regression pattern is recognized [16]. Research on nonparametric regression using truncated splines has shown better model performance compared to the Fourier series [17]. The advantage of truncated splines lies in their ability to handle data with varying patterns and adapt estimation based on the data's structure, as seen in the use of knot points that adjust the shape of the curve across different intervals. There are several types of spline functions commonly used in regression modeling, including truncated splines, B-splines, and P-splines, each with its own characteristics in terms of smoothness, flexibility, and computational efficiency. This allows researchers to select the most appropriate approach based on the nature of the data.

One of the challenges in semiparametric regression modeling is the unknown form of the relationship pattern. Model mismatch with the actual relationship pattern can lead to biased and inaccurate estimates. With this problem, the solution that can be used is the Ramsey RESET Test method [18]. This method can detect linear and nonlinear relationships between the variables being analyzed [19].

Research on truncated spline semiparametric regression has been applied in several fields, including modeling the Human Development Index (HDI) in East Java, which indicates the most suitable model using Generalized Cross Validation (GCV) is a minimum of 0.019 and a coefficient of determination (R^2) value of 99.96% [20]. Another study, conducted by researchers applying the method to modeling stunting in Indonesia, obtained a lowest value of GCV at 9.30, representing the optimal model choice with a value of 92.71% [21].

This study aims to identify and evaluate the key factors contributing to the high crime rate in East Java Province by applying truncated spline semiparametric regression. Unlike previous studies, such as [22] that employ parametric models and [23] using nonparametric approaches, this study explicitly addresses methodological gaps that have often been overlooked. Prior research generally did not test for the presence of nonlinear relationships using diagnostic tools such as the Ramsey RESET test, nor did it optimize the selection of knot points in spline-based models. By incorporating these steps, this study not only ensures that model specifications are more robust but also provides a more accurate and interpretable representation of crime determinants in East Java.

2. RESEARCH METHODS

2.1 Crime in East Java

The term criminality can be defined as crime, referring to actions that violate the law. Crimes can be committed by anyone, regardless of gender or age. Criminality refers to wrongful acts committed by individuals that lead to social problems and public unrest. Fig. 1 illustrates the number of crimes reported in East Java Province from 2019 to 2023.

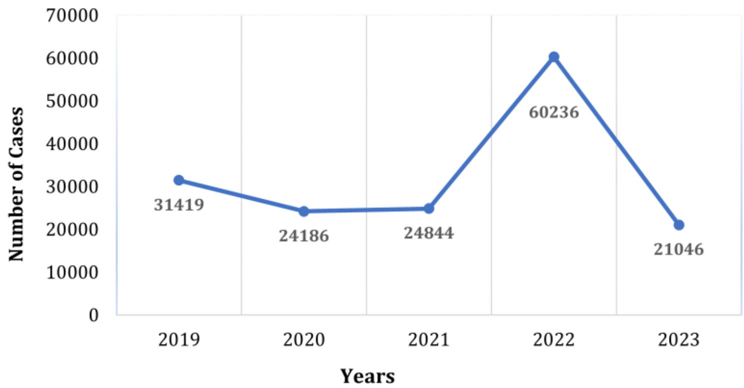


Figure 1. Number of Criminality Reported in 2019–2023
Data Sources: *(Central Bureau of Statistics, 2024)*

Fig. 1 shows that the number of crimes in East Java from 2019 to 2023 remained relatively high each year. The number of criminal cases appears to fluctuate, as the graph displays varying increases and decreases over the years. In 2023, a significant decline was observed, with a total of 21046 reported criminal cases. However, this decrease does not imply that criminal activities have completely ceased; rather, it offers an objective overview of the phenomenon, highlighting the importance of further investigation [24].

2.2 Research and Analysis Data

This research represents a quantitative study that utilizes secondary data sourced from the East Java, Indonesia's national statistical agency, in 2023. Table 1 presents the independent variables (x) and the dependent variable (y) used in this investigation, which involves thirty-eight districts and cities across one of Indonesia's provinces, East Java.

Table 1. Research Variables

Variable	Description	Scale
y	Number of Crimes (Cases)	Ratio
x_1	Number of Poor People (People)	Ratio
x_2	Average Length of Schooling (Years)	Ratio
x_3	Gross Regional Domestic Product (Billion Rupiah)	Ratio
x_4	Population (People)	Ratio
x_5	Gini Coefficient	Ratio
x_6	Per Capita Expenditure (Thousand Rupiah)	Ratio
x_7	Open Unemployment Rate (Percent)	Ratio

Data Sources: *(Central Bureau of Statistics, 2024)*

Modeling of the count of crimes in East Java in 2023 using truncated spline semiparametric regression can be explained in the following steps:

1. Descriptive data analysis to obtain an overview of the data.
2. Identify parametric and nonparametric components by examining scatter plot patterns, followed by a linearity test using Ramsey's RESET Test. After that, determine the types of parametric and nonparametric patterns in each variable in the data.
3. Determine the model's goodness of fit by calculating the R^2.
4. Selecting the most suitable knot point using Generalized Cross Validation (GCV).
5. Creating a model of crime rates in East Java using semiparametric regression with truncated spline estimation.
6. Performing simultaneous and partial hypothesis testing.
7. Interpreting the regression model that has been obtained.

2.3 Truncated Spline Semiparametric Regression

Regression is a statistical technique for modeling the connection between predictor variable (x) and the outcome variable (y). Several forms of regression analysis exist, namely parametric, nonparametric, and semiparametric regression. Semiparametric regression is a process that combines parametric and nonparametric components [25]. Suppose we are given a paired data set (x_i, y_i, z_i) or a relationship between (x_i, z_i) and y_i, then the following semiparametric regression model can be formed [26].

$$y_i = f(x_i) + g(z_i) + \varepsilon_i \quad ; i = 1,2, \dots, n. \tag{1}$$

The variable y_i serves as the dependent variable, where $f(x_i)$ captures the parametric component and $g(z_i)$ accounts for the nonparametric element. The residual term ε_i, associated with the i −th data point, σ^2 is normally distributed characterized by a zero expected value and constant spread.

The application of parametric regression is appropriate when the structure of the regression model is predefined, allowing analysis of the relationship between x and y. The parametric component function $f(x_i)$, with s representing the number of parametric components, is presented in Eq. (2) [27].

$$\begin{aligned} f(x_i) &= \beta_0 + \beta_1 x_{i1} + \beta_1 x_{i2} + \dots + \beta_s x_{is} + \varepsilon_i \quad ; i = 1,2, \dots, n, \\ &= \beta_0 + \sum_{r=1}^{s} \beta_r x_{ir} + \varepsilon_i. \end{aligned} \tag{2}$$

Nonparametric regression is utilized as a flexible technique to explore the association between cause (independent) and effect (dependent) variables in situations where the underlying data structure is not specified [28]. One approach to a common method in nonparametric regression is the use of truncated splines. A truncated spline is defined as a piecewise polynomial function with flexible characteristics, marked by the presence of knot points [29]. The form of the nonparametric regression model is presented below [30].

$$\begin{aligned} g(z_i) &= \varphi_0 + \varphi_1 z_{i1} + \dots + \varphi_{jq} z_{ij}^{m-1} + \varphi_1 (z_{i1} - k_l)^{m-1} + \varphi_{j(m-1)+l} \left(z_{ip} - k_K\right)^{m-1} + \varepsilon_i \\ &= \sum_{j=1}^{p} \left[\sum_{q=1}^{m-1} \varphi_{jq} z_{ij}^{q} + \sum_{l=1}^{K} \varphi_{j(m-1)+l} (z_{ij} - k_l)^{m-1} + \varepsilon_i \right] \quad ; i = 1,2, \dots, n, \end{aligned} \tag{3}$$

where,

$$(z_{ij} - k_l)^{m-1} = \begin{cases} (z_{ij} - k_l)^{m-1} & ; z_{ij} \geq k_l \\ 0 & ; z_{ij} < k_l \end{cases}. \tag{4}$$

The number of nonparametric predictor variables dependent on knot points as p, K which refers to the total count of knot points, estimated variables within the model framework can be written as φ_{jq}, $(z_{ij} - k_l)^{m-1}$ defined as a truncated function, k_l is the l knot point, $l = 1,2, \dots, K$.

Based on Eqs. (2) and (3), the following expression defines the structure of the semiparametric regression model [31].

$$\begin{aligned} y_i &= f(x_i) + g(z_i) + \varepsilon_i \\ &= \beta_0 + \sum_{r=1}^{s} \beta_r x_{ir} + \sum_{j=1}^{p} \left[\sum_{q=1}^{m-1} \varphi_{jq} z_{ij}^{q} + \sum_{l=1}^{K} \varphi_{j(m-1)+l} \left(z_{ij} - k_l \right)^{m-1} + \varepsilon_i \right] \\ &; \ i = 1,2,\dots,n \\ &; \ r = 1,2,\dots,s \\ &; \ q = 1,2,\dots,m-1 \\ &; \ j = 1,2,\dots,p. \end{aligned} \tag{5}$$

The semiparametric regression model employs Ordinary Least Squares (OLS) as the method for estimating its parameters [32]. The following matrix representation also describes the semiparametric regression model.

$$\boldsymbol{y} = \boldsymbol{M\delta} + \boldsymbol{\varepsilon}, \tag{6}$$

$\boldsymbol{y}$ is the dependent variable with size $\boldsymbol{n} \times \boldsymbol{1}$. Matrix $\boldsymbol{M}$ is a semiparametric regression matrix containing parametric variables with size $n \times (s+1)$ and nonparametric components dependent on knots with size $n \times p(K + (m-1))$. The parameter vector is denoted as $\boldsymbol{\delta}$, containing parameters β and φ of size $((s+1) + p(K + (m-1)) \times 1)$, and $\boldsymbol{\varepsilon}$ denotes an error vector with dimensions $n \times 1$. Therefore, Eq. (7) is obtained regarding the result of the parameter estimation formula derivation using the OLS method.

$$\hat{\boldsymbol{\delta}} = [\boldsymbol{M}'\boldsymbol{M}]^{-1}\boldsymbol{M}'\boldsymbol{y}. \tag{7}$$

2.4 Ramsey's RESET Test

RESET (Regression Specification Error Test) serves as a method to identify the linearity or nonlinearity of data developed by Ramsey in 1969 [33]. This method can also help identify the relationship between parametric variables or those requiring a nonparametric approach, thereby enabling more accurate model selection with the following formula [34].

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 \hat{y}_i^2 + \beta_3 \hat{y}_i^3, \tag{8}$$

where f is the number of predictor variables, as in Eq. (9) below,

$$\begin{aligned} \hat{y}_i &= \hat{\beta}_0 + \sum_{f=1}^{g} \hat{\beta}_f x_f, \\ \hat{y}_i^2 &= \hat{\beta}_0 + \sum_{f=1}^{g} \hat{\beta}_f x_f^2, \\ \hat{y}_i^3 &= \hat{\beta}_0 + \sum_{f=1}^{g} \hat{\beta}_f x_f^3. \end{aligned} \tag{9}$$

Based on Eq. (9), the next step is to construct a new regression, denoted by $\hat{y}_i^*$, as shown in Eq. (10) below.

$$\hat{y}_i^* = \hat{\beta}_0^* + \sum_{f=1}^{g} \hat{\beta}_f^* x_f + \hat{\beta}_{g+1} \hat{y}_i^2 + \hat{\beta}_{g+2} \hat{y}_i^3. \tag{10}$$

The hypothesis formulated as a result of this test is as follows.

H_0 : The model obtained is linear.

H_1 : The model obtained is nonlinear.

After that, the value of $F_{statistic}$ is calculated as in Eq. (11) with the critical region for rejecting H_0, if $F_{statistic} > F_{\alpha;(d,n-k)}$ or p-value $< \alpha$ (0.05).

$$F_{statistic} = \frac{(R_{New}^2 - R_{Old}^2)/d}{(1 - R_{New}^2)/(n-k)}, \tag{11}$$

where,

$$R_{Old}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}\,)^2},$$

$$R_{New}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i^* - \hat{y}_i^*)^2}{\sum_{i=1}^{n}\left(y_i^* - \bar{y}^*\,\right)^2}.$$

R_{Old}^2 as the initial regression equation value containing the predictor and response variables without additional new variables. R_{New}^2 as the value of the new regression equation, meaning that the regression model has been expanded to include additional variables ($\hat{y}_i^2$ and $\hat{y}_i^3$) as well as independent variables. The dependent variable observed in the initial regression is symbolized by y_i, the predicted value in the initial regression is $\hat{y}_i$, and $\bar{y}$ is the median of the initial regression. The collected observation values associated with the response variable during the new regression are denoted as y_i^*, the predicted values during the new regression are $\hat{y}_i^*$, and $\bar{y}^*$ is the mean of the new regression. The number of additional variables is denoted as d, while k indicates the number of parameters in the updated model, and n is the number of data points.

2.5 Generalized Cross Validation (GCV)

In the context of spline regression techniques, determining the right knot points really affects how good the model is. One way to do this is with the GCV method [35]. GCV is measured by taking the smallest value, like in Eq. (12).

$$\boldsymbol{GCV} = \frac{\boldsymbol{MSE(k)}}{(\boldsymbol{n^{-1}trace[I - A(k)]})^{\boldsymbol{2}}} \tag{12}$$

where,

$$\boldsymbol{A} = [\boldsymbol{M}][\boldsymbol{M'M}]^{-\boldsymbol{1}}[\boldsymbol{M'}]$$

$$\boldsymbol{MSE(k)} = \boldsymbol{n^{-1}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}}(\boldsymbol{y_i} - \boldsymbol{\hat{y}_i})^{\boldsymbol{2}}\ ;\ \boldsymbol{i = 1, 2, \dots, n}$$

where $\boldsymbol{I}$ is an $\boldsymbol{n \times n}$ identity matrix, $\boldsymbol{k = (k_1, k_2, \dots, k_l)}$ is the number of knots, and $\boldsymbol{A}$ is a matrix containing the semiparametric regression matrix ($\boldsymbol{M}$).

2.6 Parameters Significance Testing

2.6.1 Simultaneously

The F-test is used to assess the statistical significance of the results for the entire regression model. The hypothesis statement is outlined below.

$H_0 : \beta_0 = \beta_1 = \cdots = \beta_s = \varphi_1 = \varphi_2 = \cdots = \varphi_{p+K} = 0$ (the independent variables had no meaningful impact on the dependent variable overall).

H_1 : There is at least one $\beta_r \neq 0$ or $\varphi_j \neq 0$; $r = 1,2,\dots,s$; $j = 1,2,\dots,p+K$ (a minimum of one explanatory variable that shows an overall impact on the dependent variable).

The basis for rejecting H_0, when $F_{statistic} > F_{\alpha(s+p+K,(n-(s+p+K)-1))}$ or p-value $< \alpha$ (0.05) with this test statistic, can be calculated using Eq. (13) below [36].

$$F_{statistic} = \frac{MSR}{MSE}, \tag{13}$$

where,

$$MSR = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{s + p + K},$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n - (s + p + k) - 1}.$$

MSR stands for Mean Squared Regression, MSE stands for Mean Squared Error, s stands for the number of parametric components, p stands for the number of nonparametric variables, K stands for how many knots are included, and n stands for the number of data points.

2.6.2 Partial

The T-test is applied to determine the statistical relevance of each predictor variable's effect on the response variable. The hypotheses used for both the parametric (β_s) and nonparametric (φ_p) tests can be formulated as follows.

$H_0 : \beta_r = 0$ or $\varphi_j = 0$ (each predictor variable does not affect the response variable)

$H_1 : \beta_r \neq 0$; $r = 1,2,\dots,s$ or $\varphi_j \neq 0$; $j = 1,2,\dots,p+K$ (each of the independent variables has a significant influence on the dependent variable)

The critical region for rejecting H_0 is when $|t_{statistic}| > t_{\frac{\alpha}{2};(n-(s+p+K)-1)}$ or when the p-value $< \alpha$ (0.05). The statistical test calculation is performed as in Eq. (14) [37].

$$t_{statistic} = \frac{\hat{\beta}_s}{SE(\hat{\beta}_s)} \tag{14}$$

$$t_{statistic} = \frac{\hat{\varphi}_p}{SE(\hat{\varphi}_p)}$$

$\hat{\beta}_s$ is denoted as a parametric estimator, $\hat{\varphi}_p$ is a nonparametric estimator, SE is the standard error that can be approximated by the value of var.

2.7 The Goodness of the Model

Throughout this research, the statistical indicator R^2 served as the method to measure the extent to which the model fits the data, reflecting the extent to which the predictor variables explain the variation in the response variable. A higher R^2 value (closer to 1) reflects more effective and accurate the formulated approach is in interpreting the variation in the data [38]. In addition to R^2, the Root Mean Square Error (RMSE) was also used as a complementary measure, since RMSE evaluates the average of prediction errors in the same unit as the response variable, where a smaller RMSE value indicates better predictive performance [39]. The formula used for this test is presented below.

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{15}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

The symbol y_i refers to the actual response value, $\hat{y}_i$ denotes the predicted value, $\bar{y}$ stands for the average of all response values, n that is data observation, SSE (Sum of Squared Errors) represents the unexplained variation the model, while SST (Total Sum of Squares) denotes the total variation of the data around its mean.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistical Analysis

A general overview of the studied variables can be obtained through descriptive statistics, which present the data in a more accessible manner. Information regarding the characteristics of crime data and the factors suspected to influence it is presented in Table 2. This table compares each variable based on its average, standard deviation (St.Dev), minimum, and maximum values.

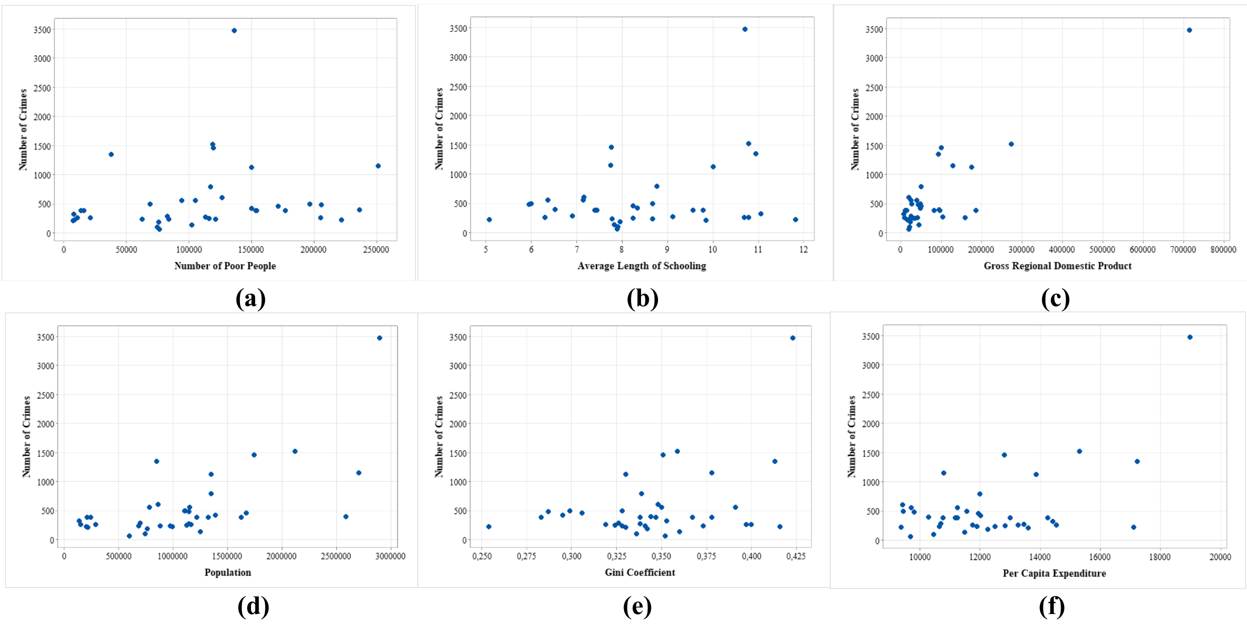
Table 2. Descriptive Statistics Results

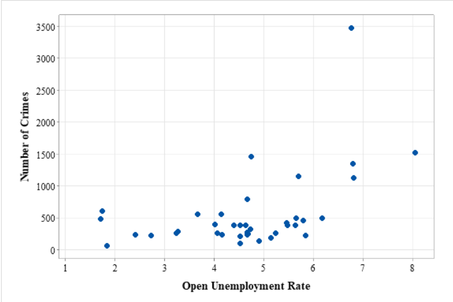
Variable	Mean	St.Dev	Minimum	Maximum
y	553.8	610.8	64	3473
x_1	110233	68319	7100	251360
x_2	8.376	1.658	5.070	11.820
x_3	77768	121046	8039	715295
x_4	1089905	678616	135414	2893698
x_5	0.3458	0.0374	0.2540	0.4230
x_6	12287	2263	9369	18977
x_7	4.663	1.429	1.710	8.050

Based on Table 2, the crime rate variable obtained an average of 553.8 cases with a standard deviation of 610.8 cases. This shows high variation between regions, with the highest number of cases occurring in the city of Surabaya, at 3,473 cases, and the lowest in Pacitan Regency, at 64 cases. This gap can be attributed to differences in population density, urban characteristics, and socioeconomic conditions between the two regions. The number of poor people is an average of 110233 people with a standard deviation of 68319 people, ranging from 7100 people in Batu City to 251360 people in Malang Regency. The average length of schooling is 8 years with a standard deviation of 2 years, indicating educational disparities between regions, such as Madiun City at 12 years and Sampang at 5 years. Furthermore, the GDP variable has an average of 77768 billion rupiahs with a high standard deviation of 121406 billion, ranging from a minimum of 8039 billion in Mojokerto City to a maximum of 715295 billion in Surabaya City. The population varies from 135414 people in Mojokerto City to 2893698 people in Surabaya City, with an average of 1089905 people and a standard deviation of 678616 people, indicating population diversity between regions. The Gini coefficient has an average of 0.3458 with a standard deviation of 0.0374, indicating economic inequality differences between regions, ranging from 0.2540 in Sampang to 0.4230 in Surabaya City. The average per capita expenditure was 12287 thousand rupiahs, with a standard deviation of 2263 thousand rupiahs, ranging from 9363 thousand rupiahs in Sampang to 18977 thousand rupiahs in Surabaya City, reflecting minor differences in consumption patterns. The open unemployment rate has a mean of 4.663% with a standard deviation of 1.429%, indicating a relatively even distribution, ranging from 1.710% in Sumenep to 8.050% in Sidoarjo.

3.2 Identification of Data Relationship Patterns

Scatter plots are used to identify correlation patterns between each independent and dependent variable, with the aim of distinguishing between parametric and nonparametric elements. The results of scatter plot identification for each data set are shown in Fig. 2.





(g)

Figure 2. Scatter Plot of Each Predictor Variable with Response Variable, (a) Scatter Plot of y vs x_1, (b) Scatter Plot of y vs x_2, (c) Scatter Plot of y vs x_3, (d) Scatter Plot of y vs x_4, (e) Scatter Plot of y vs x_5, (f) Scatter Plot of y vs x_6, (g) Scatter Plot of y vs x_7
*(**Source**: Data Processing Result from Minitab)*

Based on the image above, it is difficult to identify a clear relationship pattern, as each variable does not exhibit a distinct or definitive relationship. This makes it challenging to determine the type of relationship that occurs. Therefore, an additional method is needed to check and confirm the form of the relationship, namely, using Ramsey's Reset Test. Table 3 below provides a summary of the test findings.

Table 3. Ramsey's Reset Test Results

Variable	$F_{statistic}$	*P-Value*
x_1	0.3982	0.6746
x_2	0.3302	0.721
x_3	0.3037	0.74
x_4	4.5584	0.0176
x_5	4.2136	0.0231
x_6	13.316	5.362×10^{-5}
x_7	4.2608	0.0223

The test results are said to reject H_0 if $F_{statistic} > F_{0.05(2,34)}$ or *p-value* < 0.05 with a value of $F_{0.05(2,34)} = 3.28$. Based on Table 4.10, a decision to reject H_0 is obtained for variables x_4, x_5, x_6, and x_7, while the decision is to fail to reject H_0 for variables x_1, x_2, and x_3. Therefore, it can be concluded that there are three variables with parametric components and four variables containing nonparametric components, as shown in Table 4.

Table 4. Parametric and Nonparametric Component Results

Variable	Description	Component
x_1	Number of Poor People	Parametric
x_2	Average Length of Schooling	Parametric
x_3	Gross Regional Domestic Product	Parametric
x_4	Population	Nonparametric
x_5	Gini Coefficient	Nonparametric
x_6	Per Capita Expenditure	Nonparametric
x_7	Open Unemployment Rate	Nonparametric

3.3 Evaluation of Model Goodness

Model accuracy in semiparametric regression is measured using the coefficient of determination (R^2). The calculation is based on Eq. (16) shown below.

$$R^2 = 1 - \frac{SSE}{SST}. \tag{16}$$

$$= 1 - \frac{883008.9}{13805227}$$

$$= 0.9360 \approx 93.60\%.$$

Based on these calculations, an R^2 value of 93.60% was obtained, which means that the number of poor people, average length of schooling, gross regional domestic product, population, Gini ratio, per capita expenditure, and open unemployment rate can explain 93.60% of the crime rate in East Java. Meanwhile, the remaining 6.40% is explained by other variables not included in this study.

Table 5. Comparison of Spline Function Estimation Results

Function Estimation	R^2	RMSE
Spline Truncated	93.60%	1.53
B-Spline	91.25%	3.75

The comparison results in Table 5 indicate that the truncated spline function achieved a higher R^2 value compared to the B-spline, reflecting better model performance. This outcome was obtained through a consistent modeling procedure applied to both methods using the same dataset, ensuring a fair and objective comparison. Therefore, the truncated spline produced a lower RMSE value. These findings consistently demonstrate that the truncated spline approach not only provides a more accurate fit but also offers greater flexibility in capturing the variation in crime rates in East Java.

3.4 Selecting the Best Regression Model

A good regression model will certainly produce the smallest error value. This modeling uses knot trials ($k = 1,2,3,4$) assisted by looking at the minimum GCV value. The table below presents a comparison of the values of each node trial.

Table 6. Comparison of GCV Values

Model	GCV
One knot point	77916.06
Two knot point	77654.35
Three knot point	52010.49
Four knot point	49644.34
Combination of Knots (4,2,4,3)	49636.60

Referring to Table 6, the semiparametric truncated spline regression model with four knot points produced the lowest GCV value of 49,636.60 from the knot point combination (4,2,4,3). Table 7 then displays the parameter estimation results based on this optimal combination.

Table 7. Comparison of GCV Values

Variable	Parameters	Estimate
-	β_0	-0.0351
x_1	β_1	-0.00009
x_2	β_2	-0.3265
x_3	β_3	0.0018
z_1	φ_1	0.0002
	φ_{11}	-0.0022
	φ_{12}	0.0124
	φ_{13}	-0.0194
	φ_{14}	0.0151
z_2	φ_2	-0.0113
	φ_{21}	0.0002
	φ_{22}	-0.0004
z_3	φ_3	0.0021
	φ_{31}	0.3853
	φ_{32}	-5.1713
	φ_{33}	12.3139
	φ_{34}	-13.2114
z_4	φ_4	-0.2658
	φ_{41}	0.0255
	φ_{42}	0.0173
	φ_{43}	-0.0625

Next, a truncated spline semiparametric regression model can be found by adding the estimated values and optimal knot points as follows.

$$\begin{aligned}\hat{y}_i = & -0.0351 - 0.00009x_{i1} - 0.3625x_{i2} + 0.0018x_{i3} + 0.0002z_{i1} - 0.0022(z_{i1} - 1361318) \\ & + 0.0124(z_{i1} - 1616715) - 0.0194(z_{i1} - 2076429) + 0.0151(z_{i1} - 2587222) - 0.0113z_{i2} \\ & +0.0002(z_{i2} - 0.2946) - 0.0004(z_{i2} - 0.4104) + 0.0021z_{i3} + 0.3853(z_{i3} - 13635.89) \\ & -5.1713(z_{i3} - 14526.07) + 12.3139(z_{i3} - 16128.41) - 13.2114(z_{i3} - 17908.78) \\ & -0.2658z_{i4} + 0.0255(z_{i4} - 5.2322) + 0.0173(z_{i4} - 6.1714) - 0.0625(z_{i4} - 6.9933)\end{aligned}$$

3.5 Significance Testing of Model Parameters

Parameter testing is divided into two stages: simultaneous testing and partial testing. Both tests share the same objective: to determine whether the predictor variables have an impact on the response variables. The results of simultaneous testing can be explained using the following Analysis of Variance (ANOVA) table.

Table 8. Results of the ANOVA test

Source	Degree of Freedom	Sum of Square (SS)	Mean Squared (MS)	$F_{statistic}$	*P-Value*
Regression	20	12922218.1	6464110.9050	12.4392	$\mathbf{1.4453 \times 10^{-6}}$
Error	17	883008.9	51941.70		
Total	37	13805227			

According to the data in Table 8, obtained that $F_{statistic}$ and p-value are 12.4392 and 1.4453×10^{-6}. respectively, with $F_{0.05(20,17)}$ value of 2.16. Therefore, the decision is to reject H_0, since $F_{statistic} > F_{0.05(20,17)}$ or *p-value* < 0.05, indicates that there is at least one predictor that has a statistically significant effect on the model. The next step is to conduct partial testing of each variable, as presented in Table 9, with the following test results.

Table 9. Partial Test Results

Variable	Parameters	$t_{Statistic}$	*P-Value*	Decision
-	β_0	−4.2808	0.0001	Reject H_0
x_1	β_1	−0.0814	0.9355	Failed to Reject H_0
x_2	β_2	−4.2876	0.0001	Reject H_0
x_3	β_3	1.8865	0.0670	Failed to Reject H_0
z_1	φ_1	1.3929	0.1719	Failed to Reject H_0
	φ_{11}	−2.5866	0.0137	Reject H_0
	φ_{12}	4.2328	0.0001	Reject H_0
	φ_{13}	−4.6328	4.359×10^{-6}	Reject H_0
	φ_{14}	4.1477	0.0001	Reject H_0
z_2	φ_2	−4.2919	0.0001	Reject H_0
	φ_{21}	2.4650	0.0184	Reject H_0
	φ_{22}	−4.2747	0.0001	Reject H_0
z_3	φ_3	0.2110	0.8340	Failed to Reject H_0
	φ_{31}	1.6723	0.1028	Failed to Reject H_0
	φ_{32}	−4.2943	0.0001	Reject H_0
	φ_{33}	4.2730	0.0001	Reject H_0
	φ_{34}	−4.1823	0.0001	Reject H_0
z_4	φ_4	−4.2824	0.0001	Reject H_0
	φ_{41}	4.1663	0.0001	Reject H_0
	φ_{42}	4.2520	0.0001	Reject H_0
	φ_{43}	−4.2818	0.0001	Reject H_0

Referring to the data in Table 9, it can be concluded that the test results indicate a total of twenty parameters in the model, with five of them not significantly influential or failing to reject H_0, namely β_1, β_3, φ_1, φ_3, and φ_{31}. These results suggest that a total of five variables have a significant influence: average length of schooling, population size, Gini coefficient, per capita expenditure, and open unemployment rate.

Although five parameters are statistically insignificant, they remain relevant to retain in the model. Their influence may be indirect and, if excluded, could introduce omitted variable bias. Thus, their insignificance reflects weaker effects rather than irrelevance in explaining crime dynamics.

3.6 Model Interpretation

After learning about the truncated spline semiparametric regression model, subsequently, we interpret each dependent variables that show a statistically significant impact on the independent variable, as detailed below.

1. The effect of the average length of schooling on crime rates, assuming constant predictor variables.

$$\hat{y}_i = -0.0351 - 0.3625x_2.$$

The estimation results show that an additional year in the average length of schooling can reduce crime rates by 0.3625 cases. This reflects that higher educational attainment provides skills, knowledge, and better job opportunities, thereby lowering the likelihood of criminal involvement. In practice, regions with higher levels of education tend to exhibit stronger legal awareness and social control, resulting in lower crime rates.

2. The effect of population size on crime rates, assuming constant predictor variables.

$$\hat{y}_i = \begin{cases} 0.0002z_1 & ; z_1 < 1361318 \\ 2994.8996 - 0.002z_1 & ; 1361318 \leq z_1 < 1616715 \\ -17052.3664 + 0.010z_1 & ; 1616715 \leq z_1 < 2076429 \\ 23230.3562 - 0.009z_1 & ; 2076429 \leq z_1 < 2587222 \\ -15836.696 + 0.0061z_1 & ; z_1 \geq 2587222 \end{cases}.$$

The interpretation model results indicate that when the population is below 1361318 people, each additional person increases crime by 0.0002 cases. However, during the interval from 1361318 to 1616715, a 0.002-case reduction in crime is observed with an increase in population. This suggests that in areas with relatively small to medium populations, population growth can be absorbed through sufficient employment, social facilities, and effective social control, thereby preventing a direct rise in crime. In contrast, when the population ranges between 1616715 and 2076429 people, each additional person increases crime by 0.0104 cases, reflecting the effect of population density that creates economic competition, limited resources, and social pressure. Yet, in the interval of 2076429 to 2587222, population growth reduces crime by 0.009 cases, which may indicate the role of community adaptation and stronger supervision. Finally, when the population exceeds 2587222 people, each additional person increases crime by 0.0061 cases, consistent with densely populated areas where issues such as unemployment, inequality, and weak social control tend to heighten crime risks.

3. The effect of the Gini ratio on crime rates, assuming constant predictor variables.

$$\hat{y}_i = \begin{cases} -0.0113z_2 & ; z_2 < 0.2946 \\ -0.00005 - 0.0111z_2 & ; 0.2946 \leq z_2 < 0.4104 \\ 0.0001 - 0.0115z_2 & ; z_2 \geq 0.4104 \end{cases}.$$

The model indicates that when the Gini ratio is less than 0.2946 and increases by one unit, the number of crimes decreases by 0.0113 cases. This implies that in regions with relatively low levels of income inequality, incremental increases in inequality do not directly trigger crime but instead correspond to a slight decline. Furthermore, when the Gini ratio lies within the interval 0.2946 to 0.4104, a one-unit increase still reduces crime by 0.0111 cases. This condition suggests that in areas with moderate inequality, rising inequality may coincide with other social dynamics, such as strengthened community solidarity or broader economic opportunities that help maintain crime at a manageable level. Meanwhile, when the Gini ratio is greater than or equal to 0.4104, each one-unit increase leads to a decrease of 0.0115 cases. This reflects the reality that in regions with relatively high inequality, the presence of stronger law enforcement, governmental interventions, or community adaptation mechanisms plays a role in ensuring that rising inequality does not escalate crime, but rather accompanies a downward trend in criminal incidents.

4. The effect of per capita expenditure on crime rates, assuming constant predictor variables.

$$\hat{y}_i = \begin{cases} 0.0021z_3 & ; z_3 < 13635.89 \\ -5253.9084 + 0.3874z_3 & ; 13635.89 \leq z_3 < 14526.07 \\ 69864.7572 - 4.7839z_3 & ; 14526.07 \leq z_3 < 16128.41 \\ 268468.3842 - 17.0978z_3 & ; 16128.41 \leq z_3 < 17908.78 \\ 505068.4403 - 30.3092z_3 & ; z_3 \geq 17908.78 \end{cases}. \quad (21)$$

The estimation results indicate that when per capita expenditure is less than 13635.89 rupiah, a 1000-rupiah increase is associated with a rise in crime of approximately 0.0021 cases. This suggests that in low-income groups, additional expenditure is insufficient to substantially improve welfare and therefore does not contribute to reducing crime. Furthermore, when per capita expenditure lies within the range of 13635.89 to 14526.07 rupiah, a 1000 rupiah increase is estimated to raise crime by 0.3874 cases. This reflects persistent social inequality, where income growth is not evenly distributed, thus creating tensions that may stimulate criminal activity. In addition, when per capita expenditure falls within the range 14526.07 to 16128.41 rupiah, a 1000 rupiah increase is associated with a reduction in crime of about 4.7839 cases. This finding is consistent with the notion that as basic needs are increasingly met, economic pressures decline, resulting in lower crime levels. The downward trend becomes more pronounced in the range 16128.41 until 17908.78 rupiah, where an additional 1000 rupiah is linked to a decrease of approximately 17.0978 cases. If per capita expenditure exceeds 17908.78 rupiah, crime is estimated to decline by 30.3092 cases. Although this figure appears relatively large in statistical terms, substantively it can be understood as reflecting the conditions of high-income groups, where greater welfare and stability contribute to lower crime rates. Overall, the results indicate a general pattern in which higher levels of per capita expenditure, particularly in the middle to upper ranges, are associated with reductions in crime, while the effect remains limited at the lower levels.

5. The effect of the open unemployment rate (OUR) on the number of crimes, assuming constant predictor variables.

$$\hat{y}_i = \begin{cases} -0.2658z_4 & ; z_4 < 5.2322 \\ -0.1334 - 0.2403z_4 & ; 5.2322 \leq z_4 < 6.1714 \\ -0.2401 - 0.223z_4 & ; 6.1714 \leq z_4 < 6.9933 \\ 0.1968 - 0.2855z_4 & ; z_4 \geq 6.9933 \end{cases}. \quad (22)$$

When the open unemployment rate is below 5.2322% and increases by 1%, crime decreases by 0.2658 cases. This suggests that in regions with relatively low unemployment, employment opportunities remain sufficiently available, so a modest increase in unemployment does not directly lead to higher crime rates. Many individuals can still be absorbed in both formal and informal sectors. Similarly, when unemployment ranges from 5.2322% to 6.1714%, a 1% increase continues to reduce crime by 0.2403 cases. This suggests that although unemployment rises, social mechanisms such as family support or alternative productive activities reduce the likelihood of engaging in criminal behavior. Furthermore, when the unemployment rate ranges between 6.1714% and 6.9933%, a 1% increase still reduces crime by 0.223 cases. This reflects that under moderate unemployment levels, communities adapt by creating opportunities in informal trade or small-scale services, limiting the impact on crime rates. Finally, when unemployment exceeds 6.9933%, a 1% increase decreases crime by 0.2855 cases. This shows that in areas with relatively high unemployment, social solidarity and survival strategies through non-criminal means, such as micro-businesses or temporary jobs, serve as a barrier to crime. Thus, the link between unemployment and crime is shaped more by socio-economic dynamics than by unemployment figures alone.

These findings are align with previous studies that have identified per capita expenditure and open unemployment rate as important determinants of crime. For example, a study by [40] using spatial regression analysis found that regions with lower per capita expenditure and higher unemployment tend to exhibit higher crime rates, with an R^2 value of 85.18%, indicating a strong explanatory power of the model. In addition, another study employing panel data regression revealed that population size and education level are also significant predictors of crime, with a R^2 of 93.21% [41]. Thus, the results of this study further support the notion that both economic vulnerability and demographic-social factors are closely linked to criminal activity.

4. CONCLUSION

The analysis results of crime modeling in East Java using truncated spline semiparametric regression showed that the optimal model was achieved with a knot point combination of (4,2,4,3), yielding a GCV value of 49636.60 and an R^2 value of 93.60%. Based on the overall hypothesis testing results, it can be concluded that predictor variables collectively have a significant impact on the response variable. Meanwhile, the partial or specific results indicate that the average length of schooling (x_2), population size (x_4), Gini ratio (x_5), per capita expenditure (x_6), and open unemployment rate (x_7) significantly influence the number of crimes in East Java Province. The findings of this study highlight that crime in East Java is strongly influenced by socioeconomic factors, including education, population density, inequality, expenditure, and unemployment. These results suggest that government efforts to reduce crime should go beyond law enforcement and focus on addressing root causes through policies that expand access to education, reduce income disparities, increase employment opportunities, and improve urban management. In the current context of persistent regional inequalities, integrated development programs remain essential to mitigating crime across East Java.

This study still has several limitations. Among these, outlier detection in the scatter plot has not been conducted, and further exploration of higher polynomial orders and variations in knot placement within the regression model has not been examined, which could potentially improve model accuracy. For future research, the inclusion of additional predictor variables that are more closely related to crime data is recommended to capture broader influencing factors. Moreover, experimenting with higher orders and more diverse knot placements could lead to better model performance. These suggestions are expected to help future researchers build upon this work and enhance its relevance and usefulness in supporting strategies to address crime problems more effectively.

Author Contributions

Yahya Vigo Tri Saputra: Conceptualization, Formal Analysis, Methodology, Visualization, Writing–Review and Editing. Moh. Hafiyusholeh: Data Curation, Resources, Validation. Hani Khaulasari: Writing – Original Draft, Project Administration. Yuniar Farida: Project Administration, Software. Putroue Keumala Intan: Supervision, Resources. All authors reviewed and approved the final manuscript.

Funding Statement

This research received financial support from the academic advisor, to whom we extend our sincere gratitude.

Acknowledgment

The authors sincerely thank all parties who contributed to the completion of this research, especially the supervisors who assisted with data processing and analysis, manuscript editing, and provided valuable input during the writing process. Thanks are also extended to the reviewers for their constructive suggestions and corrections, which helped improve this article.

Declarations

This research is carried out jointly, with each author following their assigned tasks, ensuring no conflict of interest between authors.

Declaration of Generative AI and AI-assisted technologies

Generative AI tools (e.g., ChatGPT) were used solely for language refinement (grammar, spelling, and clarity). The scientific content, analysis, interpretation, and conclusions were developed entirely by the authors. The authors reviewed and approved all final text.

REFERENCES

[1] D. R. Ningsih, P. K. Intan, and D. Yuliati, "PEMODELAN TINDAK PIDANA KRIMINALITAS DI KOTA TANGERANG MENGGUNAKAN METODE REGRESI LASSO," *Estimasi J. Stat. Its Appl.*, vol. 4, no. 1, pp. 64–77, 2023, doi:

https://doi.org/10.20956/ejsa.vi.24853.

[2] Badan Pusat Statistik, *PROVINSI JAWA TIMUR DALAM ANGKA 2024*. Surabaya: Badan Pusat Statistik Provinsi Jawa Timur, 2024.

[3] B. C. Boeky, "FAKTOR PENYEBAB DAN UPAYA PENANGGULANGAN PENGULANGAN DALAM TINDAK PIDANA PENCURIAN DENGAN KEKERASAN (RESIDIV) DI WILAYAH HUKUM KEPOLISIAN RESOR KUPANG KOTA," *COMSERVA J. Penelit. dan Pengabdi. Masy.*, vol. 3, no. 8, pp. 3147–3158, 2023, doi: https://doi.org/10.59141/comserva.v3i08.1090.

[4] W. Pratama, "TIGA PENCURI SEPEDA DI PERUMAHAN ELITE SURABAYA AKHIRNYA DIBEKUK POLISI," 2022. https://www.suarasurabaya.net/kelanakota/2022/tiga-pencuri-sepeda-di-perumahan-elite-surabaya-akhirnya-dibekuk-polisi/ (accessed Mar. 09, 2025).

[5] K. Rosidin, "PEMBUNUHAN SADIS DI MENGANTI, BEGINI KRONOLOGI KEJADIAN DAN PENANGKAPAN TERSANGKA," 2023. https://infogresik.id/pembunuhan-sadis-di-menganti-begini-kronologi-kejadian-dan-penangkapan-tersangka/ (accessed Mar. 09, 2025).

[6] M. A. F. Hakim and A. G. AE, "LIMA KEJAHATAN DI KEDIRI YANG CURI PERHATIAN PUBLIK SEPANJANG FEBRUARI 2024," *Kompas.com*, 2024. https://surabaya.kompas.com/read/2024/03/08/111812278/lima-kejahatan-di-kediri-yang-curi-perhatian-publik-sepanjang-februari-2024?page=all (accessed May 08, 2025).

[7] R. Septaria and Siti Mutmainnah Zulfaridatulyaqin, "TINGKAT KRIMINALITAS DI KOTA BANJARMASIN DENGAN PENDEKATAN EKONOMI," *JIEP J. Ilmu Ekon. dan Pembang.*, vol. 4, no. 1, pp. 50–64, 2021, doi: https://doi.org/jiep.ulm.ac.id/index.php/jiep/article/view/2188.

[8] A. I. Effendi and A. Julia, "FAKTOR EKONOMI YANG MEMPENGARUHI KEJAHATAN PROPERTI DI PULAU JAWA TAHUN 2014-2019," *J. Ris. Ilmu Ekon. dan Bisnis*, vol. 1, no. 1, pp. 41–47, 2021.

[9] A. A. C. Cindy and M. Ratna, "PEMODELAN KEJAHATAN DI INDONESIA DENGAN METODE REGRESI NONPARAMETRIK SPLINE TRUNCATED," *J. Sains dan Seni ITS*, vol. 12, no. 2, pp. D167–D174, 2023, doi: https://doi.org/10.12962/j23373520.v12i2.112958.

[10] Y. Febriani, "PENGARUH ASPEK SUMBER DAYA MANUSIA TERHADAP JUMLAH KRIMINALITAS DI SUMATERA SELATAN TAHUN 2019," *J. Media Wahana Ekon.*, vol. 18, no. 1, pp. 146–156, 2021, doi: https://doi.org/10.31851/jmwe.v18i1.5601.

[11] D. A. Sulistiani, Y. Farida, S. K. Sari, and D. C. R. Novitasari, "MODELING CRIMINALITY IN SOUTH SULAWESI USING GEOGRAPHICALLY WEIGHTED REGRESSION (GWR) METHOD," *AIP Conf. Proc.*, vol. 3142, no. 1, p. 20068, Mar. 2025, doi: https://doi.org/10.1063/5.0262022.

[12] D. Kuciswara, F. Muslihatinningsih, and E. Santoso, "PENGARUH URBANISASI, TINGKAT KEMISKINAN, DAN KETIMPANGAN PENDAPATAN TERHADAP KRIMINALITAS DI PROVINSI JAWA TIMUR," *JAE J. Akunt. dan Ekon.*, vol. 6, no. 3, pp. 1–9, 2021, doi: https://doi.org/10.29407/jae.v6i3.16307.

[13] M. Arynta, I. N. Budiantara, and M. Ratna, "PEMODELAN FAKTOR-FAKTOR YANG MEMPENGARUHI (CURANMOR) DI JAWA TIMUR MENGGUNAKAN REGRESI NONPARAMETRIK SPLINE TRUNCATED," *J. Sains dan Seni ITS*, vol. 8, no. 2, pp. D246–D251, 2020, doi: https://doi.org/10.12962/j23373520.v8i2.44436.

[14] L. Mahfiroh and Y. Farida, "SPLINE NONPARAMETRIC REGRESSION TO IDENTIFY FACTORS AFFECTING GENDER EMPOWERMENT MEASURE (GEM) IN EAST JAVA," *CAUCHY J. Mat. Murni dan Apl.*, vol. 7, no. 1, pp. 105–117, 2021, doi: https://doi.org/10.18860/ca.v7i1.12993.

[15] F. Ubaidillah, A. A. R. Fernandes, A. Iriany, N. W. S. Wardhani, and S. Solimun, "TRUNCATED SPLINE PATH ANALYSIS MODELING ON IN COMPANY X WITH THE GOVERNMENT 'S ROLE AS A MEDIATION VARIABLE," *J. Stat. Appl. Probab.*, vol. 794, no. 3, pp. 781–794, 2022, doi: https://doi.org/10.18576/jsap/110303.

[16] W. B. Altukhaes, M. Roozbeh, and N. A. Mohamed, "FEASIBLE ROBUST LIU ESTIMATOR TO COMBAT OUTLIERS AND MULTICOLLINEARITY EFFECTS IN RESTRICTED SEMIPARAMETRIC REGRESSION MODEL," *AIMS Math.*, vol. 9, no. 11, pp. 31581–31606, 2024, doi: https://doi.org/10.3934/math.20241519.

[17] A. T. R. Dani and N. Y. Adrianingsih, "PEMODELAN REGRESI NONPARAMETRIK DENGAN ESTIMATOR SPLINE TRUNCATED DAN DERET FOURIER," *Jambura J. Math.*, vol. 3, no. 1, pp. 26–36, 2021, doi: https://doi.org/10.34312/jjom.v3i1.7713.

[18] C. Christodoulou-Volos and D. Tserkezos, "SENSITIVITY OF THE RAMSEY'S REGRESSION SPECIFICATION ERROR TERM TEST ON THE DEGREE OF NONLINEARITY OF THE FUNCTIONAL FORM," *J. Appl. Econ. Sci.*, vol. XVIII, no. 1, pp. 2–7, 2023, doi: https://doi.org/10.57017/jaes.v18.1(79).01.

[19] M. R. H. Nurdin, A. A. R. Fernandes, E. Sumarminingsih, and M. O. Ullah, "DEVELOPMENT OF RAMSEY RESET TO IDENTIFY THE POLYNOMIALS ORDER OF SMOOTHING SPLINE WITH SIMULATION STUDY," *JTAM (Jurnal Teor. dan Apl. Mat.*, vol. 9, no. 1, pp. 175–189, 2025, doi: https://doi.org/10.31764/jtam.v9i1.26785.

[20] S. A. Mudawamah, G. T. Swastika, R. Narendra, and M. Qomarudin, "PEMODELAN REGRESI SEMIPARAMETRIK DENGAN PENDEKATAN SPLINE TRUNCATED PADA INDEKS PEMBANGUNAN MANUSIA (IPM) DI JAWA TIMUR," *J. Stat.*, vol. 22, no. 2, pp. 183–194, 2022, doi: https://doi.org/10.29313/statistika.v22i2.1433.

[21] R. D. Fadlirhohim, Sifriyani, and A. T. R. Dani, "MODELING STUNTING PREVALENCE IN INDONESIA USING SPLINE TRUNCATED SEMIPARAMETRIC REGRESSION," *BAREKENG J. Math. Its Appl.*, vol. 18, no. 3, pp. 2015–2028, 2024, doi: https://doi.org/10.30598/barekengvol18iss3pp2015-2028.

[22] Y. Wulandari and Yusnida, "PENGARUH PENDIDIKAN, JUMLAH PENDUDUK, DAN PENGANGGURAN TERHADAP KRIMINALITAS DI PROVINSI SUMATRA BARAT PERIODE 2020-2022," *Ranah Res. J. Multidiscip. Res. Dev.*, vol. 7, no. 3, pp. 2015–2024, 2025, doi: https://doi.org/10.38035/rrj.v7i3.1417.

[23] M. A. Juniar, A. Fania, D. Ulya, R. Ramadhan, and N. Chamidah, "MODELLING CRIME RATES IN INDONESIA USING TRUNCATED SPLINE ESTIMATOR," *BAREKENG J. Math. Its Appl.*, vol. 18, no. 2, pp. 1201–1216, 2024, doi: https://doi.org/10.30598/barekengvol18iss2pp1201-1216.

[24] C. B. Harahap and I. Sulhin, "PENGENDALIAN KEJAHATAN PADA SUB-KEBUDAYAAN GENG KLITIH (DALAM PARADIGMA KRIMINOLOGI BUDAYA)," *Deviance J. Kriminologi*, vol. 6, no. 1, pp. 86–102, 2022, doi: https://doi.org/10.36080/djk.1569.

[25] H. Nurcahyani, I. N. Budiantara, and I. Zain, "THE SEMIPARAMETRIC REGRESSION CURVE ESTIMATION BY

USING MIXED TRUNCATED SPLINE AND FOURIER SERIES MODEL," *AIP Conf. Proc.*, vol. 2329, 2021, doi: https://doi.org/10.1063/5.0042870.

[26] A. T. Ampa, L. Laome, M. Ridwan, Baharuddin, and Makkulau, "MODELING GENDER DEVELOPMENT INDEX IN SOUTHEAST SULAWESI PROVINCE USING SEMIPARAMETRIC KERNEL REGRESSION," *BAREKENG J. Math. Its Appl.*, vol. 19, no. 3, pp. 1525–1536, 2025, doi: https://doi.org/10.30598/barekengvol19iss3pp1525-1536.

[27] A. S. Suriaslan, I. N. Budiantara, and V. Ratnasari, "TRUNCATED SPLINE REGRESSION FOR BINARY RESPONSE: A COMPARATIVE DTUDY OF NONPARAMETRIC AND SEMIPARAMETRIC APPROACHES," *Commun. Math. Biol. Neurosci.*, vol. 2025, no. 51, 2025, doi: https://doi.org/10.28919/cmbn/9209.

[28] R. Putra, M. G. Fadhlurrahman, and Gunardi, "DETERMINATION OF THE BEST KNOT AND BANDWIDTH IN GEOGRAPHICALLY WEIGHTED TRUNCATED SPLINE NONPARAMETRIC REGRESSION USING," *MethodsX*, vol. 10, p. 101994, 2023, doi: https://doi.org/10.1016/j.mex.2022.101994.

[29] N. Chamidah *et al.*, "CONSISTENCY AND ASYMPTOTIC NORMALITY OF ESTIMATOR FOR PARAMETERS IN MULTIRESPONSE MULTIPREDICTOR SEMIPARAMETRIC," *Symmetry (Basel).*, vol. 14, no. 2, p. 336, 2022, doi: https://doi.org/10.3390/sym14020336.

[30] A. S. Suriaslan, I. N. Budiantara, and V. Ratnasari, "NONPARAMETRIC REGRESSION ESTIMATION USING MULTIVARIABLE TRUNCATED SPLINES FOR BINARY RESPONSE DATA," *MethodsX*, vol. 14, p. 103084, 2025, doi: https://doi.org/10.1016/j.mex.2024.103084.

[31] B. Lestari, N. Chamidah, I. N. Budiantara, and D. Aydin, "DETERMINING CONFIDENCE INTERVAL AND ASYMPTOTIC DISTRIBUTION FOR PARAMETERS OF MULTIRESPONSE SEMIPARAMETRIC REGRESSION MODEL USING SMOOTHING SPLINE ESTIMATOR," *J. King Saud Univ. - Sci.*, vol. 35, no. 5, p. 102664, 2023, doi: https://doi.org/10.1016/j.jksus.2023.102664.

[32] A. M. Sadek and L. A. Mohammed, "EVALUATION OF THE PERFORMANCE OF KERNEL NON-PARAMETRIC REGRESSION AND ORDINARY LEAST SQUARES REGRESSION," *Int. J. Informatrics Vis.*, vol. 8, no. 3, pp. 1352–1360, 2024, doi: https://doi.org/10.62527/joiv.8.3.2430.

[33] H. Prabowo, Suhartono, and D. D. Prastyo, "THE PERFORMANCE OF RAMSEY TEST, WHITE TEST AND TERASVIRTA TEST IN DETECTING NONLINEARITY," *INFERENSI*, vol. 3, no. 1, pp. 1–12, 2020, doi: https://doi.org/10.12962%2Fj27213862.v3i1.6876.

[34] F. H. Junianto, A. A. R. Fernandes, and Solimun, "STRUCTURAL EQUATION MODELING SEMIPARAMETRIC IN MODELING THE ACCURACY OF PAYMENT TIME FOR CUSTOMERS OF CREDIT BANK IN INDONESIa," *JTAM (Jurnal Teor. dan Apl. Mat.*, vol. 8, no. 4, pp. 1082–1095, 2024, doi: https://doi.org/10.31764/jtam.v8i4.23668.

[35] A. T. R. Dani, V. Ratnasari, and I. N. Budiantara, "OPTIMAL KNOTS POINT AND BANDWIDTH SELECTION IN MODELING MIXED ESTIMATOR NONPARAMETRIC REGRESSION," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1115, p. 012020, 2021, doi: https://doi.org/10.1088/1757-899X/1115/1/012020.

[36] H. Luepsen, "ANOVA WITH BINARY VARIABLES: THE F-TEST AND SOME ALTERNATIVES," *Commun. Stat. - Simul. Comput.*, vol. 52, no. 3, pp. 745–769, 2023, doi: https://doi.org/10.1080/03610918.2020.1869983.

[37] G. F. Mumtaz, Suyitno, and Sifriyani, "GEOGRAPHICALLY WEIGHTED PANEL REGRESSION MODELLING OF POVERTY RATES IN TROPICAL RAINFOREST AREAS OF KALIMANTAN," *BAREKENG J. Math. Its Appl.*, vol. 19, no. 2, pp. 903–916, 2025, doi: https://doi.org/10.30598/barekengvol19iss2pp903-916.

[38] D. Chicco, M. J. Warrens, and G. Jurman, "THE COEFFICIENT OF DETERMINATION R-SQUARED IS MORE INFORMATIVE THAN SMAPE , MAE , MAPE , MSE AND RMSE IN REGRESSION ANALYSIS EVALUATION," *PeerJ Comput. Sci.*, pp. 1–24, 2021, doi: https://doi.org/https://peerj.com/articles/cs-623/.

[39] I. Sulistiana, M. Al Hafiz, E. R. Supangadi, J. I. Situmorang, and M. Ikhsan, "PEMODELAN REGRESI NONPARAMETRIK MENGGUNAKAN PENDEKATAN POLINOMIAL LOKAL PADA INFLASI DI INDONESIA," *MATHunesa J. Ilm. Mat.*, vol. 13, no. 2, pp. 545–551, 2025, [Online]. Available: https://ejournal.unesa.ac.id/index.php/mathunesa/article/view/67350

[40] Y. Farida, M. Farmita, P. K. Intan, H. Khaulasari, and A. T. Wibowo, "MODELLING CRIME IN EAST JAVA USING SPATIAL DURBIN MODEL REGRESSION," *BAREKENG J. Math. Its Appl.*, vol. 18, no. 3, pp. 1497–1508, 2024, doi: https://doi.org/10.30598/barekengvol18iss3pp1497-1508.

[41] Nurwahyuni, Z. W. Baskara, and N. A. Purnamasari, "MODEL REGRESI DATA PANEL PADA TINGKAT KRIMINALITAS DI INDONESIA NUSA TENGGARA BARAT DENGAN MENGGUNAKAN FIXED EFFECT MODEL," *Var. J. Stat. Its Appl.*, vol. 5, no. 2, pp. 169–184, 2023, doi: https://doi.org/10.30598/variancevol5iss2page169-184.