Zero : Jurnal Sains, Matematika, dan Terapan
E-ISSN : 2580-5754; P-ISSN : 2580-569X
Volume 10, Number 1, 2026
DOI: 10.30829/zero.v10i1.26671
Page: 80 - 91

Integrating Clustering and Hybrid Nonparametric Path Modeling for Waste Management Behavior in Batu City

[1] Moh Zhafran Hidayatulloh
Departement of Statistics, Universitas Brawijaya, Malang, Indonesia

[2] Solimun
Departement of Statistics, Universitas Brawijaya, Malang, Indonesia

[3] Adji Achmad Rinaldo Fernandes
Departement of Statistics, Universitas Brawijaya, Malang, Indonesia

[4] Anggun Fadhilla Rizqia
Departement of Statistics, Universitas Brawijaya, Malang, Indonesia

[5] Fachira Haneinanda Junianto
Departement of Mathematics, Universitas Brawijaya, Malang, Indonesia

Article Info

Article history:

Accepted 25 January 2026

Keywords:

Fuzzy C-Means;
Fourier Series;
Path Analysis;
Truncated Spline;
Waste Management Behavior

ABSTRACT

This study formulates a hybrid nonparametric path analysis utilizing truncated spline-Fourier series integrated with Fuzzy C-Means clustering methodologies to investigate nonlinear behavioral patterns in community-based waste management. The model was utilized on survey data from 210 respondents in Bumiaji District, Batu City, encompassing Environmental Quality, Use of Waste Banks, Use of the 3R Principles, and Economic Benefits from Waste. Two distinct behavioral groups with varying emphasis on Use of Waste Banks and Environmental Quality were identified by the analysis. The hybrid model obtained a high coefficient of determination (0.891) and captures nonlinear relationships more successfully than the traditional linear approach. These results emphasize how crucial nonlinear behavioral dynamics are in influencing waste management behavior/decisions. The proposed framework helps local governments create more focused and efficient waste-management plans by offering a useful and adaptable analytical for understanding community behavior.

Corresponding Author:

Moh Zhafran Hidayatulloh,
Department of Statistics
Universitas Brawijaya, Indonesia
Email: zhafranhidayat@student.ub.ac.id

1. INTRODUCTION

Waste management remains a major challenge for many Indonesian cities, including Batu City, East Java, where rapid population growth and economic expansion have substantially increased waste generation while infrastructure capacity remains limited [1], [2]. Although policy interventions such as recycling programs, public awareness campaigns, and community-based initiatives have been implemented [3], [4], [5], community responses remain highly diverse. This behavioral heterogeneity suggests that uniform analytical approaches are insufficient to capture the complexity of waste management behavior [6], [7], [8].

In multivariate behavioral studies, relationships are frequently nonlinear, irregular, and influenced by interacting social and environmental factors [9], [10]. Classical path analysis, which depends on linear and additive assumptions, becomes inadequate under these conditions [12], [13], [14]. When the functional form is unknown, nonparametric path analysis offers a flexible alternative by allowing the data to determine the relationship structure [15], [16], [17], [18]. However, existing models typically rely on a single estimator, most commonly truncated spline or Fourier series applied uniformly across variables.

This single estimator strategy introduces restrictive structural assumptions because behavioral variables often exhibit different patterns, including abrupt shifts, smooth trajectories, and periodic fluctuations. Truncated splines effectively capture local changes through knot points, whereas Fourier series model oscillatory behavior [19], [20]. Imposing one estimator across all relationships may therefore reduce flexibility and increase estimation bias, particularly in heterogeneous behavioral data [21]. Beyond functional form, population heterogeneity also necessitates segmentation. Measuring multivariate characteristics rarely depends on a single variable, making clustering essential for uncovering latent structures [22]. Fuzzy C-Means (FCM) is especially advantageous because it assigns partial membership and is more robust to outliers than K-Means.

Prior studies have advanced nonparametric modeling but remain methodologically fragmented. Study [23] employed truncated spline-based nonparametric path analysis, while [24] demonstrated the effectiveness of Fourier series in capturing recurring patterns. Separately, [25] applied FCM for classification tasks. Nevertheless, these approaches have not been integrated, limiting the capacity to simultaneously model nonlinear relationships and behavioral heterogeneity.

This study addresses that limitation by developing a hybrid nonparametric path framework that integrates truncated spline and Fourier series with dummy variables derived from Fuzzy C-Means clustering. The proposed approach enhances model adaptiveness by allowing different functional representations across variables while explicitly accounting for latent population structure. In addition, a modified Ramsey RESET test is introduced to improve the detection of nonlinear relationships generated by hybrid estimators.

This study investigates: (1) the construction of a hybrid truncated spline–Fourier path estimator with FCM-based segmentation, and (2) the performance of its hypothesis testing using t-statistics with jackknife resampling. The framework is expected to provide a more flexible methodological foundation for modeling behavioral complexity and to contribute to the advancement of nonparametric path analysis. To maintain analytical focus, the truncated spline is restricted to linear order with one and two knot points, the Fourier component is limited to a maximum of three oscillations, and the FCM algorithm is configured to form two clusters. These constraints define the modeling space while preserving interpretability.

2. RESEARCH METHOD

2.1 Data Sources

This study uses primary data collected through a survey on public mindset toward the economic benefits of waste in Batu City. The population consists of residents of Bumiaji District, with a sample of 210 respondents selected using quota sampling based on the proportion of residents in each village engaged in community-based waste-management activities. The minimum sample size was determined using the Slovin formula for an estimated population of 440 households participating in waste-management programs, with a 5% margin of error, as presented in Equation (1):

$$n = \frac{N}{1 + Ne^2} = \frac{440}{1 + 440(0.05)^2} \approx 210 \tag{1}$$

Although formal power analysis for nonlinear models is not straightforward, a sample size of 210 is considered adequate for stable nonparametric estimation, GCV-based model selection, and capturing nonlinear patterns across behavioral clusters. Latent variables were measured using multiple Likert-scale items, which were rescaled using a summated rating scale approach and averaged to form composite scores. All instruments satisfied validity and reliability criteria. To meet the input requirements of Fourier terms, composite variables were normalized to the 0–1 range using min–max scaling. No missing data were observed, and therefore no imputation was required.

2.2 Research Design

This study involves two exogenous variables, namely Environmental Quality (X_1), and Use of Waste Banks (X_2), one intervening variable, namely the Use of the 3R Principles (Y_1), and one endogenous variable, namely the Economic Benefits from Waste (Y_2). The model used in this study is presented in Figure 1. All data processing, model estimation, and visualization were carried out using the RStudio software, which supports nonparametric modeling spline–Fourier analysis.

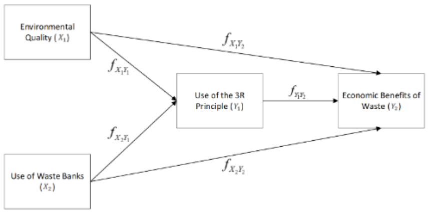

Figure 1. Research Path Diagram

2.3 Research Step

The procedures in this research are outlined as follows:

a. Develop a path diagram that illustrates the relationships among exogenous, intervening endogenous, and purely endogenous variables identified in the study.
b. Input the secondary data with $n = 210$.
c. Perform Fuzzy C-Means clustering.
d. Construct a dummy variable matrix based on the Fuzzy C-Means clustering results to distinguish between clusters.
e. Evaluate the linearity assumption of the relationships among variables by employing Ramsey's Regression Specification Error Test (RESET) and the modified Ramsey RESET for nonparametric relationships.
f. Develop estimators for the combined nonparametric path functions of truncated spline and Fourier series with dummy variables obtained from the Fuzzy C-Means clustering results.
g. Assess the validity of the model by utilizing the coefficient of determination.
h. Calculate the significance of the estimated path functions from the best model using the t-test statistic through the jackknife resampling method.
i. Interpret the results of the waste management behavior data analysis.

2.4 Fuzzy C-Means

According to [25], Fuzzy C-Means (FCM) is a widely used clustering algorithm that assigns data points to multiple clusters with varying degrees of membership through an iterative optimization process. The number of clusters is specified in advance, and cluster centers are used to determine membership levels for each observation. The main stages of the FCM procedure are outlined below [25].

a. Determine the number of clusters (in this study, two clusters), the weighting exponent (in this study, $w = 2$), the maximum number of iterations, smallest error ($\varepsilon = 10^{-9}$), the initial objective function ($P_0 = 0$), and the initial iteration.
b. Generate random numbers u_{ik}
c. Calculate the centroids of the two groups using the following Equation (2).

$$v_{kj} = \frac{\sum_{i=1}^{n}(u_{ik})^w x_{ij}}{\sum_{i=1}^{n}(u_{ik})^w} \tag{2}$$

d. Determine the stopping criterion for the iteration using the following Equation (3).

$$\mathrm{U}^l = \sum_{i=1}^{n}\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right](u_{ik})^w\right) \tag{3}$$

e. Calculate the degree of membership of each data point in each cluster using the following Equation (4).

$$u_{ik} = \frac{\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{-\frac{1}{w-1}}}{\sum_{k=1}^{c}\left(\left[\sum_{j=1}^{m}\left(x_{ij} - v_{kj}\right)^2\right]^{-\frac{1}{w-1}}\right)} \tag{4}$$

f. Get cluster results.

2.5 Ramsey RESET

This study requires a linearity test to determine whether parametric path analysis is appropriate, that is, whether the linearity assumption is satisfied. The Ramsey RESET test is employed to examine the functional form of the relationship between exogenous and endogenous variables. The hypotheses are formulated as follows:
H_0 : $\beta_2 = \beta_3 = 0$
H_1 : There is at least one $\beta_j \neq 0$; $j = 2,3$
Equation (5) is an equation used to perform a linearity test.

$$\begin{gathered} \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} \\ \hat{Y}_i^* = \hat{\beta}_0^* + \hat{\beta}_1^* X_{1i} + \hat{\beta}_2 \hat{Y}_i^2 + \hat{\beta}_3 \hat{Y}_i^3 \end{gathered} \tag{5}$$

From Equation (5), the coefficient of determination of each model is then calculated to obtain R_1^2 and R_2^2. Subsequently, a linearity test is performed using the F-test as shown in Equation (6).

$$F_{calculated} = \frac{(R_2^2 - R_1^2)/m}{(1 - R_2^2)/(n - p - 1 - m)} \sim F_{m,n-p-1-m} \tag{6}$$

Description:
p : Number of initial predictor variables
m : Number of additional predictors

If the p-value is greater than 0.05, the relationship is considered linear; if it is less than 0.05, the relationship is considered nonlinear. While the classical Ramsey RESET test detects model misspecification, it does not identify the true functional form. Therefore, this study employs a modified RESET test using a flexible nonparametric basis to capture a wide range of nonlinear patterns in a more principled manner.

2.6 Modified Ramsey RESET

The Ramsey RESET test is modified by incorporating truncated spline and Fourier-based nonparametric path equations into the auxiliary regression, enabling direct detection of nonlinear misspecification arising from piecewise and oscillatory relationships beyond conventional parametric forms.

The first modification of the Ramsey RESET aims to detect nonparametric relationships modeled using a linear-order truncated spline with one knot point. The hypotheses used are as follows.
$H_0: \beta_2 = 0$
$H_1 : \beta_2 \neq 0$
The equation used is presented in the following Equation (7).

$$\begin{gathered} \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} \\ \hat{Y}_i^* = \hat{\beta}_0^* + \hat{\beta}_1^* X_{1i} + \hat{\beta}_2 (X_{1i} - K_{11})_+ \end{gathered} \tag{7}$$

If the p-value obtained from the Ramsey RESET test for the linear truncated spline with one knot is less than 0.05, this indicates that the identified relationship follows a linear-order truncated spline pattern with one knot. The same interpretation applies to the case with two knot points.

The second modification of the Ramsey RESET test is designed to detect nonparametric relationships modeled using the Fourier series with one oscillation. The hypotheses for this test are formulated as follows:
$H_0: \gamma_1 = 0$
$H_1 : \gamma_1 \neq 0$
The equation used is presented in the following Equation (8).

$$\begin{gathered} \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} \\ \hat{Y}_i^* = \hat{\beta}_0^* + \hat{\beta}_1^* X_{1i} + \gamma_1 Cos X_{1i} \end{gathered} \tag{8}$$

If the p-value obtained from the Ramsey RESET test for the Fourier series at oscillation level 1 is less than 0.05, this indicates that the identified relationship follows a Fourier series pattern with one oscillation. The same interpretation applies to oscillation levels 2 and 3.

The p-value is calculated using the same procedure as in the classical Ramsey RESET test, by computing the coefficient of determination for each equation and then obtaining the p-value from the F-test.

2.7 Nonparametric Path Analysis with Truncated Spline

A truncated spline is a polynomial function segmented at a certain order m, within which there are points known as knots. Generally, define a truncated spline function of order m with knots $u = 1,2, \dots, r$ is any function that can be expressed in the form of Equation (9)[26].

$$f(X_{1i}) = \sum_{k=0}^{m} \beta_k X_{1i}^k + \sum_{u=1}^{r} \beta_{m+u} (X_{1i} - K_u)_+^m \tag{9}$$

The + sign in $(X_{1i} - K_u)_+^m$ indicates that the expression has two possible values, just as shown in Equation (10):

$$(X_{1i} - K_u)_+^m = \begin{cases} (X_{1i} - K_u)^m & ; X_{1i} \geq K_u \\ 0 & ; X_{1i} < K_u \end{cases} \tag{10}$$

Description:
$f(X_{1i})$: The estimated spline function evaluated at X_{1i}
X_{1i} : The value of the predictor variable for observation i.
β_k : Coefficients for the polynomial basis terms of degree k.
k $= 1,2, \dots, m$; m: Order spline
u $= 1,2, \dots, r$; r: Number of knot points
K_u : The u-th knot point

Subsequently, the estimator of β is obtained using the Weighted Least Squares (WLS) method, which determines the parameter values by minimizing the weighted sum of squared errors, as shown in Equation (11).

$$\hat{\underset{\sim}{\beta}} = \left(\mathbf{X}[\underset{\sim}{K}]' \hat{\mathbf{\Sigma}}^{-1} \mathbf{X}[\underset{\sim}{K}]\right)^{-1} \mathbf{X}[\underset{\sim}{K}]' \hat{\mathbf{\Sigma}}^{-1} \underset{\sim}{y} \tag{11}$$

In the nonparametric truncated spline path, identifying the optimal knot point plays a crucial role, as the choice of this point significantly influences the shape of the resulting regression curve [27]. One method commonly used to select the optimal knot point is the Generalized Cross Validation (GCV) method. The GCV method can be expressed by the following Equation (12).

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

2.8 Nonparametric Path Analysis of Fourier Series

The Fourier series, as a trigonometric polynomial, possesses a high degree of flexibility that enables it to effectively capture and adapt to the local characteristics of the data [28]. One of its main advantages is its ability to model data exhibiting trigonometric patterns, such as sine and cosine functions [29]. In the context of nonparametric regression utilizing the Fourier series, the model's performance largely relies on the oscillation parameter. This parameter indicates the number of cosine wave oscillations incorporated within the model. As stated in [30], the Fourier series estimator for the curve f is expressed in Equation (13).

$$\hat{f}(X_i) = \frac{1}{2}\hat{a}_0 + \hat{b}_1 X_i + \sum_{l=1}^{O} \gamma_l Cos l X_i \tag{13}$$

Description:

γ_l : Fourier coefficient for the l-th cosine component
$\hat{b}_1$: Regression coefficient for variable x
$\hat{a}_0$: Constant coefficient for the zero-order cosine component
l = $1, 2, \dots, O$; O = Number of oscillations

2.9 Combined Nonparametric Path of Truncated Spline and Fourier Series

In nonparametric path analysis, a single estimator is often applied across all functions, implicitly assuming similar data patterns. In practice, however, different relationships may exhibit distinct functional characteristics. To address this, this study combines Truncated Spline and Fourier Series within a unified nonparametric path framework.

Let the data be given as $(x_{1i}, x_{2i}, y_{1i}, y_{2i})$, and the relationships among the exogenous, intervening, and endogenous variables are assumed to follow a nonparametric path model. assumed that the regression curves $f_{11}(X_{1i})$ and $f_{22}(X_{2i})$ follow truncated spline regression, while $f_{12}(X_{2i})$, $f_{21}(X_{1i})$, and $f_{23}(Y_{1i})$ follow Fourier series regression. The estimator of the combined nonparametric path model is as follows Equation (14):

$$\begin{aligned} \hat{f}_1(X_{1i}, X_{2i}) &= \hat{f}_{11}(X_{1i})_{trunc} + \hat{f}_{12}(X_{2i})_{fourier} \\ \hat{f}_2(X_{1i}, X_{2i}, Y_{1i}) &= \hat{f}_{21}(X_{1i})_{fourier} + \hat{f}_{22}(X_{2i})_{trunc} + \hat{f}_{23}(Y_{1i})_{fourier} \end{aligned} \tag{14}$$

Equation (16) presents one hybrid configuration of truncated spline and Fourier series within the proposed nonparametric path framework. Estimator selection for each structural path is data-driven, based on comparisons of truncated spline and Fourier models using the lowest Generalized Cross Validation (GCV) value and the highest coefficient of determination (R^2). This procedure ensures flexible adaptation to diverse data patterns and avoids arbitrary model specification.

2.10 Development of the Dummy Approach in Combined Nonparametric Path Analysis

Fuzzy C-Means was employed to account for the ambiguous and overlapping nature of community waste management behavior, enabling partial membership across clusters rather than rigid classification. The resulting membership degrees were incorporated as dummy variables within the truncated spline Fourier nonparametric path model to assess the effects of exogenous variables on endogenous outcomes. When two clusters were identified, the corresponding path specification involving two exogenous variables, one intervening variable, and one endogenous variable is presented in Equation (15).

$$\begin{aligned} \hat{f}_{1i} &= \hat{\beta}_{10} + \hat{\beta}_{11}D_i + \hat{\beta}_{12}X_{1i} + \hat{\beta}_{13}D_iX_{1i} + \hat{\beta}_{14}(X_{1i} - K_{11})_+ + \hat{\beta}_{15}(X_{1i} - K_{11})_+D_i \\ &\quad + \hat{\beta}_{16}X_{2i} + \hat{\beta}_{17}D_iX_{1i} + \hat{\beta}_{18}(X_{2i} - K_{12})_+ + \hat{\beta}_{19}(X_{2i} - K_{12})_+D_i \\ \hat{f}_{2i} &= \hat{\beta}_{20} + \hat{\beta}_{21}D_i + \hat{\beta}_{22}X_{1i} + \hat{\beta}_{23}D_iX_{1i} + \hat{\beta}_{24}(X_{1i} - K_{21})_+ + \hat{\beta}_{25}(X_{1i} - K_{21})_+D_i \\ &\quad + \hat{\beta}_{26}X_{2i} + \hat{\beta}_{27}D_iX_{2i} + \hat{\beta}_{28}(X_{2i} - K_{22})_+ + \hat{\beta}_{29}(X_{2i} - K_{22})_+D_i + \hat{\beta}_{210}Y_{1i} + \hat{\beta}_{211}D_iY_{1i} + \\ &\quad \hat{\beta}_{212}(Y_{1i} - K_{23})_+ + \hat{\beta}_{213}(Y_{1i} - K_{23})_+D_i \end{aligned} \tag{15}$$

When it is found that the number of clusters obtained is two, the cluster model with a nonparametric Fourier series path for two exogenous variables, one intervening variable, and one endogenous variable is written in Equation (16).

$$\begin{aligned} \hat{f}_{1i} &= \hat{\beta}_{10} + \hat{\beta}_{11}D_i + \hat{\beta}_{12}X_{1i} + \hat{\beta}_{13}D_iX_{1i} + \hat{\gamma}_{11}cosX_{1i} + \hat{\gamma}_{12}cosX_{1i}D_i + \hat{\beta}_{14}X_{2i} + \\ &\quad \hat{\beta}_{15}D_iX_{2i} + \hat{\gamma}_{13}cosX_{2i} + \hat{\gamma}_{14}cosX_{2i}D_i \\ \hat{f}_{2i} &= \hat{\beta}_{20} + \hat{\beta}_{21}D_i + \hat{\beta}_{22}X_{1i} + \hat{\beta}_{23}D_iX_{1i} + \hat{\gamma}_{21}cosX_{1i} + \hat{\gamma}_{22}cosX_{1i}D_i + \hat{\beta}_{24}X_{2i} + \\ &\quad \hat{\beta}_{25}D_iX_{2i} + \hat{\gamma}_{23}cosX_{2i} + \hat{\gamma}_{24}cosX_{2i}D_i + \hat{\beta}_{26}Y_{1i} + \hat{\beta}_{27}Y_{1i}D_i + \hat{\gamma}_{25}cosY_{1i} + \hat{\gamma}_{26}cosY_{1i}D_i \end{aligned} \tag{16}$$

The model equation formed is as in Equation (14), where two nonparametric approaches are contained within one model with the addition of a dummy variable. A nonparametric path model that includes two groups can be represented with two dummy values (D_i; i = 1,2). D_1 for cluster 1 is assigned a value of 0 and D_2 for cluster 2 is assigned a value of 1.

The proposed hybrid nonparametric path model integrates truncated spline and Fourier components through dummy-variable activation to capture cluster-specific nonlinear patterns within a unified estimation framework. Identifiability is ensured by avoiding multicollinearity among hybrid bases, while consistency follows from standard regularity conditions of spline and Fourier nonparametric regression.

2.11 Model Validity

A model is considered valid when its underlying assumptions are met, and one way to assess this validity is through the total coefficient of determination [14]. This measure reflects the proportion of data variability explained by the model, indicating how well the model captures the relationships among variables. The computation of total explained variability is given in Equation (17).

$$R^2_{T,adj} = 1 - \left(1 - R^2_{1,adj}\right)\left(1 - R^2_{2,adj}\right)\cdots\left(1 - R^2_{k,adj}\right) \quad (17)$$

Where:

$$R^2_{k,adj} = 1 - \left(\frac{\sum_{i=1}^{n}(y_{ki} - \hat{f}_{ki})^2/(n-p-1)}{\sum_{i=1}^{n}(y_{ki} - \bar{y}_k)^2/(n-1)}\right) \quad (18)$$

Description:

$R^2_{T,adj}$: Adjusted total coefficient of determination

$R^2_{k,adj}$: Adjusted coefficient of determination for the k-th equation; $k = 1,2$

Next, 5-fold cross-validation is applied to evaluate model performance by dividing the dataset into five equal subsets, providing robust performance assessment and reducing overfitting through repeated training–testing across folds [31], [32], [33].

2.12 Hypothesis Testing with Resampling Jackknife

According to [34], resampling is the process of drawing repeated samples from an existing sample in order to obtain a new sample. The new sample is derived from the original sample of size n, drawn randomly either with replacement or without replacement. Before the bootstrap was introduced, simple resampling methods such as the jackknife had already been widely used. The jackknife reduces estimation bias by systematically removing observations and examining the resulting changes in parameter estimates. It is also used to estimate standard deviation. In its simplest form, the delete-1 jackknife removes one observation at a time and repeats the process for all n observations. Hypothesis testing using the *t*-test statistic as shown in Equation (19).

$$t = \frac{\bar{\hat{\beta}}^*_{jp}}{SE_{\hat{\beta}^*_{jp}}} \sim t_{n-1} \quad (19)$$

3. RESULT AND ANALYSIS

This chapter presents the results and discussion of the analysis, in which Fuzzy C-Means clustering is used to form two clusters that are incorporated as dummy variables into a hybrid truncated spline–Fourier nonparametric path model to capture data-driven nonlinear relationships without imposing a predetermined functional form [30].

3.1 Fuzzy C-Means Results

In this study, 210 observations on four variables were analyzed using the Fuzzy C-Means method to form two clusters. Each observation was not only assigned to a specific cluster but also given a membership value indicating the degree of association of the data with each cluster. The initial step in the Fuzzy C-Means analysis involved generating random u_{ik}values. The next step was to calculate the cluster centers or centroids for the objects in each cluster using Equation (1). The results of the centroid calculations for each cluster are presented in Table 1.

Tabel 1. Centroid Cluster

Variable	Cluster 1	Cluster 2
Environmental Quality (X1)	0.683	0.553
Use of Waste Banks (X2)	0.513	0.594
Use of the 3R Principles (Y1)	0.671	0.517
Economic Benefits from Waste (Y2)	0.612	0.556

Table 1 show that cluster 1 has higher average values for Environmental Quality (X1), Use of the 3R Principles (Y1), and Economic Benefits from Waste (Y2) compared to Cluster 2. Meanwhile, cluster 2 shows a higher average value for Use of Waste Bank (X2) compared to Cluster 1. The members of each cluster are presented in Table 2.

Table 2. Clustering Results of Fuzzy C-Means

Cluster	Number of Members
1	119
2	91

Table 2 shows that cluster 1 consists of 119 respondents, while cluster 2 consists of 91 respondents. After determining the members of each cluster, a single dummy variable can be created, where the dummy variable takes a value of 0 when the dataset belongs to cluster 1 and a value of 1 when it belongs to cluster 2.

3.2 Results of the Ramsey RESET and Modified Ramsey RESET

Relationships with a linear form are resolved using a parametric approach, while relationships that are not linear can be addressed using a nonparametric approach. The linearity test in this study uses Ramsey's RESET. The results of the linearity test are presented in Table 3.

Table 3. Results of the Ramsey RESET Linearity Test

Relationship between Variables	P-value	Results
X1→Y1	<0.001	Not linear
X2→Y1	0.0018	Not linear
X1→Y2	<0.001	Not linear
X2→Y2	0.0271	Not linear
Y1→Y2	0.0037	Not linear

Table 3 show that for all relationships between variables, the p-values are less than α (0.05), so H_0 is rejected. Thus, at a 5% significance level, it can be inferred that the relationships between variables are nonlinear and that the specific form of nonlinearity is either unknown or has not yet been established. The results of the linearity test revealed that the relationship between the variables was nonlinear, but the specific form of this nonlinearity remains unknown. The analysis therefore continued using Ramsey's modified RESET model. In this study, the modified RESET model was applied to five conditions in order to identify the appropriate nonparametric form. The results of these five conditions are presented below.

Table 4. Results of the Modified Ramsey RESET

Relationship between Variables	P-value ($R^2_{adj,auxiliary}$)					Results
	RRT1K	RRT2K	RRF1O	RRF2O	RRF3O	
X1→Y1	0.049 (0.543)	0.031 (0.748)	0.044 (0.601)	0.037 (0.653)	0.041 (0.633)	RRT2K
X2→Y1	0.019 (0.671)	0.047 (0.601)	0.002 (0.796)	0.022 (0.648)	0.031 (0.611)	RRF1O
X1→Y2	0.045 (0.555)	0.011 (0.740)	0.024 (0.696)	0.038 (0.691)	0.045 (0.556)	RRT2K
X2→Y2	0.045 (0.630)	0.036 (0.657)	<0.001 (0.812)	0.003 (0.779)	0.014 (0.766)	RRF1O
Y1→Y2	0.013 (0.716)	0.029 (0.699)	0.041 (0.654)	0.049 (0.643)	0.017 (0.706)	RRT1K

Table 4 shows that functional form selection is based on the smallest p-value and the highest auxiliary adjusted R^2. The relationships between Environmental Quality (X_1) and both Use of the 3R Principles (Y_1) and Economic Benefits from Waste (Y_2) are best modeled by a linear truncated spline with two knots (RRT2K). For the relationships between Use of Waste Banks (X_2) and both Y_1 and Y_2, the first-order Fourier series (RRF1O) provides the best fit. Meanwhile, the relationship between Use of the 3R Principles (Y_1) and Economic Benefits from Waste (Y_2) is optimally represented by a linear truncated spline with one knot (RRT1K), based on both criteria.

3.3 Best Model Selection

The optimal model is selected as the one with the smallest GCV value, which reflects the best trade-off between model accuracy and complexity. Additionally, the model with the highest coefficient of determination (R^2) is preferred, as it indicates a stronger ability to explain data variability. The outcomes of the best model selection, including the optimal knot points, are presented in Table 5.

Table 5. Best Model Result with Optimal Knot Points

Relationship	Optimal Knot	GCV	$R^2_{T,adj}$
$X_1 \rightarrow Y_1$	$K_{11} = 0.22$ $K_{12} = 0.59$	0.0165	0.891
$X_2 \rightarrow Y_1$	-		
$X_1 \rightarrow Y_2$	$K_{21} = 0.14$		

Relationship	Optimal Knot	GCV	$R^2_{T,adj}$
	K_{22} =0.48		
$X_2 \rightarrow Y_2$	-		
$Y_1 \rightarrow Y_2$	K_{31} =0.34		

Table 5 show that the selection of optimal knot points for each relationship between variables in the model are presented. The optimal knot points were determined using the Generalized Cross Validation (GCV) criterion, where a smaller GCV value indicates a model with better fit and appropriate complexity. After obtaining the optimal knot points, it was found that the GCV value is 0.0165 and the adjusted R^2 is 0.891. This means that the constructed model can explain 89.1% of the variability in the Economic Benefits from Waste variable, while the remaining 10.9% is explained by factors outside the model.

3.4 Model Validation Results

Model validation ensures that the proposed hybrid nonparametric path model yields stable and generalizable performance beyond the estimation sample. The process used five-fold cross validation with a 70% training and 30% testing split in each iteration, allowing every subset to serve as the test set. This approach provides a comprehensive assessment of the model's consistency and its ability to generalize to unseen data.

Table 6. K-Fold Cross Validation Results

Fold	Adjusted R^2_{total}	GCV Model
1	0.873	0.0282
2	0.903	0.0173
3	0.887	0.0145
4	0.916	0.0221
5	0.893	0.0164
Mean	0.894	0.0197

Table 6 indicates that the hybrid truncated spline–Fourier nonparametric path model demonstrates stable and reliable performance under k-fold cross-validation. The Adjusted R^2_{total} values are consistently high (0.873–0.916, average 0.894), while GCV values remain low and stable (0.0145–0.0282, average 0.0197), confirming strong predictive accuracy and good generalization without overfitting.

3.5 Results of Path Analysis Modeling Using the Combined Truncated Spline–Fourier Approach with Dummy Variables

The estimation of the nonparametric path function combining truncated spline and Fourier series was conducted after the optimal knot points had been determined based on the Generalized Cross Validation (GCV) criterion. The results of the estimation of the combined nonparametric path function with truncated spline, Fourier series, and dummy variables are presented as follows Equation (20).

$$\begin{aligned}
\hat{f}_{1i} = {} & 0.431 + 0.317D_i + 0.241X_{1i} + 0.211X_{1i}D_i - 0.492(X_{1i} - 0.22)_+ - 0.145(X_{1i} - 0.22)_+D_i \\
& + 0.494(X_{1i} - 0.59)_+ + 0.218(X_{1i} - 0.59)_+D_i - 0.168X_{2i} + 0.582X_{2i}D_i \\
& + 0.288CosX_{2i} - 0.374CosX_{2i}D_i \\
\hat{f}_{2i} = {} & 0.336 + 0.265D_i - 0.194X_{1i} - 0.127X_{1i}D_i + 0.262(X_{1i} - 0.14)_+ + 0.192(X_{1i} - 0.14)_+D_i \\
& - 0.365(X_{1i} - 0.48)_+ + 0.513(X_{1i} - 0.48)_+D_i + 0.122X_{2i} + 0.453X_{2i}D_i \\
& - 0.674CosX_{2i} + 0.387CosX_{2i}D_i - 0.496Y_{1i} - 0.174Y_{1i}D_i + 0.278(Y_{1i} - 0.34)_+ \\
& + 0.257(Y_{1i} - 0.34)_+D_i
\end{aligned} \tag{20}$$

The path equation for Cluster 1, which represents the group with a stronger focus on Environmental Quality, the application of the 3R Principles, and the economic benefits derived from waste management, can be seen in the following Equation (21).

$$\begin{aligned}
\hat{f}_{1i} = {} & 0.431 + 0.241X_{1i} - 0.492(X_{1i} - 0.22)_+ + 0.494(X_{1i} - 0.59)_+ - 0.168X_{2i} + \\
& 0.288CosX_{2i} \\
\hat{f}_{2i} = {} & 0.336 - 0.194X_{1i} + 0.262(X_{1i} - 0.14)_+ - 0.365(X_{1i} - 0.48)_+ + 0.122X_{2i} - \\
& 0.674CosX_{2i} - 0.496Y_{1i} + 0.278(Y_{1i} - 0.45)_+
\end{aligned} \tag{21}$$

Based on Equation (21), the truncated spline results reveal nonlinear direct effects among variables. Environmental Quality (X_1) has a positive effect on the Use of the 3R Principles (Y_1) below 0.22, a negative effect between 0.22 and 0.59, and a renewed positive effect above 0.59. For Economic Benefits from Waste (Y_2), Environmental Quality (X_1) shows a negative effect below 0.14, a positive effect between 0.14 and 0.48, and a declining effect beyond 0.48. The relationship between Use of the 3R Principles (Y_1) and Economic Benefits from Waste (Y_2) exhibits a threshold at 0.34, with a negative effect below this value and a stronger positive effect above it.

The path equation for Cluster 2, which represents the group that is more actively involved in waste bank programs, can be seen in the following Equation (22).

$$\hat{f}_{1i} = 0.748 + 0.452X_{1i} - 0.637(X_{1i} - 0.22)_+ + 0.712(X_{1i} - 0.59)_+ + 0.414X_{2i} - 0.086CosX_{2i}$$
$$\hat{f}_{2i} = 0{,}601 - 0.321X_{1i} + 0.454(X_{1i} - 0.14)_+ + 0.148(X_{1i} - 0.48)_+ + 0.575X_{2i} - 0.287CosX_{2i} - 0.670Y_{1i} + 0.535(Y_{1i} - 0.34)_+ \quad (22)$$

Based on Equation (22), the truncated spline results indicate nonlinear direct effects among variables. Environmental Quality (X_1) has a slight positive effect on the Use of the 3R Principles (Y_1) below 0.22, a negative effect between 0.22 and 0.59, and a stronger positive effect above 0.59. For Economic Benefits from Waste (Y_2), Environmental Quality (X_1) shows a negative effect below 0.14, a positive effect above 0.14, and a stronger positive effect beyond 0.48. The relationship between Use of the 3R Principles (Y_1) and Economic Benefits from Waste (Y_2) exhibits a threshold at 0.34, with a negative effect below this value and a stronger positive effect above it.

3.6 Result of Hypotesis Testing Using Resampling Jackknife

Significance testing of the best model was conducted using the t-test, starting with the jackknife resampling procedure. In this research, the jackknife resampling process involved randomly excluding five observations at each iteration. This resampling step was repeated 1,000 times. The results of the hypothesis testing using resampling jackknife, which was performed 1,000 times, are presented in Table 7.

Table 7. Results of Hypothesis Testing for the Best Model

Relationship	Cluster 1			Cluster 2		
	Coefficient	Estimator	P-value	Coefficient	Estimator	P-value
$X_1 \to Y_1$	$\beta_{12}X_{1i}$	0.241	0.058	$(\beta_{12}+\beta_{13})X_{1i}$	0.452	<0.001*
	$\beta_{14}(X_{1i}-K_{11})_+$	-0.492	<0.001*	$(\beta_{14}+\beta_{15})(X_{1i}-K_{11})_+$	-0.637	<0.001*
	$\beta_{16}(X_{1i}-K_{12})_+$	0.494	0.048*	$(\beta_{16}+\beta_{17})(X_{1i}-K_{12})_+$	0.712	0.029*
$X_2 \to Y_1$	$\beta_{18}X_{2i}$	-0.168	0.054*	$(\beta_{18}+\beta_{19})X_{2i}$	0.414	<0.001*
	$\gamma_{11}CosX_{2i}$	0.288	0.009*	$(\gamma_{11}+\gamma_{12})CosX_{2i}$	0.086	0.034*
$X_1 \to Y_2$	$\beta_{22}X_{1i}$	-0.194	0.002*	$(\beta_{22}+\beta_{23})X_{1i}$	-0.321	0.003*
	$\beta_{24}(X_{1i}-K_{21})_+$	0.262	<0.001*	$(\beta_{24}+\beta_{25})(X_{1i}-K_{21})_+$	0.454	<0.001*
	$\beta_{26}(X_{1i}-K_{22})_+$	-0.365	0.001*	$(\beta_{26}+\beta_{27})(X_{1i}-K_{22})_+$	0.148	0.066
$X_2 \to Y_2$	$\beta_{28}X_{2i}$	0.122	0.025*	$(\beta_{28}+\beta_{29})X_{2i}$	0.575	<0.001*
	$\gamma_{21}CosX_{2i}$	-0.674	<0.001*	$(\gamma_{21}+\gamma_{22})CosX_{2i}$	-0.287	<0.001*
$Y_1 \to Y_2$	$\beta_{210}Y_{1i}$	-0.496	0.004*	$(\beta_{210}+\beta_{211})Y_{1i}$	-0.670	<0.001*
	$\beta_{212}(Y_{1i}-K_{31})_+$	0.278	0.033*	$(\beta_{212}+\beta_{213})(Y_{1i}-K_{31})_+$	0.535	<0.001*

Notes: *Significant with $\alpha = 0.05$

Table 7 shows that in Cluster 1, 11 function components have p-values < 0.05 and are therefore statistically significant, while one component is not significant (p-value > 0.05). Despite the presence of an insignificant component, the exogenous variables remain significant in the model because each has at least one significant functional component.

Similarly, in Cluster 2, 11 function components are significant (p-value < 0.05), with one component not significant. Nonetheless, the exogenous variables continue to exhibit significant effects for the same reason. Overall, Environmental Quality (X_1) and Use of Waste Banks (X_2) significantly affect the Use of the 3R Principles (Y_1), while Environmental Quality (X_1), Use of Waste Banks (X_2), and Use of the 3R Principles (Y_1) significantly influence the Economic Benefits from Waste (Y_2).

3.7 Discussion

The clustering results reveal clear structural differentiation in waste management behavior, indicating heterogeneous community responses. Cluster 1 exhibits consistently higher centroids for Environmental Quality (X_1), Use of the 3R Principle (Y_1), and Economic Benefits of Waste (Y_2), reflecting an integrated behavioral system in which environmental awareness, behavioral commitment, and economic incentives mutually reinforce each other. In contrast, Cluster 2 shows higher Use of Waste Banks (X_2) but weaker performance in 3R practices and economic outcomes, suggesting that waste bank participation functions more as a transactional activity rather than as part of an internalized pro-environmental behavior system.

From a modeling perspective, this segmentation highlights the importance of incorporating fuzzy clustering, as assuming population homogeneity would mask structurally distinct behavioral regimes and bias parameter estimation. The existence of multiple clusters further implies that transitions toward sustainable practices occur through different behavioral pathways rather than a single linear process.

The relationship between Environmental Quality (X_1) and Use of the 3R Principle (Y_1) exhibits non-monotonic patterns in both clusters, with turning points around 0.22 and 0.59, indicating behavioral thresholds. These thresholds suggest that improvements in environmental conditions initially yield limited behavioral change until a critical level is reached, after which adoption accelerates. The consistently higher trajectory in Cluster 1 indicates stronger behavioral resilience, aligning with environmental behavior theory.

Similarly, the relationship between Environmental Quality (X_1) and Economic Benefits of Waste (Y_2) demonstrates nonlinear dynamics characterized by an initial decline followed by growth, reflecting short-term adjustment costs before economic returns materialize. The divergence between clusters indicates asymmetric benefit realization, with early saturation in Cluster 1 and continued gains in Cluster 2, consistent with differences in behavioral and institutional maturity.

The stepwise relationship between Use of the 3R Principle (Y_1) and Economic Benefits of Waste (Y_2), with a turning point near 0.34, further shows that economic outcomes depend more on behavioral quality than on participation intensity alone. The sharper post-threshold increase in Cluster 1 suggests greater efficiency in converting behavioral commitment into economic value, while the flatter response in Cluster 2 points to structural constraints.

Finally, the fluctuating relationships between Use of Waste Banks (X_2) and both behavioral and economic outcomes indicate that participation mechanisms alone do not guarantee broader environmental engagement. These irregular patterns justify the inclusion of Fourier components to capture oscillatory behavioral responses that are not well represented by linear or spline-only models.

Overall, the integration of fuzzy clustering and hybrid truncated spline–Fourier estimation provides a statistically grounded framework for capturing heterogeneous behavioral structures, offering a more coherent interpretation of the interactions among environmental awareness, participation mechanisms, and economic incentives in waste management behavior.

4. CONCLUSION

This study establishes that waste management behavior is characterized by nonlinear and heterogeneous structures, rendering conventional linear path models and single estimator nonparametric approaches inadequate. Beyond the empirical findings, the primary contribution of this research lies in the development of an integrated hybrid nonparametric path framework that combines truncated spline and Fourier series estimators with Fuzzy C-Means, based dummy variables to explicitly accommodate behavioral heterogeneity. In addition, this study proposes a modified Ramsey RESET test tailored to hybrid nonparametric bases, enabling more comprehensive detection of nonlinear functional forms.

The proposed hybrid estimator demonstrates strong performance, achieving an adjusted total coefficient of determination of 0.891 and a GCV value of 0.0165, confirming its ability to flexibly capture diverse nonlinear patterns across different structural relationships. Methodologically, this result indicates that allowing variable-specific functional bases significantly improves model adaptiveness and reduces structural bias compared to purely linear models and nonparametric models relying on a single estimator.

The fuzzy clustering results further validate the necessity of integrating segmentation into the path structure. Two distinct behavioral regimes were identified: Cluster 1, characterized by high Environmental Quality orientation, strong 3R implementation, and higher economic benefits, and Cluster 2, dominated by intensive waste bank participation with weaker integration of 3R practices. This differentiation illustrates how the proposed framework can uncover latent behavioral structures that would remain hidden in aggregated models.

Although the empirical application is focused on waste management, the methodological contributions are not domain-specific. The hybrid estimator, modified RESET procedure, and fuzzy dummy integration are directly applicable to other fields involving heterogeneous populations and nonlinear relationships, such as public health behavior, educational participation, consumer decision-making, agricultural adoption behavior, and socio-economic policy analysis. This generalizability enhances the theoretical and practical relevance of the proposed framework.

Several limitations remain, including restrictions on spline order, knot selection, and Fourier oscillation levels. Future research may extend this framework by incorporating higher order spline structures, alternative nonparametric bases, adaptive knot selection, and multi-cluster configurations, as well as conducting comparative studies across different application domains to further assess identifiability, consistency, and robustness properties.

ACKNOWLEDGEMENT

This research received no external funding. The authors sincerely thank the Department of Statistics, Brawijaya University, as well as the participating partners in Batu City, for their time, cooperation, and valuable contributions that made this research possible.

5. REFERENCES

[1] A. Istiqamah, T. Noor, and P. Sukowati, “Study of household scale waste management in Batu City, Indonesia: Implementation of Batu mayor’s regulation on household waste management policy,” *International Journal of Advanced Science Research and Engineering*, vol. 8, no. 2, pp. 23–29, 2022, doi: 10.31695/IJASRE.2022.8.2.3

[2] I. Y. Prastya, “Co-production in waste management through waste bank programs in Batu City,” *Journal of Local Government Issues*, vol. 2, no. 2, pp. 149–160, 2024, doi: 10.22219/logos.vol2.no2.149-167.

[3] Y. Zahrah, J. Yu, and X. Liu, “How Indonesia’s cities are grappling with plastic waste: An integrated approach toward sustainable plastic waste management,” *Sustainability*, vol. 16, no. 10, pp. 3921, 2024, doi: 10.3390/su16103921.

[4] H. Kamaruddin, Maskun, F. Patittingi, H. Assidiq, S. N. Bachril, and N. H. A. Mukarramah, “Legal aspect of plastic waste management in Indonesia and Malaysia: Addressing marine plastic debris,” *Sustainability*, vol. 14, no. 12, p. 6985, 2022, doi: 10.3390/su14126985.

[5] E. Sembiring, M. R. Fenitra Rakotoarisoa, A. R. Dangkua, *et al.*, “Improving household waste management in Indonesia: A mixed-methods approach for waste sorting,” *Recycling*, vol. 10, no. 10, p. 100185, 2024, doi: 10.1016/j.clwas.2024.100185.

[6] P. Pongpunpurt, P. Muensitthiroj, and P. Pinitjitsamut, “Studying waste separation behaviors and environmental impacts toward sustainable solid waste management,” *Sustainability*, vol. 14, no. 9, p. 5040, 2022, doi: 10.3390/su14095040.

[7] U. Rohma, A. A. R. Fernandes, S. Astutik, and Solimun, “Development of nonparametric path function using hybrid truncated spline and kernel for modeling waste-to-economic value behavior,” *Barekeng: Jurnal Ilmu Matematika dan Terapan*, vol. 19, no. 1, pp. 331–344, 2025, doi: 10.30598/barekengvol19iss1pp331-344.

[8] A. Budiyarto, C. Beverley, and K. Ross, “Overview of waste bank application in Indonesian regencies,” *Waste Management & Research: The Journal for a Sustainable Circular Economy*, vol. 43, no. 3, pp. 1298–1310, 2025, doi: 10.1177/0734242X241242697.

[9] M. Alaghemandi, “Sustainable solutions through innovative plastic waste recycling technologies,” *Sustainability*, vol. 16, no. 23, p. 10401, 2024, doi: 10.3390/su162310401.

[10] A. Budiyarto, “Enhancing solid waste management system through waste banks: A strategy for increasing household waste recycling in Indonesia,” *Environmental Research Communications*, vol. 4, no. 3, 2023.

[11] M. T. Rahman, S. Ghosh, and P. Singh, “Modeling determinants of waste sorting behavior: A path analysis approach,” *Journal of Environmental Management*, vol. 307, no. 1, p. 114539, 2022, doi: 10.1016/j.jenvman.2022.114539.

[12] H. S. Kwon, J. H. Lee, and S. Y. Kim, “Applying path analysis to examine factors influencing pro-environmental behavior in urban communities,” *Sustainability*, vol. 13, no. 24, p. 13677, 2021, doi: 10.3390/su132413677.

[13] D. A. Adyatama, S. Solimun, S. Efendi, N. Nurjannah, and M. B. T. Mitakda, “Estimated path analysis parameters using weighted least square to overcome heteroskedasticity at various sample sizes,” *Journal of Statistics Applications & Probability*, vol. 12, no. 2, pp. 345–356, 2023, doi: 10.18576/jsap/120223.

[14] Solimun, "Analisis multivariat pemodelan struktural metode partial least square (PLS)". Malang, Indonesia: CV Citra, 2010.

[15] A. Hermawati, “Marketing strategy with path analysis in increasing competitive advantage in tourism industry SMEs in East Java,” *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 20, pp. 5935–5944, 2022.

[16] A. A. R. Fernandes and Solimun, "Analisis regresi dalam pendekatan fleksibel: ilustrasi dengan paket program R". Malang: Universitas Brawijaya Press, 2021.

[17] A. Iriany and A. A. R. Fernandes, “Hybrid Fourier series and smoothing spline path non-parametric modelling,” *Frontiers in Applied Mathematics and Statistics*, vol. 9, pp. 1–10, 2023, doi: 10.3389/fams.2022.1045098.

[18] R. Ruliana, R. Maru, Z. Rais, and A. Saleh Ahmar, “Integration of SEM and nonparametric spline in spatial data modeling and visualization for analysis of CO_2 reduction through green space in Makassar City,” *International Journal on Informatics Visualization*, vol. 9, no. 3, pp. 1319-1329, 2025, doi: 10.62527/joiv.9.3.3807.
[19] M. Fidino, A. Hargrove, M. P. Anderson, and J. L. Martin, “Using Fourier series to estimate periodic patterns in dynamic occupancy models,” *Ecological Applications*, vol. 27, no. 6, pp. 1827–1834, 2017, doi: 10.1002/ecs2.1944.
[20] K. A. Yasa, I. N. Suparta, and A. A. N. G. S. Sapteka, “Pemodelan multiple pulse width modulation berbasis Fourier series untuk jumlah pulsa sebanyak tiga,” *Jurnal Simetrik Politeknik Negeri Madiun*, vol. 14, no. 2, pp. 873–876, 2024, doi: 10.31959/js.v14i2.2669.
[21] M. Yasir, U. K. Zaheer, H. U. Rehman, and S. Z. Ali, “Machine learning–assisted efficient demand forecasting in supply chains: role of exogenous indicators,” *International Journal of Production Research*, vol. 62, no. 21, pp. 6671-6688, 2024, doi: 10.1080/13675567.2022.2100334.
[22] Solimun, A. A. R. Fernandes, and Nurjannah, "Metode statistika multivariat: pemodelan persamaan struktural (SEM) pendekatan WarpPLS". Malang: UB Press, 2017.
[23] A. A. Puspitasari, A. A. R. Fernandes, A. Efendi, S. Astutik, and E. Sumarminingsih, “Development of nonparametric truncated spline at various levels of autocorrelation of longitudinal generating data,” *Journal of Statistical and Applied Probability (J. Stat. Appl. Pro)*, vol. 12, no. 2, pp. 757–766, 2023, doi: 10.18576/jsap/120234.
[24] L. P. Inasari, M. Irwan, W. Abidin, and W. Alwi, “Model regresi nonparametrik dengan pendekatan deret Fourier pada kasus tingkat kemiskinan di Sulawesi Selatan,” *Jurnal MSA (Matematika dan Statistika serta Aplikasinya)*, vol. 12, no. 2, pp. 79–87, 2024, doi: 10.24252/msa.v12i2.51107.
[25] S. P. Aditiyo, E. Sumarminingsih, and R. Fitriani, “Fuzzy C-Means in content-based document clustering for grouping general websites based on their main page contents,” *ComTech: Computer, Mathematics and Engineering Applications*, vol. 14, no. 2, pp. 87–96, 2023, doi: 10.21512/comtech.v14i2.9732.
[26] 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: 10.1016/j.mex.2024.103084.
[27] M. A. Juniar, A. Fania, D. Ulya, R. Ramadhan, and N. Chamidah, “Modelling crime rates in Indonesia using truncated spline estimator,” *Barekeng: Journal of Mathematics and Its Applications*, vol. 18, no. 2, pp. 1201–1216, 2024, doi: 10.30598/barekengvol18iss2pp1201-1216.
[28] N. I. R. Wisisono, A. I. Nurwahidah, and Y. Andriyana, “Regresi nonparametrik dengan pendekatan deret Fourier pada data debit air Sungai Citarum,” *Jurnal Matematika MANTIK*, vol. 4, no. 2, pp. 75–82, 2018, doi: 10.15642/mantik.2018.4.2.75-82.
[29] L. R. Khairunnisa, A. Prahutama, and R. Santoso, “Pemodelan regresi semiparametrik dengan pendekatan deret Fourier (studi kasus: Pengaruh indeks Dow Jones dan BI rate terhadap indeks harga saham gabungan),” *Jurnal Gaussian*, vol. 9, no. 1, pp. 50–63, 2020, doi: 10.14710/j.gauss.14.2.302-313.
[30] M. Zulfadhli, I. N. Budiantara, and V. Ratnasari, “Nonparametric regression estimator of multivariable Fourier series for categorical data,” *MethodsX*, vol. 13, p. 102983, 2024, doi: 10.1016/j.mex.2024.102983.
[31] Y. N. Fuadah, I. D. Ubaidullah, N. Ibrahim, F. F. Taliningsing, N. K. Sy, and M. A. Pramuditho, “Optimasi convolutional neural network dan k-fold cross validation pada sistem klasifikasi glaukoma,” *ELKOMIKA: Jurnal Teknik Energi Elektrik, Teknik Telekomunikasi, dan Teknik Elektronika*, vol. 10, no. 3, p. 728, 2022, doi: 10.26760/elkomika.v10i3.728.
[32] A. Nugroho and A. Amrullah, “Evaluasi kinerja algoritma k-NN menggunakan k-fold cross validation pada data debitur KSP Galih Manunggal,” *JINTEKS: Jurnal Informatika Teknologi dan Sains*, vol. 5, no. 2, pp. 294–300, 2023, doi: 10.51401/jinteks.v5i2.2506.
[33] J. White and S. D. Power, “K-fold cross-validation can significantly over-estimate true classification accuracy in common eeg-based passive bci experimental designs: An empirical investigation,” *Sensors*, vol. 23, no. 13, p. 6077, 2023. doi: 10.3390/s23136077
[34] R. Caro-Carretero, A. Carnicero, J. R. Jiménez-Octavio, and D. Cousineau, “Utilizing jackknife and bootstrap to understand tensile stress to failure of an epoxy resin,” *Quality Engineering*, vol. 36, no. 4, pp. 726–742, 2024. doi: 10.1080/08982112.2023.2286500.