Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2025, 2025:38
https://doi.org/10.28919/cmbn/9168
ISSN: 2052-2541

IMPLEMENTATION OF DIFFERENTIAL EVOLUTION ALGORITHM FOR MARS MODELING OPTIMIZATION ON INDONESIA COMPOSITE INDEX DATA

NUR ISRA, ANNA ISLAMIYATI*, ERNA TRI HERDIANI

Department of Statistics, Faculty of Mathematics and Natural Sciences, Hasanuddin University, Makassar, 90245, Indonesia

Abstract: Nonparametric regression is applied as an alternative to parametric regression if the data to be studied does not meet all the assumptions of parametric regression. This study aims to model using the Multivariate Adaptive Regression Splines model integrated with the Differential Evolution algorithm on the Indonesia Composite Index data. Indonesia Composite Index reflects the overall performance of stocks on the Indonesia Stock Exchange, which is influenced by various macro and microeconomic factors that lead to erratic patterns. Thus, the MARS method was chosen because of its ability to capture nonlinear relationships and interactions among the independent variables. At the same time, the Differential Evolution algorithm was implemented to optimize the selection of model parameters. The MARS model is found by combining Basis Function (BF), Maximum Interaction (MI), and Minimum Observation (MO), with concern to the minimum value of Generalized Cross-Validation (GCV). The study results using MARS-DE indicate that the optimal combination of models is BF = 35, MI = 1, and MO = 2 with a GCV value of 0.0068. In this study, 6 independent variables were used. The study results showed that the monthly exchange rate (USD-IDR), inflation, interest rates in Indonesia, the Dow Jones stock index, and the Shanghai Stock Exchange Composite index are independent variables that affect the Indonesia Composite Index (ICI). However, the Nikkei 225 stock index does not influence the Indonesia Composite Index. The resulting model is based on the

*Corresponding author
E-mail address: annaislamiyati701@gmail.com
Received February 01, 2025

smallest Mean Squared Error value of 0.0034.

Keywords: differential evolution; GCV; IDX composite; multivariate adaptive regression splines (MARS).

2020 AMS Subject Classification: 62G08.

1. INTRODUCTION

Regression is an analytical technique that is often used in various fields to model relationships between variables. There are three approaches to regression: parametric, semiparametric, and nonparametric [1]. Parametric regression is applied when the form of the regression curve is known. In addition to fulfilling the assumption related to the curve's shape, parametric regression must also fulfill the criteria of being the Best Linear Unbiased Estimator [2]. However, not all data used in the research meet all assumptions of parametric regression, so parametric regression methods become less effective.

Nonparametric regression is an alternative to parametric regression because it does not require strict assumptions. This method is used to determine the relationship between the response and predictor variables without knowing the form of the regression curve or the form of its function [3]. Some estimators in nonparametric regression include spline [4], spline logistic [5], local polynomial [6], kernel [7], Fourier Series [8], and MARS [9]. Previous research has shown that nonparametric regression can overcome the problem of the complexity of relationships between variables that parametric regression cannot handle. However, the challenge in nonparametric regression lies in the selection of appropriate parameters, such as the determination of the optimal knot point in spline regression, bandwidth parameters in local polynomial and kernel, as well as the number of oscillations/waves in Fourier Series analysis, which often affect the accuracy of the model.

Multivariate Adaptive Regression Splines (MARS) is one of the most effective approaches in dealing with nonlinear relationships and interactions between variables. The advantage of the MARS method is that the determination of knots can be done automatically using stepwise forward and backward algorithms, which are based on the minimum GCV value, which makes it more adaptive than other methods [10]. MARS is a nonparametric regression development method combining the Recursive Partitioning Regression model and the Spline method [11]. This method results in accurate predictions with a continuous model at knots, which means that the regression lines are always connected, and each knot is connected to its basis function [12].

Although MARS has the advantage of handling nonlinear relationships and interactions between variables, its performance is highly dependent on proper parameter selection, such as the number and location of knots. MARS modeling is conducted through trial and error based on the number of basis functions, maximum interactions, and minimum observations [12]. Therefore, to make it easier to find these combinations, research was carried out to combine the method with other more sophisticated and innovative techniques. One of them is by integrating the Differential Evolution (DE) algorithm. The DE algorithm is one of the most popular evolutionary algorithms with simplicity of operation and excellent performance, and it proved to be effective in finding optimal solutions to non-linear and complex optimization problems [13].

Some studies show that combining MARS and DE results in better prediction models. Application of MARS-DE to predict river water flow and evaluate the model's performance using Mean Absolute Percentage Error (MAPE) [14]. However, using MAPE has a significant drawback, as it can produce infinite or undefined values when the actual value is zero or near zero [15]. So, to overcome the problem in MAPE, a model performance evaluation method using Mean Square Error is proposed. The research using MARS and DE focuses on the hydrological, geotechnical, and environmental sectors, which tend to be more stable. Meanwhile, the stock market has a much more dynamic characteristic, with high volatility and patterns influenced by various macroeconomic factors, market sentiment, and other external factors. Stock price prediction is a challenge because of its complex nature and uncertainty. The Indonesia Composite Index is one of the indices that investors frequently consider when investing in the Indonesia Stock Exchange. The Indonesia Composite Index is influenced by macro aspects such as interest rates, rupiah exchange rates, inflation, and microeconomic elements, which lead to uncertain patterns. In addition, the economic conditions of developed countries have the power to influence the global economy, so stock indices in developed countries, such as the SSEC (China), the Nikkei 225 (Japan), and the Dow Jones Index (America), impact stock indices in developing countries [16]. The relationships between these variables are often non-linear and complex. Thus, the MARS method is the right choice in ICI data analysis because it captures nonlinear relationships and interactions between variables and does not require linearity or normality assumptions that are difficult to meet in analyzing stock data. Therefore, this study aims to apply MARS optimization with DE to model the Indonesia Composite Index. It is hoped that with the use of DE, the MARS model can overcome challenges in parameter optimization and improve prediction accuracy, which is crucial for investors in making investment decisions.

2. METHODOLOGY

2.1 Experimental Dataset

The data used in this research is secondary data obtained from previously published sources. Monthly Indonesia Composite Index data, Dow Jones (DJIA) stock index, Nikkei 225 (N225) stock index, and Shanghai Stock Exchange Composite (SSEC) index (finance.yahoo.com). Exchange rate and inflation data (www.bi.go.id). Monthly interest rate data in Indonesia (www.bps.go.id). The data used is from the Indonesia Composite Index for the period 2017 to 2022.

In this study, there are 72 data with 1 dependent variable (Y), which is the Indonesia Composite Index and 6 independent variables (X), including the monthly Exchange Rate (US Dollar-Rupiah) (X_1), monthly inflation (X_2), monthly Interest Rate in Indonesia (X_3), DJIA (X_4), N225 (X_5), and SSEC (X_6). The following is a graph of the Indonesia Composite Index data for 2017–2022 in Figure 1.

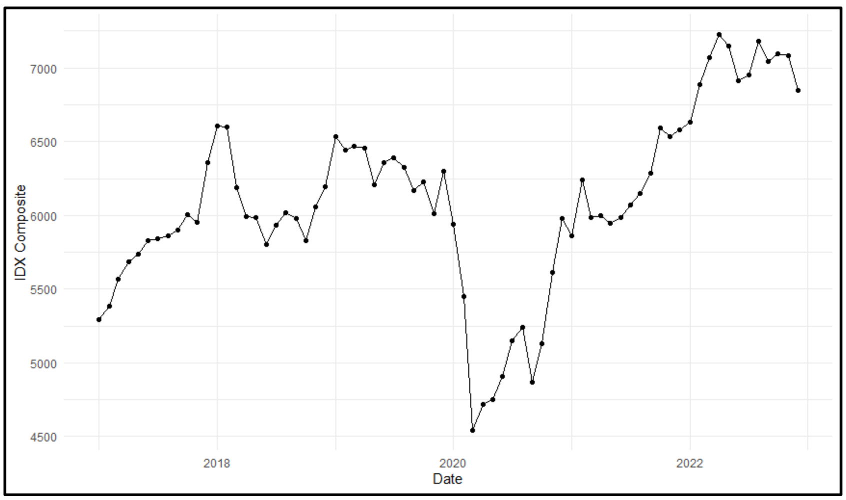

Figure 1. ICI Chart for 2017-2022

From January 2017 to December 2022, the Indonesia Composite Index value fluctuated significantly, as shown in Figure 1. The effects of the Covid-19 pandemic caused a sharp decline in early 2020. In 2021, the ICI peaked in October 2021, indicating a strong economic recovery after the decline at the pandemic's beginning. The year 2022 shows a further increase followed by some fluctuations but remains higher than in previous years. During the observed period, the ICI showed several significant fluctuations, reflecting the market's response to various factors. Factors affecting the ICI are the Exchange Rate, Inflation, Interest Rates, Dow Jones Stock Index, Nikkei 225, and SSEC Index [16].

2.2 Multivariate Adaptive Regression Spline-*Differential Evolution*

The MARS is a non-parametric regression approach that assumes the function shape of the relationship between dependent-independent variables is unknown and has a flexible function form, as proposed by Friedman [17]. The MARS method produces accurate response predictions and overcomes the limitations of Recursive Partition Regression by making a continuous method at the knots [12]. The result of modifying the RPR model with a combination of splines is the estimation of the MARS model [18] as follows:

$$\hat{f}(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} \left[s_{km} . \left(x_{v(k,m)} - t_{km}\right)\right]_+ \quad (1)$$

$$= a_0 + \sum_{m=1}^{M} \alpha_m B_m(x)$$

where a_0 is the coefficient of the parent basis function or constant, a_m is the coefficient for the m-th basis function, M denotes the maximum basis function, K_m indicates the maximum degree of interaction, s_{km} is the sign of the basis function at the k-th interaction and the m-th basis function with a value of ± 1 if the data is located to the right of the knot point and -1 if it is to the left, $x_{v(k,m)}$ corresponds to the v-th independent variable, t_{km} is the knot value of the independent variable $x_{v(k,m)}$, v represents the number of independent variables, m is the number of basis functions and k is the number of interactions.

Equation (1) of the MARS model can be written in matrix form in the following way [19]:

$$y = Ba + \varepsilon \quad (2)$$

where y is the dependent variable, B is the basis function = $\left[1, \left(x_{v(k,m)} - t_{km}\right)_1^K\right]$, a denotes the coefficient of the basis function and ε represents the error of the random variable which is assumed to be independent and normally distributed with a mean of zero and a certain variance σ^2.

with, $y = (y_1, \ldots, y_n)^T$, $a = (a_0, \ldots, a_M)^T$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^T$, and

$$B = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} \left[s_{k1}\left(x_{v1(k,1)} - t_{k1}\right)\right] & \cdots & \prod_{k=1}^{K_M} \left[s_{kM}\left(x_{v1(k,M)} - t_{kM}\right)\right] \\ 1 & \prod_{k=1}^{K_1} \left[s_{k1}\left(x_{v2(k,1)} - t_{k1}\right)\right] & \cdots & \prod_{k=1}^{K_M} \left[s_{kM}\left(x_{v2(k,M)} - t_{kM}\right)\right] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} \left[s_{k1}\left(x_{vn(k,1)} - t_{k1}\right)\right] & \cdots & \prod_{k=1}^{K_M} \left[s_{kM}\left(x_{vn(k,M)} - t_{kM}\right)\right] \end{bmatrix}$$

To obtain the estimator $\hat{\alpha}$ can use the Ordinary Least Square [20], as follows:

$$\min_{\alpha \in R^{M+1}} (\varepsilon' \varepsilon) = \min_{\alpha \in R^{M+1}} \left((y - B\alpha)'(y - B\alpha) \right) \tag{3}$$

The solution to equation (3) is obtained by partially decreasing the result of $\varepsilon'\varepsilon$ to α and equating the result to zero, thus obtaining an estimate for $\{\alpha_i\}_{i=1}^{m}$ as follows:

$$\hat{\alpha} = (B'B)^{-1} B' y \tag{4}$$

Two approaches are used to select the best model for MARS: the first approach is using forward stepwise to obtain the number of basis functions, the selection of BF by minimizing the Average sum of Square Residual [21], and the second approach is the backward stepwise approach to simplify the resulting model by eliminating basis functions that contribute less than the forward stepwise by minimizing the GCV value [22]. The forward procedure will produce a model with many basis functions. In practice, there is a limit to the maximum number of basis functions in the model. The same applies to the degree of interaction, often limited to only three degrees [23].

The selection of the combination between the number of basis functions, maximum interaction, and minimum observation is the basis of MARS modeling. Therefore, the DE algorithm is used to help find the optimal parameters of MARS. The DE algorithm is a mathematical optimization approach for multidimensional functions and belongs to the group of evolutionary algorithms. The Differential Evolution algorithm is one of the best genetic algorithms that improve the shortcomings of other evolutionary algorithms by using a simple optimization strategy for a fast optimization process (fast calculation time with few iterations to find the optimal global solution) [13]. This method is an evolution of the Genetic Algorithm by replacing logical operators with mathematical operators. The following are the processing stages in the DE algorithm for MARS modeling [24] as follows:

a. Define the Fitness Functions

In MARS modeling, the model with the lowest Generalized cross-validation value is the best. So that the fitness function used is GCV, as follows:

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^{N} \left[y_i - \hat{f}_M(x_i) \right]^2}{\left[1 - \frac{\tilde{C}(M)}{N} \right]^2} \tag{5}$$

with $\tilde{C}(M) = C(M) + d.M$, $C(M) = tr[B(B^T B)^{-1} B^T] + 1$, $\hat{f}_M(x_i)$ is the estimated value of the response variable on M-basis functions at x_i, N represents the sample size and $\tilde{C}(M)$ is the complex model function = $C(M) + d.M$, where d denotes the smoothing parameter.

The optimum value is in the range of $2 \leq d \leq 4$, $C(M)$ indicates the number of parameters estimated, M is the number of BF defined, d is the optimal value of the basis function $2 \leq d \leq 4$, y_i is the dependent variable, and x_i is the independent variable [18].

b. Initialization

In this stage, MARS parameters (BF, MI, and MO) are initialized, where random numbers are generated. The parent population (x) will be generated at this stage. The following equation expresses the initialization process:

$$x_{ij} = rand(0,1) \times (U_b - L_b) + L_b \tag{6}$$

The random number is generated based on a uniform distribution in the range (0,1) or in other words, $0 \leq rand < 1$. Where Ub is the upper bound and Lb is the lower bound.

c. Mutation

The next stage is mutation, producing a mutant population (v). Mutant individuals were formed by combining three randomly selected individuals from the initial population with F as the differentiating scale factor. The first individual value is selected once for all individuals in the same population, and the second and third are selected for each individual to be formed. The formation is carried out as much as the population size. The equation formed for the mutant vector is:

$$v_i^{g+1} = x_{r1}{}^g + F \times (x_{r2}{}^g - x_{r3}{}^g) \tag{7}$$

d. Crossover

The individual values of the initial population will be exchanged with the individual values of the mutant population by using a comparison of CR values and random numbers to form the trial population.

$$u_{ij}^{g+1} f(x) = \begin{cases} v_{ij}^{g+1} \ if \ rand(0,1) \leq CR \ or \ j = r_n(i) \\ \quad x_{ij}^{g} \ \ otherwise \end{cases} \tag{8}$$

where j denotes the jth gene of an individual, and $r_n(i)$ is a randomly generated number in the range $[1, D\,]$, used to ensure that at least one dimension of the experimental individuals comes from an individual that has undergone mutation [14]. The probability of crossover, $CR \in (0,1)$ is the value used to control the fraction of variable values copied from mutants. The CR value influences individuals who will experience crossovers. The higher the score, the more individuals will experience crossovers.

e. Selection

At this stage, replication is carried out to determine which vector will survive in the next generation. To determine whether a vector becomes a member of the generation $g+1$, the test vector u_i^{g+1} is compared with the target vector x_i^g. If the vector u_i^{g+1} produces a smaller fitness value than x_i^g then x_i^{g+1} will be set to u_i^{g+1}, and otherwise, the old x_i^g value is retained. The explanation is shown in the following equation:

$$x_i^{g+1} = \begin{cases} u_i^{g+1} \; if \; f(u_i^{g+1}) < f(x_i^g) \\ x_i^g \quad otherwise \end{cases} \tag{9}$$

The iteration is stopped after going through the above steps until the optimal condition is reached [13]. When all criteria are met, the MARS-DE process is terminated. Otherwise, the model will continue to be the next generation.

The criterion used to evaluate model performance is the Mean Square Error value [23], which is calculated using the following equation:

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} \tag{10}$$

Where y_i is the i-th dependent variable, $\hat{y}_i$ is the estimated value of the i-th dependent variable, and n is the amount of data.

3. MAIN RESULTS

3.1 Scatter Plot

This study uses monthly Indonesia Composite Index data from 2017 to 2022. Before conducting the MARS model analysis using the DE algorithm, the relationship pattern between the Indonesia Composite Index and the variables that affect it is identified, and a scatter plot is presented. Then, data normalization is carried out using min-max normalization, which uses minimum and maximum values to balance one data set and another in the same range. The DE algorithm will then use the normalized data in the MARS modeling.

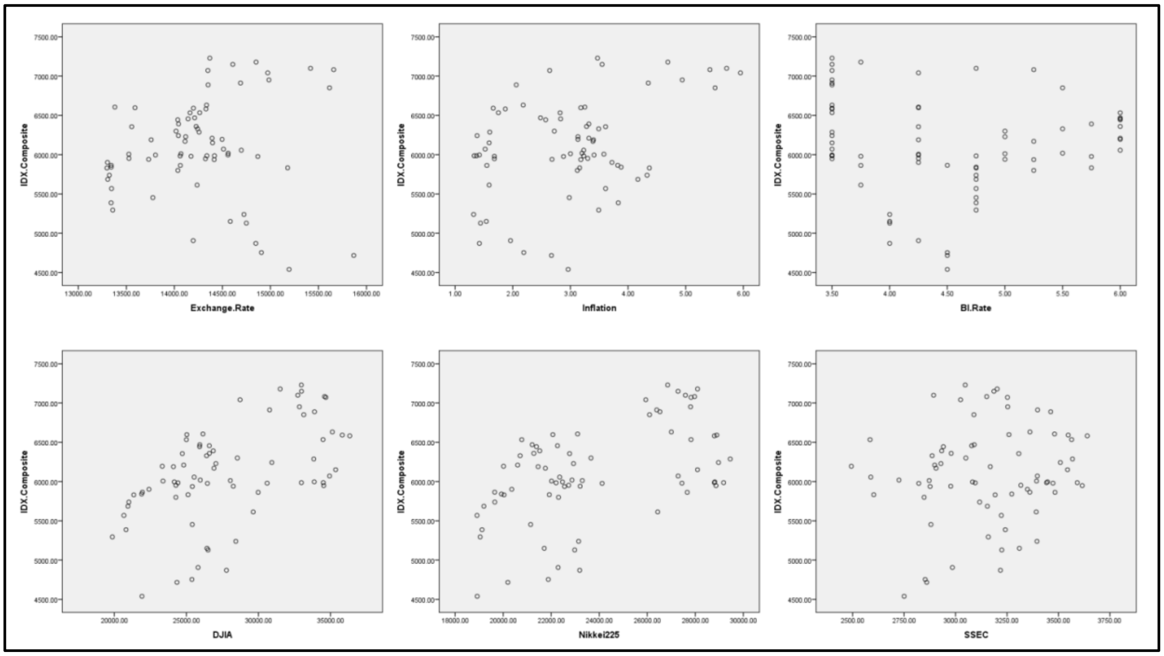

Figure 2. Scatter Plot of Dependent Variables with Independent Variables

The data pattern of the independent variable on the dependent variable does not follow a specific pattern, and the plot does not indicate a clear pattern (Figure 2). Limited information about relationship patterns and changes in regression curve behavior in several variables in the study is one reason for using a nonparametric regression approach, such as the MARS method used in research to model data patterns.

3.2 MARS Modeling with Differential Evolution Algorithm

The initial stage is the initialization process of MARS parameters (BF, MI, and MO); in this case, each member of the initial population is represented as a parameter vector corresponding to the MARS model. This initialization is carried out randomly with the constraints given to each parameter.

Table 1. Search for each MARS parameter in the DE procedure

MARS parameters	Lower bound	Upper bound
Maximum of basis functions	12	60
Maximum Interactions	1	3
Minimum Observations	0	3

Table 1 shows that the lower bound used is (12,1,0), while the upper bound is (60,3,3). Lower and upper bounds determine the minimum and maximum value limits for each parameter of MARS. The differential evolution algorithm has several components for performing optimization.

This research uses R-Studio software in the Differential Evolution algorithm stage. The results of the Differential Evolution algorithm stages are shown in Table 2 as follows:

Table 2. Components of the Differential Evolution Algorithm

DE *Settings*	
Population Size	10×the number of MARS parameters
Crossover Probability	0.5
Scaling Factor	0.8

Table 2, Population Size, aims to find the right population size to obtain optimal results by referring to the results of the best fitness value. The population size generated in this study is 10×the number of MARS parameters with 200 iterations. Crossover probability is a value determined to set the proportion of components copied from the mutant vector [13]. The higher the Cr value, the more points or vectors that crossover. Therefore, the Cr used in this study is 0.5. The mutation factor has a scale $F \in (0, 1)$ of positive real value, which has a function to control the population growth rate. Although there is no upper bound for the F value, the most effective value is between 0 and 1. This Differential Evolution algorithm uses a value of 0.8 as a scaling factor.

In the differential evolution algorithm, fitness value is a value that expresses the goodness of an individual and is a reference for obtaining the optimum value. In the Differential Evolution Algorithm stages, fitness values play a role in determining the results of mutations, crossovers, and selection. Based on the results obtained, the best valid value is 0.0068, which indicates that The Differential Evolution algorithm found a solution with a fitness value of around 0.0068 in each iteration. The smaller the fitness value, the better the solution generated by the Differential Evolution Algorithm. Table 3 shows the fitness value of the individual representation with 200 iterations.

Table 3. Fitness value

Iteration	Fitness value	Best parameters		
1	0.0069	14.7796	2.8697	0.4861
2	0.0069	14.7796	2.8697	0.3665
3	0.0069	14.7796	2.3307	0.3665
4	0.0069	48.8520	1.3806	2.3820
5	0.0069	48.8520	1.3806	2.3820

6	0.0069	48.6095	1.2359	2.4295
7	0.0068	25.3797	1.4304	2.6289
8	0.0068	32.1418	1.2790	2.8697
⋮	⋮	⋮	⋮	⋮
200	0.0068	35.4219	1.4551	2.5210

Based on Table 3, the fitness value from the first iteration was 0.0069; then in the 200th iteration, the fitness value became 0.0068; this shows that the fitness function in the analysis process aims to minimize the GCV value in each generation. In the process of iteration of the differential evolution algorithm, the fitness value continues to be optimized until the best value is obtained. Thus, it can be concluded that the differential evolution algorithm performs exploration to find the optimal objective function value. The differential evolution algorithm runs and tries to get the best solution for the given objective function with predefined constraints. Each iteration produces the best value found in that iteration, as well as the parameter values that correspond to that best value. This process continues until the defined termination criteria or the maximum iteration limit is reached. The optimal value is obtained at the maximum iteration limitation, namely at iteration 200, with a fitness value of 0.0068 with a solution vector of [35, 1, 2]. The following combinations are obtained by searching for combinations of BF, MI, and MO of MARS modeling using the differential evolution algorithm in Table 4.

Table 4. Search with DE for each MARS parameter

MARS hyperparameters	Lower limit	Upper limit	MARS-DE
Basis Function	12	60	35
Maximum Interactions	1	3	1
Minimum Observations	0	3	2
GCV		0.0068	

Using the Differential Evolution algorithm to find MARS parameters based on the results in Table 4, the minimum GCV value of 0.0068 is obtained with the combination of BF, MI, and MO, respectively 35, 1, and 2. This means that the model generated from the combination is the best for determining the model on the Indonesia Composite Index data.

In the MARS model using the DE Algorithm, 200 iterations were performed. Iteration begins by iteratively updating the parameters and recording the best value found. The best fitness value obtained is 0.006763. The solution that fulfills the MARS model uses the Differential Evolution Algorithm in the following equation:

$$Y = 0.124 - 0.422BF_1 + 2.181BF_2 - 1.656BF_3 - 0.992BF_4 + 1.437BF_5 - 0.946BF_6 + 0.614BF_7 + 0.222BF_8 + 3.931BF_9 - 4.976BF_{10}$$

where $BF_1 = h(0, X_1 - 0.2$, $BF_2 = h(0, X_2 - 0.073), BF_3 = h(0, X_2 - 0.251), BF_4 = h(0, X_3 - 0.200)$, $BF_5 = h(0, X_3 - 0.400)$, $BF_6 = h(0, 0.261 - X_4)$, $BF_7 = h(0, X_4 - 0.261)$, $BF_8 = h(0, X_6 - 0.330)$, $BF_9 = h(0, X_6 - 0.827)$, and $BF_{10} = h(0, X_6 - 0.855)$.

Based on the modeling results, the independent variables that influence the dependent variable in the MARS model, based on the smallest GCV value, are the Exchange Rate (X_1), Inflation (X_2), Interest Rate (X_3), Dow Jones stock index (X_4), and SSEC index (X_6). While the Nikkei 225 (N225) index (X_5) does not influence the Indonesia Composite Index, so the independent variable of the Nikkei 225 stock index (X_5) is not included in the best model equation obtained. Meanwhile, 10 basis functions contribute to the formation of the model, namely BF1, BF2, BF3, BF4, BF5, BF6, BF7, BF8, BF9, and BF10. In the model, there is no interaction between the independent variables that affect the dependent variable.

3.3 Evaluation of MARS-Differential Evolution Modeling

Model evaluation aims to assess how close the obtained model is to the actual size. The MSE value is used in the model evaluation process.

Table 5. MSE values of MARS-DE models

BF	MI	MO	GCV	MSE
35	1	2	0.0068	0.0034

Based on the combination of MARS parameters (BF, MI, and MO), the optimal model is obtained for each combination, and the optimal model is selected based on the criteria of the smallest GCV value. Table 5 shows the MSE value for the MARS model using the Differential Evolution Algorithm, where the model produces a MSE value of 0.0034. MSE is an evaluation metric that measures how much difference there is between the actual test data values and the model's predicted values. The lower the Mean Square Error value, the better the model's performance; this low MSE indicates that the model has captured the pattern in the data well.

To calculate the estimation results of the Indonesia Composite Index data using the MARS method with the DE algorithm, the best model that has been obtained is used, namely:

$$\hat{Y}_i = 0.124 - 0.422 * h(0, X_1 - 0.287) + 2.181 * h(0, X_2 - 0.073) - 1.656 * h(0, X_2 - 0.251) - 0.992 * h(0, X_3 - 0.200) + 1.437 * h(0, X_3 - 0.400) - 0.946 * h(0, 0.261 - X_4) + 0.614 * h(0, X_4 - 0.261) + 0.222 * h(0, X_6 - 0.330) + 3.931 * h(0, X_6 - 0.827) - 4.976 * h(0, X_6 - 0.855)$$

From the best model that has been obtained, the calculation of the $\hat{Y}_i$ estimator is carried out where $i = 1, 2, \dots, 72$.

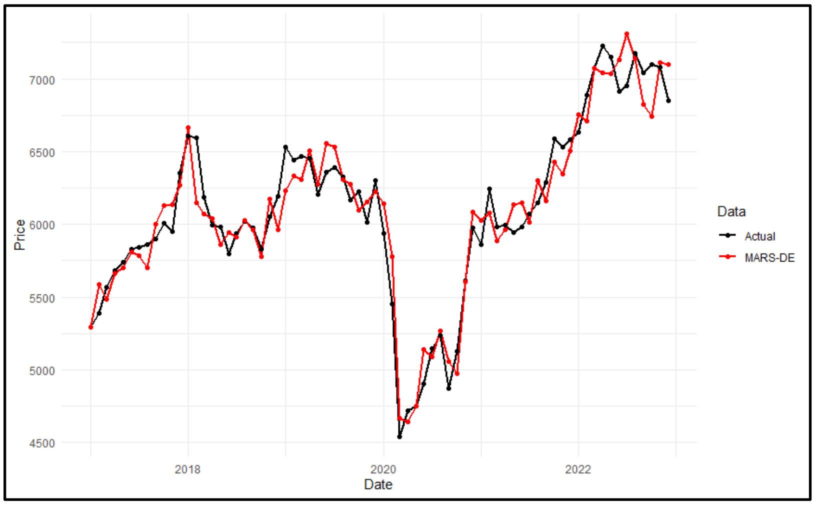

Figure 3. Plot of Actual Data and MARS-DE Estimation

Figure 3 shows that the actual and predicted values of the Indonesia Composite Index with MARS-DE are not much different. The black line in the figure is the actual Indonesia Composite Index, and the red line is the estimation result using MARS-DE. There are several Indonesia Composite Index that are suspected to have a considerable difference in capturing the actual Indonesia Composite Index, such as in February 2018, December 2018, January 2019, January 2020, February 2020, June 2020, June 2022, July 2022, September 2022, October 2022, December 2022. The model is able to predict the actual Indonesia Composite Index well in January 2017, March to July 2017, December 2017, January 2018, April 2018, July to October 2018, April to May 2019, August 2019, December 2019, April to May 2020, July to August 2020, November 2020, April 2021, July 2021, December 2021, March 2022, August 2022, November 2022 with a smaller difference to the actual ICI data. This indicates that the estimated value of the MARS-DE model tends to be close to the actual value.

4. CONCLUSIONS

According to the results and discussion, it is concluded that the optimal model using

MARS-DE is the combination of BF = 35, MI = 1, and MO = 2, where the GCV is 0.0068. In addition, the independent components that exerted significant influence were the Exchange Rate (X_1), Inflation (X_2), Interest Rate (X_3), the Dow Jones Stock Index (X_4), and the SSEC Index (X_6). Meanwhile, the Nikkei 225 index (X_5) does not influence the Indonesia Composite Index. By integrating the DE algorithm, the MARS model produced a good model in forecasting the actual ICI with the MSE value of 0.0034. Therefore, investors can consider using the MARS model with the DE algorithm as a guide in making investment decisions while conducting a thorough analysis before making a final decision.

CONFLICT OF INTERESTS

The author(s) declared that there is no conflict of interest.

REFERENCES

[1] H. Husain, I.N. Budiantara, I. Zain, Mixed Estimator of Spline Truncated, Fourier Series, and Kernel in Biresponse Semiparametric Regression Model, IOP Conf. Ser.: Earth Environ. Sci. 880 (2021), 012046. https://doi.org/10.1088/1755-1315/880/1/012046.

[2] M. Ramli, I.N. Budiantara, V. Ratnasari, A Method for Parameter Hypothesis Testing in Nonparametric Regression with Fourier Series Approach, MethodsX 11 (2023), 102468. https://doi.org/10.1016/j.mex.2023.102468.

[3] A. Islamiyati, Raupong, A. Kalondeng, et al. Estimating the Confidence Interval of the Regression Coefficient of the Blood Sugar Model through a Multivariable Linear Spline with Known Variance, Stat. Transit. New Ser. 23 (2022), 201–212. https://doi.org/10.2478/stattrans-2022-0012.

[4] A. Islamiyati, Spline Longitudinal Multi-Response Model for the Detection of Lifestyle-Based Changes in Blood Glucose of Diabetic Patients, Curr. Diabetes Rev. 18 (2022), e171121197990. https://doi.org/10.2174/1573399818666211117113856.

[5] A. Islamiyati, Anisa, M. Zakir, et al. The Use of the Binary Spline Logistic Regression Model on the Nutritional Status Data Of Children, Commun. Math. Biol. Neurosci. 2023 (2023), 37. https://doi.org/10.28919/cmbn/7935.

[6] M.D. Cattaneo, M. Jansson, X. Ma, Simple Local Polynomial Density Estimators, J. Amer. Stat. Assoc. 115 (2020), 1449–1455. https://doi.org/10.1080/01621459.2019.1635480.

[7] T.H. Ali, H.A.A.M. Hayawi, D.S.I. Botani, Estimation of the Bandwidth Parameter in Nadaraya-Watson Kernel Non-Parametric Regression Based on Universal Threshold Level, Commun. Stat. - Simul. Comput. 52 (2023), 1476–1489. https://doi.org/10.1080/03610918.2021.1884719.

[8] M.A.D. Octavanny, I.N. Budiantara, H. Kuswanto, et al. Modeling of Children Ever Born in Indonesia Using Fourier Series Nonparametric Regression, J. Phys.: Conf. Ser. 1752 (2021), 012019. https://doi.org/10.1088/1742-6596/1752/1/012019.

[9] Z. Qayyum, T. Mehmood, L.A. Al-Essa, Prediction of Inhibition Activity of Dihydrofolate Reductase Inhibitors With Multivariate Adaptive Regression Splines, IEEE Access 11 (2023), 50595–50604. https://doi.org/10.1109/ACCESS.2023.3272231.

[10] M.S. Abed, F.J. Kadhim, J.K. Almusawi, et al. Utilizing Multivariate Adaptive Regression Splines (MARS) for Precise Estimation of Soil Compaction Parameters, Appl. Sci. 13 (2023), 11634. https://doi.org/10.3390/app132111634.

[11] S.D.P. Yasmirullah, B.W. Otok, J.D.T. Purnomo, et al. Multivariate Adaptive Regression Spline (MARS) Methods with Application to Multi Drug-Resistant Tuberculosis (MDR-TB) Prevalence, AIP Conf. Proc. 2329 (2021), 060019. https://doi.org/10.1063/5.0042145.

[12] B.W. Otok, R.Y. Putra, Sutikno, et al. Bootstrap Aggregating Multivariate Adaptive Regression Spline for Observational Studies in Diabetes Cases, Syst. Rev. Pharm. 11 (2020), 406–413.

[13] G. Sun, C. Li, L. Deng, An Adaptive Regeneration Framework Based on Search Space Adjustment for Differential Evolution, Neural Comput. Appl. 33 (2021), 9503–9519. https://doi.org/10.1007/s00521-021-05708-1.

[14] Z.A. Al-Sudani, S.Q. Salih, A. Sharafati, et al. Development of Multivariate Adaptive Regression Spline Integrated with Differential Evolution Model for Streamflow Simulation, J. Hydrol. 573 (2019), 1–12. https://doi.org/10.1016/j.jhydrol.2019.03.004.

[15] C. Banciu, A. Florea, R. Bogdan, Monitoring and Predicting Air Quality with IoT Devices, Processes 12 (2024), 1961. https://doi.org/10.3390/pr12091961.

[16] A. Lesmana, The Effect of the World Exchange Indexes (DJIA, SSEC and NIKKEI 225) on the Indonesia Stock Exchange Index (IDX), J. Asian Multicult. Res. Econ. Manag. Study 3 (2022), 1–9. https://doi.org/10.47616/jamrems.v3i1.230.

[17] M.B Adiguzel, M.A. Cengiz, Model Selection in Multivariate Adaptive Regressions Splines (MARS) Using Alternative Information Criteria, Heliyon 9 (2023), e19964. https://doi.org/10.1016/j.heliyon.2023.e19964.

[18] M. Meilisa, B.W. Otok, J.D.T. Purnomo, Estimation Curve of Multivariate Adaptive Biresponse Fuzzy Clustering Means Regression Splines Approach to Stunting and Wasting Cases in Southeast Sulawesi, MethodsX 12 (2024), 102775. https://doi.org/10.1016/j.mex.2024.102775.

[19] S.D.P. Yasmirullah, B.W. Otok, J.D.T. Purnomo, et al. Modification of Multivariate Adaptive Regression Spline (MARS), J. Phys.: Conf. Ser. 1863 (2021), 012078. https://doi.org/10.1088/1742-6596/1863/1/012078.

[20] S. Zurimi, Darwin, Analysis of Multivariate Adaptive Regression Spline (MARS) Model in Classifying Factors Affecting on Student the Study Period at FKIP Darussalam University of Ambon, J. Phys.: Conf. Ser. 1463 (2020), 012005. https://doi.org/10.1088/1742-6596/1463/1/012005.

[21] B. Triandi, S. Efendi, H. Mawengkang, Regression-Based Analytical Approach for Speech Emotion Prediction Based on Multivariate Additive Regression Spline (MARS), Int. J. Adv. Sci. Eng. Inf. Technol. 13 (2023), 2213–2218. https://doi.org/10.18517/ijaseit.13.6.18603.

[22] H. Venkateswarlu, S. Sharma, A. Hegde, Performance of Genetic Programming and Multivariate Adaptive Regression Spline Models to Predict Vibration Response of Geocell Reinforced Soil Bed: A Comparative Study, Int. J. Geosynth. Ground Eng. 7 (2021), 63. https://doi.org/10.1007/s40891-021-00306-6.

[23] D. Devianto, P. Permathasari, M. Yollanda, et al. The Model of Artificial Neural Network and Nonparametric MARS Regression for Indonesian Composite Index, IOP Conf. Ser.: Mater. Sci. Eng. 846 (2020), 012007. https://doi.org/10.1088/1757-899X/846/1/012007.

[24] P.J. García-Nieto, E. García-Gonzalo, J.P. Paredes-Sánchez, et al. Modelling Energy Performance Using a New Hybrid DE/MARS-Based Approach for Fossil-Fuel Thermal Power Stations, Environ. Sci. Pollut. Res. 28 (2021), 4417–4429. https://doi.org/10.1007/s11356-020-10725-z.