International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 8 Issue 6 June - 2024, Pages: 72-80

Modeling of Human Development Index (HDI) in Indonesia Using Multivariate Adaptive Regression Spline Method

M. Aqil Zaidan Kamil[1], Ardi Kurniawan[2*], Dita Amelia[3]

[1,2,3,4]Department of Mathematics, Faculty of Science and Technology, Campus C, Airlangga University, Surabaya, Indonesia

Corresponding Author: [2]ardi-k@fst.unair.ac.id

Abstract: *Human development is one of the important aspects for the development of a country. Indonesia has a human development index value of 71.92, which is in the upper middle category in the standard category set by the Human Development Index. Thus, Indonesia still has to improve its human development index because it is still not in the high category and is still inferior to other countries. In the sustainable development goals, the Human Development Index is included in the 3rd and 4th goals, namely a healthy and prosperous life and quality education. The purpose of this study focuses on describing and modeling the Human Development Index in Indonesia, as well as interpreting the results of the best model obtained. The method used is a method with a nonparametric regression approach, namely Multivariate Adaptive Regression Spline. The results showed that the best model obtained was in a combination of 14 basis functions, a maximum interaction of 1, and a minimum observation between knots of 1. From this model, the highest level of importance was obtained for the health complaint variable with an importance level of 100%. The best MARS model produced a Generalized Cross Validation value of; 8,736* R^2 *of 0.688 and a Mean Square Error of 5,402. The research is expected to be a consideration for efforts to improve the Human Development Index and support the Sustainable Development Goals (SDGs) action plan.*

Keywords : Human Development Index, Spline, Modeling

1. INTRODUCTION (*Heading 1*)

Indonesia is a developing country that has the desire to become a developed country in the world. In 2013 to 2020, Indonesia experienced a population growth of 1.17% (population census) and in 2023 Indonesia was ranked 4th most populous in the world with a total of 277 million people (BPS, 2023). The Human Development Index is a combination of the health index, expenditure index and education index. The human development index can also be used as an indicator of the status of a country's progress (Idris, 2014). The HDI is used as an indicator to assess the quality aspects of development and to classify whether a country is a developed country, a developing country, or an underdeveloped country (Mahuze, 2022).

The human development index (HDI) or commonly known as the Human Development Index (HDI) was first introduced by the United Nations Development Program (UNDP) in 1990 with the aim of emphasizing that society and their abilities should be the main criteria for assessing a country's development, not just economic growth. The Human Development Index according to is a summary of the average achievement measures in the main dimensions of human development: long and healthy life, broad knowledge and a decent standard of living (UMDP, 1990). The concept of human development basically includes a very broad definition. The human development index is important to discuss because the welfare of society is not only seen from how much economic income a country has, but also how good the quality of human resources the country has for the development process. Meanwhile, if the quality of a country's human resources is poor, it will hinder the HDI development process in accordance with the goals of sustainable development (SDGs), the Human Development Index in accordance with goals 3 and 4, namely a decent and prosperous life and quality education. The HDI is able to describe how the population can access development results in obtaining income, health, education, and so on (BPS, 2019).

One method that can be used to analyze the variables that affect the Human Development Index is regression analysis. Regression can be divided into parametric regression and nonparametric regression. The parametric approach is carried out to assume the form of an existing model and the nonparametric approach does not depend on a particular curve. Since nonparametric regression does not rely on a particular curve or assumption, it can provide greater flexibility (Budiantara, 2006). Multivariate Adaptive Regression Spline (MARS) is a nonparametric regression model that does not assume a functional relationship between the response variable and the predictor variable. MARS was first introduced by Friedman in 1991. The advantage of the MARS method is that it can see the interaction between predictor variables (Friedman, 1991). MARS is a nonparametric method that focuses on overcoming high-dimensional problems in predictor variables so as to produce accurate response variable predictions.

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 8 Issue 6 June - 2024, Pages: 72-80

From the description above, a study was conducted on modeling the level of human development index (HDI) in Indonesia using the Multivariate Adaptive Regression Spline (MARS) method approach. The novelty of this study lies in the research being conducted for all of Indonesia, not just in one region or province. This study has the benefit of providing suggestions to related parties in the form of what factors need to be improved in order to increase the Human Development Index.

2. LITERATURE REVIEW

2.1 Human Development Index (HDI)

The United Nations Development Programme (UNDP) explains that the Human Development Index (HDI) can be interpreted as a summary of the average achievements in the main dimensions of human development such as: long life and healthy body, having broad knowledge and having a decent standard of living. The HDI was created with the aim of emphasizing that society and their capabilities should be the main criteria for assessing a country's development, not just based on economic growth alone. The HDI is the geometric average of the normalized index for each of the three dimensions. In this case, the dimensions in question are health, education and economy.

Based on the Human Development concept developed by the UN (United Nations), it ranks human development performance on a scale of 0.0 – 100.0 with the following categories:

a. High: HDI more than 80.0

b. Upper Middle: HDI between 66.0 – 79.9

c. Lower Middle: HDI between 50.0 – 65.9

d. Low: HDI less than 50.0

2.2 Pearson Corelation

The commonly used correlation test is Pearson Correlation. Pearson Correlation is a simple correlation that only involves one response variable and one predictor variable. The hypothesis for the Pearson Correlation test is as follows.

H_0 : There is no linear relationship between the response variable and the predictor variable

H_1: There is a linear relationship between the response variable and the predictor variable

The following is the Pearson correlation calculation formula (Sudjana, 1996).

$$r_{xy} = \frac{n\sum_{i=1}^{n} X_iY_i - (\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{\sqrt{n(\sum_{i=1}^{n} X_i^2) - (\sum_{i=1}^{n}(X_i)^2)(n\sum_{i=1}^{n}(Y_i)^2) - (\sum_{i=1}^{n}(Y_i)^2)}}$$

The critical area in correlation testing using Pearson correlation is obtained by comparing the p-value with the significance level (α). If the p-value is less than alpha (α), then the decision is to reject the null hypothesis (H_0), so that it is concluded that there is a relationship between the response variable and the predictor variable.

2.3 Pearson Corelatio

According to Friedman (1991), states that spline is a piecewise polynomial of order q and has continuous derivatives with knots up to order $q-1$. Simply put, spline is a modified result of polynomial regression (Tripena, 2011). In univariate spline with K knot has the following basis functions.

$$1, \{x^j\}_1^q, \{(x-t_k)_+^q\}_1^K$$

The spline regression model can be written as follows (Wicaksono et al., 2014)

$$f(x) = \beta_0 + \beta_1 x + \cdots + \beta_q x^q + \sum_{k=1}^{K} \gamma_k (x - t_k)_+^q$$

$$(x - t_k)_+^q = \begin{cases} (\mathrm{x} - \mathrm{t_k})^{\mathrm{q}}; \ \mathrm{x} - \mathrm{t_k} > 0 \\ 0; \ \mathrm{x} - \mathrm{tk} \le 0 \end{cases}$$

2.4 Multivariate Adaptive Regression Spline

The Multivariate Adaptive Regression Spline method was first introduced by Friedman in 1991. The MARS method is a complex combination of the spline method with Recursive Partitioning Regression (RPR) to produce continuous regression function estimates (Shafana & Gunawan, 2021). Selecting the best model in MARS is done by combining basis functions (BF), maximum interaction (MI), and minimum observations between knots (MO). The criteria for selecting the best model that is the main priority is the minimum Generalized Cross Validation (GCV) value (Wicaksono et al, 2014). If there are two or more models that have the same minimum GCV value, it can be seen from the maximum R^2 value and the minimum Mean Square Error (MSE) value. The general MARS model according to Friedman (1991) is as follows.

$$f(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [S_{km}(x_{v(k,m)} - t_{v,k,m})]$$

$$f(x) = a_0 + \sum_{K_m=1} f_i(x_i) + \sum_{K_m=2} f_{ij}(x_i, x_j) + \sum_{K_m=3} f_{ijk}(x_i, x_j, x_k) + \cdots + \varepsilon$$

Model selection in MARS uses forward stepwise and backward stepwise algorithms. The first model selection is using the forward stepwise stage which is carried out to obtain the maximum number of basis functions. Furthermore, to fulfill the concept of parsimony, the backward stepwise stage is carried out. The backward stepwise stage aims to select the basis functions produced by the forward stepwise stage by minimizing the Generalized Cross Validation (GCV) value (Lembang et al., 2019). The form of Generalized Cross Validation (GCV) as a criterion for selecting the best model is as follows (Pintowati and Otok, 2012).

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

2.5 Simultaneous Testing

Coefficient testing carried out simultaneously on the basis function in the MARS model can use the F test statistic or Fisher Test (Risambessy et al., 2022). The hypotheses used in simultaneous testing are as follows.

$H_0: a_1 = a_2 = \cdots = a_M = 0$

H_1 : There must be at least one $a_m \neq 0; m = 1,2, \dots, M$

$$F_{hitung} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})/M}{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 /N-M-1}$$

2.6 Partial Testing

Partial Testing Partial coefficient testing of the basis function in the MARS model can use the t-test statistic (Risambessy et al., 2022). The hypotheses used in partial testing are as follows.

$H_0: a_m = 0; m = 1,2, \dots, M$

$H_1: a_m \neq 0; m = 0,1,2, \dots, M$

$$t_{count} = \frac{\hat{a}_m}{Se(\hat{a}_m)}$$

$$Se(\hat{a}_m) = \sqrt{var(\hat{a}_m)}$$

The critical region is obtained by comparing the t_{count} value with $t_{\frac{\alpha}{2}(N-M)}$ or comparing the p-value with the significance level (α). If $|t_{count}| > t_{\frac{\alpha}{2}(N-M)}$ or the p-value is less than alpha (α), then the decision is to reject the null hypothesis (H_0) .

2.7 Normality test

A good regression model is one that has normally distributed residual values. Whether the residual value is normally distributed or not can be known through the normality test (Mardiatmoko, 2020). The normality test aims to determine whether the collected data is normally distributed or taken from a normal population (Fahmeyzan et al., 2018). Currently, there are various ways to test normality. One of the normality tests is the Kolmogorov-Smirnov test.

In the Kolmogorov-Smirnov test, the hypothesis proposed is as follows (Usmadi, 2020).

H_0: Residual data is normally distributed

H_1 : Residual data is not normally distributed

The steps of the Kolmogorov-Smirnov test are as follows.

1. Determine the mean and standard deviation of the data
2. Arrange the data starting from the smallest followed by the respective frequencies, cumulative frequencies (F) of each score. The Z value is determined by the formula;
$$Z\ skor = \frac{X - \bar{X}}{\sigma}$$
3. Determine the probability below the Z value that can be seen in the Z(P ≤ Z) table
4. Determine the difference value of each row $\frac{F}{n} = F_z$ with $P \leq Z\ (a_2)$ and the difference of each $\frac{F}{n}$ with $a_2\ (a_1)$
5. Then compare the highest value of a_1 with the Kolmogorov-Smirnov table
6. Then the test criteria are:
   Accept H_0 if $a_1\max \leq D_{table}$
   Reject H_0 if $a_1\max > D_{table}$

2.8 Glejser test

The Glejser test is a method of testing heteroscedasticity. Heteroscedasticity is a condition where there is inequality of residual variance for all observations in the regression model. The test is done by regressing the independent variables against the absolute value of the residual. The residual is the difference between the value of the Y variable and the predicted value of the Y variable, and the absolute is its absolute value (all positive). If the significance value between the predictor variable and the absolute residual is > 0.05 then there is no heteroscedasticity. The hypothesis for the Glejser test is as follows.

$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2 = \sigma^2$

H_1: There is at least one $\sigma_i^2 \neq \sigma^2; i = 1, 2, \dots, n$

$$F_{hitung} = \frac{\sum_{i=1}^{n}((|\hat{e}_i| - |\bar{e}|)^2)/(v-1)}{\sum_{i=1}^{n}((|e_i| - |\hat{e}_i|)^2)/(n-v)}$$

The critical area in testing heteroscedasticity using the Glejser test is obtained by comparing F_{count} with $F_{(\alpha,v-1,n-v)}$. In addition, you can use the p-value which is compared to the significance level (α). If F_{count} is less than $F_{(\alpha,v-1,n-v)}$ or the p-value is more than alpha (α), then the decision is to fail to reject the null hypothesis (H_0), so it is concluded that there are no symptoms of heteroscedasticity. Likewise, vice versa, if F_{count} is more than $F_{(\alpha,v-1,n-v)}$ or the p-value is less than alpha (α), then the decision is to reject the null hypothesis (H_0) so it is concluded that there are symptoms of heteroscedasticity. (Mardiatmoko, 2020)

2.9 Sofware MARS

MARS software is a statistical application or program that can be used to model data using the Multivariate Adaptive Regression Spline (MARS) method. MARS software is software focused on overcoming high-dimensional problems in predictor variables to produce accurate response variable predictions. MARS software is made with the Salford system by an American company, Minitab Inc. The advantage of this

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 8 Issue 6 June - 2024, Pages: 72-80

software is that it is able to see non-linear curve shapes and interactions that occur in the model (Mintab, 2021). MARS software has an easy-to-use Graphical User Interface (GUI), thus helping users control the variables, functional forms, and interactions that will be used. In addition, MARS software can also construct best-fitting models, either gradually or directly. The procedure in this software is to use the backward and forward methods. The results of modeling using MARS software can display Generalized Cross Validation (GCV), r-square, and Mean Square Error (MSE) values. The priority criteria in determining the best model is the minimum Generalized Cross Validation (GCV) value because it is more conservative and reliable than the r-square value (Steinberg, 1999).

3. RESULT AND DISCUSSION

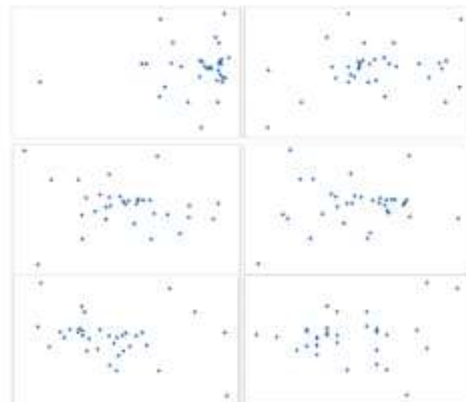
3.1 DESCRIPTIVE STATISTICS

Descriptive statistics is the initial step in data analysis that aims to describe and explain the research object in general. The use of descriptive statistics includes various forms of data presentation such as bar charts, line charts, pie charts, histograms, polygons, and scatterplots, which are selected according to the needs of the researcher and the purpose of the analysis.

The highest human development index value is in DKI Jakarta province with a value of 81.65 which is classified as a very high HDI value. This condition is because DKI Jakarta is the capital city where all types of development are centered in the area starting from the economy, education and health. The condition of the HDI itself is influenced by economic growth and DKI Jakarta is the province with the highest average growth rate in Indonesia (Lumbantoruan & Hidayat, 2014). Meanwhile, the lowest HDI value is in Papua province with a value of 61.9, this is in accordance with the conditions in Papua where the level of poverty, GRDP, government spending on health, and government spending on education affect the HDI in Papua Province, making Papua the province with the lowest HDI value in Indonesia (Suhendi & Astuti, 2023) and for the average Indonesia gets a Human Development Index value of 71.96 where this value is included in the good category according to the United Nations Development Program (UMDP).

Variabel	Nama Variabel	Skala Data
Y	Human development index (IPM)	Rasio
X_1	Literacy rate (persen)	Rasio
X_2	Regional income and expenditure figures for education (persen)	Rasio
X_3	Unmet need for health services (Persen)	Rasio
X_4	health complaints (persen)	Rasio
X_5	Labor Force Participation Rate (persen)	Rasio
X_6	Gini Ratio (Persen)	Rasio

Scatterplot and Pearson Correlation Test between Response Variable and Each Predictor Variable



Based on Figure , it can be seen that there is no particular data distribution pattern (trend) between the response variable of the human development index and all predictor variables. Therefore, these variables can be analyzed using a nonparametric regression approach.

Variabel	*Pearson Correlation*	*Sig. (2-tailed)*	Decision
Y and X_1	0.093	0.599	Significant
Y and X_2	0.22	0.211	Significant
Y and X_3	-0.166	0.348	Significant
Y and X_4	0.101	0.57	Significant
Y and X_5	-0.301	0.083	Significant
Y and X_6	0.157	0.374	Significant

The nonparametric approach is further strengthened by the results of the Pearson correlation test between the response variable Human Development Index and the predictor variable School Participation Rate. It is known that the p-value is more than alpha (0.05). So the decision to fail to reject H_0 is obtained. This indicates that the response variable with the predictor variable does not form a linear model or a certain pattern.

3.2 MARS Model Parameter Estimation

1. Combination of basis functions

   The calculation is done with the help of the MARS application by combining the Basis Function (BF), Maximum Interaction (MI), and Minimum

Observation between knots (MO). The Basis Function (BF) value is 12 to 24, Maximum interaction (MI) is 1 and 2 and Minimum Observation between knots (MO) is 0, 1, 2, and 3. The criteria for the best model is the lowest Generalized Cross Validation (GCV) value. the best model produced is at the value of the fourteen basis function. The results of the combination of the fourteen basis functions MI 1 and MO 1 with a Generalized Cross Validation (GCV) value of 8,736, the R^2 value of 0.688 and the Mean Square Error (MSE) of 5,402 from the combination results. The R^2 value of 0.866 means that the diversity of the response variable values of the Human Development Index (Y) that can be explained by the predictor variable (X) is 68.8 percent.

2. Estimation of basis functions

Fungsi Basis (BF)	Estimasi Parameter
$BF_2 = \max(0, 16.760 - X4)$	-3.062
$BF_4 = \max(0, 4.140 - X3)$	3.549
$BF_5 = \max(0, X6 - 0.418)$	322.921
$BF_8 = \max(0, 73.660 - X5)$	0.334

Based on the following table, the MARS model for estimating the Human Development Index in Indonesia is

$$\hat{Y} = 69.866 - 3.062\ BF2 + 3.549\ BF4 + 332.921\ BF5 + 0.334\ BF8;$$

$$\hat{Y} = 69.866 - 3.062(16.760 - X4) + 3.549(4.140 - X3) + 322.921(X6 - 0.418) + 0.334(73.660 - X5);$$

3.3 Residual Assumption Test

Residual assumption testing is essential for further inference. The residuals are assumed to be normally distributed and the residuals from each observation are assumed to have constant variance and have a value of σ^2. The results of the residual assumption test are as follows

1. Residual Normality Assumption Test

The hypothesis used in the residual assumption test is as follows.

H_0 : Residuals are normally distributed

H_1: Residuals are not normally distributed

The test was carried out using software assistance. The results of the statistical test calculation using the Kolmogorov-Smirnov normality test are p-values of more than >0.15, which is more than the significance level (α = 0.05). Therefore, a decision was made to fail to reject H_0, so the conclusion that can be drawn is that the residuals are normally distributed.

2. Test of Constant Variance σ^2 Assumption

This test is conducted to detect symptoms of heteroscedasticity in the residuals. This test is conducted using the Glejser test. The hypothesis used in this test is as follows.

$H_0: {\sigma_1}^2 = {\sigma_2}^2 = \cdots = {\sigma_{34}}^2 = \sigma^2$
H_1: There is at least one ${\sigma_i}^2 \neq \sigma^2 ; i = 1,2, \dots ,34$

The results of the heteroscedasticity detection test using the Glejser test are a p-value of 0.173, this value exceeds the significance level (α = 0.05). Therefore, a decision to fail to reject H_0 was obtained, so the conclusion is that there is no case of heteroscedasticity.

3.4 MARS Model Significance Test

1. Simultaneous Test of Basis Function Coefficients of the MARS Model

Test statistics	value
Statistik Uji F	15.988
P-value	0.607399×10^{-6}

The critical area of simultaneous testing is to reject H_0 if F_{count} is more than $F_{(0.05;\, 4.29)}$ or the p-value is less than the significance level ($\alpha = 0.05$). Based on the table, the F_{count} value is 15.988, which is more than $F_{(0.05;\, 4.29)}$ = 2.7. In addition, the resulting p-value is 0.607399×10^{-6} which is less than the significance level ($\alpha = 0.05$).

2. Partial Test of Basis Function Coefficients of MARS Model

The partial test aims to determine whether each basis function coefficient in the MARS model partially shows an influence on the response variable. The hypothesis used in the partial testing of the basis function coefficient is as follows.

$H_0: a_m = 0, m = 2,4,5,8$
$H_1: a_m \neq 0, m = 2,4,5,8$

Parameter	Estimate	S.E.	P-value	Decision
Constant	69.866	0.801	$0{,}999201.10^{-15}$	reject H_0
BF_1	-3.062	0.547	0.47989×10^{-5}	reject H_0

International Journal of Academic and Applied Research (IJAAR)
ISSN: 2643-9603
Vol. 8 Issue 6 June - 2024, Pages: 72-80

BF_4	3.549	0,808	$0,135563 \times 10^{-3}$	reject H_0
BF_5	332.921	77.834	$0,110385.10^{-4}$	reject H_0
BF_{15}	0.334	0.126	0.013	reject H_0

The critical region of partial testing is to reject H_0 if $|t_{count}| > t_{(0.025.29)}$ or the p-value is less than the significance level (α=0.05). Based on the Table, the value of $|t_{count}|$ of each basis function in the model is more than $t_{(0.025;\,29)} = 2.048$. In addition, the p-value of each basis function in the model is less than the significance level ($\alpha = 0.05$). Therefore, the decision obtained is to reject H_0, so the conclusion is that there is at least one a_m that is not equal to zero, with m=2,4,5,8. This can be interpreted that the model obtained shows a relationship between the basis function coefficient and the response variable.

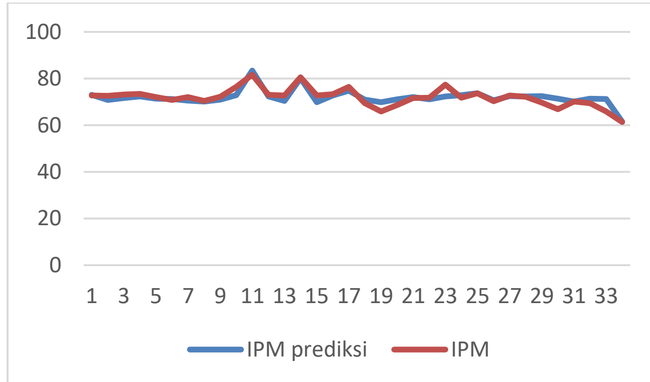
3. Variabel important

The level of variable importance is one of the outputs produced by the MARS application. The level of variable importance is used to sort the predictor variables that affect the response variable. The level of variable importance in modeling human development index data in Indonesia is as follows.

Variable	Name Variable	level of important	gcv reduction
X_4	health complaints	100%	15.470
X_3	Unmet need for health services (Persen)	73.61%	12.384
X_6	Gini Ratio	70.91%	12.122
X_5	Labor Force Participation Rate	27.32%	9.238
X_1	Literacy rate	0%	8.732
X_2	Regional income and expenditure figures for education	0%	8.732

3.5 Best MARS Model Interpretation

The best MARS model has been obtained with a combination of fourteen basis functions, a maximum interaction of one, and a minimum observation between knots of one. After obtaining the best model and testing significant variables, as well as assumptions on the residuals, the response variables and their estimated results can be plotted to compare the two values. The plot is as follows



1. Basis two function (BF_2)

$$BF_2 = \begin{cases} 16.76 - X_4\ ; & untuk\ X_4 < 16.76 \\ 0\ ; & untuk\ X_4\ yang\ lain \end{cases}$$

Interpretation of the value of the basis function two (BF_4) with a coefficient of -3.062 in the best model in equation (4.1) means that every one unit increase in (BF_4) will decrease the HDI value by 3.062 with the basis functions (BF_4), (BF_5), and (BF_6) considered constant. In addition, this means that provinces that have a health complaint score of less than 16.67 have a significant influence, namely a decrease in the HDI value. It can be seen that there are 3 provinces, namely Maluku, North Maluku and Papua, which have a value at (BF_4) while the remaining 31 provinces do not have a value at (BF_4). This means that there are three provinces that have a low level of health complaints. This is in line with what happened where the provinces of Maluku, North Maluku and Papua have quite high health complaint scores and affect the Human Development Index in the province in the form of a decrease in the HDI value. The reduction in value is in accordance with the basis of the human development index, one of which is influenced by health factors, in this case public health complaints.

2. Basis four function (BF_4)

$$BF_4 = \begin{cases} 4.14 - X_3\ ; & for\ X_3 < 4.14 \\ 0\ ; & for\ another\ X_3 \end{cases}$$

Interpretation of the value of the three basis functions BF_4 means that every one unit increase in BF_4 will increase the HDI by 3,549 with the basis functions $BF_2, BF_5\ and\ BF_8$ considered constant. In addition, this means that provinces that have an unmet need for health services of less than 4.14 will have a significant impact, namely an increase in the HDI value. It can be seen that the provinces in Indonesia that have a value of BF_4 are 5 provinces, namely DKI Jakarta, Riau Islands, Bali, Maluku and Papua which do not have BF_4 as many as 29 provinces. This shows that health services in the province are good because only a few people whose health service needs are not met, thus providing a positive impact in the form of an increase in the HDI value. The addition of value to the model is in accordance with the basis of the human development index, one of which is influenced by health factors in this case is the unmet need for health services.

3. Basis five function

$$BF_5 = \{ \begin{matrix} X_6 - 0.418 & ; & untuk\ X_6 > 0.418 \\ 0 & ; & untuk\ X_6\ yang\ lain \end{matrix}$$

Interpretation of the value of the five basis function BF_5 with a coefficient of 332.921 means that every one unit increase in BF_5 will increase the HDI by 332.921 with the basis functions BF_2, $BF_4, and\ BF_8$ considered constant. In addition, this means that provinces that have a gini ratio value of more than 0.418 will have a significant effect, namely an increase in the HDI value. It can be seen that the provinces in Indonesia that have a value of BF_5 are 3 provinces, namely DKI Jakarta, Yogyakarta and Gorontalo and those that do not have BF_5 are 31 provinces. From these results, it can be seen that the 3 provinces actually have a very high gini ratio value, this means that there is quite a large social inequality in the province. However, in the model, a high Gini ratio value will increase the HDI value. This is not appropriate because the higher the Gini ratio value, the lower the HDI value should be. This may occur because the model is not appropriate and the economic index should not be represented by the Gini ratio.

4. Basis eight function

$$BF_8 = \{ \begin{matrix} 73.660 - X_5 & ; & untuk\ X_5 < 73.660 \\ 0 & ; & untuk\ X_5\ yang\ lain \end{matrix}$$

Interpretation of the value of the five basis function (BF_8 with a coefficient of 0.334 means that every one unit increase in BF_8 will decrease the HDI by 0.334 with the basis functions $BF_2, BF_4\ and\ BF_5$ considered constant. In addition, this means that provinces that have a labor force participation rate value of less than 73.66 will have a significant effect, namely an increase in the HDI value. it can be seen that the provinces in Indonesia that have a value of BF_8 are 29 provinces and those that do not have BF_8 are 5 provinces, namely Yogyakarta, East Java, Bali, East Nusa Tenggara and Papua. This means that provinces that have a BF_8 value do not have a fairly good TPAK value but in the model experience an increase in the HDI value. This could happen due to the imperfection of the model and the TPAK value which is less able to represent the economic index or decent living standards in the Human Development Index.

4. CONCLUSION

Based on the results of research that has been conducted related to the Human Development Index value in 34 provinces in Indonesia using the MARS method, several things can be concluded as follows:

1. In the description of the research variables, data descriptions were carried out using bar charts and statterplots. The value of the Human Development Index in Indonesia shows a value of 71.9, meaning it is included in the fairly high category. In addition, it can be seen that the highest value of the Human Development Index is in the province of DKI Jakarta with a value of 81 and the lowest is in the province of Papua with a value of 61.3. This means that development in Indonesia is still too centered in DKI Jakarta as the capital city.

2. The best model obtained using the Multivariate Adaptive Regression Spline (MARS) method is a combination of fourteen basis functions, a maximum interaction of one, and a minimum observation between knots of one. The model produces a Generalized Cross Validation (GCV) value of 8,736, R^2 of 0688, and a Mean Square Error (MSE) of 5,402. The following is the best model obtained using the MARS method.

3. The interpretation of the best model in this study is as follows:

   a. Based on the best model that has been obtained, it can be seen that the most influential predictor variable on the response variable is the Health Complaints variable with an importance level of 100%. The predictor variables that influence the response variable based on the order of importance are the unmet need for health services variable with an importance level of 73.605%, the Gini ratio variable with a percentage of 70.991% and the labor force participation rate with an importance level of 27.319%. Meanwhile, the other two predictor variables, namely the APBD for education and literacy rates, have an importance level of 0%

   b. Based on the best model that has been obtained, it is known that the basis functions that have significant values are BF_2, BF_4, BF_5, and BF_8. BF_2 means that every one unit increase in BF_2 will decrease the Human Development Index by 3,062 with other basis functions considered constant. BF_4 means that every one unit increase in BF_4 will increase the value of the Human Development Index by 3,549 with other basis functions considered constant. BF_5 means that every one unit increase in BF_4 will decrease the Human Development Index by 332.921 with other basis functions considered constant. BF_8 means that every one unit increase in BF_8 will decrease the Human Development Index by 0.334 with other basis functions considered constant.

5. REFERENCES

[1] Aristiyowati, E. S., & Falianty, T. A. (2018). Peranan Perkembangan Inovasi Finansial Sistem Pembayaran Dalam Mempengaruhi Permintaan Uang Di Indonesia. *EKUITAS (Jurnal Ekonomi dan Keuangan)*, *2*(3), 404-426.

[2] Azmi, N., & Perdana, H. (2021). Pemodelan Akreditasi Sekolah Sma/Ma Di Kalimantan Barat Dengan Pendekatan Multivariate Adaptive Regression Spline. *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, *10*(1).

[3] Budiantara, I. N., Suryadi, F., Otok, B. W., & Guritno, S. (2006). Pemodelan B-Spline dan MARS Pada Nilai Ujian Masuk terhadap IPK Mahasiswa Jurusan Disain Komunikasi Visual UK. Petra Surabaya. Jurnal Teknik Industri, 8(1), 1-13.

[4] Deichmann, J., Eshghi, A., Haughton, D., Sayek, S., & Teebagy, N. (2002). Application of multiple adaptive regression splines (MARS) in direct response modeling. Journal of Interactive Marketing, 16(4), 15-27.

[5] Eubank, R. L. (1999). Nonparametric regression and spline smoothing. CRC press.

[6] Fajri, R. H. (2021). Analisis Faktor-Faktor Yang Mempengaruhi Indeks Pembangunan Manusia Di Provinsi Riau. ECOUNTBIS: Economics, Accounting and Business Journal, 1(1), 212-222.

[7] Friedman, J.H. (1991). Multivariate Adaptive Regression Splines. The Annals of Statistiks. Vol. 19 No. 1.

[8] Gunandri F.& Panggabean, G.F.(2023). Penerapan Multivariate Adaptive Regression Splines pada Laju Produk Domestik Regional Bruto Menurut Lapangan Usaha diProvinsiSumatera Utara. Jurnal Sains dan Teknologi,Vol. 2 No. 1

[9] Hardle, W. (1994). Applied Nonparametric Regression. New York: Cambrige University Press

[10] Hardle,W.(1990). Applied Nonparametric Regession. Cambridge: Cambridge University Press.

[11] Idris, M. K. (2014) IPM indikator status kemajuan negara. (20 January 2014). in Utusan Malaysia

[12] Insany, A. N. (2019). Penerapan Multivariate Adaptive Regression Spline (MARS) Pada Kasus Indeks Pembangunan Manusia Di Kawasan Timur Indonesia (Doctoral dissertation, Universitas Tadulako).

[13] Lembang, F. K., Patty, H. W. M., & Maitimu, F. (2019). Analisis Kemiskinan di Kabupaten Maluku Tenggara Barat Menggunakan Pendekatan Multivariate Adaptive Regression Spline (MARS). Media Statistika, 12(2), 188-199.

[14] Mahuze, W., Masinambow, V., & Lapian, A. (2022). Pengaruh Pengeluaran Pemerintah Pada Bidang Pendidikan dan Kesehatan Terhadap Indeks Pembangunan Manusia di Kabupaten Boven Digoel. Jurnal Berkala Ilmiah Efisiensi,22(3).

[15] Nuryanto, N., Achsani, N.A.& Ilhami,S. (2014). Analisis Pengaruh Anggaran Pendidikan terhadap Indeks Pembangunan Manusia di Indonesia.Tesis Institut Pertanian Bogor.

[16] Pintowati, W., & Otok, B. W. (2012). Pemodelan Kemiskinan di Propinsi Jawa Timur dengan Pendekatan Multivariate Adaptive Regression Splines Ensemble. Jurnal Sains dan Seni ITS, 1(1), D283-D288.

[17] Rahmatin, U. Z., & Soejoto, A. (2017). Pengaruh Tingkat Kemiskinan Dan Jumlah Sekolah Terhadap Angka Partisipasi Sekolah (Aps) Di Kota Surabaya. JPEKA: Jurnal Pendidikan Ekonomi, Manajemen Dan Keuangan, 1(2), 127–140.

[18] Risambessy, S., Aulele, S. N., & Lembang, F. K. (2022). Misclassification Analysis of Elementary School Accreditation Data in Ambon City Using Multivariate Adaptive Regression Spline. Jurnal Matematika, Statistika dan Komputasi, 18(3), 394-406.

[19] Shafana, N. R., & Gunawan, G. (2021). Analisis Faktor yang Mempengaruhi Kesadaran Penduduk dalam Vaksin Covid-19 Menggunakan Metode 62 Multivariate Adaptive Regression Spline. Jurnal Riset Matematika, 1(2), 154- 162.

[20] Sirait, S. C. (2017). Tanggung Jawab Pemerintah Untuk Memberikan Pendidikan Kepada Anak Terlantar Dalam Perspektif Undang-Undang Perlindungan Anak. De Lega Lata: Jurnal Ilmu Hukum, 2(1), 158-182.

[21] Sudjana, N. (1996). Statistik Dasar. Tarsito. Bandung. Sumaryadi, I. N. (2010). Efektifitas Implementasi Otonomi Daerah. Jakarta: Citra Utama

[22] Todaro, Michael. P. 1994. Ekonomi untuk negara berkembang. Edisi ketiga. Jakarta : Bumi Aksara

[23] Tripena, A. (2011). Analisis Regresi Spline Kuadratik. In dipresentasikan dalam Seminar Nasional Matematika dan Pendidikan Matematika, Jurusan Pendidikan Matematika, FMIPA UNY. 92-102
Wahyudi, H., Naufal, F., Pongoh, F. W., & Amelia, R. (2023). Analisis Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kalimantan Timur. In Seminar Nasional Teknologi Informasi dan Matematika (SEMIOTIKA) (Vol. 2, No. 1, pp. 121-135).

[24] Wibowo, A. (2019). Multivariate Adaptive Regression Splines Modeling for Household Food Security in Central Borneo Province 2017. Global Science Education Journal, 1(1), 39-47.

[25] Wicaksono, W., Wilandari, Y., & Suparti, S. (2014). PEMODELAN MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) PADA FAKTOR-FAKTOR RESIKO ANGKA KESAKITAN DIARE (Studi Kasus: Angka Kesakitan Diare Di Jawa Tengah, Jawa Timur Dan Daerah Istimewa Yogyakarta Tahun 2011). Jurnal Gaussian, 3(2), 253-262.

[26] Zain, I. & Melliana,A.(2014). Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten/Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel." Jurnal Sains dan Seni ITS, 2(2).