

Classification Of Perceptions Of The Covid-19 Vaccine Using Multivariate Adaptive Regression Spline

Rizki Fitri Ananda¹, Lisa Harsyah², Muhammad Rijal Alfian³

^{1,2,3}University of Mataram, Indonesia

Article Info

Article history:

Received : 01-02-2022

Revised : 05-07-2023

Accepted : 05-16-2023

Keywords:

Classification;

Covid-19;

MARS;

Perseption;

Vaccine.

ABSTRACT

Indonesia is one of the countries infected with the covid-19 virus. One of the government's efforts is the covid-19 vaccination. However, the covid-19 vaccination caused controversy for some people because many people refused to be vaccinated. Public perception of the covid-19 vaccine can be categorized into two, namely positive and negative, based on survey from Indonesia ministry of health about acceptance of covid-19 vaccine state that this can be influenced by many factors. These factors are important to know as an effort to increase acceptance of covid-19. Multivariate Adaptive Regression Splines (MARS). The purpose of this study is to determine the classification model of public perception of the covid-19 vaccine and the factors that influence it. The method used in this study is Multivariate Adaptive Regression Splines (MARS). This method is appropriate classification method to be applied to categorical response variable data, The outcomes demonstrate that the optimum mars model is produced by combining $BF = 24$, $MI = 3$, $MO = 1$, and $GCV = 0.07340546$. The resulting classification level is 91.5% with influencing factors yaitu gender (x_1), age (x_2), last education (x_4), willingness to vaccinate (x_6), education (x_8). Based on the results obtained, the government can consider these factors for socialization



Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020

DOI: <https://doi.org/10.30812/varian.v6i2.2639>

Corresponding Author:

Rizki Fitri Ananda,
Departement of Mathematics, Universitas Mataram, Indonesia.
Email: rizkyfit309@gmail.com

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



A. INTRODUCTION

The city of Wuhan in China has reported the emergence of the corona virus since December 2019, later named Severe Acute Respiratory Syndrome Coronavirus 2 (Sars-Cov-2). Sars-Cov-2 is a virus that produces a group of atypical pneumonia diseases that spread rapidly throughout the world and are known as coronavirus disease 2019 (covid-19) (Kim et al., 2020). The Covid-19 emergency was designated a public health emergency on January 30, 2020 when the World Health Organization (WHO) stated that the pandemic was of international concern (PHEIC). The World Health Organization (WHO) formally classified covid-19 a pandemic on March 11, 2020. Covid-19 symptoms include a cough, fever, diarrhea, shortness of breath, myalgia, sore throat, headache, and tiredness (Vkovski et al., 2021). More than 114 countries had been affected by the covid-19 infection as of October 25, 2020 and there had been more than 43,140,173 confirmed cases and more than 1,155,235 fatalities as a result (Ozkara et al., 2020).

One of the nations with the covid-19 infection is Indonesia. In March 2020, the covid-19 pandemic was first reported. The covid-19 vaccine is one of the strategies the state has put in place. The covid-19 vaccine effort, however, did not go as planned. Some people were upset by this, including some community groups who opposed vaccinations. People's opinions of the covid-19 vaccination will change as a result of the propagation of false information, which will also change how they act. Based on information from the internet, particularly social media, decisions and choices are made (Rizma et al., 2020).

Perception is a process of selection, arrangement, and completion by individuals who interpret information as meaningful logical images. Perception occurs when a person imitates external stimuli and is captured by others and then enters the brain. Perception is

the process of using sensory tools to find information to be understood (Duri Kartika et al., 2015). Indonesian people's perceptions of the covid-19 vaccine can be categorized into two, namely positive and negative (Priadi, 2017), this can be influenced by many factors which of course can be different for each individual. These factors can be analyzed using statistical methods.

In statistical methods, there are classification techniques used to organize data systematically. A good classification method produces a minimal number of errors. One of the method for data classification is Multivariate Adaptive Regression Spline (MARS). This method is created by Friedman (1991). MARS is commonly used to solve two types of statistical problems: continuous and categorical response variables. MARS is a flexible method for determining nonlinear relationships between response variables and predictor variables that does not depend on model assumptions from the regression method. The MARS technique produces continuous knot models based on the smallest generalized cross-validation (GCV) values, which can be used to overcome the difficulties of high-dimensional data and produce accurate predictions of response variables (Addini et al., 2023). Data with many predictor variables is referred to as high-dimensional data. High-dimensional data is data that has a number of predictor variables of $3 \leq n \leq 20$ (Zurimi et al., 2020). Based on this explanation, the MARS method is very appropriate to use on public perception data of the covid-19 vaccine, where this data is high-dimensional data with a total of 8 variables and this data is categorical data with perception categories, namely positive and negative. The purpose of this study was to analyze public perceptions of the covid 19 vaccine where this problem is still very new and no one has discussed it using the MARS method.

1. Multivariate Adaptive Regression Spline (MARS)

Multivariate adaptive regression splines (MARS) is a nonparametric regression technique that can be used to identify nonlinear relationships and interactions between response and predictor variables. The MARS method is based on a statistical approach, dependent and independent variable data with a series of splines different slopes (Friedman, 1991). The endpoint of splines (knots) mark the end and beginning of another dataset. resulting in piecewise curves called basis functions or hinge functions (Qureshi et al., 2022). The base function's maximum is 2-4 times the quantity of predictor variables. The analysis of MARS model use Earth package (Milborrow, 2021). The following equation can be used to represent the general MARS model.

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

Description:

- a_0 = regression constant of the base function
- a_m = base function's m -th coefficient, where $m = 1, 2, \dots, M$
- M = maximum base function
- K_m = interactions in the m -th base function in terms of number
- S_{km} = ± 1 if the data is to the right or left of the knot point
- $X_{v(k,m)}$ = the v -th predictor variable, the-select and km subregion
- t_{km} = the knot value of the predictor variable $X_{v(k,m)}$ (Kharisma et al., 2021)

2. MARS Parameter Estimation

The estimation method used in this study is Ordinary Least Square (OLS). The MARS model can be written in the form of a matrix, namely (Zurimi et al., 2020):

$$Y = B\alpha + \varepsilon \quad (2)$$

with:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$B = \begin{pmatrix} 1 & \prod_{k=1}^{K_1} [S_{1m}X_{1(1,m)} - t_{1m}] & \dots & \prod_{k=1}^{K_M} [S_{Mm}X_{1(M,m)} - t_{Mm}] \\ 1 & \prod_{k=1}^{K_1} [S_{1m}X_{2(1,m)} - t_{1m}] & \dots & \prod_{k=1}^{K_M} [S_{Mm}X_{2(M,m)} - t_{Mm}] \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} [S_{1m}X_{V(1,m)} - t_{1m}] & \dots & \prod_{k=1}^{K_M} [S_{Mm}X_{V(M,m)} - t_{Mm}] \end{pmatrix}$$

$$\varepsilon = Y - B\alpha$$

In order to get an estimator α , the OLS method is carried out by minimizing the square of error by squaring the equation as a linear regression. So that α_{OLS} is obtained as follows:

$$\alpha_{OLS} = B\alpha + \varepsilon \tag{3}$$

3. Choosing the Best MARS Model

The forward stepwise and backward stepwise algorithms are used to determine knots on MARS, and they are based on the least Generalized Cross Validation (GCV) value. In other words, the knot point that was chosen has a minimum GCV value. This demonstrates that the best way to choose a model is to look at its GCV value when it was built using the value of a particular basis function. The model with the lowest or minimum GCV value among the others is the one that is the best. The following equation gives the definition of the minimum GCV function (Addini et al., 2023):

$$GCV(M) = \frac{\frac{1}{n} \sum_{i=0}^n [y_i - f_M(x_i)]^2}{\left[1 - \frac{\widehat{C}(M)}{n}\right]} \tag{4}$$

Description:

- M = maximum base function
- $f_M(x_i)$ = the response variable's estimated value in the M base function
- y_i = i -th response variable
- n = number of observations
- \widehat{C} = complexity cost function $\widehat{C}(M) = C(M) + dM$
- $C(M)$ = $trace [B(B'B)^{-1}B' + 1]$
- d = function when the base function reaches optimal $2 \leq d \leq 4$

4. Significance Test of MARS Model

1. Simultaneous testing

Hypothesis formulation (Azmi & Perdana, 2021):

- H_0 = $a_1 = a_2 = \dots = a_m = 0$ (model is not significant)
- H_1 = there is at least one $a_m \neq 0$ dengan $m = 1, 2, \dots, M$ (significant model)

Statistic test :

$$F_{hitung} = \frac{SSE/k}{SSE/(n - k - 1)} \tag{5}$$

Source: (Addini et al., 2023)

Critical area

Reject H_0 if value $F_{count} > F_{\alpha/2}(k; n - k - 1)$ or $p - value < \alpha$

2. Partial Test

Hypothesis formulation:

- H_0 = $a_m = 0$ (a_m coefficient has no effect on the model)
- H_1 = every $a_m \neq 0$ dengan $m = 1, 2, \dots, M$ (a_m coefficient affects the model)

Statistic test :

$$t_{hitung} = \frac{\widehat{a}_m}{Se(\widehat{a}_m)} \tag{6}$$

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

$$Se(\hat{a}_m) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1} C_{ij}}$$

The entries on the major diagonal of the matrix $(B^T B)^{-1}$ are denoted as C_{ij} (Azmi and Perdana, 2021).

Critical area

Reject H_0 if value $|t_{hitung}| > t_{\frac{\alpha}{2}}(n - k)$ or $p - value < \alpha$.

5. Multicollinearity Assumption

A multicollinearity test determines whether or not predictor variables in a regression model have a high correlation. If the predictor variables have a high correlation, the relationship between the predictor variables and the response variables will be disrupted. The value of Tolerance and VIF (Variance Inflation Factor) indicates multicollinearity. A regression model is said to be multicollinear if it has a VIF value of < 10 and a tolerance number greater than 0.10. The following equation can be used to calculate the VIF value:

$$VIF = \frac{1}{(1 - R_i^2)} \quad (7)$$

Source: (Kim, 2019)

6. Classification Accuracy

The Apparent Error Rate can be used to calculate the model prediction error on grouping results (APER). APER is a measurement metric used to determine the possibility of a classification function's misclassification (Kharisma et al., 2021). The APER value represents the proportion of samples incorrectly classified by the classification function. In this study, a binary response variable was used so that the classified error could be calculated using the Table 1:

Table 1. Classification Table

Results Observation	Predictions		Results
	y_1	y_2	
y_1	n_{11}	n_{12}	
y_2	n_{21}	n_{22}	

Source: (Utami et al., 2020)

Information:

y_1 : response variable category 1

y_2 : response variable category 2

n_{11} : the number of y_1 observations that are correctly classified as y_1

n_{12} : the number of y_1 observations that are incorrectly classified as y_1

n_{21} : the number of y_2 observations that are incorrectly classified as y_1

n_{22} : the number of y_2 observations that are correctly classified as y_1

$$APER(\%) = \frac{n_{21} + n_{12}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\% \quad (8)$$

B. RESEARCH METHOD

This study makes use of primary data from surveys that were distributed in NTB with ages 18-59 years. The number of respondents is 390 respondents. The study's response variable is the public's view of the covid-19 vaccination, with categories 1 and 0 denoting favorable and negative assessment, respectively. While the predictor variables are gender (x_1), age (x_2), employment status (x_3), last education (x_4), insurance ownership status (x_5), willingness to be vaccinated (x_6), history of non-communicable diseases (x_7), and education on covid-19 vaccines (x_8).

Table 2. Definitions of Response Variables and Predictors

	Variable	Description	Encoding	Scale data
Response Variables	Y	Persepsi masyarakat terhadap vaksin covid 19	Positif (1) Negatif(0)	Nominal
Predictors variables	X ₁	Gender	Men Wowan	Nominal
	X ₂	Age	18-25 26-40 41-59	Ordinal
	X ₃	employment status	Working No/Not yet Working	Nominal
	X ₄	last education	Elementary schoolg Junior high school Senior high school D3/S1/S2/S3	Ordinal
	X ₅	Insurance ownership status	BPJS/Private There isn't any	Nominal
	X ₆	Willingness to vaccinate	Yes Not	Nominal
	X ₇	Non contagious diseases	There is There isn't any Never	Nominal
	X ₇	education	Never Sometimes Often	Nominal

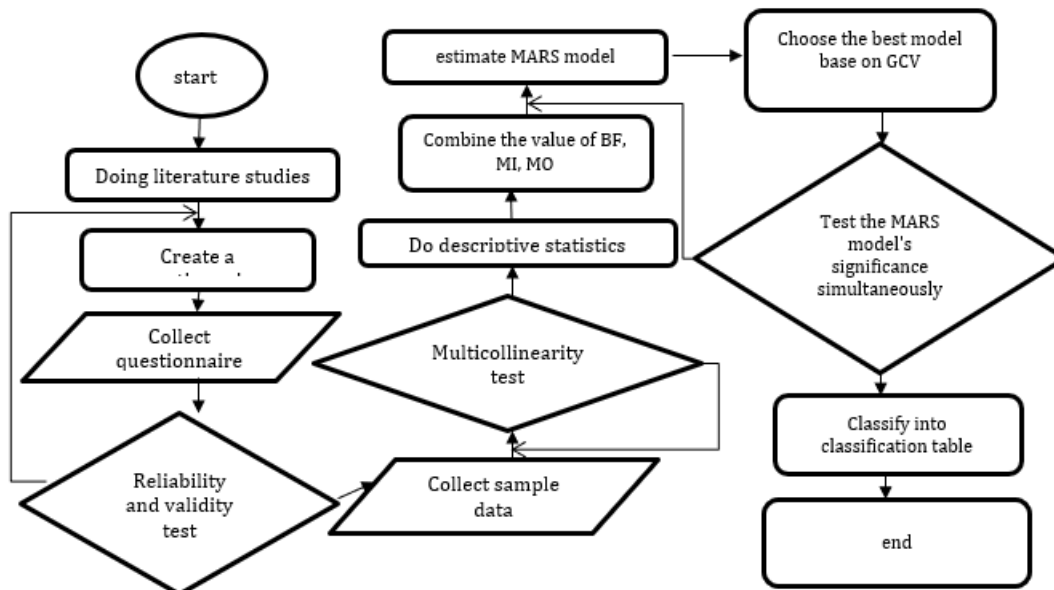


Figure 1. Research Flow Chart

Analysis method

1. Create a questionnaire
2. Testing the validity and reliability of the questionnaire
3. Carry out data collection
4. Doing descriptive statistics with SPSS
5. To determine whether the predictor variables in a regression model have a high correlation, run a multicollinearity test.
6. Perform MARS analysis with the following analysis steps:
 - (a) The MARS model was developed by experimenting by combining BF (basic function), MI (maximum interaction),

and MO (minimum observation). The number of BF is 2-4 times the independent variable so that 16, 24, and 32 basis functions are obtained. The quantity for MI are 1,2 and 3 because if it is more then the model will be more complex, the number of MO is 0, 1, 2, and 3 (Wibowo and Ridha, 2020).

- (b) Based on the model with the lowest GCV, choose the best MARS model. The model with the lowest or smallest GCV value among the models created is the best model.
- (c) Explaining the public opinion model of the covid-19 vaccination and the factors that affect it.
- (d) Test the MARS model's significance simultaneously and in partial.
- (e) Create a confusion matrix for the classifications. The apparent error rate (APER) can be used to calculate the model prediction error on the grouping results.

C. RESULTS AND DISCUSSION

1. Data Description

Descriptive statistics aims to describe the object of research taken from a sample or population to produce useful information. In this study, descriptive statistics are useful for knowing the characteristics of people's perceptions of the covid-19 vaccine.

Table 3. Respondent Distribution According to Response Variables

No	Variable	Information	Frequency	Percentage (%)
1	Perception	Positive	224	57.4%
		Negative	166	42.6%
Total			390	100%

Furthermore, the distribution table of respondents based on predictor variables is as follows:

Table 4. Respondent Distribution Determined by Predictor Variables

No	Variable	Information	Frequency	Percentage (%)
1	Gender (x_1)	Men	165	42.3%
		Woman	225	57.7%
2	Age(x_2)	18-25	203	52.1%
		26-40	103	26.4%
		41-59	84	21.5%
3	employment status (x_3)	Working	179	45.9%
		No/Not yet Working	211	54.1%
4	last education (x_4)	Elementary school	0	0%
		Junior high school	19	4.9%
		Senior high school	176	45.1%
		D3/S1/S2/S3	195	50.01%
5	Insurance ownership status (x_5)	BPJS/Private	272	69.7%
		There isn't any	118	30.3%
6	Willingness to vaccinate (x_6)	Yes	350	89.7%
		Not	40	10.3%
7	Non contagious disease (x_7)	There is	25	6.4%
		There isn't any	365	93.6%
8	education (x_8)	Never	18	4.6%
		Sometimes	128	32.8%
		Often	244	62.6%
Total			390	100%

Table 2 shows that 57.4% of the covid-19 vaccination is viewed favorably by the public and 42.6% have a negative perception while table 3 shows that most of the respondents are female with a percentage of 57.7%, mostly aged 18-25 years with a percentage of 52.1%, most are no/not yet working with a percentage of 54.1%, most of the latest education is D3/S1/S2/S3 with a percentage of 50.0%, most use BPJS/private with a percentage of 69.7% , most are willing to be vaccinated with a percentage of 89.7%, most have no history of infectious diseases with a percentage of 93.6%, and most often receive vaccine education with a percentage of 62.6%.

2. Test for Multicollinearity

A multicollinearity test can determine whether the predictor variables in a regression model have a high degree of correlation. If a regression model's VIF score is less than 10 and its tolerance number is greater than 0.10, it can be said to be multicollinearity-free.

Table 5. Multicollinearity Test

No	Variable	VIF
1	Gender	1.045360
2	Age	2.649418
3	employment status	2.228228
4	last education	2.282086
5	Insurance ownership status	1.173017
6	Willingness to vaccinate	1.117267
7	Non-contagious disease	1.076097
8	education	1.386753

It is clear from the results in Table 4 that there is no multicollinearity in the data because every variable has a VIF value that is less than 10.

3. Constructing the Best Mars Model

The MARS model will then be chosen based on the least *GCV* value. Combining the number of basis functions (*BF*), maximum interaction (*MI*), and minimum observation (*MO*) yields the minimum *GCV* value through trial and error. The value of *BF* is 16, 24, 32. While the value of *MI* of 1, 2, and 3. The values of *MO* are 0, 1, 2, and 3. The following is a table for calculating the *GCV* of all possible MARS models.

Table 6. MARS Model Selection Results Using *GCV* Kriteria Criteria

No	Model	BF	MI	MO	GCV	MSE
1	1	16	1	0	0.08113386	0.07581508
2	2	16	1	1	0.08113386	0.07581508
3	3	16	1	2	0.08113386	0.07581508
4	4	16	1	3	0.08113386	0.07581508
5	5	16	2	0	0.07427355	0.06830432
6	6	16	2	1	0.07427355	0.06830432
7	7	16	2	2	0.07427355	0.06830432
8	8	16	2	3	0.07427355	0.06830432
9	9	16	3	0	0.07427355	0.06830432
10	10	16	3	1	0.07412638	0.06726067
11	11	16	3	2	0.07412638	0.06726067
12	12	16	3	3	0.08113386	0.06726067
13	13	24	1	0	0.08113386	0.07581508
14	14	24	1	1	0.08113386	0.07581508
15	15	24	1	2	0.08113386	0.07581508
16	16	24	1	3	0.07412638	0.07581508
17	17	24	2	0	0.07347060	0.06488321
18	18	24	2	1	0.07347060	0.06488321
19	19	24	2	2	0.07347060	0.06488321
20	20	24	2	3	0.07347060	0.06488321
21	21	24	3	0	0.07347060	0.06488321
22	22	24	3	1	0.07340546	0.06571309
23	23	24	3	2	0.07340546	0.06571309
24	24	24	3	3	0.07340546	0.06571309
25	25	32	1	0	0.08113386	0.07581508
26	26	32	1	1	0.08113386	0.07581508
27	27	32	1	2	0.08113386	0.07581508
28	28	32	1	3	0.08113386	0.07581508
29	29	32	2	0	0.07347060	0.06488321
30	30	32	2	1	0.07347060	0.06488321

31	31	32	2	2	0.07381915	0.06519102
32	32	32	2	3	0.07381915	0.06519102
33	33	32	3	0	0.07347060	0.06488321
34	34	32	3	1	0.07340546	0.06571309
35	35	32	3	2	0.07340546	0.06571309
36	36	32	3	3	0.07340546	0.06571309

The MARS model with the minimum *GCV* value is the best one. The number of basis functions (*BF*), maximum interaction (*MI*), and minimum observation (*MO*) are combined to get the *GCV* value. Based on the table above, there are 36 models, and the combination of $BF = 24, 32$, $MI = 3$, and $MO = 1, 2, 3$ results in the best model. The next step is to find the lower MSE value if there are many best models based on the same *GCV* value. The three models all have the same lowest MSE of 0.06571309 for $BF = 24, 32$ with $MO = 1, 2, 3$, hence the value of classification accuracy will be examined next. For the three models, the classification accuracy value is 91.53%. Therefore, the smallest combination of models, $BF = 24$, $MI = 3$, and $MO = 1$ with a *GCV* value of 0.07340546, is examined in the final stage to find the optimal model. Consequently, the resulting MARS model is :

$$\hat{f}(x) = 0.788993 + 0.2003294BF_1 - 0.6303254BF_2 - 0.1023236BF_3 - 0.1267894BF_4 - 0.1918498BF_5 \\ + 0.6741352BF_6 - 0.1242206BF_7 - 0.714493BF_8$$

with:

$$\begin{aligned} BF_1 &= (SV) \\ BF_2 &= \max(0, 1 - U) \\ BF_3 &= JK * \max(0, 1 - U) \\ BF_4 &= \max(0, 1 - U) * SA \\ BF_5 &= SV * \max(0, 1 - E) \\ BF_6 &= \max(0, 1 - U) * \max(0, P - 2) \\ BF_7 &= \max(0, 1 - U) * \max(0, E - 1) \\ BF_8 &= \max(0, 1 - U) * \max(0, P - 2) * \max(0, 1 - E) \end{aligned}$$

Information :

$$\begin{aligned} JK &: \text{gender variable } (x_1) \\ U &: \text{age variable } (x_2) \\ P &: \text{last education variable } (x_4) \\ SA &: \text{variable of willingness to vaccinate } (x_6) \\ E &: \text{covid-19 vaccine education variable } (x_8). \end{aligned}$$

Based on the MARS model which is formed from 8 variables, there are 6 variables that influence covid-19 vaccine perception in the general public, namely last education (x_4), age (x_2), education (x_8). insurance status (x_5), willingness to vaccine (x_6) and gender (x_1). The following is an explanation of the model formed in the equation:

1. $BF_1 = (SV)$ with a coefficient of 0.2003294 in the model means that if the variable willingness to be vaccinated is equal to 1, namely, willing to be vaccinated, it will increase positive perceptions by 0.2003294
2. $BF_2 = \max(0, 1 - U)$ with a coefficient of -0.6303254 in the model means that if the age variable is more than smaller than 1, namely, 18-25 years old, it will reduce positive perceptions of the covid-19 vaccine of - 0.6303254 .
3. $BF_3 = JK * \max(0, 1 - U)$ with a coefficient of -0.1023236 in the model means that if the sex variable is equal to 1, that is, males interact with the age variable smaller than 1, namely 18-25 years, it will reduce positive perceptions against the covid-19 vaccine is -0.1023236 .
4. $BF_4 = \max(0, 1 - U) * SA$ with a coefficient of -0.1267894 in the model means that if the age variable is smaller than 1, namely 18-25 years, the insurance ownership status variable is equal to 1, namely BPJS/Private it will reduce the positive perception of the covid-19 vaccine amounted to - 0.1267894.
5. $BF_5 = SV * \max(0, 1 - E)$ with a coefficient of -0.1918498 in the model means that if the variable willingness to be vaccinated is equal to 1, that is, the willingness to be vaccinated, interacts with the educational variable smaller than 1, i.e. never, it will reduce positive perceptions of covid-19 vaccine amounted to - 0.1918498.

- 6. $BF_6 = \max(0, 1 - U) * \max(0, P - 2)$ with a coefficient of 0.6741352 in the model means that if the age variable is smaller than 1, namely 18-25 years, it interacts with the last education variable greater than two, namely D3 /S1/S2/S3 will increase positive perceptions of the covid-19 vaccine of 0.7056162.
- 7. $BF_7 = \max(0, 1 - U) * \max(0, E - 1)$ with a coefficient of -0.1242206 in the model means that if the age variable is smaller than 1, namely 18-25 years, it interacts with education variables greater than one or often it will reduce the positive perception of the covid-19 vaccine amounted to - 0.1242206.
- 8. $BF_8 = \max(0, 1 - U) * \max(0, P - 2) * \max(0, 1 - E)$ with a coefficient of - 0.714493 in the model means that if the age variable is smaller than 1, namely 18-25 years old interact with the last education variable being greater than two, namely D3/S1/S2/S3 and the education variable being less than one or never , it will reduce positive perceptions of the covid-19 vaccine. amounted to - 0.714493.

4. SIGNIFICANCE TEST OF MARS MODEL

1. Simultaneous Test

Hypothesis formulation :

- H_0 : $a_1 = a_2 = \dots = a_m = 0$ (model is not significant)
- H_1 : there is at least one $a_m \neq 0$ with $m = 1, 2, 4, 5, 6, 8$ (significant model)

The values F_{count} based on the MARS model on the data are:

$$F_{count} = \frac{69,71554307/8}{25,62810491/38} = 129.5531898$$

Values F_{table} as follows:

$$F_{\alpha}(k; n - k - 1) = F_{0,05}(8;381) = 1.96$$

Based on the value F_{count} obtained, the decision is rejected H_0 because the value of $F_{count} > F_{\alpha}(k; n - k - 1)$ or $129.5531898 > 1.96$ so that the conclusion is a significant model.

2. Partial Test

Hypothesis formulation :

- H_0 : $a_m = 0$ (the coefficient a_m has no effect on the model)
- H_1 : $a_m \neq 0$ with $m = 1, 2, 4, 5, 6, 8$ (coefficient a_m effect on the model)

The following table shows the calculated t value:

Table 7. T Count Score

T Count	Score
$t_{(x_1)}$	25.08
$t_{(x_1)}$	-93.8
$t_{(x_1)}$	-22.98
$t_{(x_1)}$	-8.67
$t_{(x_1)}$	19.0
$t_{(x_1)}$	-2.28

Values t_{table} as follows:

$$t_{\alpha/2}(n - k) = t_{0.025}(382) = 1.96$$

Based on the value t_{count} obtained, the decision is rejected H_0 because the value for all $|t_{count}| > t_{\alpha/2}(n - k)$ so that in conclusion each coefficient has an a_m effect on the model.

5. MARS CLASSIFICATION

The following table shows the results that were obtained:

Table 8. MARS Classification

Group updates	Predicted group	
	0	1
0	160	6
1	27	197

The APER value can be calculated in the following equation:

$$\begin{aligned} APER(\%) &= \frac{n_{21} + n_{12}}{n_{11} + n_{12} + n_{21} + n_{22}} \\ &= \frac{6 + 27}{160 + 27 + 6 + 197} \\ &= \frac{33}{390} \\ &= 0.085 \\ &= 8.5\% \end{aligned}$$

So the value of classification accuracy is $100\% - 8.5\% = 91.5\%$

Based on the APER value obtained, it shows that the MARS method in this case obtained good results with a classification level of 91.5%, so that these results can be used as a reference for government in increasing acceptance of the covid-19 vaccine and for further researchers. The difference between this study and previous research is in the cases used. The case that will be analyzed in this study is public perception of the Covid-19 vaccine, where previously no one has discussed this case using statistical methods, especially the MARS method.

D. CONCLUSION AND SUGGESTION

Based on the results, the conclusion obtained is the MARS model used to classify peoples perceptions of the covid-19 vaccine is a model with a combined value of $BF = 24$, $MI = 3$, and $MO = 1$ because it has a GCV of at minimum 0.07340546. From the model obtained, there are 6 variables that affect the model, namely age (x_2), last education (x_4), willingness to vaccinate (x_6), insurance ownership status (x_5), education (x_8) and gender (x_1) with level of classification accuracy obtained is 91.5%. so that the government can determine further efforts to increase vaccine acceptance by considering these factors. The next researcher is advised to expand on this research by increasing the number of predictor factors and observed response variables, or by applying MARS methods or by applying the MARS method using different cases.

REFERENCES

- Addini, P. F., Hadi, W., and Harahap, P. M. R. (2023). Application of the multivariate adaptive regression spline (mars) method in analyzing misclassification of elementary school accreditation data in the city of tebing tinggi. *Jurnal Scientia*, 12(01):617–620.
- Azmi, N. and 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).
- Kharisma, R. D. L., Juhari, J., and Rosa, R. A. (2021). Multivariate adaptive regression splines and bootstrap aggregating multivariate adaptive regression splines of poverty in central java. *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, 6(4):238–245.
- Kim, D., Lee, J.-Y., Yang, J.-S., Kim, J. W., Kim, V. N., and Chang, H. (2020). The architecture of sars-cov-2 transcriptome. *Cell*, 181(4):914–921.
- Kim, J. H. (2019). Multicollinearity and misleading statistical results. *Korean journal of anesthesiology*, 72(6):558–569.
- Milborrow, S. (2021). Notes on the earth package. pages 1–69. <http://www.milbo.org/doc/Earth-Notes.Pdf>.
- Ozkara, A., Lambert, K., Baser, D. A., and Erbas, B. (2020). Effective implementation of unprecedented measures for the protection from covid-19 syndrome/covid-19 sendromuna karsi alinan olaganustu onlemlerin etkin uygulaması. *Bezmialem Science*, 8(2S1):63–67.
- Priadi, A. (2017). Hubungan antara pola asuh dan konsep diri dengan perilaku lingkungan mahasiswa. *Jurnal Ilmiah Pendidikan Lingkungan Dan Pembangunan*, 18(1):1–13.
- Qureshi, M. U., Mahmood, Z., and Rasool, A. M. (2022). Using multivariate adaptive regression splines to develop relationship between rock quality designation and permeability. *Journal of Rock Mechanics and Geotechnical Engineering*, 14(4):1180–1187.

- Rizma, S., Adlia, M., et al. (2020). Pengetahuan terkait usaha pencegahan coronavirus disease (covid-19) di indonesia. *HIGEIA (Journal of Public Health Research and Development)*, pages 333–346.
- Utami, I. T., Fadjryani, F., and Daniaty, D. (2020). Perbandingan klasifikasi status pendonor darah dengan menggunakan regresi logistik dan k-nearest neighbor. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 12(1):1–12.
- Vkovski, P., Kratzel, A., Steiner, S., Stalder, H., and Thiel, V. (2021). Coronavirus biology and replication: implications for sars-cov-2. *Nature Reviews Microbiology*, 19(3):155–170.
- Wibowo, A. and Ridha, M. R. (2020). Comparison of logistic regression model and mars using multicollinearity data simulation. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, 4(1):39–48.
- Zurimi, S. et al. (2020). Analysis of multivariate adaptive regression spline (mars) model in classifying factors affecting on student the study period at fkip darussalam university of ambon. In *Journal of Physics: Conference Series*, volume 1463, page 012005. IOP Publishing.

