JSDS: JOURNAL OF STATISTICS AND DATA SCIENCE
VOLUME 4, No 2, October 2025
e-ISSN: 2828-9986
https://ejournal.unib.ac.id/index.php/jsds/index

Factors Affecting The Open Unemployment Rate in West Sumatera Province Using The Multivariate Adaptive Regression Spline Method

Amesha Putri Pratama[1*], Devni Prima Sari [1]
[1] Department of Mathematics, Padang State University, Indonesia
* Corresponding Author: ameshaputripratama@gmail.com

Article Info
Article History:
Received: mm dd yyyy
Revised: mm dd yyyy
Accepted: mm dd yyyy
Available Online: mm dd yyyy

Key Words:
Open Unemployment Rate
MARS
Generalized Cross Validation

Abstract
The Open Unemployment Rate (OUR) is the percentage value of the ratio of the number of open unemployed to the total labor force. During the 2017-2022 period, the OUR faced fluctuations that showed significant numbers that could hurt economic growth and community welfare. The purpose of this study is to determine the factors that affect OUR in West Sumatra Province in 2017-2022 using the Multivariate Adaptive Regression Splines (MARS) method. This research is an applied research with data in the form of secondary data obtained from the Central Bureau of Statistics (CBS) of West Sumatra Province. The analysis carried out produces the best model at the minimum Generalized Cross Validation (GCV) value of 1.31675 from a combination of BF = 28, MI = 2 and MO = 1 so that the factors that influence the percentage of OUR are gross regional domestic product at current prices, population, population aged 15 years and over who are in school, population aged 15 years and over who work with the main job, average monthly per capita expenditure on food, average monthly per capita expenditure on non-food and monthly minimum wage with R2 adj of 81.6 percent.

1. INTRODUCTION

Unemployment is a condition where a person is included in the labor force who wants to get a job but has not been able to find one. Someone who does not want to work, but is not actively looking for work, is not included as unemployed [1]. One important factor that determines the welfare of a society is its level of income. Community income reaches its maximum/top if a sufficient level of labor utilization can be achieved. Unemployment can reduce people's income, and this reduces the level of prosperity they obtain. If the unemployment situation in a country is very bad, it will have an impact on per capita income, economic growth, economic development, as well as political and social chaos which has an adverse effect on the country's economy as well as the welfare of the community and the prospects for economic development in the long term [2].

One type of unemployment that quite often becomes a critical problem in terms of economic development is open unemployment because open unemployment is more relevant to the impact on productivity and community income, so that it causes other social problems, such as poverty [3]. Open unemployment is part of a labor force unit that is not working, is looking for work, or is planning a job venture, but is unable to get a job, so they are not looking for work, and those who have gotten a job but have not started working [4]. In addition, open unemployment can hamper economic growth because the potential of the existing workforce is not optimally utilized. The inability of the labor market to absorb the available labor leads to a decrease in people's purchasing power and reduced consumption, which in turn can result in a slowdown in economic growth. Open unemployment can also affect the psychological condition of affected individuals, thus impacting overall social welfare [5].

The issue of open unemployment is not only a concern at the national level, but also at the provincial level in Indonesia [6]. One of the provinces facing this challenge is West Sumatra Province [1]. To measure the condition of open unemployment, it can be seen through the value of the Open Unemployment Rate (OUR). OUR is the percentage value of the ratio of the number of unemployed to the labor force [7]. According to data from the BPS-Statistics

Indonesia, the OUR in West Sumatra Province during the 2017-2022 period showed fluctuations that reflected varying economic and social conditions. The data can be seen in Figure 1.

Figure 1. Open Unemployment Rate in West Sumatra 2017-2022

In Figure 1, it can be seen that the open unemployment rate in West Sumatra Province in 2017-2022 experienced fluctuations, which means that it went up and down every year. In the last 6 years, namely in 2018 the open unemployment rate increased by 0.08% from the previous year, then in 2019 the open unemployment rate decreased by 0.28% from the previous year, then in 2020 the open unemployment rate experienced a drastic increase of 1.5% from the previous year, then in 2021 it decreased by 0.36% from the previous year and decreased again in 2022 by 0.24% from the previous year.

The open unemployment rate can be influenced by factors related to the economic, social, and other factors. One of the previous studies in [7] modeled the open unemployment rate in Sumatra with the Multivariate Adaptive Rergessions Splines (MARS) approach with an R-square of 24.12 percent with influencing variables, namely the population aged 15 and over who worked a week ago with the main occupation of agriculture, plantations, forestry and fisheries, the population aged 15 and over who worked a week ago with the main occupation of the processing industry, the number of people aged 15 and over who are in school, the average monthly per capita expenditure for non-food, and gross regional domestic product at current prices. Then in research [8] which modeled the open unemployment rate in the Java Island Region with the MARS approach with an R-square of 86.73 percent with influencing variables namely population, population aged 15 and over who worked a week ago with the main occupation, monthly minimum wage, population aged 15 and over who went to school, average monthly per capita expenditure on non-food and gross domestic product at current prices.

This research will look at the influence of several aspects on the open unemployment rate, such as population on the total population, the population aged 15 years and over who worked a week ago with the main field of work and the population aged 15 years and over who are in school also affect the number of labor force [9]. Furthermore, it will be seen from the economic aspect in the form of a minimum wage per month, population expenditure factors such as monthly per capita expenditure on food, and non-food also have an influence, and Gross Regional Domestic Product at current prices [4]. An analysis is needed to see the overall influence of many influencing factors [10].

To examine the factors that are expected to affect the open unemployment rate problem, regression analysis is used. Regression analysis in this analysis technique can be done parametrically or nonparametrically. The parametric approach is applied if all classical assumptions are met and has a function and pattern known to the shape of the curve or follows a certain pattern model while the nonparametric approach is applied if the regression has a function and pattern that is not known to the shape of the curve or does not follow a certain pattern model on the relationship pattern between variables [11].

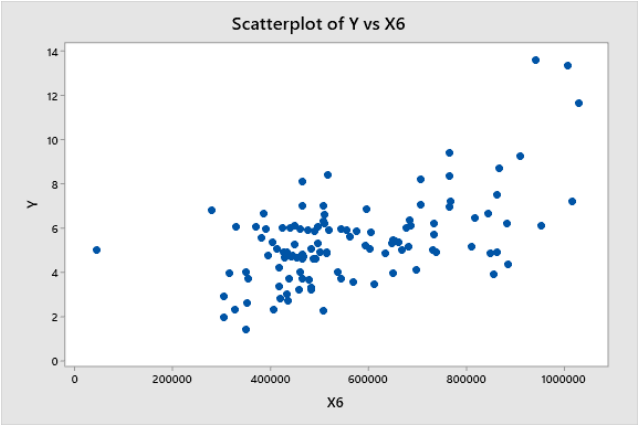

Figure 2. Data Plot Between the Value of the Open Unemployment Rate of West Sumatra Province and One of the Factors Suspected of Influence

Figure 2 is a data plot for the open unemployment rate with one of the factors estimated to affect it, namely the average monthly per capita expenditure on non-food items, which shows that the relationship pattern between variables does not follow a certain pattern model, thus indicating that a nonparametric approach is taken in this study. The nonparametric approach has one of the advantages, namely the existence of conditions that do not require the data to fulfill the assumptions of parametric models. In nonparametric cases, there is one method that is able to solve multivariable data cases, namely the MARS method. In [12], MARS can be used on data with as many observational data samples as there are independent variables. The MARS method is obtained from the results of the expansion of two method approaches, namely the Recursive Partition Regression (RPR) approach combined with the Spline method approach [11].

Based on the description above, this research will examine the analysis of the formation of the open unemployment rate model in West Sumatra and determine the factors that affect the open unemployment rate by applying the MARS method. Then the problem limitation in this research is the Open Unemployment Rate in the regency/city of West Sumatra province in 2017-2022. Thus, this research is titled "Factors Affecting the Open Unemployment Rate in West Sumatra Using the Multivariate Adaptive Regression Spline (MARS) Method".

2. THEORETICAL BASIC

2.1 Multivariate Adaptive Regression Splines (MARS)

Multivariate adaptive regression spline (MARS) is a nonparametric regression method developed by Jerome H. Friedman in 1991. MARS is designed to model the nonlinear relationship between the dependent variable (Y) and the independent variable (X) and capture the interaction between independent variables. In contrast to linear or polynomial regression, MARS does not require the assumption of a particular form of relationship between Y and X. Instead, MARS adaptively builds a nonlinear relationship between Y and X. Instead, MARS adaptively builds a model based on the data, using a combination of hinge functions to capture local patterns in the data [12]. In MARS, it is expected that each knot between regions (regression lines) has continuity from the basis function. The MARS model developed by Friedman (1991) is as follows:

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

$$\left(x_{v(k,m)} - t_{km}\right)_+ = \begin{cases} \left(x_{v(k,m)} - t_{km}\right)_+ ; x_{v(k,m)} - t_{km} > 0 \\ 0 \qquad\qquad\qquad\quad ; x_{v(k,m)} - t_{km} \leq 0 \end{cases} \tag{1}$$

Where, a_0 : constant, a_m : coefficient of the m-th basis function, M : maximum of the basis function, K_m : degree of interaction at the m-th basis function, S_{km} : sign of the knot point which has a value of ±1, if the knot is located in the right part of the subregion then the value is +1 and vice versa, $x_{v(k,m)}$: independent variable, t_{km} : knot value of the independent variable $x_{v(k,m)}$, v : number of independent variables, m : number of basis functions, k : number of interactions.

2.2 Generalized Cross Validation (GCV)

Generalized Cross Validation (GCV) is a method for selecting the best model by balancing predictive ability (fit) and model complexity (simplicity). GCV is often used in algorithms such as MARS, spline regression, and other statistical methods that involve adaptive model selection [12]. The main purpose of GCV is to prevent overfitting, which is when the model is too complex to learn the noise in the data, and underfitting, which is when the model is too simple to capture important patterns in the data. GCV has the advantages of being more efficient and simpler in calculation. Asymptotically, GCV has an optimal value, does not change transformation, and information related to σ^2 is not required. The modified form of the GCV criteria used in this stage is as follows:

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

Where, y : dependent variabel, x : independent variabel, $C(M)$: $Trace\ (X(X'X)^{-1}X') + 1$, indicating many estimated parameters, $\tilde{C}(M) : C(M) + d \cdot M$, $\hat{f}_M(x_i)$: estimated value of the dependent variables in M basis function x_i, M : estimated value is not constant in the basis function, d : the value of each basis function when $(2 \leq d \leq 4)$, N : many observation.

3. METHODS

This type of research is applied research. According to [13], applied research deals with practical realities, the application and development of knowledge produced by basic research in real life. Applied research is more often used to make it easier to provide solutions to problems at a practical stage. The data to be used in this research is secondary data obtained from the official website of the Central Bureau of Statistics in the form of data obtained from the National Labor Force Survey conducted in all regions of West Sumatra Province, entitled The State of the Labor Force in West Sumatra Province. The data to be studied includes data from 19 regencies/cities from West Sumatra Province from 2017 to 2022. Based on the research problem, the data will be analyzed using Salford Predictive Modeler 8.0 (SPM 8.0) software and Minitab 19. The work steps that will be carried out in analyzing the data are as follows:

1. Make descriptive statistics on the data of each variable that has been collected, both dependent and independent variables. These descriptive statistics are useful to show the characteristics of the data for each variable based on each district/city.
2. Making a scatterplot to show or display the pattern of data behavior between variables and in the form of a description of the relationship between variable Y and each variable X.
3. Determining the maximum number of basis functions (BF), or the maximum basis function allowed, is two to four times the number of independent variables used. In this study, the maximum number of basis functions used is 14, 21, and 28, based on the 7 independent variables used.
4. Determining the maximum number of interactions (MI), in this study, will be used as 1, 2, and 3 because if more than 3 interaction values are used, it will result in a very complex model interpretation.
5. Determining the minimum value of observation (MO) between knots. Performed by trial and error, this is because there is no fixed determination or foundation for determining the minimum observation between knots.
6. Obtain the best model based on the minimum CGV criteria obtained from a combination of possible BF, MI, and MO values, and perform parameterization of the model.
7. Perform a significance test on the MARS model by a simultaneous test and a partial test.
8. Interpret the model obtained and draw conclusions on the results of each test that has been fulfilled.

4. RESULTS AND DISCUSSION

4.1 Data Descriptions

The data description of this research is conducted to obtain an overview of the form of information contained in the open unemployment rate data and other factors that are thought to influence it.

Table 1. Descriptive Results for Research Data

Variable	Average	Standard Deviation	Minimum	Maximum
Y	5.43	1.97	1.39	13.64
X_1	288597	212528	52422	950871
X_2	20902	22189.8	3549	112324
X_3	129928	94382.4	12124	424213
X_4	2547372	355679	1894513	3339847
X_5	636298	95200.6	55505	873722
X_6	563372	18417	43628	1029923
X_7	12890.44	12855.4	29743.5	72978.95

In Table 1, it can be seen that the standard deviation for the open unemployment rate variables Y, X_2, X_5, and X_6 is relatively low, indicating that the data distribution on these variables spreads around the average, or there is no considerable difference in the data distribution of each variable. The variables X_1, X_3, X_4 and X_7 have relatively large standard deviation values, indicating that there is considerable diversity in the distribution of data or the data spreads far from the average value.

4.2 Results

Knot Data

At this stage, it is important to see the estimated presence of knot points with smoother lines on the data distribution and changes in the regression line on each segmented independent variable (X).

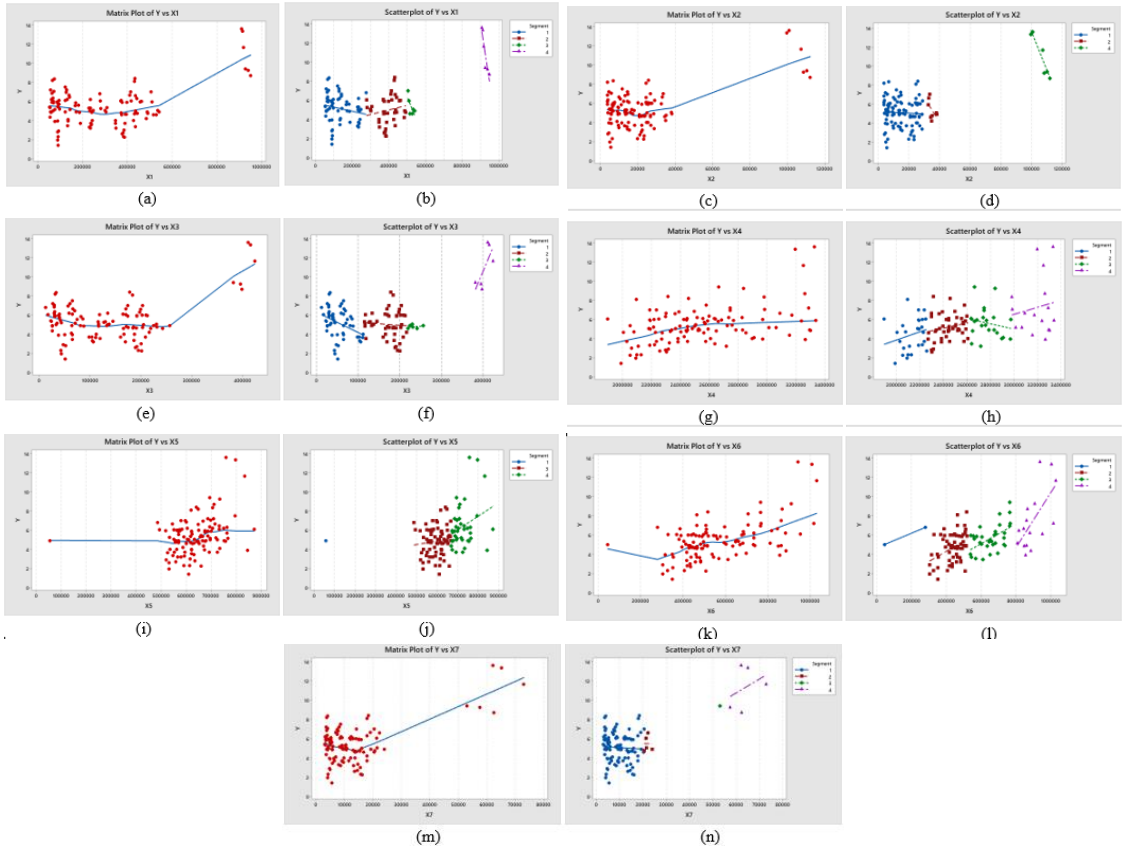

Figure 3. Smoother Lines and Graph of Changes in Regression Line Data that has been Segmented

Figure 3 is a scatterplot containing smoother lines and changes in the regression line for the value of the open unemployment rate (Y) as the vertical axis against each independent variable (X) as the horizontal axis. From the results in Figure 3, it can be seen that the data pattern is spread or does not show a tendency to follow a certain pattern as seen with the distribution of data on the relationship between the value of the open unemployment rate (Y) and each independent variable (X), so it can be suspected that there are knots in the data which are also clarified by the shape of each smoother line in parts (a), (c), (e), (g), (i), (k), and (m) which are not in the form of straight or linear lines. In parts (b), (d), (f), (h), (j), (l), and (n), the horizontal axis or data of each segmented variable X forms a regression line that changes or is not connected to each other following a certain pattern and the data is on different slopes.

Model Formation of Open Unemployment Rate with MARS Method

The formation of the open unemployment rate model in Sumatra is obtained from the combination of the maximum number of Basis Functions (BF), Maximum Interaction (MI), and minimum observation (MO) that have been determined. The basis function obtained will be in the form of the following equation:

$$B_m^{(q)}(x) = \prod_{k=1}^{K_m} \left[S_{km} \cdot \left(x_{v(k,m)} - t_{km}\right)\right]_+^q \tag{3}$$

Where, q : polynomial order, B_m : m-th basis function on variabel X, K_m : degree of interaction in the m-th basis function, S_{km} : The sign of the knot point has a value of ±1, if the knot is located in the right part of the subregion then the value is +1 and vice versa, $x_{v(k,m)}$: $v(k,m)$-th independent variable, t_{km} : knot value of independent variable $x_{v(k,m)}$, v : number of independent variables, m : number of base functions and k : number of interactions.

The basis functions used are 14, 21 and 28 basis functions based on the 7 independent variables used. The maximum number of interactions (MI) used is 1 to 3 interactions and the Minimum Observations (MO) used are 0, 1, 2, 3, 4 and 5 minimum observations because the minimum Generalized Cross Validation (GCV) value after trial and error is found at these values. The best model is obtained from the combination of BF = 28, MI = 2 and MO = 1 with GCV value = 1.31675 as the minimum GCV value and R^2 adj value = 0.816 which shows that the percentage of open unemployment rate can be explained by the independent variables $X_7, X_1, X_2, X_3, X_5, X_6$ and X_1 by 81.6 percent with the remaining 18.4 percent can be explained by other factors outside the observation, the form of the model obtained is as follows:

$$\begin{aligned} Y = {} & 3.61546 - 9.95102 \times 10^{-10} \times FB4 - 0.000399319 \times FB5 + 7.02303 \times 10^{-11} \times FB9 \\ & + 2.89267 \times 10^{-8} \times FB15 - 2.73448 \times 10^{-11} \times FB16 + 3.60182 \times 10^{-10} \times FB17 \\ & + 1.28127 \times 10^{-9} \times FB18 - 2.11137 \times 10^{-8} \times FB19 - 7.90304 \times 10^{-5} \times FB20 \\ & + 4.18187 \times 10^{-5} \times FB22 + 3.2243 \times 10^{-5} \times FB24 + 8.6633 \times 10^{-10} \times FB27 \\ & + 2.64556 \times 10^{-10} \times FB28; \end{aligned}$$

with the Basis Function (BF) in the model as follows:

$FB1 = max(0, X_7 - 14603.2);$
$FB2 = max(0, 14603.2 - X_7);$
$FB4 = max(0, 2.49433 \times 10^6 - X_4) \times FB2;$
$FB5 = max(0, X_2 - 100941);$
$FB6 = max(0, 100941 - X_2);$
$FB9 = max(0, X_6 - 351786) \times FB6;$
$FB15 = \max(0, 4793.15 - X_7) \times FB6;$
$FB16 = max(0, X_4 - 2.02804 \times 10^6) \times FB6;$
$FB17 = max(0, 2.02804 \times 10^6 - X_4) \times FB6;$
$FB18 = max(0, X_1 - 73438) \times FB2;$
$FB19 = max(0, 73438 - X_1) \times FB2;$
$FB20 = max(0, X_3 - 144718);$

$FB22 = max(0, X_4 - 3.2655 \times 10^6)$;
$FB24 = max(0, X_1 - 275728)$;
$FB27 = \max(0,191709 - X_3) \times FB2$;
$FB28 = max(0, X_5 - 55505) \times FB1$;
and the level of each variable to the model is 100 percent for X_7, 82.01 percent for X_1, 70.01 percent for X_2, 67.74 percent for X_3, 62.49 percent for X_5, 62.41 percent for X_6, 47.74 percent for X_4.

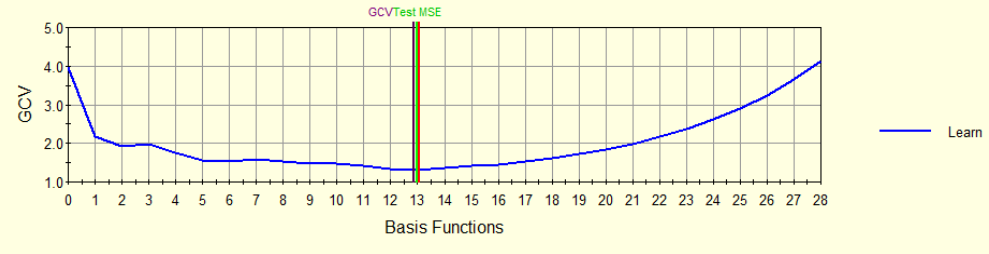

Figure 4. MARS Result

In Figure 4, it can be seen that 28 maximum BFs are 13 BFs that contribute to generating GCV.

Model Significance Test

The following are the results of the significance test of the open unemployment rate model obtained to see the significance of the best model parameters.

Concurrent Test (F Test)

The simultaneous test is conducted with the following hypothesis:

$$H_0 : a_4 = a_5 = a_9 = a_{15} = a_{16} = a_{17} = a_{18} = a_{19} = a_{20} = a_{22} = a_{24} = a_{27} = a_{28} = 0$$

H_1:there is $a_m \neq 0$,

with m = 4,5,9,15,16,17,18,19,20,22,24,27 and 28
the significance level (α) used is 0.05.

Table 2. Simultaneous Test Results

F-Statistic	39.46039
P-Value	0.00000
[MDF,NDF]	[13,100]
S.E. of Regression	0.85019
Residual Sum of Squares	72.28220
Regression Sum of Squares	370.79691

Based on Table 2, the results of the simultaneous test on the best model parameters with a significance level $\alpha = 0.05$ obtained F-Statistic value = 39.46039 and $p - value = 0.00000$ and based on the F table obtained $F_{0.05(13;100)} = 1.82$. It can be concluded that there is a rejection of the null hypothesis (H_0) or the model is significant, because the p-value $< \alpha$ or $F_{count} > F_{0.05(13;100)}$.

Partial Test (t Test)

Partial tests are carried out with the following hypothesis:

$H_0 : a_m = 0$

H_1:there is $a_m \neq 0$,

with m = 4,5,9,15,16,17,18,19,20,22,24,27 and 28
the significance level (α) used is 0.05

Table 3. Partial Test Results

Parameter	Estimate	S.E.	T.Ratio	P-Value
Constant	3.61546	0.39163	9.23174	0.00000
Basis Function 4 (BF4)	-0.00000	0.00000	-6.93972	0.00000
Basis Function 5 (BF5)	-0.00040	0.00008	-5.01467	0.00000
Basis Function 9 (BF9)	-0.00000	0.00000	7.02730	0.00000
Basis Function 15 (BF15)	-0.00000	0.00000	5.27896	0.00000
Basis Function 16 (BF16)	-0.00000	0.00000	-5.12418	0.00000
Basis Function 17 (BF17)	-0.00000	0.00000	4.17802	0.00000
Basis Function 18 (BF18)	-0.00000	0.00000	4.15642	0.00000
Basis Function 19 (BF19)	-0.00000	0.00000	-5.69722	0.00000
Basis Function 20 (BF20)	-0.00008	0.00001	-7.58853	0.00000
Basis Function 22 (BF22)	-0.00004	0.00001	4.29584	0.00000
Basis Function 24 (BF24)	-0.00003	0.00000	8.68860	0.00000
Basis Function 27 (BF27)	-0.00000	0.00000	3.01621	0.00000
Basis Function 28 (BF28)	-0.00000	0.00000	7.03543	0.00000

Based on the partial test results on the model parameters, it can be seen that each BF included in the model has a p-value < α. This indicates the rejection of H_0 and it can be concluded that the coefficient a_m for $m = 4,5,9,15,16,17,18,19,20,22,24,27,$ and 28 has an effect on the model.

4.3 Discussion

Interpretation of the coefficients of the best model obtained by meeting the minimum GCV criteria using the MARS method is as follows:

1. $FB4 = max(0,2.49433 \times 10^6 - X_4) \times FB2$ with coefficient -9.95102×10^{-10}
   $FB2 = max(0,14603.2 - X_7)$
   So $FB4 = max(0,2.49433 \times 10^6 - X_4) \times max(0,14603.2 - X_7)$
   This means that each increase in $FB4$ by one unit will reduce the percentage value of OUR (Y) by 9.95102×10^{-10} if the value of the minimum monthly wage (X_4) is less than 2.49433×10^6 and the value of GRDP at current prices (X_7) is less than 14603.2.
2. $FB5 = max(0, X_2 - 100941)$ with a coefficient of -0.000399319; meaning that each increase in $FB5$ by one unit will reduce the percentage value of OUR (Y) by 0.000399319 if the value of the number of people aged 15 years and over who are in school (X_2) is more than 100941.
3. $FB9 = max(0, X_6 - 351786) \times FB6$ with coefficient 7.02303×10^{-11}
   $FB6 = max(0,100941 - X_2)$
   So $FB9 = max(0, X_6 - 351786) \times max(0,100941 - X_2)$
   This means that each increase in $FB9$ by one unit will increase the value of the percentage of OUR (Y) by 7.02303×10^{-11} if the value of the average monthly per capita expenditure on non-food (X_6) is more than 351786 and the number of people aged 15 years and over who are in school (X_2) is less than 100941.
4. $FB15 = max(0,4793.15 - X_7) \times FB6$ with coefficient 2.89267×10^{-8}
   $FB6 = max(0,100941 - X_2)$
   So $FB15 = max(0,4793.15 - X_7) \times max(0,100941 - X_2)$
   This means that each increase in $FB15$ by one unit will increase the percentage value of OUR (Y) by 2.89267×10^{-8} if the value of GDP at current prices (X_7) is less than 4793.15 and the number of people aged 15 years and over who are in school (X_2) is less than 100941.
5. $FB16 = max(0, X_4 - 2.02804 \times 10^6) \times FB6$ with coefficient -2.73448×10^{-11}
   $FB6 = max(0,100941 - X_2)$
   So $FB16 = max(0, X_4 - 2.02804 \times 10^6) \times max(0,100941 - X_2)$
   This means that each increase in $FB16$ by one unit will reduce the percentage value of OUR (Y) by 2.73448×10^{-11} if the value of the minimum monthly wage (X_4) is more than 2.02804×10^6 and the number of people aged 15 years and over who are in school (X_2) is less than 100941.
6. $FB17 = max(0,2.02804 \times 10^6 - X_4) \times FB6$ with coefficient 3.60182×10^{-10}
   $FB6 = max(0,100941 - X_2)$
   So $FB17 = max(0,2.02804 \times 10^6 - X_4) \times max(0,100941 - X_2)$

This means that each increase in $FB17$ by one unit will increase the percentage value of OUR (Y) by 3.60182×10^{-10} if the value of the minimum monthly wage (X_4) is less than 2.02804×10^6 and the value of GRDP at current prices (X_7) is less than 100941.

7. $FB18 = max(0, X_1 - 73438) \times FB2$ with a coefficient of 1.28127×10^{-9}
   $FB2 = max(0,14603.2 - X_7)$
   So $FB18 = max(0, X_1 - 73438) \times max(0,14603.2 - X_7)$
   This means that every one unit increase in $FB18$ will increase the value of the percentage of OUR (Y) by 1.28127×10^{-9} if the value of population (X_1) is more than 73438 and the value of GRDP at current prices (X_7) is less than 14603.2.
8. $FB19 = max(0,73438 - X_1) \times FB2$ with coefficient -2.11137×10^{-8}
   $FB2 = max(0,14603.2 - X_7)$
   So $FB19 = max(0,73438 - X_1) \times max(0,14603.2 - X_7)$
   This means that for every one unit increase in $FB19$, the percentage value of OUR (Y) will decrease by 2.11137×10^{-8} if the value of total population (X_1) is less than 73438 and the value of GRDP at current prices (X_7) is less than 14603.2.
9. $FB20 = max(0, X_3 - 144718)$ with a coefficient of -7.90304×10^{-5}; meaning that each increase in $FB20$ by one unit will reduce the percentage value of OUR (Y) by 7.90304×10^{-5} if the value of the number of people aged 15 years and over who worked a week ago with the main job (X_3) is more than 144718.
10. $FB22 = max(0, X_4 - 3.2655 \times 10^6)$ with a coefficient of 4.18187×10^{-5}; meaning that each increase in $FB22$ by one unit will increase the percentage value of OUR (Y) by 4.18187×10^{-5} if the value of the minimum monthly wage (X_4) is more than 3.2655×10^6.
11. $FB24 = max(0, X_1 - 275728)$ with a coefficient of 3.2243×10^{-5}; meaning that each increase in $FB24$ by one unit will increase the percentage value of OUR (Y) by 3.2243×10^{-5} if the value of total population (X_1) is more than 275728.
12. $FB27 = max(0,191709 - X_3) \times FB2$ with a coefficient of 8.6633×10^{-10}
    $FB2 = max(0,14603.2 - X_7)$
    So $FB27 = max(0,191709 - X_3) \times max(0,14603.2 - X_7)$
    This means that for every one unit increase in $FB27$ will increase the percentage value of OUR (Y) by 8.6633×10^{-10} if the value of the number of people aged 15 years and over who worked a week ago with the main job (X_3) is less than 191709 and the value of GRDP at current prices (X_7) is less than 14603.2.
13. $FB28 = max(0, X_5 - 55505) \times FB1$ with coefficient 2.64556×10^{-10}
    $FB1 = max(0, X_7 - 14603.2)$
    So $FB28 = max(0, X_5 - 55505) \times max(0, X_7 - 14603{,}2)$.
    This means that it can be interpreted that for every increase in BF28 by one unit, it will increase the percentage of OUR (Y) by 2.64556×10^{-10}or the percentage of open unemployment rate (Y) will increase by 2.64556×10^{-10} if the value of the average monthly per capita expenditure on food (X_5) is more than 55505 and the value of GRDP at current prices (X_7) is more than 14603.2.

If we look closely, all of the estimated coefficients, both positive and negative, are very small or close to zero. This is normal because the explanatory variables in this study have a fairly large scale, for example, minimum wage and GRDP, which can reach thousands to millions. The large scale makes the coefficient values appear very small numerically. In other words, the direction of the influence of the variables can still be interpreted, whether it increases or decreases OUR, but the magnitude of the influence is relatively weak quantitatively compared to other variables with larger coefficients.

5. CONCLUSION

Based on the results and discussion in this research analysis, the following conclusions are obtained:

1. The best model for the open unemployment rate in West Sumatra Province in 2017-2022 with the MARS method is obtained in the combination of BF = 28, MI = 2, and MO = 1 because it has a minimum GCV value of 1.31675 with an R^2 of 0.83686 and an R^2 adj value on the model of 0.81566. The form of the model equation is as follows:

$$Y = 3.61546 - 9.95102 \times 10^{-10} \times FB4 - 0.000399319 \times FB5 + 7.02303 \times 10^{-11} \times FB9 + 2.89267 \times 10^{-8} \times FB15 - 2.73448 \times 10^{-11} \times FB16 + 3.60182 \times 10^{-10} \times FB17 + 1.28127 \times 10^{-9} \times FB18 - 2.11137 \times 10^{-8} \times FB19 - 7.90304 \times 10^{-5} \times FB20 + 4.18187 \times 10^{-5} \times FB22 + 3.2243 \times 10^{-5} \times FB24 + 8.6633 \times 10^{-10} \times FB27 + 2.64556 \times 10^{-10} \times FB28;$$

with the Basis Function (BF) in the model as follows:

$FB1 = max(0, X_7 - 14603.2);$
$FB2 = max(0,14603.2 - X_7);$
$FB4 = max(0,2.49433 \times 10^6 - X_4) \times FB2;$
$FB5 = max(0, X_2 - 100941);$
$FB6 = max(0,100941 - X_2);$
$FB9 = max(0, X_6 - 351786) \times FB6;$
$FB15 = \max(0,4793.15 - X_7) \times FB6;$
$FB16 = max(0, X_4 - 2.02804 \times 10^6) \times FB6;$
$FB17 = max(0,2.02804 \times 10^6 - X_4) \times FB6;$
$FB18 = max(0, X_1 - 73438) \times FB2;$
$FB19 = max(0,73438 - X_1) \times FB2;$
$FB20 = max(0, X_3 - 144718);$
$FB22 = max(0, X_4 - 3.2655 \times 10^6);$
$FB24 = max(0, X_1 - 275728);$
$FB27 = max(0,191709 - X_3) \times FB2;$
$FB28 = max(0, X_5 - 55505) \times FB1;$

2. The factors that significantly influence the percentage of open unemployment rate in West Sumatra Province in 2017-2022 are gross regional domestic product at current prices (X_7) with a variable importance level of 100 percent, total population (X_1) with a variable importance level of 82.01 percent, population aged 15 years and over who are in school (X_2) with a variable importance level of 70.01 percent, population aged 15 years and over who worked a week ago with the main occupation (X_3) with an importance level of 67.74 percent, average monthly per capita expenditure on food (X_5) with a variable importance level of 62.49 percent, average monthly per capita expenditure on non-food (X_6) with a variable importance level of 62.41 percent and minimum wage per month (X_4) with an importance level of 47.74 percent. So it can be said that all variables in this study affect the level of open unemployment in West Sumatra Province in 2017-2022.
3. The factor that has the most influence on the Open Unemployment Rate (OUR) in West Sumatra Province during the 2017-2022 period is Gross Regional Domestic Product (GRDP) at current prices (X_7). This variable has an importance level of 100%, which means that the contribution of GRDP is very dominant in influencing changes in the OUR in West Sumatra Province in 2017-2022. In other words, an increase or decrease in GRDP in West Sumatra in 2017-2022 has the greatest impact on the unemployment rate in the region, compared to other factors studied.

REFERENCES

[1] BPS, *Perkembangan Pembangunan Provinsi Sumatera Barat 2017-2022*. Badan Pusat Statistik Provinsi Sumatera Barat, 2022.
[2] Veritia, I. Lubis, I. A. Priatna, dan Susanto, *Teori Ekonomi Makro*, no. 1. Banten: UNPAM PRESS, 2019.
[3] S. S. Atmodjo dan I. A. Sumbogo, *Pengantar ekonomi makro*, no. May. Jakarta: PT Kreasi Skrip Dijital, 2023.
[4] BPS, *Keadaan Angkatan Kerja di Provinsi Sumatera Barat Agustus 2019*. Badan Pusat Statistik Provinsi Sumatera Barat, 2019.
[5] J. Voßemer, M. Gebel, K. Täht, M. Unt, B. Högberg, dan M. Strandh, "The Effects of Unemployment and Insecure Jobs on Well-Being and Health: The Moderating Role of Labor Market Policies," *Soc. Indic. Res.*, vol. 138, no. 3, hal. 1229–

1257, 2018, doi: 10.1007/s11205-017-1697-y.
[6] BPS, *Keadaan Angkatan Kerja di Provinsi Sumatera Barat 2018*. Badan Pusat Statistik Provinsi Sumatera Barat, 2018.
[7] L. Andani, "Analisis Faktor-Faktor yang Mempengaruhi Tingkat Pengangguran Terbuka di Sumatera Menggunakan Metode Multivariate Adaptive Regression Spline (MARS)," *UNPjoMath*, vol. 3, no. 3, hal. 61–67, 2020.
[8] R. P. Sari, "Faktor-faktor yang Mempengaruhi Pengangguran Terbuka di Wilayah Pulau Jawa Menggunakan Metode Multivariate Adaptive Regression Spline (MARS)," Universitas Negeri Padang, 2022.
[9] BPS, *Booklet SAKERNAS*. Badan Pusat Statistik, 2023.
[10] T. J. Wardani, "Faktor-faktor yang Mempengaruhi Ketidakmerataan Jumlah Penduduk di Indonesia Menggunakan Analisis Faktor," *UNPjoMath*, vol. 2, no. 4, hal. 39–44, 2019.
[11] J. Kloke dan J. W. Mckean, *Nonparametric Statistical Methods Using R*. CRC Press, 2015.
[12] J. Friedman, "Multivariate Adaptive Regression Splines," *Tha Ann. Stat.*, vol. 19, hal. 1–141, 1991.
[13] Sudaryono, *Metode Penelitian*. Jakarta: Rajawali Press, 2017.