

MODELING THE LEVEL OF OPEN UNEMPLOYMENT IN CENTRAL JAVA WITH MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS) APPROACH

Erma Oktania Permatasari¹, Firda Nasuha², Carlos L. Prawirosastro³

^{1,2} Statistics, Institut Teknologi Sepuluh Nopember, Surabaya 60111, Indonesia

³ Universitas Hang Tuah, Surabaya 60111, Indonesia

e-mail: erma.oktania@gmail.com

ABSTRACT

The level open unemployment is a value that shows the number of working-age population who looking for work, is preparing a business, feels impossible to get a job or already have a job but have not started working and often used for measured employment. Like at Central Java has increasing the total population at 2014 and have high total investment whereas should be can getting more employment, but actually still give high unemployment about 996.344 population at 2014. So that, in this research used nonparametric regression approach which multivariate adaptive regression splines (MARS) for modeling the level open unemployment in Central Java at 2014 because the level open unemployment in Central Java predicted influence by some factors. This research resulted in the best modeling for level of open unemployment in Central Java Province with value of GCV minimum that obtained at 0,396 with R-square at 86,5 percent as well as the predictor variables were entered into the model as much as three, namely the total population with interest rate of 100 percent, the minimum wage with interest rate of 41,955 percent, and the total working population with interest rate of 39,547 percent.

Keywords: MARS, Nonparametric regression, Level of open unemployment

1. INTRODUCTION

Indonesia as one of the developing countries have national development goals to be achieved, namely building a complete Indonesian people both materially and spiritually. One of the indicators used to assess the improvement of the quality of human resources is through work ethic possessed by each individual. But in this era of globalization, they are faced with a wide range of employment issues are characterized by employment growth rate is not comparable with the rate of growth of the labor force, low quality, competitiveness and labor productivity and investment climate was not conducive [1]. The highest number of populations in Indonesia is located on the island of Java. As is the case in Central Java, which increased total population of 258.324 inhabitants in 2014 with the highest contribution in Central Java is a textile investment [2]. But the problem today is that high total investment which can absorb a large amount of manpower, in fact they provide the number of high unemployment. The definition of the level open unemployment is a value that indicates the number of working-age population who are looking for work, or preparing for business, or find it impossible to get a job, or already have a job but have not started working.

Previously had done some research on the level open unemployment rate. Among them are studies on the classification of districts / cities in East Java based the level open unemployment rate to approach MARS [3], the research on cluster analysis and discriminant analysis to classify the unemployment rate in the province of East Java [4], the research on modeling of unemployment in East Java uses regression spline [5], the research on modeling of unemployment in East Java using Fourier series [6], the research on the classification of unemployment based on the factors affecting the open unemployment provinces in North Sulawesi with methods CART [7], research on statistical analysis of unemployment rates East Java in 2012 [8] and the latest research on modeling the open unemployment rate in Central Java using regression spline [9]. Therefore, in this study will be used nonparametric methods, multivariate adaptive regression splines (MARS) to perform modeling of the level open unemployment in the districts / cities in Central Java in 2014 due to the level open unemployment in Central Java, allegedly influenced by several factors. In this case, it will also do a prediction based on the best model MARS obtained.

2. LITERATURE REVIEW

2.1 Multivariate Adaptive Regression Splines (MARS)

Multivariate Adaptive Regression Splines (MARS) is one approach to multivariate nonparametric regression [10]. MARS models focused to overcome the problems of high dimension, has many variables, as well as a large sample size so that the necessary complex calculations. That is why, MARS method can be used on high-dimensional data. Some things to consider in using MARS include knots and basis function [11]. Knot is a point that separates the end of a data area with the

start of a data area to another. The number of observations on each knot is called minimum observation (MO). MO used are 0, 1, 2 and 3. Basis function (BF) is a function that is separated by dots knots which describes the relationship between the predictor variables and the response variable. Methods MARS form the basis function that states the number of BF is 2 to 4 times the number of predictor variables, while the maximum amount of interaction (MI) was 1, 2 and 3 with consideration if MI > 3 will generate complex models and interpretation will be almost the same.

2.2 Modeling on MARS

Then obtained estimator for MARS models as follows:

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

Where:

| | |
|--------------|--|
| a_0 | = constant coefficient of the basis function B_0 |
| a_m | = coefficient of the basis function to- m |
| m | = number of the basis function |
| K_m | = number of interactions on the basis function all- m |
| k | = number of interactions |
| s_{km} | = value +1 (knots on the right) and the value -1 (knots on the left) |
| $x_{v(k,m)}$ | = predictor variable |
| t_{km} | = value knots of predictor variables |

In general, equation (2) can also be written as follows:

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

In equation (3) shows that the first sum covers all basis for single-variable function, the second summation function covers all basis for the interaction between two variables, the third summation function covers all basis for the

interaction between the three variables and so on.

From equation (1) and (2), if written in the form matrix be as follows:

$$\mathbf{y} = \mathbf{B}\mathbf{a} + \boldsymbol{\varepsilon}$$

With,

$$\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)^T, \mathbf{a} = (\mathbf{a}_0, \dots, \mathbf{a}_M)^T, \boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n)^T$$

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} s_{1m}(x_{1(1,m)} - t_{1m}) & \dots & \prod_{k=1}^{K_M} s_{km}(x_{1(k,m)} - t_{km}) \\ 1 & \prod_{k=1}^{K_1} s_{1m}(x_{2(1,m)} - t_{1m}) & \dots & \prod_{k=1}^{K_M} s_{km}(x_{2(k,m)} - t_{km}) \\ \dots & \dots & \dots & \dots \\ 1 & \prod_{k=1}^{K_1} s_{1m}(x_{n(1,m)} - t_{1m}) & \dots & \prod_{k=1}^{K_M} s_{km}(x_{n(k,m)} - t_{km}) \end{bmatrix}$$

In MARS, some things need to be considered in the selection of the most optimum model (the best) is that if the value of the GCV model has

the lowest value (minimum) among other models [12].

Minimum GCV function is defined in the following equation:

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

Where:

- x_i = the value of the predictor variables on the observation of the i
- y_i = the value of the variable responses to observations i
- n = number of observations or the amount of data
- $C(M)$ = the number of parameters in the model = trace $\left[\mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T\right] + 1$
- $\hat{f}(x_i)$ = the estimated value of the variable responses to observations i

2.3 Labor Force

In Indonesia, the labor force is the working age population (15 years and over) who work, or have a job but temporarily absent from work and unemployment.

2.4 The Level Open Unemployment

The level open unemployment is a value that indicates the number of working-age population who are looking for work, or preparing for business, or find it impossible to get a job, or already have a job but have not started working.

2.5 Factors Influencing the Level Open Unemployment

The main problem further and fundamental in employment in Indonesia is the problem of low wages and a direct and indirect effect on the unemployment rate is high. Determination of the level of wages by the government of a country have an impact on the level of unemployment that exist. The higher the wage set by the government will result in a decrease in the number of people working in the country [14]. Further economic indicators that affect the level of unemployment is economic

growth. Increased economic growth is expected to absorb labor in the country. Due to the increase in national income can increase production capacity. This indicates that the decrease in the rate of economic growth can be attributed to the high number of unemployed in a country [15].

3. RESEARCH METHODOLOGY

3.1 Data Source

The data used in this research is secondary data from districts / cities in Central Java province, as many as 29 districts and 6 cities in 2014 were obtained from the Central Statistics Agency of Central Java Province and the Office of Manpower, Transmigration and Population Central Java Province.

3.2 Variable

The variables used in this research as follows:

- a. The response variable (Y) is the percentage of the level open unemployment at 2014 from districts / cities in Central Java.
- b. The predictor variables (X) consisting of variables supposed to influence the

response variable. Here is the predictor variables were used.

- 1) The minimum wage each district / city (X_1)
- 2) The working population of each county / city (X_2)
- 3) Regional economic growth each district / city (X_3)
- 4) The population aged 15 years and over with at least one high school and over who worked each district / city (X_4)
- 5) Number of Indonesian Workers in each district / city (X_5)
- 6) The population of each district / city (X_6).

3.3 Step of Analysis Data

The following step of analysing in this research is:

- a. Conduct descriptive analysis of the response variable (Y) and the predictor variable (X).
- b. Plot matrix formed between the response variable (Y) with each predictor variable (X) as an early detection patterns of relationships between variables.

- c. Determine the maximum basis function (BF), which is 2 to 4 times the number of predictors to be used.
- d. Determine the maximum amount of interaction (MI) of 1.2, and 3 with the assumption that if $MI > 3$ will produce increasingly complex models.
- e. Determine the minimum number of observations of each knot (MO), ie 0, 1, 2, and 3.
- f. Establish best model with the minimum value of GCV.
- g. Grouping basis function based on the predictor variables that enter the model.
- h. Interpret the level of contributions that have an interest in grouping the response variable.
- i. Doing the predictions of the model obtained the best MARS.

4. RESULT AND ANALYSIS

4.1 Analysis Descriptive Variables

Prior to modeling the level open unemployment in Central Java with MARS, first performed a descriptive analysis of each variable in this research. Here is an analysis descriptive of each variable.

Table 1. Descriptive Variables

| Variables | Mean | Standard Deviation | Minimum | Maximum | Lowest | Highest |
|-------------------------|-----------|--------------------|---------|------------|--------------------|-----------------|
| Y (%) | 2,86 | 1,69 | 0,41 | 8,07 | Salatiga City | Brebes District |
| X ₁ (rupiah) | 1.069.457 | 110.434 | 910.000 | 1.423.500 | Purworejo District | Semarang City |
| X ₂ (people) | 472.877 | 189.116 | 59.628 | 820.317 | Magelang City | Semarang City |
| X ₃ (rupiah) | 2.500.805 | 2.159.547 | 559.523 | 12.479.508 | Magelang City | Semarang City |

Table 2. Descriptive Variables

| | | | | | | |
|-------------------------|---------|---------|---------|-----------|---------------|------------------|
| X ₄ (people) | 127.124 | 82.538 | 32.163 | 495.322 | Magelang City | Semarang City |
| X ₅ (people) | 2.645 | 3.345 | 45 | 16.001 | Magelang City | Cilacap District |
| X ₆ (people) | 957.790 | 411.681 | 120.373 | 1.773.379 | Magelang City | Brebes District |

Table 1 provides information that Brebes District has the highest level open unemployment in Central Java, which amounted to 8,07 percent. The high level open unemployment in Brebes could occur due to the agricultural sector as the most dominant sector in Brebes has not been able to absorb the amount of the labor force, with a total labor

force 844.001 people and a total population employed in Brebes district 763.581 people. It is also occur because Brebes has the highest number of inhabitants in Central Java. While Salatiga City which has the smallest population in Central Java and the smallest area has the lowest unemployment rate, which amounted to 0,41 percent. Salatiga City able to explore

some other sectors such as agriculture, industry and processing sectors to absorb labor in terms of reducing the level open unemployment.

For variables suspected to affect the level open unemployment, Semarang City always had the highest value except for the variables X_5 and X_6 . This could happen because Semarang City is the capital of Central Java province, so that all aspects of community life centered in Semarang City as aspects economy, education, and employment which has related with variables X_1 , X_2 , X_3 , and X_4 . Purworejo District has the lowest variable X_1 amounted of 910,000 rupiah. The low minimum wage is based on ability of the company and the agreement between the government and employers in the area to provide wages to workers or employees. While Magelang City always have the lowest values except for the variable X_1 . Magelang City has the lowest population in Central Java amounted of 120.373 people. So, it becomes the influence of Magelang City has the lowest contributor to these variables. The highest for variable X_5 owned by Cilacap District and the highest of variable X_6 is Brebes District which also has the

highest level open unemployment in Central Java.

In addition, it can be seen that the standard deviation from data the level open unemployment in Central Java is relatively low. This indicates that each district / city in Central Java has a percentage of the level open unemployment is almost the same or not much different between one region and another region with an average the level open unemployment 2,86 percent.

4.2 Modelling the Level Open Unemployment With MARS

From Figure 1 below shows has a different pattern and do not show a clear pattern. In addition, due to a lack of information form and the function do not show a clear pattern from relation between the response variable with predictor variables, in this research using a nonparametric regression approach. In this case, nonparametric regression approach is used to modelling the level open unemployment in Central Java with variables suspected is MARS method because of high dimensional data.

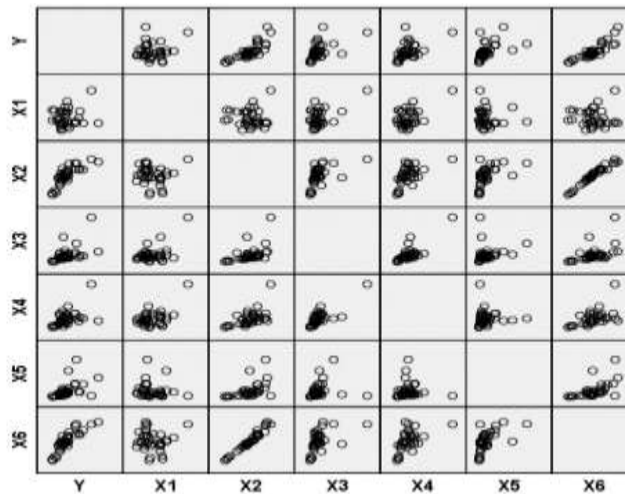


Figure 1. Modelling the Level Open Unemployment With MARS

In this research, the number of predictor variables are supposed to influence the level open unemployment are six variables, combinations of basis function (BF) that used in the modelling are 12, 18, and 24. The result of the best model in this research is based on the concept of parsimony (simple model) from combination basis function (BF) 12 with $MO = 1$

and $MI = 3$. The value of minimum GCV is 0,396 with R-square is 86,5 percent as well as predictor variables that entered into the model as much as three, namely X_1 , X_2 , and X_6 . Based on the procedures that have done, then the best MARS model from the level open unemployment in Central Java are:

$$\hat{f}(\mathbf{x}) = 3,776 + 0,213599 \times 10^{-4}BF_1 - 0,206779 \times 10^{-5}BF_2 + 0,260585 \times 10^{-4}BF_3 + 0,495982 \times 10^{-15}BF_7$$

With basis function is obtained as follows:

$$BF_1 = \max(0, X_6 - 1.620.918);$$

$$BF_2 = \max(0, 1.620.918 - X_6);$$

$$BF_3 = \max(0, X_2 - 737.931);$$

$$BF_4 = \max(0, 737.931 - X_2);$$

$$BF_5 = \max(0, X_6 - 1.106.328)BF_4;$$

$$BF_7 = \max(0, X_1 - 910.000,063)BF_5;$$

It will be explain the interpretation of each basis function from equation (10) as follows:

$$BF_1 = \max(0, X_6 - 1.620.918) = \begin{cases} (X_6 - 1.620.918), & \text{jika } X_6 > 1.620.918 \\ 0, & \text{jika } X_6 \leq 1.620.918 \end{cases}$$

Each increase of one unit basis function (BF₁) will increase the percentage of the level open unemployment in district / city Central Java amounted of $0,213599 \times 10^{-4}$ percent if the population of each district / city

more than 1.620.918 people, with other basis function are included in the model considered constant.

$$BF_2 = \max(0, 1.620.918 - X_6) = \begin{cases} (1.620.918 - X_6), & \text{jika } X_6 < 1.620.918 \\ 0, & \text{jika } X_6 \geq 1.620.918 \end{cases}$$

Each increase of one unit basis function (BF₂) will decrease the percentage of the level open unemployment in district / city Central Java amounted of $0,206779 \times 10^{-5}$ percent if the population of each district / town

less than 1.620.918 people, with other basis function are included in the model considered constant.

$$BF_3 = \max(0, X_2 - 737.931) = \begin{cases} (X_2 - 737.931), & \text{jika } X_2 > 737.931 \\ 0, & \text{jika } X_2 \leq 737.931 \end{cases}$$

Each increase of one unit basis function (BF₃) will increase the percentage of the level open unemployment in district / city Central Java amounted of $0,260585 \times 10^{-4}$ percent if the total working population in district / city more than 737.931 people with other basis functions are included in the model considered constant. The population in each district / city Central Java has increased every year. In 2014, the

population in Central Java has increased by 258.324 people from the previous year. Thus, increasing the number of working populations as follow as increasing the population in Central Java. Increasing the number of people in area, it will cause the trend of rising unemployment in the area.

$$BF_4 = \max(0, 737.931 - X_2) = \begin{cases} (737.931 - X_2), & \text{jika } X_2 < 737.931 \\ 0, & \text{jika } X_2 \geq 737.931 \end{cases}$$

$$BF_5 = \max(0, X_6 - 1.106.328)BF_4 = \begin{cases} (X_6 - 1.106.328), (737.931 - X_2) & \text{jika } X_6 > 1.106.328, X_2 < 737.931 \\ \square & \\ 0, & \text{jika } X_6 \leq 1.106.328 \\ & \text{atau } X_2 \geq 737.931 \end{cases}$$

$$BF_7 = \max(0, X_1 - 910.000,063)BF_5 = \begin{cases} (X_1 - 910.000,063)(X_6 - 1.106.328)(737.931 - X_2) & \text{jika } X_1 > 910.000,063, X_6 > 1.106.328, X_2 < 737.931 \\ \square & \\ 0, & \text{jika } X_1 \geq 910.000,063 \\ & \text{atau } X_6 \leq 1.106.328 \\ & \text{atau } X_2 \geq 737.931 \end{cases}$$

Each increase of one unit basis function (BF₇) will increase the percentage of the level open unemployment in district / city Central Java amounted of $0,495982 \times 10^{-15}$ percent if the working population from each district / city less than 737.931 people, the population of each district / city more than 1.106.328 people, and the minimum wage each district / city more than 910.000,063 rupiah with other basis function in the model is assumed to be

constant. For variable minimum wage each district / city is higher potentially influence the level open unemployment there. The higher wage that set by the government can decrease the working population in the country. So that the level open unemployment will increase. Table 3 below shows that the importance rate predictor variables on the best model.

Table 3. Importance Rate Predictor Variables

| Variables | Importance Rate Predictor Variables |
|----------------|-------------------------------------|
| X ₆ | 100,00 % |
| X ₁ | 41,955 % |
| X ₂ | 39,547 % |
| X ₃ | 0,00 |
| X ₄ | 0,00 |
| X ₅ | 0,00 |

It can be seen that X₆ is the most important variable in MARS model with importance rate amounted of 100 percent, followed by X₁ amounted of 41,955 percent, and X₂ amounted of 39,547 percent. While the other three variables have no importance rate (0 percent) because it is already represented by three variables earlier. After MARS model found, it

can do predictions on percentage the level open unemployment in the 35 districts / cities in Central Java. The following will be displayed plot between actual data the percentage of level open unemployment and the percentage of data predicted the level open unemployment to know the difference between the two data.

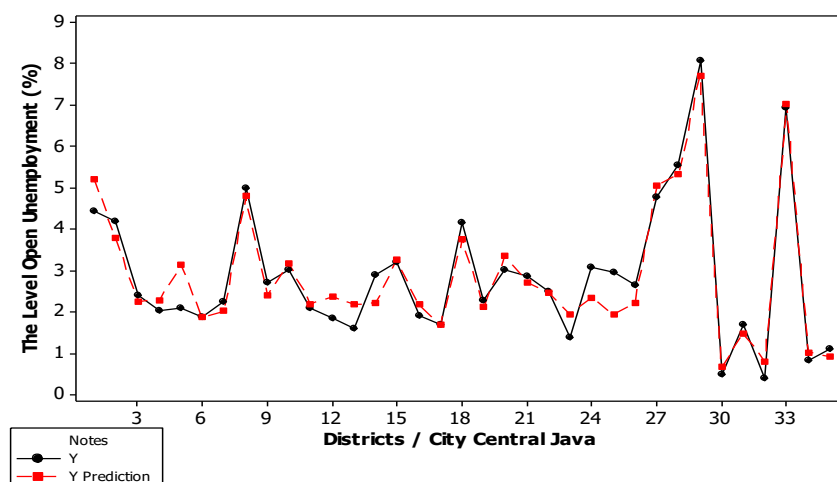


Figure 2. Plot Data Actual and Predicted Data Percent Unemployment Rate

Based on Figure 2, it can be seen that prediction data of the level open unemployment have fairly accurate value with the actual data. So, in this case can be concluded that the

MARS method for this case has been good in predicting the level open unemployment.

5. CONCLUSIONS

The conclusion of the analysis in this research are the highest-level open unemployment in districts / cities in Central Java is owned by Brebes district by 8,07 percent and the lowest level open unemployment in Salatiga City by 0,41 percent. In modeling the level open unemployment in Central Java with multivariate adaptive regression splines (MARS) obtained the best model as follows $\hat{f}(x) = 3,776 + 0,213599 \times 10^{-4}BF_1 - 0,206779 \times 10^{-5}BF_2 + 0,260585 \times 10^{-4}BF_3 + 0,495982 \times 10^{-15}BF_7$ with a minimum GCV obtained at 0,396 and R-square of 86,5 percent as well as predictor variables were entered into the model as much as three, namely the population (X_6) with importance amounted of 100 percent, followed by the minimum wage (X_1) amounted of 41,955 percent, and the working population (X_2) amounted of 39,547 percent.

ACKNOWLEDGEMENT

The authors greatly acknowledge the support from Institut Teknologi Sepuluh Nopember and Universitas Hang Tuah Surabaya Indonesia for providing the necessary resources to carry out this research work. The authors are also grateful to Institut Teknologi Sepuluh Nopember and Universitas Hang Tuah, Surabaya, Indonesia who always give me opportunity to improve my research. And the last the authors would like to say thank to the anonymous reviewers and journal editorial board for their many insightful comments, which have significantly improved this article.

REFERENCES

- Badan Koordinasi Sertifikasi Provinsi Jawa Tengah. (2014). *Membangun Infrastruktur Kompetensi*. Pemerintah Provinsi Jawa Tengah.
- Badan Pusat Statistik Provinsi Jawa Tengah. (2015). *Data Ketenagakerjaan Jawa Tengah*. Katalog BPS.
- Friedman, J.H. (1991). *Multivariate Adaptive Regression Splines*. *The Annals of Statistics*, Vol. 19 No.1 (Maret 1991).
- Friedman, J.H. dan Silverman, B.W. (1989). *Flexible Parsimony Smoothing and Additive Modelling*. *Technometrics*, 31.
- Kaufman, Bruce E. dan Hotchkiss, Julie L. (1999). *The Economic Labor Markets*. USA: Georgia State University.
- Mankiw, N. Gregory. (2000). *Teori Makro Ekonomi*. Jakarta: Erlangga.
- Nash, M.S. dan Bradford, D.F. (2001). *Parametric and Non-Parametric Logistic Regression for Prediction of Precense/Absence of an Amphibian*. Las Vegas, Nevada.
- Otok, B.W., Guritno, S., & Subanar. (2008). *Asimtotik Model Multivariate Adaptive Regression Spline*. *Jurnal Natur Indonesia*.
- Prahutama, Alan. (2013). *Model Regresi Nonparametrik dengan Pendekatan Deret Fourier Pada Kasus Tingkat Pengangguran Terbuka di Jawa Timur*, Seminar Nasional Statistika, Dosen Jurusan Statistika Universitas Diponegoro.
- Pratiwi, Febti E. dan Zain, Ismaini. (2014). *Klasifikasi Pengangguran Terbuka Menggunakan CART (Classification and Regression Tree) di Provinsi Sulawesi Utara*, Volume 3, No.1, 2014. FMIPA ITS Surabaya.
- Pratiwi, Y., Rahardjo, S., Susiswo. (2013). *Analisis Kelompok Dan Analisis Diskriminan Untuk Menggolongkan Tingkat Pengangguran Di Provinsi Jawa Timur Berdasarkan Pendidikan Terakhir Yang Ditempuh*, Skripsi, Mahasiswa Jurusan Matematika Universitas Negeri Malang
- Putra, Nugroho M. dan Zain, Ismaini. (2014). *Analisis Statistik Tingkat Pengangguran Terbuka di Jawa Timur Tahun 2012*, Skripsi, Mahasiswa Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya.
- Sari, Ruli S. dan Budiantara, I.N. (2012). *Pemodelan Pengangguran Terbuka di Jawa Timur dengan Menggunakan Pendekatan Regresi Spline Multivariabel*, Volume 1, No.1, September 2012. FMIPA ITS Surabaya.
- Santoso, Noviyanti. dan Akbar, M.S. (2009). *Klasifikasi Kabupaten/Kota di Jawa Timur Berdasarkan Tingkat*

Pengangguran Terbuka Dengan Pendekatan Multivariate Adaptive Regression Spline (MARS), Skripsi, Mahasiswa Jurusan Statistika Institut Teknologi Sepuluh Nopember Surabaya.

Pengangguran Terbuka Di Jawa Tengah Menggunakan Regresi Spline, Skripsi, Jurusan Statistika FSM Universitas Diponegoro.

Utama, Seta Satria., Suparti., Rahmawati, Rita. (2015). *Pemodelan Tingkat*