Analysis of Multivariate Adaptive Regression Spline (MARS) Model in Classifying factors affecting on Student the Study Period at FKIP Darussalam University of Ambon

S Zurimi*, Darwin2.
1 Lecturer of Mathematics Education Study Program, Faculty of Teacher Training and Education (FKIP) Darussalam University of Ambon, Maluku-Indonesia
2 Lecturer of Biology Education Study Program, Faculty of Teacher Training and Education (FKIP) Darussalam University of Ambon, Maluku-Indonesia

* E-mail: Zurimifarino6@gmail.com

Abstract. This study aims to determine the applicative model that can be revealed through the Multivariate Adaptive Regression Spline (MARS) classification Model for the problems alumni study period of the FKIP Darussalam University of Ambon. To find out the applicative model that can be revealed through the MARS classification model, parameter estimation is first performed to find the best MARS model. The best MARS model is chosen based on the minimum generalized cross validation (GCV) value. The research result indicates that the parameter estimation of MARS model using Ordinary Least Square (OLS) method obtained were convergent. This is indicated by the smallest Mean Square Error (MSE). In addition, the application of the MARS model using the OLS method on data alumni study period of the FKIP Darussalam University of Ambon, result indicates that there is an influential base function, BF3 which contains two predictor variable, that is first semester achievement index and family economics conditions. This is caused by BF3 has a significance value $t_{count} > t_{table}$ so the decision taken is to reject $H_0$ which means that the BF3 basis function parameter has a significant effect on the model.

1. Introduction
One measure of learning success at the tertiary level is the level of accuracy of the study period. The standard time required to complete a Bachelor's program is 4 years or the equivalent of 8 semesters. However, in reality many students experience difficulties in completing their studies in accordance with predetermined study time standards. This is supported by data taken at the forlap feeder feeder especially the Teaching and Education Faculty (FKIP) Darussalam University of Ambon, the average length of study of undergraduate students for graduating from January 2013 to August 2018 is around 10 semesters which is equivalent to 5 years.

The Anticipate problem of the length of study of these students, an analysis can be done to find out the factors that influence the length of study of students of the S1 study program FKIP Darussalam University of Ambon. Previous research conducted [1] states that the predictor variables that most influence on the study period are the college entry point, while other factors are gender, early semester achievement index, and study program. Thus the interesting thing to know from this alumni classification problem based on the length of study is that it is influenced by many predictor variables,
involves binary response variables, and has large sample data. The characteristics of this problem are in accordance with the capabilities of a model, namely the Multivariate Adaptive Regression Spline (MARS). It is hoped that this research can help the writer to verify an applicative model through the analysis of the MARS Model in classifying the factors that influence the duration of student study. To achieve these objectives several models can be used in regression analysis. Regression analysis has several approaches in estimating a relationship between variables, namely parametric methods and nonparametric methods. MARS is a group of modern statistical models with a nonparametric regression approach that results in flexible regression modeling for data with predictor variables $3 \leq k \leq 20$ and sample size $50 \leq n \leq 1000$. [4].

This study aims to determine the applicative model that can be revealed through the MARS classification Model for the problems alumni study period of the FKIP Darussalam University of Ambon.

2. Research Methods
2.1 Research sites
The location of this research at the Darussalam University of Ambon, specifically the Faculty of Teacher Training and Education (FKIP).

2.2 Types of research
The type of research used is "ex post facto" research, which is in nature. This study only examined an incident without prior treatment of the object under study.

2.3 Population and sample
Population in this study is the data on alumni FKIP Darussalam University of Ambon consisting of four course of study such as education mathematic, education physics, chemical education and education biology. The sample used technique is a technique purposive the sampling method of. The sampling method of purposive technique to the sampling method of consideration. Samples to be taken is the data on alumni FKIP Darussalam University of Ambon who passed in august 2017 up until august 2018. , this by a due consideration the average length of study of a student the program an undergraduate degree in the year 2017 and 2018 especially in FKIP Darussalam University of Ambon is about 10 the first half which is equivalent to five years. Thus the selected sample was 470 students.

2.4 Data analysis techniques
After the data is collected and then analyzed, the data in question is a Long period of study and factors affecting the long period of study students FKIP Darussalam university of Ambon so as to what factors are known to affect the time student study so that it can verify an applicative model through the analysis of the Model of MARS.
As for measures to be used in this research is given in the following:

![Flow Chart Of The Procedure The Research Work](image)

**Figure 1. Flow Chart Of The Procedure The Research Work**

### 3. Results and Discussion

#### 3.1 Estimation of parameters in the Model of MARS using the method of Ordinary Least Square

Model MARS defined as follows:

\[
Y = \ln \left( \frac{\pi(x)}{1-\pi(x)} \right) = a_0 + a_1 B_1(x) + a_2 B_2(x) + \cdots + a_M B_M(x) + e
\]

That equation is a linear in parameter \( \alpha \), suppose that:

\[
Y = a_0 + a_1 B_1(x) + a_2 B_2(x) + \cdots + a_M B_M(x) + e,
\]

So the equation (3.1) can be used as

\[
Y = B \alpha + e
\]

where:

\[
Y = (y_1, y_2, \ldots, y_N)^T
\]
$B = \left[ \begin{array}{c} 1 \prod_{k=1}^{K_1} \left( S_{1m} \cdot (x_{1(m)} - t_{1m}) \right) \\ \vdots \\ 1 \prod_{k=1}^{K_M} \left( S_{K_M} \cdot (x_{K_M(m)} - t_{K_M(m)}) \right) \end{array} \right]$

$\alpha = (\alpha_0, \alpha_1, \alpha_2, ..., \alpha_M)^t$

$\varepsilon = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_N)^t$

Suppose given some sample and to obtain the value of parameter $\alpha$ damage done to minimize the value of $\varepsilon = Y - B\alpha$ or can be written in the form of function as follows:

$$\psi = \varepsilon^t \varepsilon = (Y - B\alpha)^t(Y - B\alpha)$$

Function $\psi$ above can be described as follows:

$$\psi = (Y - B\alpha)^t(Y - B\alpha)$$

$$\psi = Y^tY - Y^tB\alpha - B^t\alpha^tY + B^t\alpha^tB\alpha$$

That can be minimized with looking for partial first of the functions $S$ against the $\alpha$ and level zero

$$\frac{\partial \psi}{\partial \alpha} = 0$$

So obtained

$$B^tB\alpha = B^tY$$

Equation (3.4) above normal. By multiplying the two segments with $(B^tB)^{-1}$

Obtained ols estimator for parameter $\alpha$ as follows:

$$\hat{\alpha}_{OLS} = (B^tB)^{-1}B^tY$$

Equations (3.5) is the Ordinary Least Square estimates for $\alpha$

### 3.2 Application of the Model of MARS using the method of Ordinary Least Square on the Data alumni the study period of FKIP Darussalam University of Ambon.

The analysis carried out on the model MARS using a method of Ordinary Least Square in this study used a software program MARS version of 2.0 an entire model that has been obtained by based on grade gcv most minimum then model MARS best selected with the value of about that model value $BF = 15; MI = 2; MO = 2$ and 3 or BF = 20; MI = 2, and MO = 2 and 3 or BF = 20;MI = 3; and MO = 0, 2 and 3 with a value of GCV 1.181. Model of MARS through the method of Ordinary Least Square is generated as follows:

$$Y = 0.189 + 0.645^{*}BF3$$
With:

\[
\text{BF2} = (\text{family economics conditions} = 0)
\]
\[
\text{BF3} = (\text{first semester achievement index} = 0)^* \text{BF2}
\]

After the estimated parameter coefficient model to suppose \((a_0, a_1, a_2, \ldots, a_M)\). Next on MARS models do test the significance of basis functions that include concurrent test and test Individual. To test simultaneously, based on the processing MARS it can be seen that the \(F_{hitung}\) as big as 14.271, by using the \(\alpha = 0.05\), so obtained \(F_{0.05(1,468)} = 4.02\), so the critical areas namely \(F_{count} > F_{0.05(1,468)}\), then the decision taken i.e. refuse \(H_0\), meaning that at least there is one \(a_j\) is not the same with zero that can be stated also that at least there is one function of \(\alpha\) base that contains the variable predictor variable effect on the response. As for the individual test, using the \(\alpha = 0.05\) then obtained: \(t_{table} = t_{0.025;468} = 0.4901\). Critical region is \(|t_{count}| > t_{0.025;468}\) but rejected \(H_0\). based on the calculation on Table 1 the following:

| Parameter | \(t_{hitung}\) | decision |
|-----------|----------------|----------|
| constant  | 3.467          | reject \(H_0\) |
| BF3       | 3.778          | reject \(H_0\) |

The source of : manufacturing output MARS

Can be seen in Table 1 function parameters that all it has value the significance that the decision is rejected \(H_0\) function parameters which means all the bases in models have an significantly to. Model Thus, modeled on the (3.6) shows that there is an influential base 1 function to model MARS through using Ordinary Least Square method that BF3 which contains two predictor variable, that is first semester achievement index and family economics conditions.

After a further parameter estimates and testing conducted the election of the best model MSE Values, on the model of MARS the Ordinary Least Square method is 0.15682. This shows that the MARS model parameter estimation using the Ordinary Least Square method acquired convergent results.

4. Conclusions

Based on the and discussion has been concluded that applicative model that can be revealed through the MARS classification Model for the problems alumni study period of the FKIP Darussalam University of Ambon, The research result indicates that the parameter estimation of MARS model using Ordinary Least Square (OLS) method obtained were convergent. This is indicated by the smallest of value Mean Square Error (MSE). In addition, the application of the MARS model using the OLS method on data alumni study period of the FKIP Darussalam University of Ambon, result indicates that there is an influential base function, BF3 which contains two predictor variable, that is first semester achievement index and family economics conditions. This is caused by BF3 has a significance value \(t_{count} > t_{table}\) so the decision taken is to reject \(H_0\) which means that the BF3 basis function parameter has a significant effect on the model. This research could still be expanded among others review more about the MARS model parameter estimation using methods other than the ordinary least square and uses a combination of base functions, maximum and minimum interaction, observation of the other, especially in the determination of a minimum number of observations in each knot, as well as for researchers who want to investigate more about the MARS model parameter estimation needs to pay attention to the number of observations and the number of basis functions as variables Predictor, because it is a deciding factor for a good estimation so that it can be research opportunities in the future.
References

[1] A Mardiah 2015 Application of the method of multivariate adaptive regression splines (MARS) to determine the factors affecting the study period the student FMIPA UPI. *Eurekamatika* 3 1-19

[2] A Azwirda. 2005 The use of regression splines adaptiv multiple binary response data *Thesis Bogor Agricultural University Graduate School*

[3] Budiantara I N, Lestari B, Islamiyati A and Wibowo W 2009 The selection of the optimal estimator knot splines weighted regression in nonparametrik heteroskedastik longitudinal data *Proceedings of the national seminar on Statistics XI, Ten November Institute of technology, Surabaya*.

[4] Friedman J H 1991 Multivariate adaptive regression splines *The Annals of Statistics* 19 1-14

[5] Otok B W 2009 Consistency and asymptotic normality model multivariate adaptive regression splines (mars) binary response *J. Basic Science* 10 133-140

[6] Purnomo 2008 Estimation penalized least square multivariate adaptive regression splines *Proceedings of the First International Conference on Mathematics and Statistics* (IcoMS-1) Bandung-Indonesia.