DETERMINATION OF FACTORS THAT AFFECTING THE INFANT MORTALITY RATE (IMR) IN EAST NUSA TENGGARA

Erma Oktania Permatasari[1], Budi Priyono[2], Al'awwa Prima Nadia Putri[3]

[1,3] Statistics, Faculty of Science and Data Analytics (FSAD), Sepuluh Nopember of Institute Technology
[2] University of Hang Tuah Surabaya
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: erma.oktania@gmail.com

ABSTRACT

The Indonesian government always tries to make its people prosperous. The level of population welfare can be measured by nutritional status, mortality, and population morbidity. One of the public health statuses seen from population mortality is the Infant Mortality Rate (IMR) indicator. IMR is the number of infant deaths aged 0 to before one year out of 1000 live births. One of the regions in Indonesia that experienced IMR problems is the East Nusa Tenggara Province. The number of death cases continued to increase from 2016 to 2018. That is thought to be caused by various internal and external factors in the population. Analysis of the increase in the number of IMR can be identified by analysis using the nonparametric regression method because the relationship between the predictor variable and the response variable does not form a particular pattern. A nonparametric method used is the Multivariate Adaptive Regression Splines (MARS) because the predictor variables are high dimensions. The data processed is IMR data in East Nusa Tenggara Province in 2018 with district/city as research units and the factors that influence the occurrence of IMR. Based on the analysis, the highest IMR is owned by Sabu Raijua District, and the variables that have a significant effect on the MARS model are the percentage of births assisted by health workers (X_1) and the percentage of babies with low birth weight (LBW) (X_4).

Keywords : *East Nusa Tenggara, Infant Mortality Rate (IMR), Multivariate Adaptive Regression Splines*

1. INTRODUCTION

The health of Indonesian population is an important issue that receives special attention from the government. Because health is closely related to the survival of the population, the government can determine the population's health status through nutritional status, mortality, and population morbidity. One of the public health status seen from population mortality is the Infant Mortality Rate (IMR) indicator. IMR is a number that shows the number of deaths of infants aged 0 to before one year from every 1000 live births [1]. IMR in an area is used to determine children's health status and the socio-economic conditions of the environment where residents live. In general, IMR is inversely correlated with the population's economic status, so that IMR can be used as an indicator to determine changes in population health conditions [2].

The Indonesian government has launched a National Medium-Term Development Plan (RPJMN) for 2015-2019 concerning the improvement of the Indonesian people's health and nutrition status, one of which is by reducing IMR [3]. Besides, IMR was also discussed by UN member countries in September 2015 in a series of the 2030 Sustainable Development Agenda, which discusses reducing infant mortality by including it in the 17 *Sustainable Development Goals* (SDGs). The goal addressing infant mortality is the third goal, with one of the targets is to ensure a healthy life and support the welfare of all ages people so that all countries by 2030 targeted to end preventable deaths in infants and young children [4].

One of the provinces in Indonesia that has IMR problems is the East Nusa Tenggara Province, which the number of infant mortality cases from 2015-2018 had fluctuated, in 2014 there were 1280 cases (14 per 1000 live births), in 2015, it increased to 1488 cases (10 per 1000 live births), in 2016 it decreased to 704 cases (5 per 1000 live births), in 2017 it increased to 7 per 1000 live births, and in 2018 it increased to 1131 cases (11.7

per 1000 live births). Case of infant mortalities from 2016 to 2018 continued to increase. This case is very worrying considering IMR is an essential indicator in public health development. Therefore, it is necessary to know the factors that significantly influence the occurrence of IMR so that it can be used by the East Nusa Tenggara Provincial government for consideration to more focused on development planning, especially in reducing the number of IMR in East Nusa Tenggara Province.

Factors that expect to influence IMR can be analyzed using a parametric or nonparametric regression approach, it is seen from the data patterns formed between the predictor variables and the response variables. The parametric approach used for data forms a particular pattern, while the nonparametric approach when the data does not form a specific pattern. IMR data and the factors that expect to influence it do not form a specific pattern so that data processing uses a nonparametric approach. One of the nonparametric methods used is the Multivariate Adaptive Regression Splines (MARS). This method can be used because the number of predictor variables in high-dimensional research is $(3 \leq v \leq 20)$, v is the number of predictor variables [5].

Therefore, this study aims to determine the characteristics of IMR data and modeling in East Nusa Tenggara Province using a nonparametric *Multivariate Adaptive Regression Splines* (MARS) regression approach.

2. LITERATURE REVIEW

A. Descriptive Statistics

The statistical method discusses the procedures used in data collection, presentation, analysis, and interpretation. Statistical methods consist of two, they are descriptive statistics and inferential statistics. Descriptive statistics is a method related to collect and present data to provide information only about data and without conclusion [6]. Descriptive statistics used in this study are mean and variance.

B. Multivariate Adaptive Regression Splines (MARS)

The most widely used statistical method to determine the relationship and effect between predictor variables and response variables is the regression method. The regression method consists of three approach models, they are parametric, semiparametric, and nonparametric. If the predictor variable and the response variable have a shaped pattern, it can be analyzed with a parametric approach model. In contrast, if the relationship between the predictor variable and the response variable does not form a particular pattern, it is analyzed using a nonparametric approach. One of the nonparametric approaches is the *Multivariate Adaptive Regression Splines* (MARS).

The MARS method is one of the nonparametric methods introduced by Friedman in 1991. The MARS model is used to estimate high-dimensional data problems, which is data with the number of independent variables produces accurate variable predictions and produces a continuous model in knots based on the smallest value of *Generalized Cross-Validation* (GCV). Besides, the MARS method can also be used in continuous and categorical responses. MARS with the continuous response is used if the response is quantitative type, which is interval or ratio data, while MARS with the categorical response is used if the response is qualitative type, which is nominal or ordinal data [5].

MARS merupakan kombinasi kompleks antara *truncated splines* dengan *Recursive Partitioning Regression* (RPR). Berdasarkan kombinasi tersebut menjadikan MARS memiliki kelebihan apabila dibandingkan dengan *truncated splines* dan *Recursive Partitioning Regression* (RPR). Berikut merupakan beberapa kelebihan metode MARS.

MARS is a complex combination of *truncated splines* with *Recursive Partitioning Regression* (RPR). Based on this combination, MARS has specialties other than *truncated splines* and *Recursive Partitioning*

Regression (RPR). The following are some of the advantages of the MARS method.

a. *Multivariate Adaptive Regression Spline* (MARS) can accommodate additive effects and interaction effects between predictors in its modeling, while *truncated spline* only accommodates additive effects.
b. *Multivariate Adaptive Regression Spline* (MARS) can be used in regression modeling involving continuous and categorical responses, whereas *truncated spline* is generally only used for continuous responses.
c. *Multivariate Adaptive Regression Spline* (MARS) has advantages in computation time for data modeling that involves many predictors compared to the *truncated spline*. The selection of knots in MARS is made by adaptive procedures, which include forward and backward stepwise.
d. *Multivariate Adaptive Regression Spline* (MARS) compared to RPR, MARS produces a continuous model at knots, whereas in RPR, a continuous model at knots is not found.

There are some things that must be considered in modeling using the MARS method are as follows [5].

a. Knot

The knot is where the pattern or regression line changes. The minimum distance between knots or minimum observation between knots (MO) is determined by trial and error until the minimum GCV is obtained.

b. Basis function (BF)

The basis function is the interval between consecutive knots and explains the relationship between the response variable and the predictor variable. The number of basis functions is two to four times the number of predictor variables.

c. Maximum interaction (MI)

Maximum interaction (MI) is the maximum number of interactions between variables with values 1, 2, and 3.

d. Minimum observation (MO)

Minimum observation (MO) is the minimum distance between knots. The values are 0, 1, 2, and 3.

The MARS regression equation, which states the relationship between a predictor variable and a single response variable, can be written in equation 1 below.

$$\hat{f} = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{Km} \left[s_{km} \cdot \left(x_{v(k,m)} - t_{km} \right) \right]_+ \quad (1)$$

Where the function,

$$\left(x_{v(k,m)} - t_{km} \right)_+ = \begin{cases} \left(x_{v(k,m)} - t_{km} \right), x_{v(k,m)} - t_{km} > 0 \\ 0, x_{v(k,m)} - t_{km} \le 0 \end{cases} \quad (2)$$

where,

$\hat{f}$: estimation of MARS model

α_0 : the constant coefficient of the function B_0 basis

α_m : coefficient of the m function basis

$x_{v(k,m)}$: independent variable

t_{km} : knot value of the independent variable $x_{v(k,m)}$

M : the number of interactions in the m-th basis function

s_{km} : the value equal to 1 if data on the right side of knot point or -1 if data on the left side of knot point

v : the number of predictor variables

k : the number of interactions

The estimation of $\{\alpha_m\}_{m=0}^{M}$ is determined using the smallest square method (*ordinary least square or* OLS). If the form of a matrix can be written as,

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

where,

$$\mathbf{y} = \left(y_1, ..., y_n \right)^T, \boldsymbol{\alpha} = \left(\alpha_0, ..., \alpha_M \right)^T, \boldsymbol{\varepsilon} = \left(\varepsilon_1, ..., \varepsilon_n \right)^T$$

$$\mathbf{B} = \begin{bmatrix} 1 & \prod_{k=1}^{K_1} \left(s_{k1} \left(x_{1v(k,1)} - t_{k1} \right) \right)_+ & \cdots & \prod_{k=1}^{K_m} \left(s_{kM} \left(x_{1v(k,M)} - t_{kM} \right) \right)_+ \\ 1 & \prod_{k=1}^{K_1} \left(s_{k1} \left(x_{2v(k,1)} - t_{k1} \right) \right)_+ & \cdots & \prod_{k=1}^{K_m} \left(s_{kM} \left(x_{2v(k,M)} - t_{kM} \right) \right)_+ \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \prod_{k=1}^{K_1} \left(s_{k1} \left(x_{nv(k,1)} - t_{k1} \right) \right)_+ & \cdots & \prod_{k=1}^{K_m} \left(s_{kM} \left(x_{nv(k,M)} - t_{kM} \right) \right)_+ \end{bmatrix}$$

In the equation, $\mathbf{y}$ is a vector of the response variable with size of $(\mathrm{n}\times 1)$ and **B** is a function-based matrix of size $(\mathrm{n}\times(\mathrm{M}+1))$. Meanwhile $\boldsymbol{\alpha}$, the regression coefficient vector of size $((\mathrm{M}+1)\times 1)$ and $\boldsymbol{\varepsilon}$ is error vector of size $(\mathrm{n}\times 1)$.

The first MARS model's formation is to determine the points of change in the data behavior pattern or knot points. The selection of knots in MARS using forward stepwise and backward stepwise algorithms based on the smallest value of *Generalized Cross-Validation* (GCV) [5]. The forward stage is performed to obtain the number of basis functions by minimizing the *average sum of square residual* (ASR). The simple model concept can be fulfilled by the backward stage, which is by selecting the basis of the function generated from the forward stage by minimizing the value of *Generalized Cross-Validation* or GCV [7]. The best model in MARS modeling is obtained with a minimum GCV value. The minimum GCV form as a criterion for determining knots is written in the following equation.

$$GCV(M)=\frac{\frac{1}{n}\sum_{i=1}^{n}\left[y_i-\hat{f}_M(x_i)\right]^2}{\left[1-\frac{C(M)}{n}\right]^2} \tag{3}$$

where,

y_i : response variable value

C : the estimated value of the response variable in the base function M

$C(M)$: *trace* $\left[\mathbf{B}\left(\mathbf{B}^T\mathbf{B}\right)^{-1}\mathbf{B}^T\right]+1$

$\mathbf{B}$: base matrix function

If MARS has the same minimum GCV value in determining the best model, then the minimum MSE value or the highest value can be used. Below is the formula R^2 used.

$$R^2=\frac{\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i-\bar{y}_i\right)^2} \tag{4}$$

The result of MARS model, the coefficient *basis function* (BF) is tested, including simultaneous tests and individual tests. Coefficient test is performed simultaneously on the essential function in the MARS model. The coefficient test aims to determine whether the selected MARS model is generally appropriate and shows the right relationship between the predictor variable and the response variable. The hypothesis used is as follows with M is the number of predictor variables.

$H_0:\alpha_1=\alpha_2=\alpha_3=...=\alpha_M=0$

H_1: at leats one of $\alpha_j\neq 0, j=1,2,...,M$

The test statistics used is as follows

$$F_{value}=\frac{MSR}{MSE}=\frac{((SSR)/p)}{((SSE)/(N-p-1))} \tag{5}$$

where,

$$SSE=\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2 \tag{6}$$

$$SSR=\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}_i\right)^2 \tag{7}$$

Reject H_0 if F_{value} more than F_{table}, with degrees of freedom value $(df_1)=p$ and $(df_2)=N-p-1$, As a significance level value α, partial (individual) test is used to determine whether each predictor variable significantly affects the response variable based on the model's function and can describe the actual data. The hypothesis used is as follows: M is the number of predictor variables.

$H_0:\alpha_j=0$

$H_1:\alpha_j\neq 0, j=1,2,...,M$

The test statistics used is as follows.

$$t_{hitung}=\frac{\hat{\alpha}_j}{SE(\hat{\alpha}_j)} \tag{8}$$

Reject H_0 if t_{value} >$t_{\left(\frac{\alpha}{2};v_2\right)}$ or t_{value} >-$t_{\left(\frac{\alpha}{2};v_2\right)}$with degrees of freedom $v_2=N-k$ where k is the number of base functions that contribute to the model.

The MARS model has relative variable importance. The relative importance of the variable is each predictor variable's contribution that makes predictions of the response variable by considering the amount of contribution made when combined with other variables. The value of the predictor variable's importance shows the level of importance of the predictor variable on the grouping function, which is estimated by the increase of GCV value due to the transfer of the considered variables from the model [8]. The following is a formula of relative variable importance.

$$\text{Relative variable importance } = \frac{RSS}{N} \tag{9}$$

C. Infant Mortality Rate (IMR).

Infants are children aged 0 until before they reach 1 years old, while the infant mortality rate is a number that shows the number of deaths of infants aged 0 before reaching 1 years old out of every 1000 live births in a particular year in an area. The use of IMR is to reflect the state of health status in a community because newborn babies are susceptible to the conditions of the environment in which they live and the social status of their parents. If IMR decreases, it shows a high economic status in the area so that it can be used as an indicator to assess changes in the health condition of a community in a particular area [1].

The cause of infant deaths can be divided into two, they are direct cause and indirect cause. The direct cause of infant mortality are influenced by factors that the baby carries from birth and are directly related to the baby's health, including low birth weight (LBW), post-birth infections (tetanus, neonatorum, sepsis), hypothermia, and asphyxia. Meanwhile, infant mortality is not directly affected by environmental factors and maternal habits during pregnancy, such as socioeconomic factors, health services, maternal health during pregnancy, and the influence of the environment in which they live [9].

The variables that affect the survival of children are divided into two, as follows [10].

1. Variables that are considered exogenous or socioeconomic, such as economic, social, cultural, community, and regional factors. Infant mortality which is influenced indirectly by socioeconomic factors, is divided into five main factors:
   a. Maternal factors: age, parity, and birth spacing
   b. Environmental contamination: environmental pollution related to transmission to children (and mothers).
   c. Nutritional deficiency: the survival of the child is not only affected by the availability of nutrition for the child but the mother as well
   d. Accidents: accidents include physical accidents, for example, infanticide.
   e. Individual disease control: one of the components in individual disease control is the preventive action done by healthy people to prevent disease.
2. Endogenous variables or biomedical factors which include breastfeeding patterns, hygiene, sanitation, and nutrition.
   a. Individual level: household productivity, which includes education, health, and time, as well as traditions/attitudes in the environment
   b. Household level: income/wealth
   c. Regional level: ecological environment, political economy, and health system.

Child mortality, which includes infant mortality, is influenced by the demand side and the supply side. The demand side is household and individual behavior or characteristics such as sanitation, disease preventive action in the family, income, education, and parental knowledge. The supply side is government policies at the micro and macro levels and the implementation of policies, the capabilities of local governments, infrastructure and access, and also quality of health services [11].

II. RESEARCH METHODOLOGY

A. Data Source

Data used in this study is secondary data. The data was obtained from the East Nusa Tenggara Provincial Health Office's official report in East Nusa Tenggara Province Health Profile Book 2018 with 22 districts/cities as observation units in East Nusa Tenggara Province.

B. Research Variable

The research variables used in this study are as follows.

Table 1. Research Variable

Variable	Variable Description	Scale
Y	Infant Mortality Rate (IMR)	Ratio
X_1	The percentage of births assisted by health workers	Ratio
X_2	The Percentage of village/sub-district in *Universal Child Immunization* (UCI)	Ratio
X_3	The percentage of babies who are exclusively breastfed	Ratio
X_4	The Percentage of babies with low birth weight (BBLR)	Ratio
X_5	The percentage of babies getting vitamin A	Ratio
X_6	The percentage of pregnant women who received Fe_3 tablet	Ratio
X_7	The presentage of pregnant women implementing the K_4 program	Ratio

Based on the variables used in this study, the variables' operational definitions were obtained as follows.

1. Infant Mortality Rate

   Infant Mortality Rate (IMR) is the number of infant deaths aged 0 years out of 1000 every live births in 2018 [1].

2. Percentage of births assisted by a health worker

   The births assisted by a health worker is a birth or the process of giving birth to a baby assisted by a health worker with midwifery competence [12].

3. Percentage of village/sub-district in Universal Child Immunization (UCI)

   Villages/sub-district Universal Child Immunization (UCI) is a village/sub-district in each district with complete immunization scope for a group of babies. UCI scope describes the level of community or infant immunity (herd immunity) to disease transmission that can be prevented by immunization [12].

4. Percentage of babies who are exclusively breastfed

   Infants who are exclusively breastfed are boys and girls who receive breastmilk (ASI) directly during 0-6 months in East Nusa Tenggara Province [12].

5. Percentage of babies with low birth weight (LBW)

   Low birth weight babies are babies who born with weight < 2500 grams. Birth weight is the baby's weight that is weighed within the first 1 (one) hour after birth.

6. Percentage of babies getting Vitamin A

   Infants receive Vitamin A are infants aged 6-11 months who receive vitamin A capsules [12].

7. Percentage of pregnant women getting Fe_3 tablets

   The percentage of pregnant women getting Fe_3 tablets are pregnant women who receive Fe_3 tablets to reduce anemia in pregnant women [12].

8. Percentage of pregnant women implementing the K_4 program

   The percentage of pregnant women implement the K_4 program for pregnant women who receive health services by professional health workers (obstetricians and midwives, general practitioners, midwives, and nurses) such as measuring body weight and blood pressure, examining uterine fundal height, immunization against *Tetanus Toxoid* (TT) as well as giving iron tablets to pregnant women during their pregnancy according to existing antenatal care guidelines with an emphasis on promotive and preventive activities [12].

C. Analysis Step

The analysis steps used in this research are as follows.

1. Collecting IMR data of East Nusa Tenggara Province in 2018 and the factors that are assumed to be influential.
2. Perform data preprocessing.
3. Perform descriptive statistical analysis of the response variables and predictor variables.
4. Identifying the form of data patterns between the response variables, which is IMR in East Nusa Tenggara Province, with each predictor variable.
5. Set a model for IMR data in East Nusa Tenggara Province using the MARS method in software R with the following steps [5].
   a. Determine the combination of basis function (BF), maximum interaction (MI), and minimum observation (MO) as follows
      - Determine the maximum BF that is 2 to 4 times the number of predictor variables used.
      - Determine the number of MI, which are 1, 2, and 3, with the assumption that $M > 3$ will produce an increasingly complex model.
      - Determine the number of MO between knots, which are 0, 1, 2, and 3.
6. Get the best model based on the minimum GCV value.
7. Estimating the MARS model parameters.
8. Perform the significance test of the MARS model parameters.
9. Draw conclusions and suggestions.

III. ANALYSIS RESULTS

A. Exploration of IMR and Factors Assumed of Affecting IMR in NTT Province

Data exploration aims to see the data's characteristics or to get an overview as initial information from data before determining or applying the appropriate analysis method. Infant Mortality Rate (IMR) reflects the state of health status in society because newborn babies are susceptible to the conditions of the environment in which they live and their parents' social status. The following are the characteristics of the factors that are thought to influence IMR in East Nusa Tenggara in 2018, shown in Table 2.

Table 2. The mean and variance of the variables used

Var.	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Mean	89,9	61,8	69,8	10,8	94,5	96,27	76,24
Var.	68,03	582,16	597,14	309,9	473,63	619,95	135,21

Based on Table 2, it can be seen that the highest mean percentage of pregnant women who get Fe_3 (X_6) tablets is 96.27%, or it can be concluded that on average pregnant women who get Fe_3 tablets from the government during pregnancy in the districts/cities in East Nusa Tenggara Province is 96.27%. In contrast, the variable percentage of babies with low birth weight (LBW) (X_4) in East Nusa Tenggara Province has the lowest mean, which is at 10.78% compared to other variables, this number means that the total number of babies present in East Nusa Tenggara Province in 2018, on average, each district/city had babies with low birth weight at 10.78%, so that the higher the LBW indicates that the worse the health conditions of babies in the region.

The variance of percentage pregnant women variable variance who receive the highest Fe_3 (X_6) tablet, which is at 619.95, shows that the data in this variable varies between districts/cities in East Nusa Tenggara Province. The lowest variance value is owned by the variable percentage of births assisted by health workers (X_1), which is at 68,03, or it can be concluded that the data on the percentage of births assisted by health workers in East Nusa Tenggara Province in 2018 is not diverse enough. After knowing the data characteristics of the factors that influence IMR, then analyze the characteristics of IMR in East Nusa

Tenggara Province. The following is a graph of IMR in East Nusa Tenggara Province in 2018.

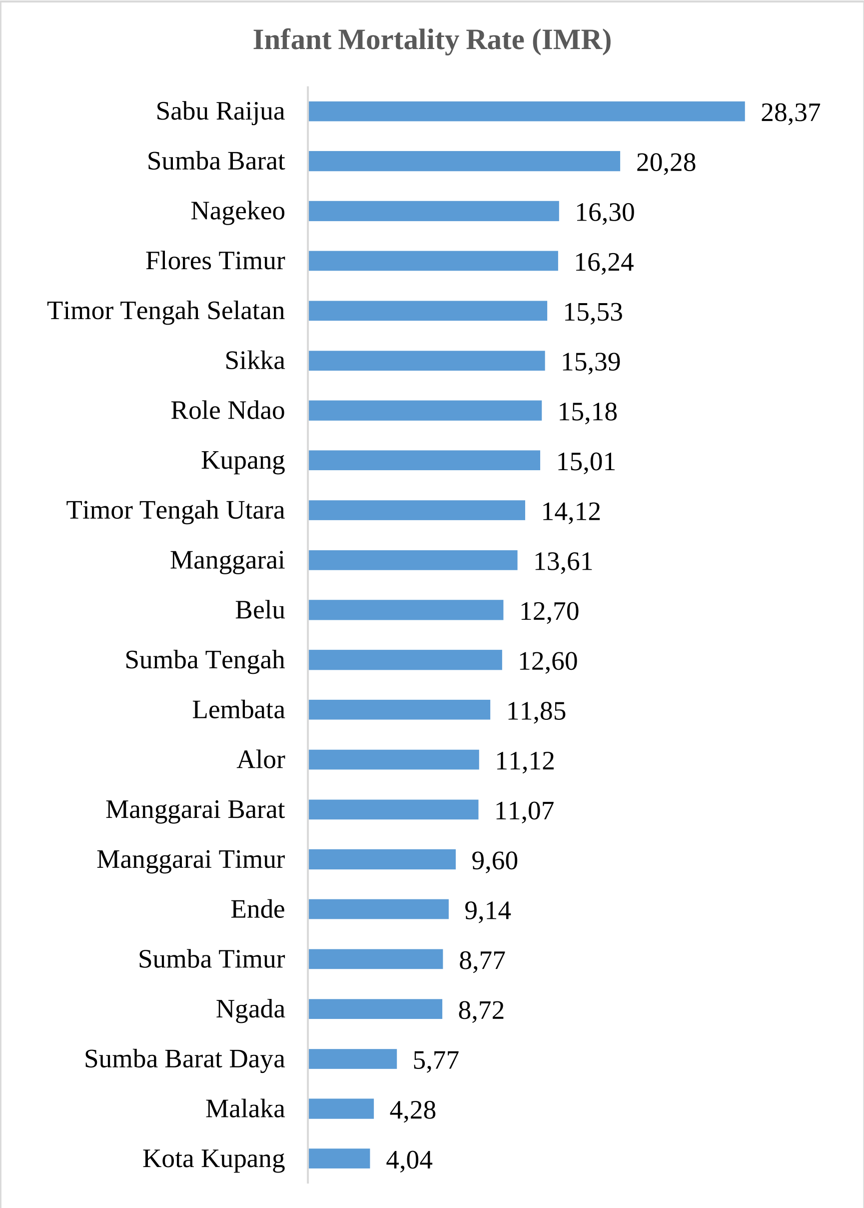

Fig 1. IMR Chart in East Nusa Tenggara Province 2018 by District/City

Figure 1 shows that Sabu Raijua District has the highest IMR. Several factors are thought to cause the high IMR in Sabu Raijua District. The percentage of births assisted by health workers in East Nusa Tenggara Province in 2018 is still relatively low compared to other districts/cities, which is occupy the second-lowest with a percentage of 74.9%. Also, compared to the overall average of births assisted by health workers in districts/cities in East Nusa Tenggara Province, the percentage of Sabu Raijua District is still lower.

Sabu Raijua District is also in the second-lowest percentage of UCI village/sub-district, which shows that there are still many villages/wards in the area that have not provided complete necessary immunization to infants so that it can be the cause of the high IMR. Besides that, the percentage of UCI village is also far below the average of overall percentage UCI villages/wards in East

Nusa Tenggara Province. The UCI villages that are owned by Sabu Raijua District is 34.9%.

The percentage of exclusively breastfed babies is also very low, ranking second to last and very few compared to the overall average in East Nusa Tenggara Province, which is only 20.5%. Breast milk is essential for babies, so it is recommended to be given exclusively. If the percentage is very low, it can lead to death in infants due to lack of nutritious intake.

Furthermore, the percentage of low birth weight (LBW) in the Sabu Raijua District is still high which is at 7.8% of all babies weighted. The percentage of pregnant women who received 90 tablets of Fe_3 tablets is still relatively low, which is at 70.1%, and is in the bottom three positions. Compared to other districts/cities, this number can also be low because it is below the average percentage of all pregnant women who receive Fe_3 tablets in East Nusa Tenggara Province.

The last variable that is likely to cause a high IMR in Sabu Raijua District is the percentage of pregnant women who implement the K_4 program is still very low or the number of pregnant women who have checked themselves with health workers since the first trimester (early pregnancy) and still very minimal so that the health of the fetus is not monitored from the beginning of pregnancy. Compared with the average percentage of pregnant women who implement the K_4 program in all districts/cities in East Nusa Tenggara Province, the figure of 57% owned by Sabu Raijua is still quite far.

B. The formation of MARS Model

The initial step for the formation of the MARS model is to determine the combination of basis functions (BF) number, maximum interaction (MI), and minimum observation (MO) between knots. The basis function is the interval between consecutive knots and explains the relationship between the response variable and the predictor variable. The number of basis functions used is two to four times the predictor variable. Maximum interaction (MI) is the maximum number of correlation relationships between variables in the model. If the MI used is 1, then there is no interaction between variables in the model. If MI is 2, then there is an interaction of 2 variables in the model, and if MI is 3, then the model will have an interaction between 3 variables. Minimum observation (MO) is the minimum number of observations between knots, MO that can be used are 0, 1, 2, and 3.

The number of predictor variables used is 7, so the number of BF used is 14, 21, and 28. The maximum interaction (MI) used are 1, 2, and 3, while the minimum observation (MO) is 0, 1, 2, and 3. The combinations obtained from BF, MI, and MO are 36 models as in Table 4 below.

Table 4. Modeling Result

No	BF	MI	MO	GCV	R^2
1	14	1	0	26,7	0,426
2	14	1	1	27,3	0,415
3	**14**	**1**	**2**	**22,5**	**0,517**
4	14	1	3	24,3	0,478
5	14	2	0	23,4	0,683
6	14	2	1	30,5	0,000
7	14	2	2	24,3	0,671
8	14	2	3	26,7	0,760
9	14	3	0	23,4	0,683
10	14	3	1	30,5	0,000
11	14	3	2	24,3	0,671
12	14	3	3	26,7	0,760
13	21	1	0	26,7	0,426
14	21	1	1	27,3	0,415
15	21	1	2	22,5	0,517
16	21	1	3	24,3	0,478
17	21	2	0	23,4	0,683
18	21	2	1	30,5	0,000
19	21	2	2	24,3	0,671
20	21	2	3	26,7	0,760
21	21	3	0	23,4	0,683
22	21	3	1	30,5	0,000
23	21	3	2	24,3	0,671
24	21	3	3	26,7	0,760
25	28	1	0	26,7	0,426
26	28	1	1	27,3	0,415
27	28	1	2	22,5	0,517
28	28	1	3	24,3	0,478
29	28	2	0	23,4	0,683
30	28	2	1	30,5	0,000
31	28	2	2	24,3	0,671
33	28	3	0	26,7	0,683
34	28	3	1	23,4	0,000
35	28	3	2	30,5	0,671
36	28	3	3	24,3	0,760

Based on Table 4, there are 36 models formed from a combination of BF, MI, and MO values. The best model is the one with the smallest GCV value. If the GCV value is the same, then it is considered with the highest R^2 value, and if the R^2 value is still the same, then it can adhere to the parsimony principle of the model, which is to consider the model that has the smallest combination of BF, MI, and MO.

The model chosen is model number 3 with the combination of BF = 14, MI = 1, and MO = 2 because it has the GCV, R^2 value, and the smallest combination of BF, MI, and MO. The model's GCV value is 22,5, and the value is 0,517 so that the MARS model can be written as follows.

$$\hat{f}(x) = 21,660 - 0,418 * BF_1 - 1,633 * BF_2 \tag{11}$$

Where basis function:

$$BF_1 = h(X_1 - 75,7)$$

$$BF_2 = h(8,6 - X_4)$$

In equation 11, it can be seen that the basis function interpretation is as follows.

a. $BF_1 = h(X_1 - 75,7) = \begin{cases} 0, X_1 \leq 75,7 \\ (X_1 - 75,7), X_1 > 75,7 \end{cases}$

The coefficient $BF_1 = -0,418$ in model means that every one unit BF_1 increase will decrease the IMR by 0.418 with basis functions being considered constant. If the percentage of births assisted by health workers (X_1) is more than 75.7%, the IMR will decrease by 0.418.

b. $BF_2 = h(8,6 - X_4) = \begin{cases} 0, X_4 \geq 8,6 \\ (8,6 - X_4), X_4 < 8,6 \end{cases}$

The coefficient $BF_2 = -1,633$ in model means that every increase of one unit BF_2 will decrease the IMR by 1.633 with basis functions considered constant. If the percentage of babies born with low birth weight (LBW) (X_4) is less than 8.6%, it will reduce infant deaths by 1.633.

In equation 11, it can be seen that there are two predictor variables included in the MARS model, they are the variable of births assisted by health workers (X_1) and the variable percentage of low birth weight (LBW) (X_4). Furthermore, coefficient test is performed to determine that the selected MARS model is appropriate and shows the relationship between the predictor variable and the response variable. The following is a hypothesis used for simultaneous test.

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = ... = \alpha_M = 0$$

H_1: at least one of $\alpha_j \neq 0, j = 1, 2, ..., M$

where rejection area: Reject H_0 if $F_{value} > F_{table}$

Table 5. Simultaneous Statistics Test

F_{value}	$F_{(0,05;7;14)}$	P-*value*	Decision
10,16	3,5219	0,0009963	Tolak H_0

Based on Table 5, it can be seen that the decision to reject H_0 because the F_{value} is more significant than F_{table} and the *p-value* is less than the value $\alpha(0,05)$, this means that there is at least one significant α_j, which means that all BF simultaneously have a significant effect on the model or the response variable.

Furthermore, partially test is perfomed to determine whether each BF has a significant effect. The following is a partial test hypothesis.

$H_0: \alpha_j = 0$

$H_1: \alpha_j \neq 0, j = 1, 2, ..., M$

where rejection area: Reject H_0 if $|T_{value}| > T_{table}$

Table 6. Partial Statistics Test

Parameter	Estimation	*Std Error*	T_{value}
α_1	-0,4180	0,1272	-3,287
α_2	-1,6325	0,4344	-3,758

The value of T_{table} with $\alpha = 0,05$ and df = 20 is 2,086 so that all parameter values in Table 6 are obtained by the decision to reject H_0, which means that each predictor variable of basis function (X_1 dan X_4) contained in the model has a significant effect on the response variable.

The importance level of each predictor variable used in the study can be seen in Table 7 below.

Table 7. The Importance level of Predictor Variable in MARS Model

Variable	Importance Level
X_4	100%
X_1	84,7%

Table 6. Partial Statistics Test

Variable	Importance Level
X_2	0
X_3	0
X_5	0
X_6	0
X_7	0

Based on Table 7, it is known that the variable percentage of babies with low birth weight (LBW) (X_4) has the highest contribution to the MARS model, which is 100%. The variable that has the second contribution is the variable percentage of births assisted by health workers (X_1) at 84,7%, while the other five variables do not have a level of importance in the MARS model because the two variables represent them are in the model.

IV. CONCLUSION AND SUGGESTION

Based on the results of the analysis, it is found that the predictor variable that has the highest mean and variance is the variable percentage of pregnant women who received Fe_3 tablets, which is at 96,27%, and the variance is 619,95. The region with the highest IMR is Sabu Raijua District at 28,369, and the lowest is Kupang City. The best model obtained is from the combination of BF=14, MI=1, and MO=2 so that the MARS model is $\hat{f}(x) = 21,602 - 0,418 * BF_1 - 1,633 * BF_2$ and the predictor variable that has the highest contribution is the variable percentage of babies with low birth weight (LBW) (X_4) at 100%, the second-highest contribution is the variable percentage of births assisted by health workers (X_1) at 84,7%.

While the suggestion for further research is that it is better to compare R software processing with MARS software results, and the government can consider the factors that have a significant effect in determining policy in suppressing IMR.

REFERENCES

[1] Badan Pusat Statistik, “Angka Kematian Bayi,” 20 Januari 2020. [Online]. Available: https://sirusa.bps.go.id/sirusa/index.php/indikator/79.

[2] A. Castro dan M. Singer, Unhealthy Health Policy, United Kingdom: Altamira Press, 2004.

[3] Kementerian Perencanaan Pembangunan Nasional, Rencana Pembangunan Jangka Menengah Nasional (RPJMN) 2014-2019, Jakarta: Badan Perencanaan Pembangunan Nasional, 2014.

[4] U. C. a. Local Governments Asia-Pacific, Tujuan Pembangunan Berkelanjutan yang Perlu diketahui oleh Pemerintah Daerah, DKI Jakarta: Pemerintah Provinsi DKI Jakarta, 2018.

[5] J. H. Friedman, “The Annals of Statistics,” *Multivariate Adaptive Regression Splines,* pp. Vol. 19, 1-141, 1991.

[6] R. E. Walpole, R. H. Myres, S. L. Myres dan K. Ye, Probability & Statistics for Engineers & Scientists, 8th penyunt., United States of America: Prentice Hall, 2007.

[7] J. H. Friedman dan B. W. Silverman, Flexible Parsimonious Smoothing and Additive Modeling, vol. 31, California: Stanford University, 1987, pp. 3-9.

[8] R. Johnson dan D. Wichern, Applied Multivariate Statistical Analysis, 5td ed. penyunt., New Jersey: Upper Saddle RIver, 2002.

[9] Kementerian Kesehatan Republik Indonesia, “Untuk Menurunkan Angka Kematian Ibu dan Kematian Bayi Perlu Kerja Keras,” 2010. [Online]. Available: https://www.depkes.go.id/article/print/793/untuk-menurunkan-angka-kematian-ibu-dan-kematian-bayi-perlu-kerja-keras.html. [Diakses 20 01 2020].

[10] W. H. Mosley dan L. C. Chen, “Population and Development Review,” *An Analytical Framework for the Study of Child Survival in Developing COuntries,* vol. 80, pp. 140-145, 1984.

[11] D. Filmer, Determinats of Health & Education Outcomes Background Note for World Development Report 2004: Making Service Work for Poor People, The World Bank, 2003.

[12] Dinas Kesehatan Provinsi Nusa Tenggara Timur, Profil Kesehatan Provinsi Nusa Tenggara Timur Tahun 2018, Kupang: Dinas Kesehatan Provinsi Nusa Tenggara Timur, 2019.

[13] W. Wicaksono, W. Yuciana dan Suparti, Pemodelan Multivariate Adaptive Regression Splines (MARS) pada Faktor-Faktor Resiko Angka Kesakitan Diare, Semarang: Universitas Diponegoro, 2014.