# Unordered Features Selection of Low Birth WeightDatain Indonesiausing the LASSO and Fused LASSO Techniques

Yenni Kurniawati[1], Khairil Anwar Notodiputro[2*], Bagus Sartono[2].
{yenni.mathunp@gmail.com[1],khairil@apps.ipb.ac.id[2],bagusco@gmail.com[2]}

Departement of Mathematics, Universitas Negeri Padang, Padang, 25171, Indonesia[1]
Department of Statistics, IPB University, Bogor, 16680, Indonesia[2]

**Abstract.**This paper aims to analyze the Low Birth Weight (LBW) data of infants in Indonesia by using the LASSO and Fused LASSO techniques. Fused LASSO is usually used to select parameters for ordered features. In this case, the features are unordered. Therefore, this research adopts three techniques in ordering features. Furthermore, all these Fused LASSO techniques and LASSO are compared. This paper utilizes data on 1,176 LBW infants collected from the 2017 Indonesian Demographic and Health Survey (IDHS). The results showed that LASSO has the sparsest solutionbased on the 5-fold cross-validation. Thefeatures that contribute to LBW are mothers' occupation, mothers' age, antenatal care, multiple birth, and birth order. However, Fused LASSO 1 has the lowest AIC and BIC valuecompared to other ordering techniques.Ordering features using the correlation between the features and the outcomes is recommended as an alternative technique to sort unordered features.

**Keywords:** Fused LASSO, LASSO, Low birth weight, Penalized Regression, Unordered features.

## 1 Introduction

Selecting parameters is one of the primary goals in analyzing a model. In classical linear regression, some techniques usually used to select significant variables are stepwise deletion and the best subset method. Unfortunately, in this method, the model estimator is unstable, where small changes in the data can end up in entirely different models [6]. Selecting with the least-squares method is not appropriate to use anymore since multicollinearity problems exist among variables. Even though the estimators are unbiased, they have a large variance. Another alternative in the selection stage is a penalized regression, such as the LASSO and Fused LASSO methods. The penalized function of the two methods shrinks some of the coefficients to be exactly zero or close to zero, and then the explanatory variable can be lessened. Both of these techniques provide a solution not only for selecting variables but also for improving the accuracy of estimating nonnull parameters [8].

The LASSO (Least Absolute Shrinkage and Selection Operator) is a technique to overcome the accuracy problems of estimation and interpretation by keeping the benefits of subset selection

techniques and ridge regression [20]. The idea for the developing LASSO comes from the non-negative garrote method which is more accurate and stable in terms of predictions [5]. LASSO minimizes the residual sum of squares subject using the constraint to be the sum of the absolute value of the coefficients being less than a constant (regularization of L1) [20]. This penalty form stimulates the solution with a coefficient equal to zero. The Fused LASSO analyzes the difference between neighboring coefficients and shrinks the value to zero [8]. Moreover, the absolute value of (almost) the same coefficient occurs when the appropriate predictor is highly correlated [23].

Fused LASSO is a technique to select parameter based on the assumptions that features $x_{ij}$ have a natural order in $j$ [21]as in the protein mass spectroscopy data [1]. However, when the features are unordered like the gene expression data [9], there are several methods to select the parameters: calculating maximum correlation value or the smallest Euclidean distance [21].Therefore, the arrangement of features in this study is based on the proximity of variables from such maximum correlation value $(r_{x_{(k)}x_{(l)}} \geq r_{x_{(k)}x_{(l)}} \geq \cdots \geq r_{x_{(k)}x_{(l)}})$, that the order of features can be formed into $X_{[1]}, X_{[2]}, \ldots, X_{[p]}$.

The three approaches used in sorting the correlation value between features are (1) the maximum order of correlation X and Y $(r_{xy})$, where $r_{x_{[1]}y} \geq r_{x_{[2]}y} \geq \cdots \geq r_{x_{[p]}y}$, (2) the sequence of correlation based on the Xi variable which has the biggest correlation with $(x_{[1]} = x_i)$, where $r_{x_iy} = max\{r_{x_jy}\}, j = 1,2 \ldots p$, and (3) the maximum correlation order between $X(r_{x_ix_j})$ for each $i, j = 1,2 \ldots p - 1$ and i $\neq$ j. The initial parameter for the three techniques is the variable that has the largest correlation value between X and $Y\left(max\{r_{x_jy}\}\right)$.

In this study, the selection of variables was conducted to analyze data on Low Birth Weight (LBW) in Indonesia. LBW is one of the causes of high infant mortality rate (IMR). Thus, it needs special attention. The national average of LBW in 2013 was 10.2%. Low birth weight prevalence spanned from 7.2% in North Sumatra to 16.9% in CentralSulawesi: an absolute difference of 9.7 percentage points[17]. Otherwise, there were around 10% ofIndonesian infants born with low birth weight in 2013 [14]. The government targets the LBW in Indonesia decreased to 8% in 2019[16].

Factors affecting LBW are widely analyzed in several countries. In Ethiopia, the common risk factors for LBW are socio-demographic, maternal/obstetric, obstetric and medical, and fetal [7]. Mothers' education, nutritional status, and socio-economic status are important factors of LBW in Bangladesh [13]. Maternal education level and antenatal care affect birth weight positively and significantly in Botswana [15], but mothers who smoke during pregnancy lead to negative and notable influences on the babies' birth weight [3], [4] and [22]. Biological conditions – multiple-births, child sex – are other causes of LBW [2], [19], and [3].

The main objectives of this study are selecting features using the LASSO technique;selecting features using the Fused LASSO technique with ordered statistics based on the cross-correlation of sequences X and Y (Fused LASSO 1); selecting the explanatory variable using the LASSO Fused technique with ordered statistics based on the correlation among 'X's (Fused LASSO 2 and Fused LASSO 3); and comparing the penalized regression results with regularization of L1. The regularization is the provisionLASSO and the three Fused LASSO techniques.

## 2 LASSO, Fused LASSO, Selection of Lambda

**LASSO.**Least Absolute Shrinkage and Selection Operator (LASSO)improve the estimation of simple linear regression techniques, with two advantages, particularly selection, and shrinkage[20]. LASSO technique can shrink the coefficient (β parameter) which correlates to zero, that is by giving a penalty called constraints $\sum_{j=1}^{p}|\beta_j| \leq t$ , $t \geq 0$ for the modeling objective function. LASSO conducts variable selection that influences outcomes as well as stability in predicting models [18].

The general equation of LASSO is formulated as follows:

$$\hat{\beta}^L = \arg\min\left\{\sum_i\left(y_i - \sum_j x_{ij}\beta_j\right)^2\right\}\text{where} \quad L_1 = \sum_j|\beta_j| \leq t \tag{1}$$

Where $y_i$is the outcomes that can be qualitative; $x_{ij}$ is the features or covariate,$i = 1,2,\dots,N$, $j = 1,2,\dots,p$, and β is the LASSO coefficient, and $t$ is the tuning parameter. Tuning parameters are selection parameters controlling the shrinkage of the LASSO coefficient. According to Tibshirani(1996), if $t < t_0 = \left|\sum\beta_j^0\right|$ where $\beta_j^0$ is the least-squares estimator, it will cause the coefficient to shrink to near zero or at zero. Then, LASSO will act as a variable selection. However, if the $t > t_0$, the estimator of the LASSO coefficient gives the same result as the least-squares estimator.

The LASSO coefficient estimator is obtained by determining the standard parameters tuning, namely $s = \frac{t}{\sum_{j=1}^{k}|\hat{\beta}_j^0|}$ with t = $\sum_{j=1}^{k}|\hat{\beta}_j|$ and $\hat{\beta}_j^0$ 0 is the ordinary least squares estimator (OLS) or the output of LASSO plot is written as |beta|/max|beta|. The optimal value of s can be obtained through cross validation (CV)[10].

**Fused LASSO.**Selection using Fused LASSO manages two constraint values, L1 and L2. L1 regularizationpushes the parameter to zero, while L2 suppresses the difference in neighboring parameters to zero. The regularization of L2 in Fused LASSO is a special fusion form introduced by Hoerl and Kennard (1970), where the penalty used is $\sum_{j=1}^{p}|\beta_j - \beta_{j-1}|^\alpha \leq t_2$ with $\alpha = 0,1,2$[11]. Fused LASSO's estimation is as follows [21]:

$$\hat{\beta}^{FL} = \arg\min\left\{\sum_i\left(y_i - \sum_j x_{ij}\beta_j\right)^2\right\}\text{where} L1 = \sum_{j=1}^{p}|\beta_j| \leq t_1$$
$$\text{and } L2 = \sum_{j=1}^{p}|\beta_j - \beta_{j-1}| \leq t_2. \tag{2}$$

**Selection of Lambda.** The least-squares method criteria penalized by regularization of L1 and L2 [21]:

$$\arg\min\left\{\sum_i\left(y_i - \sum_j x_{ij}\beta_j\right)^2 + \lambda^{(1)}\sum_{j=1}^{p}|\beta_j| + \lambda^{(2)}\sum_{j=1}^{p}|\beta_j - \beta_{j-1}|\right. \tag{3}$$

$\lambda^{(1)}$ and $\lambda^{(2)}$ are selection parameters for each L1 and L2 regularization. The selection of tuning parameters or penalty values in LASSO is performed by cross-validation (CV). Cross-validation

divides data into two parts, specifically training data and testing data. Training data is applied in determining the value or in compiling the model. Otherwise, testing data is employed to test the goodness of X. The cross-validation values obtained serve as estimators of error prediction. One type of cross-validation method is k-fold. The cross-validation should use 5-fold or 10-fold cross-validation since it produces not only CV values with high bias but also low variance. The best model is determined based on *Akaike information criterion* (AIC) or *Bayesian information criterion* (BIC). The model will be the variable that has the lowest AIC or BIC value. The AIC criterion in defined for a large class of models fit by maximum likelihood [12]. In this case, AIC is given by

$$\text{AIC} = \frac{1}{n\hat{\sigma}^2}\left(Residual\ Sum\ of\ Squares(RSS) + 2d\hat{\sigma}^2\right) \tag{4}$$

The BIC is derived from Bayesian point of view, but ends up looking similar AIC as well [12]. For the least squares model with *d* predictors, the BIC is given by,

$$\text{BIC} = \frac{1}{n}\left(RSS + \log(n)d\hat{\sigma}^2\right) \tag{5}$$

## 3 Research Method

### 3.1 Research Data

The research data are taken from the Indonesian Demographic and Health Survey in 2017 about the birth weight of babies born to women of child-bearing age. In this research, the samples were infants born under LBW conditions. The population of this study was 1176 infants.

**Table 1.** Features of the Research

| Features | Scale | Features | Scale |
|---|---|---|---|
| Y (response) Birth weight | Ratio | X7 Wealth Index | Ordinal 1 = Upper-income economies 0 = middle to lower income economies |
| X1 Mother's Age | Ratio | X8 Insurance | Nominal 1 = have insurance 0 = No insurance |
| X2 Mother's occupation | Nominal 1 = working 0 = unemployed | X9 Antenatal Care | Nominal 1 = hospital/clinic 0 = home or others |
| X3 Mother's education | Nominal 1 = education 0 = no education | X10 Smoking | Nominal 1 = Yes 0 = No |
| X4 Father's Age | Ratio | X11 Birth Order | Ratio |
| X5 Father's education | Nominal 1 = education 0 = no education | X12 Multiple Birth | Nominal 1 = Yes 0 = No |

| X6 | Nominal | X13 | Nominal |
|---|---|---|---|
| Place of | 1 = Urban | Sex | 1 = Male |
| residence | 0 = Rural | | 0 = Female |

The birth weight of the baby serves as an outcome or response variable. The features are demographic, socio-economic, biological factors, and lifestyle, as mentioned in Table 1. This study used 13 factors contributing to LBW. There are the demographic conditions of mother's and father's age, occupation, education level, place of residence, family socio-economic – economic status, ownership of health insurance, antenatal care, biological factors – multiple births, birth order, child's gender, and lifestyle – smoking.The analyzed variables have two types of scales – ratio and binary. Some variables are sorted into dummy categories (binary) that all covariates can be standardized before selecting with LASSO and Fused LASSO.

## 3.2 Model

The linear model of this study formulated as:

$$Y^* = \beta_0 + \beta_1 X_1^* + \cdots + \beta_{13} X_{13}^* + \varepsilon = X^* \boldsymbol{\beta} \tag{6}$$

Where
$Y^*$is the standardized response vector with the size of $1176 \times 1$ ;
$X_j^*$ is the $j^{th}$standardizedfeatures, $j = 1, 2, \ldots 13$ (Table 1)

Estimating coefficients in LASSO and Fused LASSO, each technique uses equations (1) and (2), so that$\hat{\beta}_0 = 0$, considering $\bar{y}^* = 0$ and$\bar{x}^* = 0$. LASSO estimator does not require the order of features, hence estimating of the LASSO coefficient is only by adjusting equation (4) to equation (1). Unlike the LASSO, the Fused LASSO estimators require an order of features. Therefore equation (2) becomes;

$$\hat{\beta}^{FL} = arg\,min\left\{ \,_i\, y_i - \sum_j x_{i[j]}\beta_{[j]} \right\}^2 \tag{7}$$

where $L1 = \,_{j=1}^{p}|\beta_{[j]}| \le t_1$ and $L2 = \,_{j=1}^{p}|\beta_{[j]} - \beta_{[k]j}| \le t_2$, and $x_{i[j]}$ is the feature $j$, the $i^{th}$ observation ;$\beta_{[j]}$ is the $j^{th}$ parameter, and $\beta_{[k]j}$ is the neighboring parameter of the feature $j$.

## 3.3 Procedure of Data Analysis

The data were analyzed by using R-studio software, and the initial step is:

1. Standardizing the outcome (Y) and each feature ($X_j$, $j = 1, 2, \ldots 13$), therefore the data frame is formed (Y * and X *), where
   $Y^* = \dfrac{Y - \bar{Y}}{\sigma_Y}$ and $X_j^* = \dfrac{X_j - \bar{X}_j}{\sigma_{X_j}}$ , $j = 1, 2, \ldots 13$

Where $\bar{Y}$= mean of the outcomes; $\bar{X}_j$= mean of the feature $j$; $\sigma_Y$ and $\sigma_{X_j}$ each of them is the standard deviation of the variables Y and $X_j$.

Calculating the Pearson's correlation between X* and Y* matrix

$$r_{X^*Y^*} = \frac{n\sum_{i=1}^n X_i^* Y_i^* - \sum_{i=1}^n X_i^* \sum_{i=1}^n Y_i^*}{\sqrt{n\sum_{i=1}^n X_i^*)^2 - \sum_{i=1}^n X_i^*)^2}\sqrt{n\sum_{i=1}^n Y_i^*)^2 - \sum_{i=1}^n Y_i^{*\ 2}}}$$

Where $X^* = \left(X_1^*, X_2^*, ..., X_p^*\right)$ is the $n \times p$ matrix of features.

2. Sorting the features using the simulation 1 technique based on the maximal correlation order of X* and Y or $r_{x_{[1]}y} \geq r_{x_{[2]}y} \geq \cdots \geq r_{x_{[p]}y}$.
3. Sorting the features using the simulation 2 technique based on the maximal correlation order of Xj, $j = 1,2 \dots p - 1$ and $i \neq j$. Xi is variable with the largest possible correlation with Y $\left(r_{x_iy} = \max\left\{r_{x_jy}\right\}\right)$.
4. Sorting the features using the simulation 3 technique based on the maximal correlation order of the X $(r_{x_ix_j})$features; for each $i,j = 1,2 \dots p - 1$ and $i \neq j$.
5. Selecting features by using the Fused LASSO technique based on the three settings of techniques in stages 2-4, see equation (2) with using model in equation (6).
6. Selecting features by using LASSO technique, see equation (1).
7. Comparing the results of selecting variables between the LASSO and the three Fused LASSO techniques.
8. Determine the best model based on AIC criterion, see equation (4) or BIC criterion, see equation (5).


## 4  Results and Discussion

### 6.1  Ordering featurestechniques

The order of the X variable can be obtained by determining the neighboring variable through Pearson's correlation value in Table 2. The correlation value between the X variables given is not strong that the value equals $0 \leq r_{x_ix_j} \leq 0.5$. The explanatory variable that has the strongest positive correlation with the Y variable (LBW) is X9 (Pregnancy Care) for 0.11. Therefore, X9 is chosen as the first variable in all three simulation techniques. The next step is to determine the order of the next features based on the 3 simulation techniques. Each method forms a structure of the X matrix data sorted based on the order of the maximal correlation among the X variables. Therefore, the X variable settings form data frame $\mathcal{X} = \left(X_{[1]}, X_{[2]}, ..., X_{[p]}\right)$, and [j] = the j$^{th}$ order feature.

Simulation 1 sets the features based on the correlation value of each X variable with the Y as outcome or response variable, called the $\mathcal{X}1$ data matrix (see the order of column 1 in Table 2).

The correlation value of $(r_{x_jy})$ is sorted from the highest to the lowest correlation, resulted in $r_{x_9y} \geq r_{x_3y} \geq \cdots \geq r_{x_{12}y}$.

**Table 2.** Correlation Value between Variables

| Cor | Y | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | 1 | 0.03 | 0.05 | 0.05 | -0.01 | -0.02 | -0.02 | 0.02 | 0.00 | **0.11** | -0.01 | -0.07 | -0.15 | -0.01 |
| X1 | 0.03 | 1 | 0.09 | -0.01 | 0.11 | 0.01 | **0.16** | -0.18 | 0.03 | 0.03 | -0.04 | -0.29 | -0.02 | 0.02 |
| X2 | 0.05 | 0.09 | 1 | -0.03 | 0.11 | -0.01 | 0.01 | -0.09 | **0.06** | 0.07 | 0.03 | 0.05 | 0.02 | 0.02 |
| X3 | 0.05 | -0.01 | -0.03 | 1 | -0.18 | **0.31** | 0.03 | -0.10 | 0.01 | 0.02 | -0.05 | -0.14 | 0.01 | -0.03 |
| X4 | -0.01 | 0.11 | **0.11** | -0.18 | 1 | -0.08 | 0.04 | 0.01 | 0.04 | 0.07 | 0.03 | **0.55** | -0.04 | 0.02 |
| X5 | -0.02 | 0.01 | -0.01 | **0.31** | -0.08 | 1 | 0.06 | -0.10 | 0.04 | -0.02 | -0.10 | -0.10 | 0.01 | -0.04 |
| X6 | -0.02 | **0.16** | 0.01 | 0.03 | 0.04 | 0.06 | 1 | -0.43 | 0.05 | 0.05 | 0.00 | -0.08 | 0.10 | 0.02 |
| X7 | 0.02 | -0.18 | -0.09 | -0.10 | 0.01 | -0.10 | -0.43 | 1 | -0.02 | -0.09 | 0.00 | 0.17 | -0.11 | -0.02 |
| X8 | 0.00 | 0.03 | 0.06 | 0.01 | 0.04 | 0.04 | 0.05 | -0.02 | 1 | 0.03 | -0.03 | 0.06 | -0.03 | -0.01 |
| X9 | **0.11** | 0.03 | 0.07 | 0.02 | 0.07 | -0.02 | 0.05 | -0.09 | 0.03 | 1 | 0.02 | 0.01 | -0.22 | 0.02 |
| X10 | -0.01 | -0.04 | 0.03 | -0.05 | 0.03 | -0.10 | 0.00 | 0.00 | -0.03 | 0.02 | 1 | 0.02 | 0.03 | -0.01 |
| X11 | -0.07 | -0.29 | 0.05 | -0.14 | **0.55** | -0.10 | -0.08 | **0.17** | 0.06 | 0.01 | 0.02 | 1 | **0.11** | 0.01 |
| X12 | -0.15 | -0. | 0.02 | 0.01 | -0.04 | 0.01 | 0.10 | -0.11 | -0.03 | -0.22 | **0.03** | 0.11 | 1 | **0.05** |
| X13 | -0.01 | 0.02 | 0.02 | -0.03 | 0.02 | -0.04 | 0.02 | -0.02 | -0.01 | 0.02 | -0.01 | 0.01 | 0.05 | 1 |

Furthermore, the $\mathcal{X}2$ data matrix is compiled based on the order of the maximum correlation of the X9 variables with other covariate variables, so that $r_{yx_9} \geq r_{x_9x_2} \geq \cdots \geq r_{x_9x_{12}}$. The $\mathcal{X}3$ data matrix was arranged based on the proximity of variables through the maximum correlation value between the X variables, so that it obtained $r_{x_9y}, r_{x_2x_4}, \ldots, r_{x_3x_{13}}$. The order statistics of the three techniques are summarized in Table 3.

**Table 3** Ordered of the features

| Order | Order Statistics 1 | Order Statistics 2 | Order Statistics 3 |
|---|---|---|---|
| 1 | X9 | X9 | X9 |
| 2 | X3 | X2 | X2 |
| 3 | X2 | X4 | X4 |
| 4 | X1 | X6 | X11 |
| 5 | X7 | X1 | X7 |
| 6 | X8 | X8 | X10 |
| 7 | X13 | X3 | X12 |
| 8 | X10 | X13 | X6 |

| 9 | X4 | X10 | X1 |
| 10 | X5 | X11 | X8 |
| 11 | X6 | X5 | X5 |
| 12 | X11 | X7 | X3 |
| 13 | X12 | X12 | X13 |

## 6.2   Selection of Lambda

The stages of selecting parameters using LASSO begin by setting the initial coefficient of all values to 0 (zero). Estimation of the LASSO regression coefficient is obtained by determining the standardized limit, s = |beta|/max|beta|. Figure 2 shows the stages of selecting variables with LASSO. The smaller the value of s (close to zero), the more likely the parameters are shrunk towards zero. This is in contrast to the value $\lambda$ (selection parameter) in the penalized least squares method used for the LASSO and Fused LASSO.
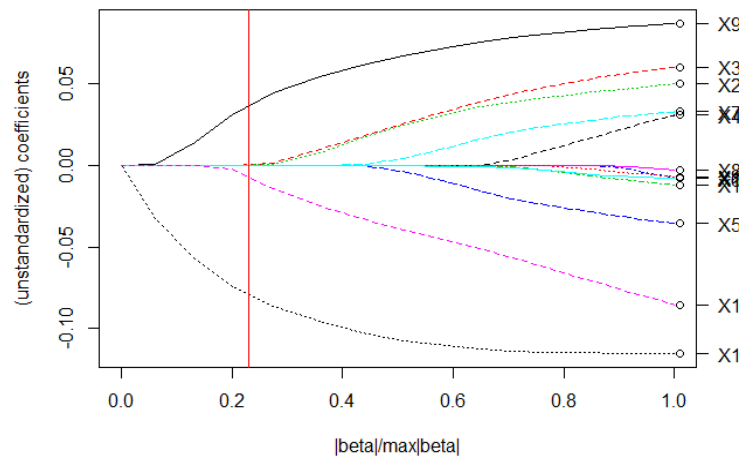


**Fig. 1.** The selection of parameters based on the standardized value of s using LASSO technique.

Determination of optimum lambda uses the candidate obtained from the plot. lqa in Figure 2, i.e {0.10, ..., 400}. Optimum lambda is obtained from the results of Cross-Validation (CV) with 5-fold so that the LASSO optimum lambda value is $\lambda_1 = 70$ with s = 0.23. Based on the optimum $\lambda_1$ and s values, the LASSO model shrinks 8 variables close to zero. The stages of selecting variables as the model with LASSO (Figure 2) are X12, X9, X11, X3, and X2. The selection of optimum

lambda in Fused LASSO 1 and 3 with 5-fold CV is $(\lambda_1, \lambda_2) = (50, 10)$, and the optimum lambda for Fused LASSO 2 is $(30, 30)$.

### 6.3 Stages of selecting parameter using LASSO

LASSO selection results in a more concise estimation (sparsity) than the three Fused LASSO techniques. LASSO selection shrinks eight features, most of which have a small effect towards zero (Figure 3) so that the remaining five features affected the LASSO model. Five features influencing the LBW condition are Mothers' Occupation (X2), Mothers' Education (X3), Antenatal care (X9), Birth Order (X11), and Multiple Birth (X12). The four features are categorized except the birth order variable (X11). Thus, the explanation of the coefficients in Table 4 is adjusted to the dummy conditions.

The positive or negative value of the explanatory coefficient explains the effect on the addition or subtraction of the Y value – the babies' weight. The coefficients of the explanatory X2, X3, and X9 variables are positive. It means that the three variables are increasing the babies' weight. Working and educatedmothers or mothers who do their pregnancy care at the hospital/clinic cause a positive influence on increasing the babies' weight. The coefficients of Birth Order (X11) and Multiple Birth (X12) variables are negative. In this case, mothers who give birth to children in higher-order and give multiple-births are at risk of lowering their babies' weight.
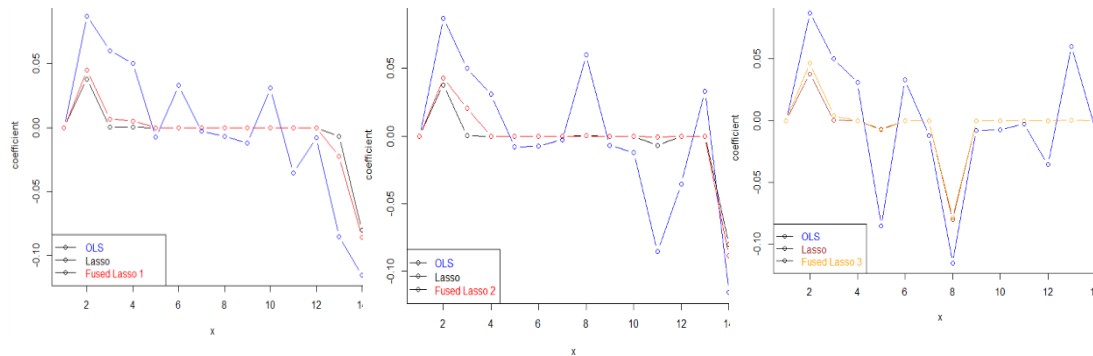


**Fig. 2.** Coefficient of OLS, *LASSO*, *Fused LASSO* 1, 2 and 3

The selection technique by using the LASSO model is the most concise compared to the three Fused LASSO models considering more features are shrunk to zero – eight features. On the contrary, the selection of Fused LASSO 1 and 2 shrink only five features towards zero and leaving 8 others influencing variables on the LBW model. Based on the Fused LASSO 2 and 3 models, the fathers' education (X5) affected the LBW cases negatively although the value was respectively close to zero. Those values are -0,0001 and -0,0003. However, in the Fused LASSO 3 model, the X5 variable is not included. The features in the Fused LASSO 1 technique covered all other features in the LASSO model. Moreover, it is added with three other variables, i.e. the mothers'

age (X1), the fathers' education (X5), and residence (X6) since the values of those additional features are close to zero (Table 4). The features with positive effects on the model are the mothers' age, the mothers' occupation, the mothers' education, and antenatal care. Otherwise, the fathers' education, living conditions, birth order, and multiple births have a negative influence on this model.

The babies' birth weight will increase when the variables that have a positive effect are high and the negative ones are low. In LBW models it turns out that the factors that strongly influenced the increasing of the babies' birth weight are the conditions when the mothers work and have formal education (school), conduct antenatal care at the Hospital/ Clinic, give the low order of birth, and the giving a single birth. In the Fussed LASSO 1 model, three other variables were added, yet it is not significantly affected the babies' birth weight, which was only |0.0001|. Besides, mothers who are mature enough in their first time of giving birth, fathers who do not go to school, and living in the rural slightly influence the babies' weight gain.

**Table 4.** The selection of variables using LASSO and Fused LASSO

| Parameter | LASSO | Fused LASSO 1 | Fused LASSO 2 | Fused LASSO 3 |
|---|---|---|---|---|
| X1 | 0 | 0.0001 | 0 | 0 |
| X2 | 0.0003 | 0.0068 | 0.0204 | 0.0041 |
| X3 | 0.0001 | 0.0052 | 0.0001 | 0.0002 |
| X4 | 0 | 0 | 0 | 0 |
| X5 | 0 | -0.0001 | -0.0003 | 0 |
| X6 | 0 | -0.0001 | 0 | -0.0001 |
| X7 | 0 | 0 | -0.0002 | 0 |
| X8 | 0 | 0 | 0 | 0 |
| X9 | 0.0378 | 0.0448 | 0.0430 | 0.0465 |
| X10 | 0 | 0 | -0.0001 | 0 |
| X11 | -0.0070 | -0.0224 | -0.0006 | -0.0073 |
| X12 | -0.0803 | -0.0858 | -0.0887 | -0.0785 |
| X13 | 0 | 0 | 0 | 0 |

In contrast to the Fused LASSO 1 model, the variables added to the Fused LASSO 2 model are the fathers' education, wealth, and smoking status. Educated fathers, withupper-income economies and smoker mothers turn out to influence the LBW. The Fussed LASSO 3 model is simpler than the other two Fused LASSO models. All variables in the LASSO model are included in the Fussed LASSO 3 model and another variable, place of residence, with small valued equals -0.0001.

Generally, variables that have a strong influence on LBW risk are mothers who do not work, do not go to school, and have never had pregnancy care at the hospital/clinic, mothers who have given birth several times and mothers who give multiple births.


## 6.4 Goodness of Fit of the Model

The model's goodness of fit statistics applied two values – AIC and BIC (see Table 5). The selection of Fused LASSO 1 presented the smallest AIC and BIC values (1143,807 and 1161,698) among the other three models – LASSO, Fused LASSO 2 and Fused LASSO 3. The ordering technique of the features by considering the correlation of the response variable contributes a better estimation than the other ordered statistics techniques. However, The Fused LASSO 3 model is simpler than the other three techniques since it considers the correlation among X's.

**Table 5.** The Goodness of Fit of the Model

| Goodness of fit | *LASSO* | *Fused LASSO* 1 | *Fused LASSO* 2 | *Fused LASSO* 3 |
|---|---|---|---|---|
| **AIC** | 1147.617 | **1143.807** | 1145.562 | 1146.612 |
| **BIC** | 1163.208 | **1161.698** | 1162.335 | 1162.995 |

## 5  Conclusion

The results of this study indicate that the results of selection using the LASSO technique produced the sparsest models compared to other techniques. However, based on the goodness of fit of the model (AIC and BIC), the Fused LASSO 1 model produces the smallest AIC and BIC values, namely 1143,807 and 1161,698, respectively. Hence, the Fused LASSO 1 technique is better than the other Fused LASSO techniques. Although the model provided is not as simple (concise) as the Fused LASSO 3 model. The orderingfeaturestechnique, based on the correlation order of X and Y, forms a better model compared to other techniques.

Based on the coefficient value considered from the results of the four models, generally, there are five feature that contribute to LBW in Indonesia. They are mothers' occupation, mothers' age, antenatal care, multiple birth, and birth order. Mothers who do not work, no education, no antenatal care at the hospital/clinic, mothers who have given birth several times, and mothers who give multiple births are at risk of having low birth weight babies.

In an unordered variable condition, we suggested for further research in developing the selection of X data frames by other techniques. One of the possible techniques is linear optimization techniques such as the Knapsack problem.

## References

[1]  Adam, B.L., Qu, Y., Davis, J. W., Ward, M. D., Clements, M. A., Cazares, L. H., Semmes, O. J., Schellhammer, P. F., Yasui, Y., Feng, Z. and Wright, Jr, G. L. W.: Serum protein fingerprinting coupled with a patternmatching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy mean. Cancer Res. pp. 3609–3614(2003)

[2]  Adamson, H.: Low birth weight in relation to maternal age and multiple pregnancies at Muhimbili National Hospital.Dar Es Salaam Medical Students' Journal 14.2. pp. 55-58(2007)

[3]  Almond, D. V., Kenneth, Y. C., and David, S.L.: Does Low Birth Weight Matter?: Evidence from the US Population of Twin Births. Center for Labor Economics, University of California, Berkeley. pp. 1-75.(2002)

[4]   Bategeka, L.,Asekenye, M.L.O., Barungi, M., Apolot, J.M.: The determinants of birth weight in Uganda. African Economic ReseachConsertium (AERC) . pp. 1-31 (2009)

[5]   Breiman, L.: Better Subset Regression Using the Nonnegative Garrote. J. Technometrics 37(4). pp. 373-384 (1995)

[6]   Breiman, L.: Heuristics of Instability and Stabilization in Model Selection. The Annals of Statistics 24. pp. 2350–2383. (1996)

[7]   Billign, N., Legesse, M., Akibu, M.: A Review of Low Birth Weight in Ethiopia: Socio-Demographic and Obstetric Risk Factors. Glob J Res Rev Vol.5. pp. 1-4 (2018)

[8]   Fan, J., and Li, R.: Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American statistical Association 96.456. pp. 1348-1360 (2001)

[9]   Golub, T., Slonim, D., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J., Coller, H., Loh, M., Downing, J., Caligiuri,M.,Bloomfield,C.andLander,E.: Molecularclassificationofcancer:classdiscoveryandclass prediction by gene expression monitoring. Science. pp. 531–536 (1999)

[10]  Heru, P. F. K, Wilandari, Y., Rusgiyono, A.: PemodelanPertumbuhanEkonomiJawa Tengah MenggunakanPendekatan Least Absolute Shrinkage and Selection Operator (*LASSO*). JURNAL GAUSSIAN, Volume 4, Nomor 4, ISSN: 2339-2541, pp. 855-864 (2015)

[11]  Hoerl, A. E., and Kennard, R.: Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12. pp. 55–67 (1970)

[12]  James, G., Witten D., Hastie, T., Tibshirani, R.: An Introduction to Statistical Learning with Applications in R. pp. 1-423. Springer, New York (2013)

[13]  Karim, M. R., Modal, M. N. I., Rana, M. M., Karmaker, H., Bharati, P., Hossain, M. G.: Maternal Factor are Important Predictor of Low Birth Weight Evidence from Bangladesh Demographic & Health Survey-2011. Malaysian Journal of Nutrition 22(2). pp. 257-265 (2016)

[14]  Kemenkes RI.: Hasil Utama Rikesdas 2018. pp. 1-200. Kemenkes RI, Jakarta (2018)

[15]  Okurut, F. N., Kepaletswe, F., and Mmopelwa.D.: Determinants of Birth weight in Botswana. International Journal of Academic Research in Economics and Management Sciences. pp. 107-117 (2013)

[16]  Ministry of Health Republic of Indonesia.: Strategic Planning Ministry of Health 2015-2019. Pp. 7-118. Ministry of Health of The Republic of Indonesia, Jakarta (2015)

[17]  Ministry of Health Republic of Indonesia, and WHO.: State of Health Inequality Indonesia, pp. 1-133. The WHO Document Production Services, Geneva (2017)

[18]  Soleh, A.M., Aunuddin.: *LASSO*: SolusiAlternatifSeleksiPeubah Dan PenyusutanKoefisien Model Regresi Linier. In *Forum Statistika dan Komputasi* Vol. 18, No. 1. pp. 21-27 (2013)

[19]  Syuichi, O.: The effect of an increase in the rate of multiple births on low-birth-weight and preterm deliveries during 1975–2008. Journal of epidemiology. pp. 480-488 (2010)

[20]  Tibshirani, R.: Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society, Ser. B, 58. pp. 267–288 (1996)

[21]  Tibshirani, R., Sounders, M.: Sparsity and smoothness via the fused lasso." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67.1. pp. 91-108 (2005)

[22]  Vahdaninia, M., Sedigheh, S. T., and Montazeri, A.: Correlates of low birth weight in term pregnancies: a retrospective study from Iran. BMC pregnancy and childbirth 8.1. pp. 8-12 (2008)

[23]  Zou, H. and Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society B 67. pp. 301–320 (2005)