Comparison of multinomial logistic discriminant analysis (mlgda) and classification and regression tree (cart) performance in classifying the impact of working children

M Z A Amri¹, I M Sumertajaya¹ and U D Syafitri¹

¹ Department of Statistics, IPB University, Bogor, 16680, Indonesia

al_amri@bps.go.id, imsjaya.stk@gmail.com

Abstract. Working children have some impacts that befall them. They are hampered by access to education, exploited in working hours and even experience the two impacts at the same time. This research focus on applying and comparing the classification performance of the impact of working children by multinomial logistic discriminant analysis (MLgDA) and classification and regression tree (CART). MLgDA is useful for dealing with cases that do not meet the linearity assumption and suitable for data containing numerical and categorical predictors. CART is not affected by outliers, collinearity and heteroscedasticity among predictors. The data sourced from National Labor Force Survey 2018. All predictors are categorical and characteristics between regions are different. More than half of working children in Indonesia face some impacts, reaching 54.31%. The best classification performance is produced by MLgDA, with an accuracy of 78.73%. MLgDA is superior to CART in almost all measures.

1. Introduction
Child labor in Indonesia has become a national important issue addressed comprehensively and continuously. Various legal regulations have been formulated by the government to deal with child labor problems. The latest issue in 2014, Indonesia through the Ministry of Manpower has issued a Roadmap “Towards a Child Labor-Free Indonesia in 2022” [1].

Five years before the target year, which is in 2017, Central Bureau of Statistics (BPS) released the proportion of working children at 7.23 percent, has increased since 2015 which only 5.99 percent [2]. They experienced various impacts of problems, including never having education, dropped out of school and various working hours, ranging from an hour to more than 40 hours a week. The 46.89 percent of working children no longer attending school and 1.02 percent working children never attending school [2]. The 58.76 percent of children aged 10-17 work more than 15 hours a week [2]. These findings are really serious because the impact of working children causes poor quality of human resources of a nation.

Iryani and Priyarsono [3] examined the factors affecting working children exploited in terms of working hours, exploited in terms of wages and exploited in terms of access to education. One finding is the education of head of household (HH) is one of the factors that cause working children to be exploited in terms of working hours and access to education. Fahlevi and Muhammad [4] identified socioeconomic factors influencing the working hours of child labor in Banda Aceh City. The results
showed that the number of dependents, the status of a single parent, the education level of child and the wage level significantly affected the children’s working hours.

The statistical problem that arises from this phenomenon is finding the right analytical method to classify the impact of working children. Several methods that can be used are discriminant analysis and classification and regression tree (CART). Discriminant analysis aims to find functions that classify an object into subpopulations/classes. When the data is a multivariate normal, homoscedastic and a variance-covariance matrices of all predictors of the same \( p \times p \) size for each class, then the appropriate discriminant function is the Fisher’s linear function. But in fact, not all cases meet the linearity assumption. Logistic discriminant analysis (LgDA) is recommended for dealing with non-linear data, where the discriminant function is the logit function [5]. Binary logistic regression (LR) as an approach for LgDA in the two-class cases and multinomial logistic regression (MLR) as an approach for LgDA in unordered multiclass cases. LgDA is more effective than linear discriminant analysis (LDA) for multiclass classification. The advantages of LgDA include: (i) suitable for data containing numerical and categorical variables; (ii) the formulation is easily understood and (iii) suitable for various distributions [6]. Next, we will utilize the terminology of multinomial logistic discriminant analysis (MLgDA) as a substitute for LgDA in unordered multiclass cases.

CART is one of nonparametric regression methods developed for classification analysis in a large dimension data cluster with numerical and categorical variables. CART produces a classification tree on categorical responses or produces a regression tree on numerical (continuous) responses. The effect of predictors and the estimation of their responses are carried out on groups of observations determined based on predictors, thus results interpretation is easier to be understood. CART is efficient in classifying new samples and has been widely tested for its implementation in software [6]. Other advantages of CART include: (i) being able to overcome missing value; (ii) not affected by outliers, collinearity and heteroscedasticity among predictors and (iii) no data transformation [7].

There are several measures in assessing the ability of a classification method to predict observations. Galar et al. [8] recommended using accuracy, true positive rate (sensitivity) and true negative rate (specificity). The results of each method are calculated in the number of correct classifications and the number of incorrect classifications. The results can be easily displayed in a classification table or confusion matrix. The best classification method is the one with the highest value of the three measurements.

Therefore, the aim of this study is to apply and compare the results of the classification of the impact of working children with MLgDA and CART based on the value of accuracy, sensitivity and specificity.

2. Methodology

2.1. Materials

The data sourced from National Labor Force Survey August 2018 collected by BPS. The definition of working children in accordance with BPS and International Labor Organization (ILO), is all people under 18 years old who work to earn or help earn income, at least one consecutive hours in the past week. These activities include the activities of unpaid workers. The number of observations in this study were 9,563 working children.

The selection of predictors based on previous research by Iryani and Priyarsono [3], Adiangga [9] and Fahlevi and Muhammad [4]. The categorization of classes with definitions refers to those used by BPS and ILO [10], Ministry of Manpower [11] as well as research conducted by Iryani and Priyarsono [3], including:

- \( Y=1 \), working children are hampered by access to education, if the working children never attending school or no longer attending school (dropped out of school).
- \( Y=2 \), working children are exploited in working hours, if: (i) children aged 10-12 work, regardless of the number of working hours per week; (ii) children aged 13-14 work >15 hours per week and (iii) children aged 15-17 work > 40 hours per week.
• Y=3, working children experience both.
• Y=0, working children do not experience both \( \leftrightarrow \) the reference class.

Table 1. Predictor information.

| Notation | Predictor | Categorization |
|----------|-----------|----------------|
| X1       | Living location | 0 = Urban* 1 = Rural |
| X2       | Number of household members (HM) | 0 ≤ 3* 1 = 4 2 = 5 3 = 6 4 ≥ 7 |
| X3       | Child’s relationship with HH | 0 = Non offspring* 1 = Offspring |
| X4       | Child’s sex | 0 = Female* 1 = Male |
| X5       | Child’s age | 0 = 10-12* 1 = 13-14 2 = 15-17 |
| X6       | Child’s education level | 0 = never attending school/not yet completed primary school 1 = Primary school 2 = Junior high school 3 = High school* |
| X7       | Child’s main activity a week ago | 0 = Working 1 = Attending school 2 = Others* |
| X8       | Child’s main employment | 0 = Agriculture 1 = Others* |
| X9       | Child’s working length on the main job (months) | 0 < 12 1 = 12-23 2 = 24-35 3 ≥ 36* |
| X10      | Child’s main employment status | 0 = Entrepreneur* 1 = Laborer 2 = Free worker/unpaid worker |
| X11      | Child’s net earnings a month | 0 = 0 1 ≠ 0* |
| X12      | HH’s sex | 0 = Female* 1 = Male |
| X13      | HH’s age | 0 < 40* 1 = 40-49 2 ≥ 50 |
| X14      | HH’s marital status | 0 = Single* 1 = Married |
| X15      | HH’s education level | 0 = never attending school/not yet completed primary school 1 = Primary – junior high school 2 = High school and above* |
| X16      | HH’s difficulties | 0 = No* 1 = Little/some 2 = Total |
| X17      | HH’s main activity a week ago | 0 = Working 1 = Housekeeping 2 = Others* |
| X18      | HH’s main employment | 0 = Agriculture 1 = Service 2 = Others* |
| X19      | HH’s working length on the main job (months) | 0 < 120 1 = 120-239 2 ≥ 240* |
| X20      | HH’s total working hours a week | 0 ≤ 40* 1 > 40 |
| X21      | HH’s main employment status | 0 = Entrepreneur* 1 = Laborer 2 = Free worker/unpaid worker |
| X22      | Financial accounting at workplace of HH | 0 = No/uninformed 1 = Yes |
Table 1. Predictor information.

| Notation | Predictor                          | Categorization                  |
|----------|-----------------------------------|---------------------------------|
| X23      | HH’s net earnings a month         | 0 = 0*                          |
|          |                                   | 1 ≠ 0                           |

Note: * = reference category

2.2. Methods
Data analysis used software SPSS 24 and R 3.5.2. The steps are as follows:
1. Data preparation.
2. Data exploration with descriptive statistics to find a general description and characteristics of working children in Indonesia 2018.
3. Calculate Cramer’s V for each predictor with the impact of working children. Cramer’s V is a measure to determine the level of power of associations between two nominal variables, one or both of which have more than two categories [12].

Table 2. Interpretation of Cramer’s V.

| Cramer’s V | Interpretation     |
|------------|--------------------|
| > 0.25     | Very strong        |
| > 0.15     | Strong             |
| > 0.10     | Moderate           |
| > 0.05     | Weak               |
| > 0        | No or very weak    |

Predictors with a very weak association level will not be included in the next stages of analysis.

4. Divided data randomly into training data and testing data with a proportion of 3:1.
5. Build a classification model of the impact of working children with MLgDA on training data and validate it on testing data. Subpopulation/classes in MLgDA are considered as responses in MLR [5]. The conditional probabilities for Y for certain X based on logistical functions according to Hosmer and Lemeshow [13] are as follows:

\[ P(Y = j|x) = \frac{e^{\beta_j(x)}}{\sum_j e^{\beta_j(x)}}, j = 1, 2, ... J \] (1)

The responses will form a binary logistic function comparing a category against the reference category by logit transformation, generally:

\[ g_j(x) = \ln \frac{P(y = j|x)}{P(y = 0|x)} = \beta_{j0} + \beta_{j1}x_1 + \cdots + \beta_{jp}x_p \] (2)

meanwhile \( g_0(x) = 0 \) because the parameter value is zero (\( \beta_0 = 0 \)). The next steps are as follows:

a. Deviance test
The deviance test is used to test the role of predictors in the model simultaneously [13].

b. Wald test
The predictor testing is done one by one using the Wald test [13].

c. Classification
The allocation of observations into a class is based on the logistic discriminant function obtained [6].

6. Build a classification model of the impact of working children with CART on training data and validate it on testing data. CART is an example of a multistage decision making process [6]. The tree structure is built through a recursive splitting algorithm for the explanatory space \( X \). The rules of splitting include [6]:

a. Defining the split
The classification tree is formed by slitting data at each node into two nodes.

b. Goodness of split
Proportion \( p_R \) to the right node and \( p_L \) to the left node. The goodness of split is seen from the decline in entropy heterogeneous function.
\[ \Delta(s, t) = i(t) - p_R i(t_R) - p_L i(t_L) \] (3)

while the entropy function for each node is obtained by formula below:

\[ i(t) = - \sum_j p(j|t) \log_2(j|t) \] (4)

c. Stop-splitting rule
d. Assigning class labels to terminal nodes

7. The results of the classification methods are evaluated through accuracy, sensitivity and specificity [8]. Accuracy shows the size of the number of observations that are precisely classified. Sensitivity is the accuracy of the positive class while specificity is the accuracy of the negative class.

| Actual class | Prediction class |
|--------------|------------------|
| Positive     | True positive (TP) | False negative (FN) |
| Negative     | False positive (FP) | True negative (TN) |

\[ \text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \] (5)
\[ \text{Sensitivity} = \frac{TP}{TP + FN} \] (6)
\[ \text{Specificity} = \frac{TN}{FP + TN} \] (7)

The best classification method is the one with the highest value on these three measures.

3. Results and discussion

3.1. Data description

The target of the National Labor Force Survey August 2018 sample were 200 000 households with a document entry rate of 187 422 households (93.17 percent), which were spread throughout Indonesia. The number of observations in this study are 9 563 working children. They are mostly HH’s children (87.66%). The majority of working children have completed primary school (37.89%) and junior high
school (37.79%). More than half of working children (55.77%) still make attending school as their main activity a week ago. They mostly work in agriculture (51.24%). More than a quarter of children have worked less than one year (29.39%) and one to less than two years (25.97%). Working children are dominated by those who are free workers/unpaid workers (80.76%). Hence they mostly do not get monthly net earnings (74.06%).

The characteristics of working children also influenced by the characteristics of HH. Working children mostly live with HH aged around 40-49 years old (45.39%). Nearly half of working children have HH who completed primary-junior high school (46.73%). The majority of working children live with HH who have worked less than 10 years (46.56%). The 61.22% of working children live with HH who have working hours > 40 hours a week. Working children mostly have a HH whose main employment status is entrepreneur (66.90%).

### 3.2. The impact of working children

The children’s problem in this study does not only about they work or not. Working children have some impacts that befell them, losing the rights they should get at the age of growth. Figure 2 shows that more than half of working children in Indonesia 2018 face some impacts, reaching 54.31%. The 15.53% of them are hampered by access to education. They are exploited in working hours by 19.89%. They are hampered by access to education while at the same time exploited in working hours by 18.89%. The 45.69% remain of them do not experience these two impacts.

![Figure 2. Percentage of the impact of working children in Indonesia, 2018.](image)

It can be said that working children who are not hampered by access to education (65.58%) are still more than those who are hampered by access to education by 34.42%.

![Figure 3. Percentage of working children according to school participation.](image)
The black cells in Table 4 show the percentage of working children in Indonesia 2018 who are exploited by working hours. They are by 38.79%. Ironically, there are children aged 10-12 years old who work (13.63%), even a small portion of them work > 40 hours a week. The 9.57% of children aged 13-14 years old work > 15 hours a week. Children aged 15-17 years old work > 40 hours a week are 15.59%. These ironic findings must become the basis for the government and related parties to tackle child labor to map the extent of the impact that children are experiencing. Furthermore, formulating other more efficient strategies to eliminate the child labor until 2022.

Table 4. Percentage of working children according to age and total working hours a week.

| Age       | Total working hours a week | Total |
|-----------|----------------------------|-------|
|           | ≤ 15 | 16 – 40 | > 40 |       |
| 10 – 12   | 9.44 | 3.82    | 0.37 | 13.63 |
| 13 – 14   | 13.37| 7.48    | 2.09 | 22.94 |
| 15 – 17   | 24.48| 23.36   | 15.59| 63.43 |
| Total     | 47.30| 34.65   | 18.05| 100.00|

3.3. Association of predictors with the impact of working children

There are five predictors that have a very strong association with the impact of working children, including child’s main activity a week ago, child’s age, child’s net earnings a month, child’s main employment status and child’s education level. The main employment of child, HH’s education level and HH’s total working hours a week have a strong association with the impact of working children. There are five predictors that have a very weak association with the impact of working children, including the number of HM, HH’s net earnings a month, HH’s main activity a week ago, HH’s sex and HH’s difficulties. Hence these five predictors are not included in the further analysis using MLgDA and CART. The remaining 10 predictors have a moderate and weak association with the impact of working children.

3.4. Selection of the best classification model

One set of data is divided into training data for 7 170 working children and testing data for 2 393 working children. The model is built using training data. The model is evaluated using testing data. Accuracy describes how accurately a model can classify observations appropriately. Sensitivity indicates what percentage of positive class observations are precisely classified by the model. Specificity indicates what percentage of negative class observations are precisely classified by the model.

All testing in MLgDA used $\alpha = 5\%$. The testing of parameter estimates simultaneously in MLgDA gave the G-test statistics of 11 300.7172 with p-value of 0.0000. This means that there is at least one predictor that significantly affects the impact of working children. The testing of parameter estimates partially gave the Wald test statistics summarized in the following table, which only contains significant parameter estimates.

MLgDA produces the estimation of logit models as follows:

$$g_1(\mathbf{X}) = 1.1382X6(0) - 1.585X6(2) + 0.1999X7(0) - 10.4442X7(1) + 0.7121X8(0) + 0.7317X9(0) - 0.5789X11(0) + 1.3153X15(0) + 1.008X15(1) - 0.2814X19(1) - 0.5129X20(1) + 0.4863X21(1)$$  

$$g_2(\mathbf{X}) = -0.771X6(2) + 2.0385X7(0) + 0.6844X7(1) - 0.7536X8(0) - 0.559X9(1) + 0.7296X10(1) - 0.9536X11(0) + 0.487X20(1)$$
\[ g(\bar{X}) = -0.5382X3(1) + 1.3037X6(0) - 1.1561X6(2) + 2.029X7(0) - 7.7419X7(1) + 0.6013X9(0) + 1.1188X10(1) - 1.9103X11(0) - 0.3795X13(2) + 1.1235X15(0) + 0.8344X15(1) + 0.6516X20(1) \] 

(10)

**Table 5.** Summary of testing in MLgDA of parameter estimates partially.

| Predictors                        | \( \hat{B} \) | SE   | Wald   | Sign. | \( \text{Exp}(\hat{B}) \) |
|-----------------------------------|---------------|------|--------|-------|---------------------------|
| Working children are hampered by access to education | X60           | 1.1382 | 0.3025 | 3.7626 | 0.0001 | 3.1213 |
|                                   | X62           | -1.5850 | 0.2241 | -7.0732 | 0.0000 | 0.2049 |
|                                   | X70           | 0.1999 | 0.1574 | 1.2702 | 0.0204 | 1.2213 |
|                                   | X71           | -10.4442 | 3.1828 | -3.2815 | 0.0001 | 0.0000 |
|                                   | X80           | 0.7121 | 0.1692 | 4.2957 | 0.0000 | 2.0786 |
|                                   | X90           | 0.7317 | 0.2008 | 6.5618 | 0.0000 | 3.7257 |
|                                   | X110          | -0.5789 | 0.2247 | -2.5768 | 0.0100 | 0.5494 |
|                                   | X150          | 1.1353 | 0.1802 | 6.5518 | 0.0000 | 3.7257 |
|                                   | X151          | 1.0080 | 0.1802 | 5.5950 | 0.0000 | 2.7401 |
|                                   | X201          | -0.5129 | 0.1239 | -4.1397 | 0.0000 | 0.5988 |
|                                   | X211          | 0.4863 | 0.2437 | 1.9952 | 0.0460 | 1.6262 |

Working children are exploited in working hours

| Predictors                        | \( \hat{B} \) | SE   | Wald   | Sign. | \( \text{Exp}(\hat{B}) \) |
|-----------------------------------|---------------|------|--------|-------|---------------------------|
|                                   | X62           | -0.7710 | 0.3607 | -2.1377 | 0.0325 | 0.4626 |
|                                   | X70           | 2.0385 | 0.3142 | 6.4888 | 0.0000 | 7.6788 |
|                                   | X71           | 0.6844 | 0.1573 | 4.2957 | 0.0000 | 2.0786 |
|                                   | X80           | -0.7536 | 0.1405 | -5.3644 | 0.0000 | 0.4707 |
|                                   | X91           | -0.5990 | 0.1573 | -3.8073 | 0.0001 | 0.5494 |
|                                   | X101          | 0.7296 | 0.2918 | 2.5002 | 0.0124 | 2.0742 |
|                                   | X110          | -0.9536 | 0.2568 | -3.7127 | 0.0002 | 0.3854 |
|                                   | X201          | 0.4870 | 0.1202 | 4.0516 | 0.0000 | 1.6274 |

Working children experience both

| Predictors                        | \( \hat{B} \) | SE   | Wald   | Sign. | \( \text{Exp}(\hat{B}) \) |
|-----------------------------------|---------------|------|--------|-------|---------------------------|
|                                   | X31           | -0.5382 | 0.1773 | -3.0351 | 0.0024 | 0.5838 |
|                                   | X60           | 1.3037 | 0.3089 | 4.2199 | 0.0000 | 3.6289 |
|                                   | X62           | -1.1561 | 0.2364 | -4.8913 | 0.0000 | 0.3147 |
|                                   | X70           | 2.0290 | 0.2077 | 9.7691 | 0.0000 | 7.6065 |
|                                   | X71           | -7.7419 | 1.4643 | -5.2872 | 0.0000 | 0.0004 |
|                                   | X90           | 0.6013 | 0.1702 | 3.5326 | 0.0004 | 1.8244 |
|                                   | X101          | 1.1188 | 0.2702 | 4.1414 | 0.0000 | 3.0613 |
|                                   | X110          | -1.9103 | 0.2166 | -8.8212 | 0.0000 | 0.1480 |
|                                   | X132          | -0.3795 | 0.1760 | -2.1568 | 0.0310 | 0.6842 |
|                                   | X150          | 1.1235 | 0.1906 | 5.8942 | 0.0000 | 3.0756 |
|                                   | X151          | 0.8344 | 0.1691 | 4.9331 | 0.0000 | 2.3035 |
|                                   | X201          | 0.6516 | 0.1249 | 5.2184 | 0.0000 | 1.9185 |

The validation results of classification by MLgDA are summarized in the following table.

**Table 6.** Confusion matrix of MLgDA.

| Actual class | Prediction class | 0  | 1  | 2  | 3  |
|--------------|------------------|----|----|----|----|
| 0            | 991              | 51 | 10 | 41 |
| 1            | 27               | 261| 1  | 83 |
| 2            | 127              | 295| 14 | 44 |
| 3            | 19               | 84 | 12 | 337|

The classification result of MLgDA produces an accuracy of 78.73%. Sensitivity class 0 of 90.67%, sensitivity class 1 of 70.16%, sensitivity class 2 of 61.97% and sensitivity class 3 of 74.56%.
Specificity class 0 of 86.69%, specificity class 1 of 92.83%, specificity class 2 of 98.80% and specificity class 3 of 91.34%.

CART produces classification tree as follows:

Figure 4. Classification tree of the impact of working children.

The validation results of classification by CART are summarized in the following table.

| Actual class | Prediction class |
|--------------|------------------|
| 0            | 0                | 1    | 2    | 3    |
| 0            | 954              | 81   | 0    | 58   |
| 1            | 0                | 243  | 0    | 129  |
| 2            | 142              | 8    | 256  | 70   |
| 3            | 0                | 77   | 0    | 375  |

The classification result of CART produces an accuracy of 76.39%. Sensitivity class 0 of 87.28%, sensitivity class 1 of 65.32%, sensitivity class 2 of 53.78% and sensitivity class 3 of 82.96%. Specificity class 0 of 89.08%, specificity class 1 of 91.79%, specificity class 2 of 100.00% and specificity class 3 of 86.76%.
Figure 5. Comparison of MLgDA and CART performance.

MLgDA is superior to CART in almost all measures. Accuracy, sensitivity class 0, sensitivity class 1, sensitivity class 2, specificity class 1 and specificity class 3 of MLgDA are higher than CART. Thus, MLgDA is more appropriate in classifying the impact of working children in Indonesia 2018 compared to CART.

4. Conclusion and suggestion

More than half of working children in Indonesia 2018 face some impacts, reaching 54.31%. The 15.53% of them are hampered by access to education. They are exploited in working hours by 19.89%. They are hampered by access to education while at the same time exploited in working hours by 18.89%. The 45.69% remain of them do not experience these two impacts.

The performance classification of working children impact is best produced by MLgDA. MLgDA obtains an accuracy of 78.73%, sensitivity class 0 of 90.67%, sensitivity class 1 of 70.16%, sensitivity class 2 of 61.97%, sensitivity class 3 of 74.56%, specificity class 0 of 86.69%, specificity class 1 of 92.83%, specificity class 2 of 98.80% and specificity class 3 of 91.34%.

Future studies are expected to continue analyzing the impact of working children using MLgDA to the interpretation of parameters and odds ratios. Further research can utilize a method that aims to train several different classifiers and combine their decisions on an output, the ensemble method. The goal is to obtain better classification results.

References
[1] Ministry of Manpower 2014 Peta Jalan (Roadmap) Menuju Indonesia Bebas Pekerja Anak Tahun 2022 (Jakarta: Ministry of Manpower)
[2] Central Bureau of Statistics 2018 Profil Anak Indonesia 2018 (Jakarta: Statistical Bureau of Statistics)
[3] Iryani B S and Priyarsono D S 2013 Eksploitasi terhadap Anak yang Bekerja di Indonesia Jurnal Ekonomi dan Pembangunan Indonesia13 177-195
[4] Fahlevi R and Muhammad S 2017 Faktor Sosial Ekonomi yang Mempengaruhi Curahan Jam Kerja Pekerja Anak di Kota Banda Aceh Jurnal Ilmiah Mahasiswa Ekonomi Pembangunan Unsyiah2 187-196
[5] Kurita T, Watanabe K and Otsu N 2009 Logistic Discriminant Analysis IEEE Transactions on System, Man and Cybernetics 2236-2241
[6] Webb A R, Copsey K D 2011 Statistical Pattern Recognition 3rd Edition (Chichester: John Wiley & Sons, Ltd)
[7] Mardika Z W, Mukid M A and Yasin H 2016 Pembentukan Pohon Klasifikasi Biner dengan
Algoritma CART (Classification and Regression Trees) (Studi Kasus: Kredit Macet di PD.BPR-BKK Purwokerto Utara) *Jurnal Gaussian* 5 583-592

[8] Galar M, Fernandez A, Barrenechea E and Bustince H 2011 A Review on Ensembles for The Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches *IEEE Transactions on System, Man and Cybernetics* 42 463-484

[9] Adiangga D, Wijayanto H and Sartono B 2015 Multivariate Adaptive Regression Spline (MARS) for Modelling of Child Labor in Jakarta *Proceeding of International Conference on Research, Implementation and Education of Mathematics and Sciences* 2 183-190

[10] Central Bureau of Statistics 2010 *Pekerja Anak di Indonesia 2009* (Jakarta: Central Bureau of Statistics)

[11] Ministry of Manpower 2005 *Modul Penanganan Pekerja Anak* (Jakarta: Ministry of Manpower)

[12] Akoglu H 2018 User’s Guide to Correlation Coefficients *Turkish Journal of Emergency Medicine* 18 91-93

[13] Hosmer D W and Lemeshow S 2000 *Applied Logistic Regression 2nd Edition* (New Jersey: John Wiley & Sons, Inc)