BAREKENG: Jurnal Ilmu Matematika dan Terapan
March 2022 Volume 16 Issue 1 Page 075–082
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol16iss1pp075-082

ORDINAL LOGISTIC REGRESSION MODEL AND CLASSIFICATION TREE ON ORDINAL RESPONSE DATA

Jajang[1*], Nunung Nurhayati[2], Suci Jena Mufida[3]

[1,2,3] *Department of Mathematics, Faculty of Mathematics and Natural Sciences, Jenderal Soedirman University, Jl. Dr. Soeparno No.61, Karangwangkal, Purwokerto, 53122, Central Java, Indonesia.*

Corresponding author e-mail: ¹ jajang@unsoed.ac.id*

Abstract. *Logistic regression (LR) is a model that associates the relationship between category-type response variables with quantitative or quantitative and qualitative predictor variables. The prediction of the LR model is in the form of probability. This research studied logistic regression (LR) models and* Classification Trees *in the case of ordinal response variable types. The data used in this research from The Central Statistics Agency (BPS). The research variables used are Human Development Index (HDI), gross enrollment rate for high school, percentage of poor people, open unemployment, and percentage of married age <17 years and some of the related predictor variables in Central Java Province in 2018. The HDI data is categorized into three levels, namely very high, high, and moderate. The results of the ordinal LR model show that there are three factors that influence the HDI, they are the gross enrollment rate for high school (GER), the percentage of the poor, and the proportion of women who married at the age of less than 17 years. Comparison of the accuracy LR model and* Classification Tree *in classification analysis shows that if the training data used is 60%-70% the LR model is better than* Classification Tree*, while the training data used is more than 70% and less than 86% then the* Classification Tree *model is better than LR.*

Keywords: *human development index, ordinal logistic model, classification tree*

Article info:

Submitted: 16th September 2021 *Accepted: 29th January 2022*

How to cite this article:
Jajang, N. Nurhayati and S. J. Mufida, "ORDINAL LOGISTIC REGRESSION MODEL AND CLASSIFICATION TREE ON ORDINAL RESPONSE DATA", *BAREKENG: J. Il. Mat. & Ter.*, vol. 16, iss. 1, pp. 075-082, Mar 2022.

1. INTRODUCTION

The Human Development Index (HDI) introduced by United Nations Development Program (UNDP) since 1990 is used as one of the measurements of human development success carried out in a region at any given time. Range of the HDI value is from 0 to 100 and the achievement values are grouped into 4 categories namely low category (HDI < 60), medium category, $(60 \leq \text{IPM} < 70)$ high category, $(70 \leq \text{IPM} < 80)$, and very high category (80).≥ [1]. The HDI in Central Java Province from 2010 to 2018 has increased. Nevertheless, the HDI value in Central Java Province still less than national level [2]. The change in HDI value is influenced by many factors. Some previous researchers mentioned that the variables that affect HDI are the gross enrollment rate of high school (GER), percentage of the poor people, open unemployment rate, and percentage of the female population whose first marriage age is less than 17 years [4]–[6]. One of many methods to analyze the factors that affect HDI is logistic regression analysis. Relevant logistic regression model When a ordinal-scale response variable is an ordinal logistic regression model [7].

Logistic regression models be able to predict the probability of events based on predictor variables. Logistic regression models can also be used for classification using the probability obtained. Several researchers have studied of the relative usefulness of logistic regression approaches, *classification* and regression *trees* (CART), *chi-squared automatic interaction detection* (CHAID), and *multi-layer perceptron neural networks* (MLPNN) in predicting the bad behavior of inmates. [8]. While in terms of the classification analysis process there is also an algorithm that combines aspects of classification and regression. [9]. On the other hand, in classification analysis there are also many methods that are often used, including classification trees and *random forests.* In the case of classification, if the data analyzed is large enough, the classification method can use *random forests. Random forest* method is commonly used for classification with high-dimensional data that is able to rank prospective predictors through measurement of the importance of its innate variables. [8]. This method can be applied to many types of regression problems including nominal response variables, metrics, and survival analysis [8].

In this study we used low-dimensional data, so we used classification tree analysis. We use classification tree for HDI data analysis. Furthermore, the result accuracy and misclassification of the classification tree model is compared to the logistics model. Thus, the purpose our research is to compare the accuracy of logistic regression models and classification tree model for ordinal data, and to analyze factors that affect HDI in Central Java in 2018.

2. RESEARCH METHODS

The method used in this study is applying the logistic regression model and classification tree model to the Human Development Index (HDI) for ordinal response data in Central Java Province in 2018. The data used is secondary data, namely Human Development Index (HDI) data in 29 districts and 6 cities in Central Java Province in 2018 which are divided into three categories and data from variables that are suspected to affect HDI. These variables include the gross enrollment rate of high school (x_1), percentage of poor people (x_2), open unemployment rate (x_3), and the percentage of women whose first marriage age is less than 17 years (x_4). The data was obtained from the central statistics agency (BPS) of Central Java Province (https://jateng.bps.go.id/). In this study, the division of HDI classification based on BPS is divided into three classifications, namely very high, high, and medium.

2.1 Ordinal Logistic Regression Model

Logistic regression model is one of the statistical methods used to analyze the relationship between response variables and predictor variables, when the response variables are categorical. If the response variable of a logistic regression model has three or more categories with a ordinal scale, then the regression model is called ordinal logistic regression. [7],[9]. Suppose Y is a variable with a J category, and that is ordinal in scale and $\mathbf{x}$ is vector of predictor variables. The ordinal logistic regression model is expressed with [10].

$$\boldsymbol{P}(\boldsymbol{Y} \leq \boldsymbol{j}|\text{x}) = \boldsymbol{\pi_1}(\text{x}) + \boldsymbol{\pi_2}(\text{x}) + \cdots + \boldsymbol{\pi_j}(\text{x}),\ \boldsymbol{j} = \mathbf{1}, \mathbf{2}, \ldots, \boldsymbol{J} - \mathbf{1}. \tag{1}$$

If we set $\boldsymbol{L_j} = \mathbf{logit}[\boldsymbol{P}(\boldsymbol{Y} \leq \boldsymbol{j}|\text{x})]$, then the model on (1) can be written by

$$\boldsymbol{L_{J-1}} = \mathbf{log}\left[\frac{\boldsymbol{P(Y \le J-1|\mathrm{x})}}{\boldsymbol{1-P(Y \le J-1|\mathrm{x})}}\right] = \boldsymbol{\beta_{0_{J-1}}} + \boldsymbol{\beta_1 x_1} + \cdots + \boldsymbol{\beta_p x_p}. \tag{2}$$

From the equation (2), we found the probability for j category

$$\boldsymbol{\pi_j}(\mathrm{x}) = \frac{e^{\boldsymbol{\beta_{0_j}+\sum_{k=1}^{p}\beta_k x_k}}}{\boldsymbol{1+}e^{\boldsymbol{\beta_{0_j}+\sum_{k=1}^{p}\beta_k x_k}}} - \frac{e^{\boldsymbol{\beta_{0_{j-1}}+\sum_{k=1}^{p}\beta_k x_k}}}{\boldsymbol{1+}e^{\boldsymbol{\beta_{0_{j-1}}+\sum_{k=1}^{p}\beta_k x_k}}}, \boldsymbol{j = 2, 3, \dots, J-1} \tag{3}$$

where $\pi_j(x)$ states the probability of category j when given the predictor variable x. An important problem in modeling is the parameter estimation. Here, we use maximum likelihood estimation (MLE) for parameter estimation in this model [11]. In the MLE method, we must find an estimator value that which maximize the likelihood function. In its application to find estimators, the likelihood function is transforms likelihood to ln(likelihood). Logarithm natural likelihood function of the logistic regression model is,

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n}\sum_{\mathrm{j}=1}^{\mathrm{J}} y_{ji} \ln[\pi_j(\mathbf{x}_i)] \tag{4}$$

It can be seen that $\pi_j(\mathbf{x}_i)$ it is a function of the parameter, and the result of the parameter estimation, then the parameter significance test is carried out. One to test significance is the likelihood ratio *test.* The likelihood ratio test *is* used to test whether the predictor variable has an overall effect on the response variable or not. This test compares the loglikelihood value of a complete model against a model that is only by constants (models without predictor variables). The test statistics used are[10]

$$G = -2\log\left[\frac{l_0}{l_1}\right], \tag{5}$$

where l_0 is the likelihood of a model without predictor variables and l_1 is the likelihood of a model involve predictor variable(s). After obtaining the results of the overall significance test, then each predictor variable is tested to identify which predictor variables have an effect. The Wald test is used to test whether each predictor variable partially has a significant on the response variable or not. The test statistics used are [10]

$$W_k = \left[\frac{\widehat{\beta}_k}{\widehat{S}E(\widehat{\beta}_k)}\right]^2 \tag{6}$$

where $\hat{\beta}_k$ is estimator of the parameter *k*, and an $\hat{S}E(\hat{\beta}_k)$ is standard deviation of $\hat{\beta}_k$, $k = 1, 2, \dots, p$.

2.2 Classification Tree Model

Suppose that there is data with response variable *Y* has *J* in the ordered category, $w_1 < w_2 < \ldots < w_J$, and there are p predictor variables, $X_1, X_2 \ldots, X_p$, for n sample unit. According to [12] The classification rule is the function d(x) defined on *X* so that for each *x*, d(x) equals one of the numbers 1, 2, ..., J. In terms of classification, the rules of *the impurity* function are important in classifying. Impurity function for category variables on[13]

$$i_{GG} = \sum_{k=1}^{J}\sum_{l=1}^{J} C(w_j|w_k)p(w_j|t)p(w_k|t), \tag{7}$$

where $C(w_j|w_k)$is the cost of misclassification of sample units assigning categories to category sample units$w_j w_k$ and $p(w_j|t)$is the proportion of units in node *t* category *j* of *Y, j=1, 2, ..., J.* $C(w_j|w_j) = 1, j = 1,2,\ldots,J$, and $C(w_j|w_k) = C(w_k|w_j), j,k = 1,2,\ldots,J$. For any split of binary value of *s* on node *t*, the units in node *t* are partitioned into child nodes t_R and t_L. The *impurity* parent node is calculated using:

$$\Delta I_{pmGG}(t,s) = p(t)i_{GG}(t) - p(t_R)i_{GG}(t_R) - p(t_L)i_{GG}(t_L), \tag{8}$$

where $p(t)$, $p(t_R)$, and $p(t_R)$ are proportion of nodes t, t_R dan t_L, respectively. Hereinafter, suppose that the set of scores $s_1 < s_2 < \ldots < s_J$ are J category. The cost of misclassification is to replace $C(w_j|w_k)$ *deviance* or also absolute deviation. If $C(w_j|w_k)$ *deviance*, $C(w_j|w_k) = (s_j - s_k)^2$, then impurity is [13]

$$i_{GG2} = \sum_{k=1}^{J}\sum_{l=1}^{J}(s_j - s_k)^2 p(w_j|t)p(w_k|t), \tag{9}$$

Furthermore, the equation in (9) can be rewritten as

$$i_{GG2} = 2\sum_{j=1}^{J}(s_j - \bar{s}(t))^2 p(w_j|t) = \frac{2}{p(t)}SS(t), \quad (10)$$

where $SS(t)$ is *deviance* of $s_1.s_2,\ldots,s_J$ scores. Furthermore, from this impurity we obtained *the split* function in the equation (11)

$$\Delta I_{pmGG}(t,s) = \frac{2}{n}[SS(t) - SS(t_R) - SS(t_L). \quad (11)$$

The equation in (11) accordance with the ANOVA separator function of a regression tree built using a scores $s_1, s_2, \ldots, s_J$ as the numerical value of the ordered response variable *Y*. For model implementation, R part package software is used [14].

3. RESULTS AND DISCUSSION

3.1. HDI Description Analysis

Central Java Province is a province consisting of 29 districts and 6 cities and is in the central part of Java Island. Central Java Province is one of the provinces in Java Island whose human development status has begun to be categorized high since 2017. In 2018, the Central Java HDI increase 0,6 points to 71,12 and according to BPS, HDI levels in each region in Central Java Province have different values for each residen and city. [1], [3], [4].

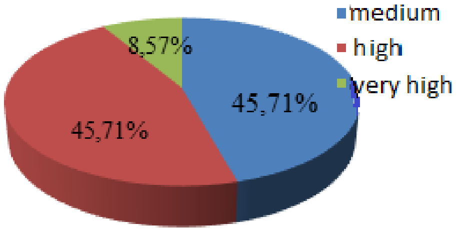

Figure 1. *Pie Chart* of HDI in Central Java Province in 2018

Based on Figure 1. It is seen that there are no residence / cities that fall into the low category. Based on Figure 1, we also see that 8,57% which is in medium category, (*Y*=1), 45,71% which is in the high category (*Y*=2), and 45,71% which is in the very high category (*Y*=3). Therefore, the HDI category in this study is divided into three, namely medium (*Y*= 1), high (*Y*= 2), and very high (*Y*= 3).

3.2. Ordinal Logistic Regression

The important thing to do in a logistic regression model is to test the significance of the model parameters with complete and partial predictor variables. This test is used to determine whether a predictor variable influences the response variable.

3.2.1 Initial model

The initial model parameter significance test is jointly performed with the likelihood ratio test. This test compares the log likelihood values of the complete model (the model with all predictor variables) against the model that is only with constants (models without predictor variables). The test statistics used are [10] $G = -2\ln\left[\frac{l_0}{l_1}\right]$ where l_0 is likelihood of model without predictor variables and l_1 is the likelihood of a model with predictor variables.

Table 1. Sequential Test of initial (1) and final (2) model

Significance test	G-value	Chi Square value	Conclusion
1	44,496	9,488	H_0 rejected
2	43,402	7,815	H_0 rejected

Based on the Table 1, initial model and the second model are significant ($G > X^2{}_{(\alpha,p)}$), so the H_0 was rejected. This means that there is at least one predictor variable affects the response variable. Next, we test the significance of the initial model partially. Significant tests of the initial model are partially used to determine the predictor variables that significantly affect the model. Variable testing by using Wald test. This statistic test is $W_k = \left[\frac{\widehat{\beta}_k}{\widehat{SE}(\widehat{\beta}_k)}\right]^2$ [1]. The result of this statistic test are Presented in Table 2 and Table 3.

Based on Table 2 and Table 3, we can see that p-value < α for GER for high school (x_1), Percentage of poor people (x_2), and Percentage of married age <17 years (x_4). Therefore, H_0 *was* rejected. However, open unemployment is not significant (p-value > 0,05). This means that gross enrollment rate of high school, percentage of poor people, and percentage of married age <17 years influence response variables.

Table 2. Partial Test of Initial Model Estimation

Predictor variable	notation	Parameter	coefficient	p-value	Conclusion
Intercept category 1		β_{01}	-4,453	-	-
Intercept category 2		β_{02}	5,370	-	-
GER for high school	x_1	β_1	0,172	0,033	Significant
Percentage of poor people	x_2	β_2	-0,825	0,014	Significant
Open unemployment	x_3	β_3	-0,350	0,317	Not significant
Percentage of married age <17 years	x_4	β_4	-0,521	0,029	Significant

Table 3. Partial Test of Final Model Estimation

Predictor variable	Notation	Parameter	coefficient	p-value	Keputusan
Intercept category 1		β_{01}	-5,496	-	-
Intercept category 2		β_{02}	3,819	-	-
GER for high school	x_1	β_1	0,139	0,038	Significant
Percentage of poor people	x_2	β_2	-0,787	0,016	Significant
Percentage of married age <17 years	x_4	β_4	-0,537	0,026	Significant

3.2.2 Final Model

In the final model, all of predictor variables (gross enrollment rate for high school, percentage of poor people, and percentage of married women aged <17 years) are significant. Furthermore, we then used three predictor variables. From the results of the estimation of parameters of the second model using R i386 3.6.1 obtained two logit function, as follows:

$$L_1 = -5{,}496 + 0{,}139x_1 - 0{,}787x_2 - 0{,}537x_4$$
$$L_2 = 3{,}819 + 0{,}139x_1 - 0{,}7874x_2 - 0{,}537x_4$$

$L_1 = L_1(Y{=}1)$ is an estimate of the logit function for a medium category and $L_2 = L_2(Y{=}2)$ is an estimate of the logit function for a high category HDI variable.

3.2.3 Goodness of Fit

Goodness of fit is a test that aims to find out whether the model fit to the data. The test used for Goodness of Fit for O_j logistic regression models is $D = 2\sum_{j=1}^{J}\left[O_j \log\left(\frac{O_j}{n\widehat{\pi}_j(\mathbf{x})}\right)\right]$, where O_j is frequency of the j category [1]. This test is called the Deviance test. Based on this statistic test, we found $D = 21{,}435 < X^2{}_{(0,05:65)} = 84{,}821$. Therefore, we concluded that the model fitted to the data.

3.2.4 Accuracy of Classification Results

The accuracy of the model is measured by comparing each category (class) the actual HDI and the predicted HDI. The classification accuracy will be tried with two scenarios scenario 1 and scenario 2. In the scenario 1, we use all observation as training data. Scenario 2 is to use data that is divided into two, namely training and testing data.

Table 4. Table of classification accuracy scenario 1 of LR model

HDI Actual	HDI predictions		
	1	**2**	**3**
1	14	2	0
2	2	13	1
3	0	1	2

Elements on the main diagonal are the result of correct classification. For example, in category 1 there are 16 regencies / cities and after the model and predicted category. It turns out that entering category 1 there are 14, and 2 other regencies / cities incorrectly classified into category 2. In category2, there are 16 regencies/ cities and after the model and predicted category. It turns out that entering category 2 there are 13, 2 other districts / cities misclassify to category 1, and 21 other districts / cities misclassify to category 3. In category 3, there are 3 regencies / cities and after obtaining the model and predicted the category, it turns out that entering category 3 there are 2, 1 other district incorrectly classified to category 2.

Based on the Table 4, the true prediction of a category or class is 14+13+2 =29, so the percentage of misclassification [(35-29)/35]x100% = 17.14%.

Table 5. Table of classification accuracy scenario 1 of Classification Tree model

HDI Actual	HDI prediction		
	1	**2**	**3**
1	14	2	0
2	3	13	0
3	0	3	0

Based on the Table 5, the true prediction of a category or class 14+13+0=27. Thus, the percentage of misclassified percentages of [(35-27)/35]x100%=22.86%. Therefore, based on this misclassified percentages, we conclude that the LR is better than Classification Tree.

The next step is to test reliability of the models based on data in scenario 2. Here, to create the model, we used data training and subsequently to test the model using data testing. The training data comprises 60% to 85% of the total observations and data testing the rest of the percentage taken by training data. Therefore, in this scenario were six sub table of classification accuracy for each model data testing. For example, if the data use 60% of observation (21), then we use 14 for testing, we can calculate classification, accuracy and so on.

The accuracy of the logistic regression and classification tree models based on scenario 2 are presented in Table 6 and Table 7, respectively.

Table 6. Table of classification accuracy scenario 2 of LR model

HDI Actual	HDI predictions																	
	60%			**65%**			**70%**			**75%**			**80%**			**85%**		
	1	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
1	6	2	0	5	2	0	4	1	0	4	0	0	3	0	0	2	0	
2	1	5	0	1	5	0	1	5	0	1	4	0	1	3	0	1	3	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
misclassification	21,43%.			23,08%			18,18%			11,11%			14,29%			16,67%		

Table 7. Table of classification accuracy scenario 2 of Classification Tree model

HDI Actual	HDI predictions																	
	60%			65%			70%			75%			80%			85%		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1	3	5	0	2	5	0	3	2	0	2	2	0	3	0	0	2	0	0
2	0	6	0	0	6	0	2	4	0	0	5	0	0	4	0	0	4	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
misclassification	35,71%.			38,46%			36,36%			22,22%			0,00%			0,00%		

For data training 60%, the accuracy LR is 21,43% and the accuracy classification tree is 35,71%. For data training 65%, the accuracy LR is 23,08% and the accuracy classification tree is 38,46%, and so on. Based on Table 6 and Table 7, we can conclude that for data training no more than 75%, the LR model is better than classification tree, vice versa.

3.1 Interpretation of LR Model

Model interpretation is intended to look at the effect of predictor variables on response variables. The data used to build the model is the overall data. Therefore, in this case, the model taken is a logistic regression model. The interpretation of parameters in a logistic regression model aims to determine the meaning of the estimated value of the parameter on the predictor variable. In ordinal logistic regression, the parameter estimation value can be used to calculate the odds ratio. The *odds* ratio is listed in Table 7.

Table 8. Odds Ratio (OR) of the LR Model

Predictor variable	OR
x_1	1,149
x_2	0,455
x_4	0,584

Based on Table 8, we can interpretate as follows. The value of $OR_j(x_1) = e^{\beta_1} = 1{,}149$. This means that for each category *j*, the increase in the gross enrolment rate of high school (GER high school) 1% can increase the value to 1,149 times. The value of $OR_j(x_2) = e^{\beta_2} = 0{,}455$ mean that for increasing in the number of poor people 1% can reduce the value to 1/0.455 or 2,198 times. The $OR_j(x_4) = e^{\beta_4} = 0{,}584$. This means that for each category *j*, the increase in the value of the married female population at the age of less than 17 years by 1% can reduce the value to 1,712 times less than if there is no increase.odds $(Y \leq j|x_k)$.

4. CONCLUSIONS

1. The predictor variables that affect HDI are GER (x_1), percentage of poor people (x_2), percentage of married women less than 17 years of age (x_4). The predictor variables that have no significant is open unemployment rates(x_3).
2. The gross enrollment rates of high school increases can increase the HDI. The percentage of the poor people and the percentage of the married female population at the age of less than 17 years negatively affect the HDI. Therefore, increase percentage of the poor people and the percentage of the married female population at the age of less than 17 years can decrease the HDI category.
3. Probability for each response category from ordinal logistic regression are

$$\pi_1(\mathbf{x})=P(Y=1|\mathbf{x}) = \frac{\exp(-5{,}496 + 0{,}139x_1 - 0{,}787x_2 - 0{,}537x_4)}{1 + \exp(-5{,}496 + 0{,}139x_1 - 0{,}787x_2 - 0{,}537x_4)}$$

$$\pi_2(\mathbf{x})=P(Y=1|\mathbf{x})=\frac{\exp(3{,}819 + 0{,}139x_1 - 0{,}787x_2 - 0{,}537x_4)}{1 + \exp(3{,}819 + 0{,}139x_1 - 0{,}787x_2 - 0{,}537x_4)} - \frac{\exp(-5{,}496 + 0{,}139x_1 - 0{,}787x_2 - 0{,}537x_4)}{1 + \exp(-5{,}496 + 0{,}139x_1 - 0.787x_2 - 0.537x_4)}$$

$$\pi_3(\mathbf{x}) = P(Y=1|\mathbf{x}) = 1 - \frac{\exp(3{,}819 + 0{,}139x_1 - 0{,}787x_2 - 0{,}537x_4)}{1 + \exp(3{,}819 + 0{,}139x_1 - 0787x_2 - 0{,}537x_4)}$$

4. The results of misclassification of logistic regression models and *CART* classification models in varying percent errors. If the training data used is less than 75%, then the ordinal logistic regression models is better, while for percent the training data is more than 75% then the CART model is better.

AKNOWLEDGEMENT

The author expressed his gratitude to LPPM Jenderal Soedirman University for the funding support in the implementation of this research.

REFERENCES

[1] BPS, *Indeks Pembangunan Manusia Metode Baru*. Jakarta: Badan Pusat Statistik, 2015.
[2] A. N. Rizqi, "Model Regresi Logistik Ordinal IPM Provinsi Jawa Tengah Periode Tahun 2007," *Jurnal Seminar Nasional Statistika*, vol. 9, pp. 1–9, 2009.
[3] BPS.c, *Indeks Pembangunan Manusia 2017*. Jakarta: Badan Pusat Statistik, 2018.
[4] BPS.a, Tran., *Indeks Pembangunan Manusia 2007-2008*. Jakarta: Badan Pusat Statistik, 2009.
[5] Y. D. P. Dwiningtias and M. Mahmudah, "Analisis Regresi Ordinal Model Logit untuk Mengidentifikasi Faktor yang Mempengaruhi Indeks Pembangunan Manusia," *Jurnal Biometrika dan Kependudukan*, vol. 8, no. 2, pp. 174–182, 2019.
[6] R. A. Kusumaningrum, "Analisis Faktor-faktor yang Mempengaruhi Indeks Pembangunan Manusia Provinsi Jawa Tengah Periode 2006-2016," 2018.
[7] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. Wiley New York, 2000.
[8] F. T. Ngo, R. Govindu, and A. Agarwal, "Assessing the predictive utility of logistic regression, classification and regression tree, chi-squared automatic interaction detection, and neural network models in predicting inmate misconduct," *American Journal of Criminal Justice*, vol. 40, no. 1, pp. 47–74, 2015.
[9] S. Kramer, G. Widmer, B. Pfahringer, and M. De Groeve, "Prediction of ordinal classes using regression trees," *Fundamenta Informaticae*, vol. 47, no. 1–2, pp. 1–13, 2001.
[10] S. Janitza, G. Tutz, and A.-L. Boulesteix, "Random forest for ordinal responses: prediction and variable selection," *Computational Statistics & Data Analysis*, vol. 96, pp. 57–73, 2016.
[11] X. Liu and H. Koirala, "Ordinal regression analysis: Using generalized ordinal logistic regression models to estimate educational data," *Journal of modern Applied Statistical methods*, vol. 11, no. 1, p. 21, 2012.
[12] A. Agresti, "Categorical Data Analysis New York," *NY: John Wiley & Sons*, 2002.
[13] L. J. Bain and M. Engelhardt, *Introduction to probability and mathematical statistics*, vol. 4. Duxbury press Belmont, CA, 1992.
[14] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "Classification and regression trees. Pacific Grove,(CA): Wadsworth and Brooks," *BreimanClassification and Regression Trees1984*, 1984.
[15] G. Galimberti, G. Soffritti, and M. Di Maso, "Classification trees for ordinal responses in R: the rpartScore package," *Journal of Statistical Software*, vol. 47, no. 1, pp. 1–25, 2012.
[16] K. J. Archer, "rpartOrdinal: An R package for deriving a classification tree for predicting an ordinal response," *Journal of Statistical Software*, vol. 34, p. 7, 2010.