JAMBURA JOURNAL OF MATHEMATICS

Jambura J. Math. Vol. 4, No. 2, pp. 177–187, July 2022

Journal Homepage: http://ejurnal.ung.ac.id/index.php/jjom

DOI: https://doi.org/10.34312/jjom.v4i2.13529

Backwards Stepwise Binary Logistic Regression for Determination Population Growth Rate Factor in Java Island

Khusnia Nurul Khikmah[1*], Indahwati[2], Anwar Fitrianto[3], Erfiani[4], Reni Amelia[5]

[1,2,3,4,5]*Department of Statistics, Faculty of Mathematic and Natural Science, IPB University, Bogor 16680, Indonesia*

[*]*Corresponding author. Email: khusniank@gmail.com*

ABSTRACT

The high population growth rate can impact various fields due to several factors. Some of the impacts of this high rate are high poverty rates, unemployment, consumption levels, inequality in education figures, gender empowerment index, and increasingly narrow land or area. Therefore, research on the rate of population growth using data on poverty, unemployment, consumption levels, education rates, gender empowerment index, and area makes sense. This data was taken from the official website of the Central Statistics Agency for six provinces on the island of Java, Indonesia. The data used contains missing data so that the missing data is presumed by using the k-nearest neighbour method. The estimated missing data values were modelled using binary logistic regression. Variables that significantly affect the rate of population growth, namely the level of consumption, gender empowerment index, and area, are obtained using the backward stepwise method and are selected based on the smallest Aikakes criterion information value or the one with the most excellent accuracy rate.

Keywords:

Aikake Information Criterion; Backwards Stepwise; Binary Logistic Regression; K-Nearest Neighbour; Population Growth Rate

How to Cite:

K. N. Khikmah, I. Indahwati, A. Fitrianto, E. Erfiani, and R. Amelia, "Backwards Stepwise Binary Logistic Regression for Determination Population Growth Rate Factor in Java Island", *Jambura J. Math.*, vol. 4, No. 2, pp. 177–187, 2022, doi: https://doi.org/10.34312/jjom.v4i2.13529

1. Introduction

The rate of population growth is the rate of population growth per year in a certain period, usually expressed as a percentage of the basic population. The high rate of population growth has an impact on many fields such as the economy [1–3], health and education [4], poverty [3], area [1, 5], unemployment [6]. Based on data published by the Indonesian Central Statistics Agency (BPS), in 2021, the population growth rate in Indonesia will be 0.98% [7]. In addition, the population on the island of Java is the most densely populated, or equivalent to 57% of the total population living in Indonesia [8]. Therefore, research on the population growth rate on the island of Java using a binary logistic regression approach makes sense.

Received: 10 February 2022 | Accepted: 17 March 2022 | Online: 8 May 2022

Binary logistic regression is a statistical method that aims to model multivariable data with binary results [9]. Multivariable data, commonly referred to as independent variables used in binary logistic regression modelling, have a significant relationship with population growth rate in regencies/cities on Java, Indonesia. The significant relationship between the independent variable and the apparent Venetian variable in this study was sought using the backward stepwise method. Backward stepwise is a method that eliminates variables that are not significant to the explanatory variables one by one [10] so that it is more effective in determining the best model. Before the modelling process, pre-processing the data is carried out by estimating the missing values in the data. The method used in predicting missing data in this study is k-nearest neighbour [11, 12]. The best results from modelling are selected based on the highest accuracy value or the smallest Aikake information criterion value.

Therefore, binary logistic regression with backward stepwise was chosen to model the population growth rate in regencies/cities on the island of Java, Indonesia, where the data used is secondary data taken from the official website of the Central Statistics Agency (BPS) for six provinces on the island of Java, Indonesia.

2. Methods

This study uses data on the population growth rate of each regencies/cities in Java, Indonesia. This data uses secondary data taken from the official website of the Central Statistics Agency (BPS) for six provinces in Java, Indonesia. There are six indicator variables wherein two of them there are missing data. The variables that were used are presented in Table 1.

Table 1. Variables used

Variable Indicator	Description
Y	Population growth rate
X1	Poverty
X2	Open unemployment rate
X3	Education
X4	Consumption rate
X5	Gender empowerment index
X6	Area

In general, the first stage of this research is data exploration, where the data used are 119 observations or 119 regencies/cities on the island of Java, Indonesia, with six independent variables, then the descriptive statistics are seen. Second, perform a correlation test to test the independence between each independent variable. Divide the data into training data with test data with a ratio of 80:20, because this ratio has the best performances according to [13]. Training data is used to form a model, while testing data is used to test the model that has been formed. The divided data is checked for missing values, then if the variable has missing, the missing data is estimated with the k-nearest neighbour with $k = 5$. Performing backwards stepwise modelling to determine the significant independent variable by looking at the lowest Aikake information criterion value or the highest accuracy value. The variables that have a significant relationship with population growth are modelled by binary logistic regression. The formed model is then evaluated and seen for its accuracy in predicting.

2.1. K-Nearest Neighbour (KNN)

K-nearest neighbour is one of the statistical methods used to predict missing data values. The method is effective in predicting missing data that has been proven in [14]. This method predicts missing data based on the value of k closest observations. Imputing missing data with k-nearest neighbour can be follows:

1. Setting the number of k, k is the closest observations from the missing data.
2. Calculating the distance by using Euclidean distance. The distance is calculating between the observation that containing missing data and not containing missing data. Where, x_{ai} with $i = 1, 2, 3, \ldots, n$ is the i-th variable that containing missing value in each observation and x_{bi} with $i = 1, 2, 3, \ldots, n$ is the i-th variable that does not containing missing value in each observation, the distance (d) can be calculating by following the equation, namely:
$$d = \sqrt{\sum_{i=1}^{n} (x_{ai} - x_{bi})^2} \tag{1}$$
3. Sorting the distance base on smallest value and determining the closest k observations
4. Imputing missing value, If it is known that the correlation between the k-th neighbour and the missing data is r_k and $\epsilon = 10^{-6}$.). So, the missing data value from the closest k observations is [15]:
$$w_k = \left(\frac{r_k^2}{(1 - r_k^2 + \epsilon)} \right)^2 \tag{2}$$

2.2. Regresi Logistik Biner

Binary logistic regression is a statistical analysis method used to model the relationship between response variables consisting of two categories with one or more explanatory variables, which can consist of categorical or continuous scales [16–18]. Binary logistic regression where $f(x)$ is logit $\pi(x)$, where $\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$, β_0 is a parameter intercept, β_1 is the slope parameter, and x is a qualitative or quantitative variable, where the parameter at $f(x)$ can be estimated using the maximum likelihood approach. Then the equation can be written as follows [17]:

$$f(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x. \tag{3}$$

The β_i coefficient was tested for difference with 0 using Wald's test. This test is carried out based on the Wald contrast, where $\widehat{\beta}$ is the estimated parameter of β_i, $SE_{\widehat{\beta}}$ is the standard error of $\widehat{\beta}$, and has the equation [19]:

$$Z_{wald} = \frac{\widehat{\beta}}{SE_{\widehat{\beta}}}. \tag{4}$$

The interpretation of the binary logistic regression modelling results in this study uses the odds ratio. Where odds itself is the probability of an event that has a ratio between the probability of the event occurring and not occurring and has the equation [19]:

$$\theta = \frac{odds_1}{odds_2} = \frac{\frac{\pi(1)}{1-\pi(1)}}{\frac{\pi(2)}{1-\pi(2)}}. \tag{5}$$

2.3. *The Goodness of Fit Test*

2.3.1. Hosmer-Lemeshow Test

The data that has been modelled with binary logistic regression is then measured for its suitability with the model. The goodness of fit test used in this analysis is the Hosmer-Lemeshow test. This test has a hypothesis, H_0, where the model used follows the data, and H_1, where the model used is not following the data. The results of the Hosmer-Lemeshow test will reject H_0 if the p-value is less than α [20].

2.3.2. Likelihood Ratio Test (LRT)

The likelihood ratio test in this study was used to test the hypothesis of no association between all explanatory variables [21]. The hypothesis in this test is H_0:$\beta_1 = \beta_2 = \cdots = \beta_k = 0$ for the model with k coefficients. The likelihood ratio test results will reject H_0 if the value of LRχ^2 is less than the p-value. The goodness of fit χ^2 with k degrees of freedom, then [22]:

$$\chi^2 = -2\log\frac{likelihood\ of\ the\ null\ model}{likelihood\ of\ the\ given\ model}\ . \tag{6}$$

2.3.3. Aikake Criterion Information (AIC)

Akaike's Information Criteria (AIC) was first introduced in 1973. This Akaike's Information Criteria (AIC) value minimizes the loss of information in model selection and provides information used to determine the best model. The best model was obtained based on the smallest Akaike's Information Criteria (AIC) value [23]. In general, Akaike's Information Criteria (AIC), where k is the model's parameter, and L is the maximum value of the likelihood function used to estimate the model, is formulated as follows [23–25]:

$$AIC = 2k - 2\ln(L)\ . \tag{7}$$

2.3.4. Predictive Performance

The classification results from modelling can be measured for accuracy by looking at the value of the confusion matrix. The confusion matrix itself shows the values of accuracy, sensitivity, and specificities of the model in the classification. Model discrimination is also seen based on the area under the receiver operating characteristic curve (ROC) [26]. The accuracy of the model with unbalanced data used in this study is seen from the balanced accuracy value is [19]:

$$Balanced\ accuracy = \frac{(sensitivity + specificity)}{2}. \tag{8}$$

3. Results and Discussion

Analysis of the population growth rate of 119 regencies/cities on the island of Java, Indonesia using binary logistic regression, where the data is labelled with a regency or cities with a value of population growth rate of less than 1 labelled 0 (low) and if more than 1 is labelled 1 (height) and are presented in Table 2.

Table 2. Class distribution of population growth rate data

Category	Class size	Distribution size
The low population growth rate	65	54.62%
The high population growth rate	54	45.38%

Data on the rate of population growth and its consequent factors, namely poverty, open unemployment rate, education, consumption level, gender empowerment index, and the area of each regencies/cities on the island of Java, Indonesia, have descriptive statistics presented in Table 3.

Table 3. Descriptive statistics of data

	X1	X2	X3	X4	X5	X6
Min	2.30	2.20	572	33.00	48.70	10.20
First quantil	7.40	4.80	8182	43.90	65.40	191.50
Median	10.10	6.30	13550	48.90	69.90	1008.40
Mean	10.30	7.10	16198	48.60	71.20	1087.70
Third quantil	12.70	9.50	19293	52.60	76.30	1481.60
Max	23.80	13.10	69802	63.00	98.20	5782.40
NA's	-	-	-	8	6	-

Source: *R software Output, results of data processing on population growth rates in regencies/cities on the island of Java, Indonesia, and the independent variables*

The data used in this research are 119 regencies/cities in Java, Indonesia. This data will be divided into training data of 80% or 96 observations and testing data of 20% or 23 observations. The data that has been split into 2 parts is then seen for lost data, lost data on training, and testing data are presented in Figure 1.

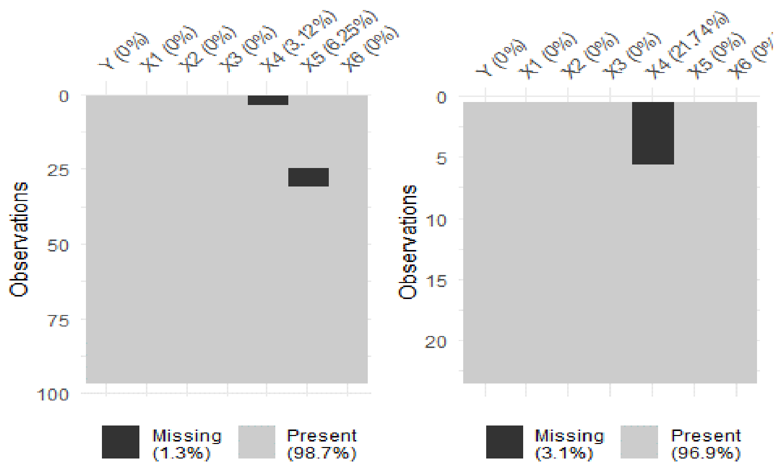

Figure 1. Left: Training data, Right: Testing data

The missing data in this study is estimated using the k-nearest neighbor (KNN) method with $k = 5$, which means that the missing data is estimated based on the five closest observations from the missing data. The estimated data in the train data are then tested for correlation using the Pearson method, and the results are presented in Figure 2.

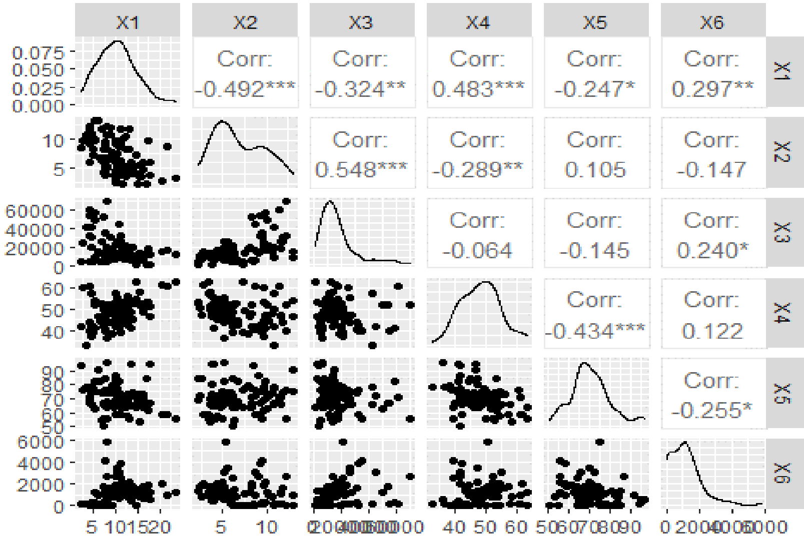

Figure 2. Correlation of explanatory variables on training data

Explanatory variables with a significant correlation coefficient are indicated by an asterisk, where the more stars, the more significant. Modelling using binary logistic regression is done by comparing several explanatory variables. The best model was selected using the backward stepwise logistic regression method. These results are in line with the research in [27]. The modelling results with backward stepwise logistic regression are presented in Table 4.

Table 4. Best Model based on Aikake information criterion (AIC) value

Model	Explanatory variables	AIC Value
1	X1, X2, X3, X4, X5, X6	138.20
2	X2, X3, X4, X5, X6	136.20
3	X2, X4, X5, X6	134.20
4	X4, X5, X6	132.21

Source: *R software output, the result of population growth rate data modelling backwards stepwise*

Based on Table 4, the best model formed using backwards stepwise logistic regression is model 4, with the lowest Aikake information criterion value or 132.21. This model contains explanatory variables, namely the level of consumption, gender empowerment index, and the area of regencies/cities on the island of Java, Indonesia. The binary logistic regression modelling results are presented in Table 5.

From Table 5, the binary logistic regression model can be written as follows:

$$f(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] \tag{9}$$
$$= -6.7745277 + 0.0790132\,(X4) + 0.0455942\,(X5) - 0.0003894\,(X6)$$

Table 5. Binary logistic regression results

	β	Std. Error	Wald	p-value	Odds Ratio
Intercept	-6.7745277	3.3870370	-2.00	0.0455	0.0011425
X4	0.0790132	0.0411979	1.92	0.0551	1.0822186
X5	0.0455942	0.0272526	1.67	0.0943	1.0466495
X6	-0.0003894	0.0002508	-1.55	0.1205	0.9996011

Source: *R software output, results of population growth rate modelling using binary logistic regression*

Based on the modelling results in Table 5 and equation (9), the odds ratio value can be interpreted as, if without paying attention to the independent variable, the probability of population growth rate of regencies/cities on the island of Java, Indonesia, is low, which is 0.0011425 times compared to the probability of population growth rates of regencies/cities on the island of Java, Indonesia is high. The consumption level of regencies/cities on the island of Java, Indonesia, with a low population growth rate, tends 1.0822186 to be higher than that of regencies/cities on the island of Java, Indonesia, with a high population growth rate assuming all other independent variables are held constant. The gender empowerment index of regencies/cities on the island of Java, Indonesia, with a low population growth rate, tends 1.0466495 to be higher than that of regencies/cities on the island of Java, Indonesia with a high population growth rate assuming all other independent variables are held constant. The area of regencies/cities on the island of Java, Indonesia, with a low population growth rate, tends 0.9996011 to be wider than regencies/cities on Java, Indonesia with a high population growth rate assuming all other independent variables are held constant.

This modelling was tested for significance level using the likelihood ratio test presented in Table 6.

Table 6. The results of the model fitting

Likelihood Ratio Test		
LR χ^2	Degree of freedom	p-value
8.21	3	0.0420

Source: *R software output, the results of fitting the population growth rate model with binary logistic regression*

The likelihood ratio test is used, the model-fitting results show that the LR χ^2 value is 8.21 compared to the p-value of 0.0420. It can be concluded that the model used is significant where with a significance level of 5%, the LR χ^2=8.21$>$ p-value=0.0420 as a result of this H_0 is rejected, which means that the response variable, namely the population growth rate on the island of Java, Indonesia, has a close relationship with the three selected explanatory variables. Furthermore, this model will be tested for the goodness of fitness test with the Hosmer-Lemeshow test. The value of x^2 from the Hosmer-Lemeshow test of 6.3769 with a significant level of α= 5% has a p-value of 0.6051. So, it means the model does not reject H_0 because the p-value of $x^2 > \alpha$ or the model used follows the data.

The binary logistic regression model formed in equation (5) is evaluated by testing the model's accuracy for training data, testing data, and visually using the ROC curve. The model's accuracy with the training data is shown in the classification results in Table 7.

Table 7. Final classification result of data training using binary logistic regression

Prediction	Actual	
	A	B
A	53	43
B	0	1

Notes:
A : The low population growth rate
B : The high population growth rate
Source:
R software output, the results of the confusion matrix on the rate of the population growth model with binary logistic regression with training data

The classification results using binary logistic regression model training data have an accuracy of 55.21% with a balanced accuracy of 51.13% and a confidence interval between 44.71% and 65.37%.

The best model formed from the training data is then tested on the testing data to measure the model's accuracy, and the results are presented in Table 8.

Table 8. Final classification result of data testing using binary logistic regression

Prediction	Actual	
	A	B
A	13	8
B	0	2

Notes:
A : The low population growth rate
B : The high population growth rate
Source:
R software output, the results of the confusion matrix on the rate of the population growth model with binary logistic regression with testing data

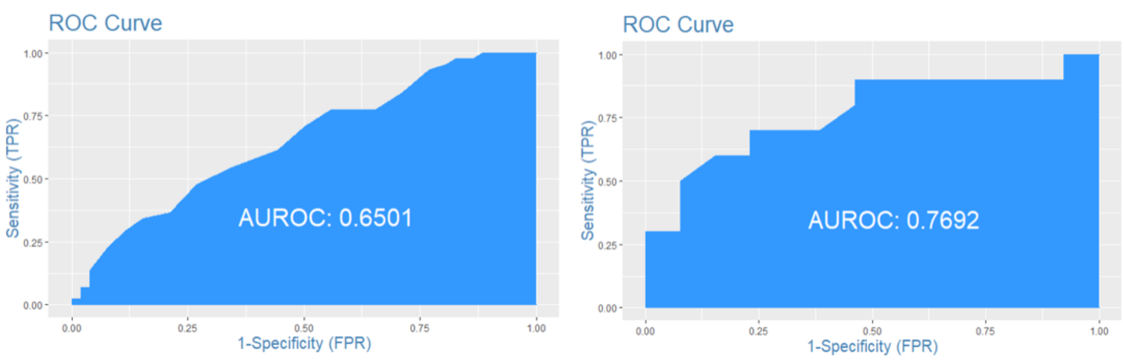

Figure 3. Left: ROC curve of training data classification model, Right: ROC Curve of testing data classification model

The classification results tested on testing data from the binary logistic regression model of training data had an accuracy of 65.22% with a balanced accuracy of 60% and a confidence interval between 42.73% and 83.62%. It can be concluded that the binary logistic regression with the best model obtained using backwards stepwise logistic regression has a good classification ability. The results of the model's accuracy in the classification are presented in the ROC curve in Figure 3.

4. Conclusion

This study provides several results based on the analysis of the binary logistic regression method. Missing data used in modelling can be estimated using the k-nearest neighbour method with $k = 5$. The best model in classifying the low rate of population growth in regencies/cities on the island of Java, Indonesia, was obtained using the backwards stepwise binary logistic regression method. The variables associated with the low rate of population growth in regencies/cities on the island of Java, Indonesia, are the level of consumption of the population of regencies/cities, the gender empowerment index of the population of regencies/cities, and the area of regencies/cities on the island of Java, Indonesia. The recommendations include promoting community empowerment programs that aim to improve and equalize welfare. Research on selecting the best variable in binary logistic regression can be done backwards stepwise and with other methods such as forward stepwise for the next.

References

[1] R. Kurniawan and S. Managi, "Economic Growth and Sustainable Development in Indonesia: An Assessment," *Bulletin of Indonesian Economic Studies*, vol. 54, no. 3, pp. 339–361, sep 2018, doi: http://dx.doi.org/10.1080/00074918.2018.1450962.

[2] J. Pegou Sibe, C. Chiatchoua, and M. Noel Megne, "The long run relationship between population growth and economic growth: a panel data analysis of 30 of the most populated countries of the world," *Análisis Económico*, vol. 31, no. 77, pp. 205–218, 2016.

[3] M. Cruz and S. A. Ahmed, "On the impact of demographic change on economic growth and poverty," *World Development*, vol. 105, pp. 95–106, may 2018, doi: http://dx.doi.org/10.1016/j.worlddev.2017.12.018.

[4] R. Efendi, S. Indartono, and S. Sukidjo, "The Relationship of Indonesia's Poverty Rate Based on Economic Growth, Health, and Education," *International Journal of Multicultural and Multireligious Understanding*, vol. 6, no. 2, pp. 323–330, may 2019, doi: http://dx.doi.org/10.18415/ijmmu.v6i2.704.

[5] R. F. Putri, S. Wibirama, S. R. Giyarsih, A. Pradana, and Y. Kusmiati, "Landuse change monitoring and population density analysis of Penjaringan, Cengkareng, and Cakung Urban Area in Jakarta Province," *E3S Web of Conferences*, vol. 76, p. 03004, jan 2019, doi: http://dx.doi.org/10.1051/e3sconf/20197603004.

[6] F. A. Kurnianto, D. Rakhmasari, F. A. Ikhsan, B. Apriyanto, and E. A. Nurdin, "The environment analysis of population growth, unemployment, and poverty level in Maesan District Bondowoso Regency," *Geosfera Indonesia*, vol. 3, no. 2, pp. 113–121, aug 2018, doi: http://dx.doi.org/10.19184/geosi.v3i2.8439.

[7] BPS, "Badan Pusat Statistik."

[8] T. Zulkarnain, Y. Hazmi, M. Nasir, F. Faisal, and D. Husin, "Dynamic Response of Dependency Ratio on Government Expenditures in Indonesia," *J. Asian Financ. Econ. Bus.*, vol. 9, no. 2, pp. 71–79, 2022.

[9] M. van Smeden, K. G. Moons, J. A. de Groot, G. S. Collins, D. G. Altman, M. J. Eijkemans, and J. B. Reitsma, "Sample size for binary logistic prediction models: Beyond events per variable criteria," *Statistical Methods in Medical Research*, vol. 28, no. 8, pp. 2455–2474, aug 2019, doi: http://dx.doi.org/10.1177/0962280218784726.

[10] P. Ranganathan, C. S. Pramesh, and R. Aggarwal, "Common pitfalls in statistical analysis:

logistic regression," *Perspect. Clin. Res.*, vol. 8, no. 3, pp. 148–151, 2017, doi: http://dx.doi.org/10.4103/picr.PICR_87_17.

[11] R. Malarvizhi and A. S. Thanamani, "K-nearest neighbor in missing data imputation," *Int. J. Eng. Res. Dev.*, vol. 5, no. 1, pp. 5–7, 2012.

[12] K. T. Do, S. Wahl, J. Raffler, S. Molnos, M. Laimighofer, J. Adamski, K. Suhre, K. Strauch, A. Peters, C. Gieger, C. Langenberg, I. D. Stewart, F. J. Theis, H. Grallert, G. Kastenmüller, and J. Krumsiek, "Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies," *Metabolomics*, vol. 14, no. 10, p. 128, oct 2018, doi: http://dx.doi.org/10.1007/s11306-018-1420-2.

[13] A. Rácz, D. Bajusz, and K. Héberger, "Effect of Dataset Size and Train/Test Split Ratios in QSAR/QSPR Multiclass Classification," *Molecules*, vol. 26, no. 4, p. 1111, feb 2021, doi: http://dx.doi.org/10.3390/molecules26041111.

[14] S. I. Khan and A. S. M. L. Hoque, "SICE: an improved missing data imputation technique," *Journal of Big Data*, vol. 7, no. 1, p. 37, dec 2020, doi: http://dx.doi.org/10.1186/s40537-020-00313-w.

[15] S. G. Liao, Y. Lin, D. D. Kang, D. Chandra, J. Bon, N. Kaminski, F. C. Sciurba, and G. C. Tseng, "Missing value imputation in high-dimensional phenomic data: imputable or not, and how?" *BMC Bioinformatics*, vol. 15, no. 1, p. 346, dec 2014, doi: http://dx.doi.org/10.1186/s12859-014-0346-6.

[16] K. J. Archer and S. Lemeshow, "Goodness-of-fit Test for a Logistic Regression Model Fitted using Survey Sample Data," *The Stata Journal: Promoting communications on statistics and Stata*, vol. 6, no. 1, pp. 97–105, feb 2006, doi: http://dx.doi.org/10.1177/1536867X0600600106.

[17] G. Nattino, M. L. Pennell, and S. Lemeshow, "Assessing the goodness of fit of logistic regression models in large samples: A modification of the Hosmer-Lemeshow test," *Biometrics*, vol. 76, no. 2, pp. 549–560, jun 2020, doi: http://dx.doi.org/10.1111/biom.13249.

[18] J. D. Canary, L. Blizzard, R. P. Barry, D. W. Hosmer, and S. J. Quinn, "A comparison of the Hosmer–Lemeshow, Pigeon–Heyse, and Tsiatis goodness-of-fit tests for binary logistic regression under two grouping methods," *Communications in Statistics - Simulation and Computation*, vol. 46, no. 3, pp. 1871–1894, mar 2017, doi: http://dx.doi.org/10.1080/03610918.2015.1017583.

[19] K. B. Watson, "Categorical Data Analysis," in *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer Netherlands, 2014, pp. 601–604, doi: http://dx.doi.org/10.1007/978-94-007-0753-5_291.

[20] M. Díaz-Pérez, Á. Carreño-Ortega, J.-A. Salinas-Andújar, and Á.-J. Callejón-Ferre, "Application of Logistic Regression Models for the Marketability of Cucumber Cultivars," *Agronomy*, vol. 9, no. 1, p. 17, jan 2019, doi: http://dx.doi.org/10.3390/agronomy9010017.

[21] B. A. L. Cengiz, E. R. Fikret, and H. Sonmez, "A review of statistical techniques for 2x2 and RxC categorical data tables in SPSS," *Journal of Pediatric Sciences*, vol. 1, 2009, doi: http://dx.doi.org/10.17334/jps.16629.

[22] H.-A. Park, "An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain," *Journal of Korean Academy of Nursing*, vol. 43, no. 2, pp. 154–164, 2013, doi: http://dx.doi.org/10.4040/jkan.2013.43.2.154.

[23] S. Ray and B. Bhattacharyya, "Availability in different source of irrigation in India: a statistical approach," *Ecosystem*, p. 109, 2015, doi: http://dx.doi.org/10.5923/c.ije.201501.16.

[24] H. de G. Acquah, "Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship," *J. Dev. Agric. Econ.*, vol. 2, no. 1, pp. 1–6, 2010.

[25] D. R. Anderson, K. P. Burnham, and G. C. White, "Comparison of Akaike information criterion and consistent Akaike information criterion for model selection and statistical inference from capture-recapture studies," *Journal of Applied Statistics*, vol. 25, no. 2, pp. 263–282, apr 1998, doi: http://dx.doi.org/10.1080/02664769823250.

[26] C.-S. Rau, S.-C. Wu, J.-F. Chuang, C.-Y. Huang, H.-T. Liu, P.-C. Chien, and C.-H. Hsieh, "Machine Learning Models of Survival Prediction in Trauma Patients," *Journal of Clinical Medicine*, vol. 8, no. 6, p. 799, jun 2019, doi: http://dx.doi.org/10.3390/jcm8060799.

[27] S. Afiqah Muhamad Jamil, M. Asrul Affendi Abdullah, K. Sie Long, N. Fazilla Mohd Jupri, and M. Mamat, "A Stepwise Logistic Regression Analysis: An application toward Poultry Farm Data in Johor," *International Journal of Engineering and Technology*, vol. 7, no. 3.28, pp. 68–71, aug 2018, doi: http://dx.doi.org/10.14419/ijet.v7i3.28.20968.