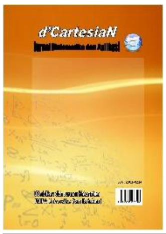


d'CartesiaN

Jurnal Matematika dan Aplikasi

p-ISSN:2302-4224 e-ISSN: 2685-1083

Journal homepage: https://ejournal.unsrat.ac.id/index.php/decartesian

Binary Logistic Regression Analysis of Variables That Influence Poverty Depth Level in West Java Province

Kezhia A. S. A. Zeekeon[1], John S. Kekenusa[1*], Deiby T. Salaki[1]

[1]Jurusan Matematika–Fakultas Matematika dan Ilmu Pengetahuan Alam–Universitas Sam Ratulangi Manado, Indonesia

[*]Corressponding author : johnskekenusa@unsrat.ac.id

ABSTRACT

Poverty continues to be a problem for countries and also regions. The province of West Java is not immune to the problem of poverty, making poverty a matter that must be considered and requires follow-up to alleviate it. This research was conducted with the aim of determining what factors influence poverty depth level in West Java Province and determining the accuracy of the classification obtained from the model. The data used in this research are secondary data in 2022, consisting of data on the poverty depth index, human development index, gini ratio, Gross Regional Domestic Product growth rate, and population growth rate taken from the Central Statistics Agency of West Java. The analysis method used in this research is binary logistic regression based on backward elimination. Factors that influence poverty depth level are the human development index (X_1), and gini ratio (X_2). The classification accuracy of the model is 81.818%, which means the model is good to use.

ARTICLE INFO

Diterima :

Diterima setelah revisi :

Tersedia *online* :

Keywords:
Binary Logistic Regression
Backward Elimination
Poverty Depth Level
West Java Province.

ABSTRAK

Kemiskinan tiada henti menjadi permasalahan bagi suatu negara dan daerah. Provinsi Jawa Barat tidak luput dari masalah kemiskinan, ini menjadikan kemiskinan sebagai suatu hal yang harus diperhatikan dan membutuhkan tindak lanjut untuk mengentaskannya. Penelitian ini bertujuan untuk menentukan faktor apa saja yang memengaruhi tingkat kedalaman kemiskinan di Provinsi Jawa Barat dan menentukan besarnya ketepatan klasifikasi yang diperoleh dari model. Data yang digunakan dalam penelitian ini ialah data sekunder tahun 2022 yaitu data indeks kedalaman kemiskinan, indeks pembangunan manusia, rasio gini, laju pertumbuhan PDRB, laju pertumbuhan penduduk, yang diambil dari Badan Pusat Statistik Provinsi Jawa Barat. Metode analisis yang digunakan adalah regresi logistik biner berbasis eliminasi mundur. Berdasarkan hasil penelitian menggunakan metode analisis regresi logistik biner maka diperoleh variabel yang berpengaruh terhadap indeks kedalaman kemiskinan adalah indeks pembangunan manusia (X_1), dan rasio gini (X_2). Ketepatan klasifikasi dari model sebesar 81.818%, yang artinya model baik digunakan.

INFO ARTIKEL

Accepted :

Accepted after revision :

Available online :

Kata Kunci:
Regresi Logistik Biner
Eliminasi Mundur
Tingkat Kedalaman Kemiskinan
Provinsi Jawa Barat

1. INTRODUCTION

Poverty is defined as a condition where a person or group of people, men and women, do not have their basic rights fulfilled to maintain and develop a decent life[1]. From this definition, it can be seen that poverty regarding a multidimensional problem that is difficult to measure, so an agreed measurement approach is needed to be used. According to the Central Statistics Agency (*Badan Pusat Statistik*, BPS), Indonesia use a poverty calculation concept namely, the concept of the ability to fulfil basic needs or basic approach [2]. Based on that concept, the broad definition of poverty experiencing narrowing because poverty will only be seen as an economic inability to fulfilled basic food and non-food needs.

Based on the data released by BPS, in 2022 the number of poor people in West Java in September increased to 48.6 thousand people compared to September 2021. As a province with a large area, West Java Province has a poor population of 4.053.620 million people. This number is considered very large if compared to other provinces in Indonesia. This situation indicates that the economic development in Indonesia, especially in West Java Province, is still not appropriate and therefore unable to deal with the existing poverty problem.

The human resources of a nation are the most determining factor in the character and speed of social and economic development of the nation concerned [3]. The benchmark used in Indonesia is the Human Development Index (HDI) which is used as evaluation material for several aspects such as level of education, health and economy or purchasing power.

The results of research conducted by Islami [4] has a positive and significant effect on poverty depth level in Indonesia. It indicates that the existence of this Gini ratio will affect economic inefficiency, social inequality, and lack of justice in society due to high inequality, so that the Gini ratio can be a benchmark for the poverty depth level for an area. Another research related to this topic has been conducted by Maulana [5]. Their research concluded that the Gini ratio has a significant influence on the poverty depth level in Indonesia. It is in line with research held by Atmojo [6].

Apart from population growth rate, HDI and Gini ratio, the growth rate of Gross Regional Domestic Product (GRDP) is also an important indicator for assessing economic development. Based on data from BPS, the GRDP growth rate in 2022 in each province on the Java Island, it was found that West Java Province has the highest GRDP growth rate, namely 5.45%. It does not necessarily lead to the conclusion that the higher the GRDP growth rate, the greater the reduction in poverty. Both research of Permana [7], also Suripto [8] concluded that increasing of the GRDP growth rate will reduce the poverty. On the other hand, Bintang [9] inferred that the increasing of GRDP growth rate would worsen/increase poverty. Thus, it is important to include the GRDP growth rate variable in this research so that it can be known whether this variable has an effect on the poverty depth level or not.

To find out factors that influence the poverty depth level in West Java Province, a statistical analysis was carried out to discover the relationship between the response variable which has two categories and several predictor variables in a model. In this research, the logistic regression is appropriate to be used because the poverty depth level is a dichotomous or binary with two categories of the poverty depth level in West Java Province namely high and low.

Furthermore, research conducted by Nurrizqi [10] shows that the classification accuracy of the model obtained using Backward Elimination Based on Binary Logistic Regression is able to predict the data correctly. Fitri [11] also inferred in their research that the accuracy of the model classification obtained using Binary Logistic Regression was good. Based on the results of previous researches, it can be seen that research using Binary Logistic Regression Based on Backward Elimination can be utilized to determine the factors that influence the poverty depth level in West Java Province.

Accuracy of Classification

The accuracy of the classification is seen from the results using the confusion matrix. Actual data and predicted data from the classification model are presented using cross tabulation, which contains information about the actual data classes represented in the matrix rows and the predicted data classes in the matrix columns. At this step, testing will be carried out by calculating accuracy, sensitivity and specificity values as the performance of the model. Accuracy is the percentage of overall classification accuracy, Sensitivity is the positive class accuracy, and Specificity is the negative class accuracy (Hapsari et al., 2022).

Table 1. Classification Table

Actual	Prediction	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Where,

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP} \times 100\% \quad (6)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \times 100\% \quad (8)$$

With,

TP = True Positive

TN = True Negative

FP = False Positive

FN = False Negative

2. RESEARCH METHOD

Data Source

The data used in this research is secondary data obtained from the official website of the BPS in West Java Province. This research uses data totaling 27 regencies/cities in West Java Province in 2022 (18 regencies, 9 cities).

Research Variables

Table 2. List of Variables

Variable	Description	Unit
Y	Poverty depth index	Y = 0 Y = 1
X_1	Human development index	%
X_2	Gini ratio	Index (0 - 1)
X_3	Growth rate of Gross Regional Domestic Product (GRDP)	%
X_4	Growth rate of Population	%

Classifying the Data

Based on research conducted by Nurrizqi *et al.* (2022), classification of poverty depth levels based on poverty depth index data for each regency/city is carried out using the following criteria:

a. If the Poverty Depth Index of the Regency/City < the Poverty Depth Index of West Java Province then it is categorized as having a low poverty depth level and is coded 0.
b. If the Poverty Depth Index of the Regency/City is ≥ the Poverty Depth Index of West Java Province then it is categorized as having a high poverty depth level and is coded 1.

After classifying the data next step, the data will be divided into two parts. The first part is called data training and the second one is called data testing. While the prior is used to obtain a binary logistic regression model the later one is utilized to determine the level of classification accuracy.

In this research the data is divided by the proportion of 60%-40%. The 60% of data (include 16 regencies/cities) is referred as data training, and the 40% of data (include 11 regencies/cities) is known as data testing. Based on the 16 regencies/cities used as data training, 11 regencies/cities were categorized into the high poverty level and 5 regencies/cities were categorized into the low poverty level. However, 3 regencies/cities of testing data were categorized into the high poverty level and 8 regencies/cities were categorized into the low poverty level.

Data Analysis Method

Data analysis in this research uses statistical analysis applications. The analysis stages are carried out as follows.

1. Perform initial data visualization to describe the data.
2. Carry out a multicollinearity test.

Some indicators in detecting multicollinearity include [12]:

- The R^2 value is too high, (more than 0.8) but there are no or few significant t-statistics.
- The F-statistic is significant, but the t-statistic of each predictor variable is not significant.
- Correlation matrix of the predictor variables. If there is a correlation coefficient of more than 0.80 then there is multicollinearity.
- Tolerance and variance inflation factor. As a rule of thumb, if the VIF of a variable exceeds 10, which

will happen if R_j^2 exceeds 0.90, that variable is said be highly collinear.

3. Model all predictor variables (X) with response variables (Y) using binary logistic regression:
   a. Perform parameter estimation.
   b. Carry out parameter testing simultaneously with the G test.

the ratio likelihood test with the G test statistic was used for this test. The following is the formula G test statistic:

$$G = -2ln\frac{L_0}{L_p}$$

With:
$L_0 =$ Maximum likelihood of the model without predictor variables
$L_p =$ Maximum likelihood of the full model or with all predictor variables

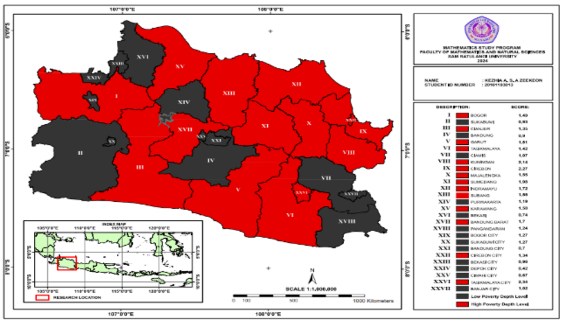
   c. Carrying out partial parameter testing with the Wald test.
4. Reduce predictor variables (X) that have no effect on the response variable (Y) using backward elimination.
   Backward elimination is a feature selection method that has been proven to increase accuracy in classification algorithms [15]. The backward elimination method is the opposite of the forward selection method, where the first step is entering all the predictor variables into the model and then eliminate one by one the variables that do not match the existing criteria. Elimination of predictor variables is based on the AIC value. The best regression model is the regression model that has the smallest AIC value [16].

5. Determine the best model.
6. Calculate the classification accuracy of the model obtained.
7. Interpret the model based on the odds ratio.
   The odds ratio value is used to show the risk/tendency of the predictor variables to the response variable which is a comparison of the relative risk levels of the two predictor variables $X_i = 1$ to $X_i = 0$ (comparison between the possibility of success and the risk of failure. The coefficient of the logit model, beta shows the change in the value of the logit function g(x) for a one-unit change in the predictor variables. Suppose 1 and 0, the relationship between the odds ratio (ϕ) and the regression coefficient can be obtained in equation (5).

$$\phi = e^{\beta_1}$$

3. RESULTS AND DISCUSSION

Visualization of Preliminary Data



Picture 1. Overview of Response Variables

Table 3. Descriptive Statistics of Response Variables

Category	N	Percentage
Category 0	13	48%
Category 1	14	52%
Total	27	100%

Table 4. Descriptive Statistics of Predictor Variables

Variable	Minimum	Maximum	Mean
X_1	65.94	82.50	72.61
X_2	0.2900	0.4800	0.3756
X_3	2.880	6.630	5.110
X_4	0.410	1.860	1.301

Multicollinearity Test

Gujarati (2009) states if the correlation between two independent variables exceeds 0.8, it indicates multicollinearity.

Table 5. Correlation Between Predictor Variables

	X_2	X_3	X_4
X_1	0.567436747	0.2086829	0.18057930
X_2		0.3244515	0.05969748
X_3			0.33705217

The correlation given in Table 5 by each predictor variables is < 0.8. It indicates that there is no multicollinearity or no linear relationship between the predictor variables. Therefore, research can be continued into the next step.

Parameters Estimation

Estimating the parameters of the binary logistic regression model using maximum likelihood estimation with the results that can be seen in Table 6:

Tabel 6. Logistic Regression Model Parameter Estimation Results

Parameters Estimation	**Estimation Value**	**Standard Error**
Intercept	79.8890	32.8651
HDI (X_1)	-1.3775	0.5707
Gini Ratio (X_2	64.6921	27.8764
Growth Rate of GRDP (X_3)	0.1772	0.9035
Growth Rate of Population (X_4)	-4.2705	2.7247

Simultaneous Testing

The results of the simultaneous test calculations in Table 7 show that the G value is 20.0471043 with $\chi^2_{(0.05,3)}$ of 7.815. From this, it can be interpreted that the G value is greater than the chi-square value, so it can be decided that H_0 rejected. Thus it can be concluded that there is at least one predictor variable that has an influence on the response variable.

Table 7. Simultaneous Parameter Test Results

G	$\chi^2_{(0.05,3)}$
20.0471043	7.815

Partial Testing

Table 8. Partial Parameter Test Results

Variable	Df	Wald Value	P-value
HDI (X_1)	1	5.8254	0.0158
Gini Ratio (X_2)	1	5.3855	0.0203
Growth Rate of GRDP (X_3)	1	0.0384	0.8446
Growth Rate of Population (X_4)	1	2.4566	0.1170

~~With $\alpha = 0.05$, $df = 1$ and chi-square table value 3.841459,~~ the results in Table 6 show that the partial parameter test with the value of variable $X_1$and X_2 has a P-value $< \alpha$ with $\alpha = 0.05$, so H_0 is rejected which indicates that the variables has a significant influence on the response variable.

Backward Elimination

Backward Elimination is a step with backwards direction by regressing all predictor variables to the response variable by looking at the AIC value they have. Table 9 are the results of the backward elimination method model selection.

Table 9. Backward Elimination Results

Model 1 Start: **AIC = 27.35** $Y \sim X_1 + X_2 + X_3 + X_4$			
	Df	Deviance	AIC
X_3	1	17.384	25.384
<none>		**17.386**	**27.384**
X_4	1	20.913	28.913
X_2	1	26.490	34.490
X_1	1	35.924	43.924
Model 2 Start: **AIC = 25.38** $Y \sim X_1 + X_2 + X_4$			
	Df	Deviance	AIC
<none>		**17.384**	**25.384**
X_4	1	21.040	27.040
X_2	1	26.526	32.526
X_1	1	36.109	42.109

The first equation model was formed by entering all predictor variables, then eliminating the predictor variable which has the largest AIC value. The prior included all predictor variables and produced a model with the largest AIC value for variable X_3 it is 25.384, thus variable X_3 was removed from the first model.

According to the table 9, the second equation model was made without the variable X_3 and obtained that the AIC value stopped at 25.38. It indicates that the second equation model obtained already has the smallest AIC value, so that the elimination results using the backward method in the final model are the best model obtained. Thus, binary logistic regression modelling is obtained using variables X_1, X_2, and X_4.

Binary Logistic Regression Model

The following are the results of binary logistic regression analysis using the backward elimination method as listed in Table 10.

Table 10. Backward Elimination Models

Variable	Estimation Value	Standard Error	P-Value
Intercept	78.8951	32.0346	0.0138
HDI (X_1)	-1.3520	0.5477	**0.0136**
Gini Ratio (X_2)	64.1938	27.5547	**0.0198**
Growth Rate of Population (X_4)	-4.0963	2.5191	0.1039

According to the Table 10, only 2 variables that has the p-value who under the significance level or in other words only 2 variables that have a significant effect. Such as, HDI and Gini Ratio. Therefore, the existing variables will be regressed to get the equation with the best model.

Table 11. Best Model Results

Variable	Estimation Value	Standard Error	P-Value
Intercept	57.4778	25.2337	0.0227
HDI (X_1)	-1.0349	0.4488	**0.0211**
Gini Ratio (X_2)	45.8850	22.1170	**0.0380**

Seen in table 11, variable X_1 and X_2 has the p-value under the significance level. It can be concluded that variable X_1 and X_2 influenced the poverty depth level. Therefore, the binary logistic regression equation formed is as seen in eq. (8),

$$\pi(x) = \frac{e^{57.4778-1.0349X_1+45.8850\,X_2}}{1+e^{57.4778-1.0349X_1+45.8850\,X_2}} \tag{8}$$

Accuracy Classification of Model

Accuracy of the classification is seen from the results using the confusion matrix. To determine the amount of classification accuracy, it is doing with using test data.

Table 12. Model Classification Accuracy

Actual	Prediction	
	Low (0)	High (1)
Low (0)	7	1
High (1)	1	2

The confusion matrix on table 12 shows that the predictive ability of the model obtained to predict a value of 0 or a low poverty depth level can be predicted correctly in 7 out of a total of 8 regency/city, and for a value of 1 or a high poverty depth level it can be predicted correctly in 2 from a total of 3 regency/city.

Table 13. Model Performance

Accuracy	Sensitivity	Specificity
81.818%	87.5%	66.667%

From the classification performance results seen in table 13, the prediction accuracy using the binary logistic regression model (accuracy) is 81.8%, while the prediction accuracy for value 0 or low poverty depth level (sensitivity) is 87.5% and prediction accuracy for value 1 or the level of depth of high poverty (specificity) is 66.667%. This indicates that the binary logistic regression model obtained is able to predict the data correctly.

Model Interpretation (Odds Ratio)

The odds ratio value is obtained from the results of Exp (β) on significant predictor variables. The results of the odds ratio calculation are shown in Table 14.

Table 14. Odds Ratio Results

Variable	Coefficient Estimation	*Odds Ratio*
HDI (X_1)	-1.03485	3.55279
Gini Ratio (X_2)	45.88505	8.46490

According to the Table 14, the HDI Odds is with a negative coefficient estimate, meaning that when the human development index is increased by 1%, the chance of a high poverty depth index will decrease by 3.55279 assuming other predictor variables are considered constant. The Gini Odds is with a positive coefficient estimate, meaning that when the gini index is increased by 0.1, the chance of a high poverty depth index will increase by 8.46490 assuming other predictor variables are considered constant

4. CONCLUSION

Based on the results of the analysis that has been carried out, the following conclusions are obtained:

1. Variables that influence the depth of poverty in WestJava province are the Human Development Index and Gini Ratio.
2. The classification accuracy of the model obtained is 81.818%.

REFERENCES

[1] Berliani, K. 2021. Pengaruh Tingkat Pengangguran, Tingkat Pendidikan dan Laju Pertumbuhan Penduduk terhadap Tingkat Kemiskinan Penduduk Provinsi Jawa Barat Tahun 2015-2020. *Jurnal Ilmiah Indonesia*. 6(2):872-882.

[2] Badan Pusat Statistik. 2022. Provinsi Jawa Barat Dalam Angka 2022. BPS, Provinsi Jawa Barat.

[3] Todaro, M. P. 1998. Pembangunan Ekonomi di Dunia. Edisi Ke-3. Erlangga, Jakarta.

[4] Islami, M. F. E, dan A. R. Fitrianto. 2023. Pengaruh Penyaluran Dana ZIS, Inflasi, Dan Gini Ratiio Terhadap Tingkat Kedalaman Kemiskinan Satu Dekade. *Jurnal Ilmiah Ekonomi Islam*. 9(**1**): 229-239.

[5] Maulana, M. A., dan Julia, A. 2022. Pengaruh Indeks Pendidikan, Gini Rasio, Jumlah Penduduk, dan Pendapatan Perkapita terhadap Tingkat Kemiskinan di Enam Provinsi Indonesia Tahun 2015-2019. *Bandung Conference Series: Economic Studies*. 2(1):17-24.

[6] Atmojo, D. 2017. Analisis Pengaruh Gini Ratio, Indeks Pembangunan Manusia (IPM), dan Jumlah Penduduk Terhadap Tingkat Kemiskinan di Provinsi Jawa Barat Tahun 2012-2016 [skripsi]. Fakultas Ekonomi dan Bisnis, Universitas Muhammadiyah, Yogyakarta.

[7] Permana, A. Y. dan F. Arianti. 2012. Analisis Pengaruh PDRB, Pengangguran, Pendidikan dan Kesehatan Terhadap Kemiskinan di Jawa Tengah tahun 2004-2009. *DIPONEGORO JOURNAL OF ECONOMICS*. 1(1):1-8.

[8] Suripto, dan L. Subayil. 2020. Pengaruh Tingkat Pendidikan, Pengangguran Pertumbuhan Ekonomi dan Indek Pembangunan Manusia terhadap Kemiskinan di D. I. Yogyakarta Periode 2010-2017. *GROWTH: Jurnal Ilmiah Ekonomi Pembangunan*. 1(2):127-143.

[9] Bintang, A. B. M. dan Woyanti, N. 2018. Pengaruh PDRB, Pendidikan, Kesehatan, dan Pengangguran Terhadap Tingkat Kemiskinan Di Jawa Tengah (2011-2015). *Media Ekonomi dan Manajemen*. 33(1):20-28.

[10] Nurrizqi, A. I., Erfiani, Indahwati, Fitrianto A., dan Amelia R. 2022. Pemodelan Regresi Logistik Berbasis *Backward Elimination* Untuk Mengetahui Faktor yang Memengaruhi Tingkat Kemiskinan di Indonesia Tahun 2021. *Jurnal Statistika dan Aplikasinya*. 6(2):160-170.

[11] Fitri, R. E. 2022. Analisis Regresi Logistik Biner Terhadap Data Indeks Kedalaman Kemiskinan Di Indonesia Tahun 2020 [skripsi]. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung. Bandar Lampung.

[12] Gujarati, D. N, D. C. Porter. 2009. Basic Econometrics. Fifth Edition. McGraw-Hill/Irwin Companies. New York.

[13] Ilham, R. S., Paendong M dan Kekenusa J. 2019. Analisis Regresi Logistik Untuk Menentukkan Kepuasan Pasien Rawat Inap Pada Kualitas Layanan Rumah Sakit Umum Pusat Prof. Dr. R. D. Kandou Manado. *d'Cartesian: Jurnal Matematika dan Aplikasi*. 8(2): 147-152.

[14] Hosmer, D. W., dan S. Lemeshow. 2000. Applied Logistic Regression. John Wiley dan Son. New York.
[15] Rizal, A. S. dan Lutfi, M. 2020. Prediksi Hasil Pemilu Legislatif Menggunakan Algoritma *K-Nearest Neighbor* Berbasis *Backward Elimination*. *Jurnal Resistor: Rekayasa Sistem Komputer*. 3(1):27-42.
[16] Widarjono, A. 2007. Ekonometrika Teori dan Aplikasi. Ekonisia FE UII. Yogyakarta.
[17] Hapsari, N. A., M. S. Chamid, dan N. Azizah. 2022. Faktor-faktor yang Mempengaruhi Berat Badan Lahir Rendah Menggunakan Regresi Logistik Biner. *Jurnal Sains dan Seni ITS*.11(1): 50-56.

Kezhia Aster Sarah Aurelia Zeekeon
(kezhiazeekeon103@student.unsrat.ac.id)

Born in Manado, North Sulawesi on July 20, 2003. She pursued higher education in the Department of Mathematics, FMIPA, Sam Ratulangi University Manado. The year 2024 is her last year of study. This paper is the result of her published thesis research.

John Socrates Kekenusa
(johnkekenusa@unsrat.ac.id)

Born in Tahuna, August 24, 1958. He obtained his Bachelor degree in Fisheries in 1982 at Sam Ratulangi University. In 1988 completed his master's degree in applied statistics at IPB Bogor. In 2006 completed his doctoral studies in the field of Mathematics / Statistics at Airlangga University. Since 2007 he has been appointed as a professor of statistics, in the Department of Mathematics FMIPA Unsrat Manado.

Deiby Tineke Salaki
(deiby.ts@unsrat.ac.id)

Born in South Minahasa on December 17, 1972. In 1998, he received his Bachelor of Mathematics degree from the Mathematics Department of IPB Bogor. In 2009 completed his master study at the Department of Mathematics IPB. In 2018 completed his doctoral study in the field of Mathematics at IPB. Currently a permanent lecturer at the Department of Mathematics FMIPA UNSRAT Manado.