REGIONAL GROUPING BASED ON POVERTY INDICATORS USING FACTOR ANALYSIS, K-MEANS CLUSTERING AND DISCRIMINANT ANALYSIS

By
Yosiana Saputri*[1], Nur Azizah[2], Dwi Endah Kusrini[3], Destri Susilaningrum[4]
[1,2,3,4]Institut Teknologi Sepuluh Nopember; Keputih, Sukolilo District, Surabaya, East Java, (031)5947274
Email: *[1]41.xb.yosianasaputri@gmail.com, [2]azizah@its.ac.id, [3]dwi_endah@statistika.its.ac.id, [4]destri_s@statistika.its.ac.id

Abstract
In 2020 Indonesia was hit by Covid-19, and one of the impacts was that the poverty rate rose. One of the goals of the SDGs is to end all forms of poverty. East Java is one of the provinces that is also affected. The purpose of this study is to describe the factors that shape Poverty Indicators in Madya districts/Cities in East Java Province, map Madya districts/Cities in East Java in 2020 based on poverty indicators, and find out the differences between the groups formed. There are 16 poverty indicator variables used in this study. The data was obtained through East Java Province in Numbers, East Java Provincial Health Statistics, East Java Provincial Education Statistics, and the website of the Central Statistics Agency. The method used is factor analysis, followed by Cluster analysis with K-Means and Discriminant Analysis. The results of the factor analysis form four factors, namely the welfare factor of education, the economic welfare factor, the factor of pln users and baduta breastfeeding, and the factor of contraceptive users (KB). Continuing with the analysis using K-Means, it produces three groups, group one is a group with moderate poverty, group two is a group with high poverty occupied by Sumenep Regency and group three is with low poverty. Followed by the Discriminant analysis, the four factors are distinguishing variables with a classification accuracy of 100 percent. The difference between this research and the research used as a reference is using non-hierarchical clusters (K-Means) while the research used as a reference uses hierarchical clusters, the results of this research analysis there are levels of poverty from the three groups formed.

Keywords: Discriminant Analysis, Factor Analysis, Poverty Indicators, K-Means Cluster, SDGs

INTRODUCTION

Poverty can be interpreted as a state in which a person's inability to meet the basic needs of life (clothing, food, and shelter). Poverty can also cause the pace of economic growth to be hampered [1]. The existence of a global action plan agreed upon by world leaders including Indonesia, the goal is to end poverty, reduce inequality, and protect the environment, the SDGs (Sustainable Development Goals) are initiated which contain 17 goals. One such goal is to end all forms of poverty everywhere [2].

In 2020 Indonesia is amid the Covid-19 pandemic so one of the impacts is poverty. East Java is one of the provinces with the largest population in Indonesia which is also affected by Covid-19. The results of the population census in 2020 showed that the total population of East Java reached 40,665,696 people. From 2011-2019 the number of poor people in East Java managed to fall, but in 2020 it has increased again due to Covid-19 [3]. According to the Central Statistics Agency in March 2020, the number

ISSN No. 1978-3787
Open Journal Systems

of poor people in East Java reached 4,419.10 thousand people (11.09%), an increase of 363.1 thousand people compared to the conditions in September 2019 of 4,056 thousand people (10.20%). One of the causes of poverty is the number of unemployed, the number of open unemployed in East Java in August 2020 was 1,301,145 people, an increase of 466,015 people from 2019 of 835,130 people [4]. Poverty management, especially in East Java, has indeed been carried out, but it is still not optimal, and evenly distributed, so it is necessary to hold groupings, to facilitate the government's efforts to maximize poverty management. The methods used in this study are factor analysis, cluster analysis with the K-Means method, and discriminant analysis.

Previous research related to poverty in East Java in 2002 and 2012 was conducted by Pratiwi [5] using factor analysis, Ward's Method, and discriminant analysis which resulted in conclusions including factor analysis in 2002 producing three factors and four factors in 2012, and 4 groups were formed. The discriminant analysis states that the differentiating variables in the 2002 poverty indicators are found in four variables, namely PLN electricity users, elementary school graduates, high school graduates, and malnutrition. Meanwhile, 2012 has 6 differentiating variables, namely high school graduates, clean water users, Infant Mortality Rate (AKB), Open Unemployment Rate (TPT), School Participation Rate (APS) in Junior High Schools (SMP), and Life Expectancy (AHH). In this study, 16 variables were used in the Poverty indicator.

The purpose of this study is to describe the factors forming the Poverty Indicators madya districts/cities in East Java Province, mapping madya districts/cities in East Java Province based on poverty indicators and analyze the differences between groups of Madya Regencies/Cities in East Java Province based on poverty indicators The benefit of this research is to obtain information related to the spread of poverty based on poverty indicators in madya districts/cities in East Java Province where it can then be used as an effort to minimize poverty. The difference between this research and the research used as a reference is using non-hierarchical clusters (K-Means) while the research used as a reference uses hierarchical clusters, the results of this research analysis there are levels of poverty from the three groups formed.

THEORETICAL BASIS

This study uses 16 variables, and the details of which variables are used will be explained in Table 1. This study uses these variables based on previous research and their relevance to data availability in 2020, besides that Ginanjar Kartasasmita, state minister/chairman of Bappenas 1996 in his book on development for the people stated that low health / low family planning. And according to the national team for accelerating poverty reduction, there are several indicators of poverty, namely in the form of infrastructure, health, and education, in this study the variables used also include these indicators.

Previous research on poverty has also been carried out by Komariyah and Akbar, in 2011 [6] with all his research The *single linkage* method is the best, another research was carried out by Abidin, in 2017 [7] the research result was a grouping formed 4 groups with the best method being *ward's method.*

The data used is secondary data, obtained through the *website* of the Badan Pusat Statistik. The data used was obtained from publication books, namely East Java Province in Numbers 2020 [8], East Java Provincial Health Statistics 2020 [9], East Java Province Education Statistics 2020 [10] and *the website* of the Badan Pusat Statistik [11] for Human Development Index data. The data used in each variable was 38 regencies/cities.

The variables used in this study are presented in Table 1.

Variable	Information	Scale	Unit
X_1	Percentage of women aged 15-49 years who are using contraceptives in madya districts/cities	Ratio	Percentage
X_2	The percentage of childbirth of women aged 15-49 years is helped by medical personnel	Ratio	Percentage
X_3	Percentage of households that have access to proper sanitation in madya districts/cities	Ratio	Percentage
X_4	Percentage of Baduta (Age 0-23 Months) according to breastfeeding in madya districts/cities	Ratio	Percentage
X_5	Percentage of the population aged 10 years and above the last education of SD in madya districts/cities	Ratio	Percentage
X_6	Percentage of the population aged 10 years and above the last education of SMP in madya districts/cities	Ratio	Percentage
X_7	Percentage of the population aged 10 years and above the last education of the SMA in in madya districts/cities	Ratio	Percentage
X_8	Literacy rate aged 10 years and above in madya districts/cities	Ratio	Percentage
X_9	School Participation Rate aged 13-15 Years	Ratio	Percentage
X_{10}	Average length of schooling of residents aged 15 years and over	Ratio	Year
X_{11}	Labor force participation rate (TPAK) in madya districts/cities	Ratio	Percentage
X_{12}	Open unemployment rate (TPT) in madya districts/cities	Ratio	Percentage
X_{13}	Percentage of poor people in madya districts/cities	Ratio	Percentage
X_{14}	Human development index in madya districts/cities	Ratio	-

ISSN No. 1978-3787
Open Journal Systems

Variable	Information	Scale	Unit
X_{15}	Percentage of households of the ownership status of owned premises	Ratio	Percentage
X_{16}	Percentage of households using PLN	Ratio	Percentage

RESEARCH METHOD

Factor analysis is used to reduce the original variable and interpret it as a new variable in the form of a formed variable. Factor analysis is used to see similarities between variables or to see variables that have correlations between variables that will later be grouped into one group or called factors. **X** is a random variable observed with p components, average **μ,** and covariance matrix A linear combination of several dependent variables that are not $\sum$, observed can be referred to as a common factor $F_1, F_2, ..., F_m$ with variance called $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$ error or specific factors. Model factor analysis can be described in the equation below [12].

$$X_1 - \mu_1 = l_{11}F_1 + l_{12}F_2 + ... + l_{1m}F_m + \varepsilon_1 \quad (1)$$

$$X_2 - \mu_2 = l_{21}F_1 + l_{22}F_2 + ... + l_{2m}F_m + \varepsilon_1 \quad (2)$$

$$\vdots$$

$$X_p - \mu_p = l_{p1}F_1 + l_{p2}F_2 + ... + l_{pm}F_m + \varepsilon_1 \quad (3)$$

or it can be written as follows.

$$\mathbf{X} - \boldsymbol{\mu} = \mathbf{L}\ \ \mathbf{F}\ + \boldsymbol{\varepsilon} \quad (4)$$

Where $X_1, X_2, ...X_p$ is the origin variable, $\mu_1, \mu_2, ...\mu_p$ is the average of the i-th variable, and $F_1, F_2, ..., F_m$ is a common factor (common factor). l_{ij} is the coefficient (loading) of the factor of the variable of origin to - i on the factor to - j; **L** is the matrix of the loading factor; starting from i 1, 2, ..., p; j starting from 1, 2, ..., m; ε_{ij} : specific factor to – i; m indicates many formed factors and p many variables. Before conducting factor analysis, an assumption test is carried out (multivariate normal distribution assumption test, Bartlett test, and data adequacy test).

KMO is an index value used to test the adequacy of data. Values greater than 0.5 to 1.0 indicate that the use of factor analysis is appropriate. Meanwhile, if the value is less than equal to 0.5, the use of analysis can be said to be incorrect. If the KMO value > 0.5 then receive H_0 so that it can be concluded that the amount of data is sufficient for analysis using factor analysis [13]. The hypotheses used are H_0: The amount of data is sufficient to be factored in and H_1: The amount of data is not enough to be factored in, with the test statistics presented in equation (5)

$$KMO = \frac{\sum_{i=1}^{p}\sum_{h=1}^{p} r_{ih}^{\ 2}}{\sum_{i=1}^{p}\sum_{h=1}^{p} r_{ih}^{\ 2} + \sum_{i=1}^{p}\sum_{h=1}^{p} a_{ih}^{\ 2}} \quad (5)$$

$$r_{ih} = \frac{n\sum x_i x_h - \sum x_i \sum x_h}{\sqrt{n\sum x_i^{\ 2} - (\sum x_i)^2}\sqrt{n\sum x_h^{\ 2} - (\sum x_h)^2}} \quad (6)$$

Where $\sum_{i=1}^{p}\sum_{h=1}^{p} a_{ih}^{\ 2} = r_{x_i x_h - x_k}$, here is the formula of the partial correlation coefficient [14].

$$r_{x_i x_h - x_k} = \frac{r_{x_i x_h} - (r_{x_i x_k})(r_{x_h x_k})}{\sqrt{[1 - r^2_{\ x_h x_k}][1 - r^2_{\ x_i x_k}]}} \quad (7)$$

r_{ih} is the coefficient of the correlation between the variables i and j; a_{ih} is the coefficient of partial correlation between variables i and j; i, h and k start from 1, 2, 3, ..., p; $r_{x_i x_h}$ shows the Correlation between the bound variable X_i and the free variable X_h; $r_{x_i x_k}$ is the Correlation between the bound variable X_i and the free variable X_h; $r_{x_h x_k}$ shows the Correlation between the bound variable X_i and the free variable X_h. The decision that is expected to proceed to factor analysis is to fail to reject H_0 with a value of KMO ≥ 0.5, so that it can be concluded that the amount of data is sufficient for factor analysis to be carried out. An assumption examination of the distribution of multivariate

normality is performed to test whether the data meets the assumption of a multivariate normal distribution [15]. The multivariate normal distribution assumption check serves to ensure the data used is normally multivariate distributed, from more than two variables. Langkah-step examination assumption of the multivariate normal distribution is [12] sort value d_j^2 from the smallest to the largest so that $d_{(1)}^2 \le d_{(2)}^2 \le ... \le d_{(n)}^2$, create a plot $d_j^2, \chi^2_{p;\left(\frac{j-0,5}{n}\right)}$ with p a free degree (i.e. the multiplicity of bound p variables). Checking The assumption of a multivariate normal distribution is also indicated by a QQ plot or χ^2. The data is said to be a normal distribution of multivariate if the plot χ^2. forming a straight line means that the data can be approached with a normal spread. The multivariate normal distribution test examination is

$$d_j^2 = \left[\mathbf{X_j}\text{-}\bar{\mathbf{X}}\right]' \mathbf{S}^{-1} \left[\mathbf{X_j}\text{-}\bar{\mathbf{X}}\right] \quad (8)$$

$$S_j = \frac{1}{n-1}\sum_{j=1}^{n}\left(\boldsymbol{X}_j\text{-}\bar{\boldsymbol{X}}\right)^2 \quad (9)$$

$$t_{proporsi} = \frac{persentase(d_j^2 < \chi_p^2)}{n} x100\% \quad (10)$$

j is the 1st to the 1st observation – n; is $f(x)$ the probability density function; d_j^2: the squared value to – j; n: many observations; p: many variables; X_j: the sample observation vector to - j measures p x 1; $\bar{\boldsymbol{X}}$: the average vector of the sample is p x 1; $\mathbf{S}^{-1}$: the inverse of the variant-covariant matrix of the sample measuring p x p; The expected decision to proceed to factor analysis is to fail to reject H_0 by d_j^2 with $\chi^2_{p\left(\frac{n-j+0,5}{n}\right)}$ being around 50%, so that it can be concluded that the amount of data is normally multivariately distributed.

The Bartlett test is used to test whether a variable is independent or dependent. Variables are said to be mutually free or independent if the correlation matrix between variables forms an identity matrix [15]. H_0: **ρ=I** (There is no correlation between variables or between independent variables) and H_1: $\boldsymbol{\rho} \neq \mathbf{I}$ (There is a correlation between variables or between dependent variables). Test statistics:

$$\chi^2 = -\left[(N-1) - \frac{(2p+5)}{6}\ln\left|\tilde{n}\right|\right] \quad (11)$$

N: Number of observations; p: Number of variables; $\left|\tilde{n}\right|$: determinants of the matrixñs correlation between variables. Rejection Area: Reject H_0 if $\chi^2 \ge \chi^2_{\frac{1}{2}p(p-1);\alpha}$ or P-value < α. The decision that is expected to proceed to factor analysis is to reject H_0 with the value $\chi^2 \ge \chi^2_{\frac{1}{2}p(p-1);\alpha}$ or P-value < α so that it can be concluded Between correlated variables or are dependent. After the factor analysis is carried out, it is continued with cluster analysis with K-Means

The K-Means method is one of the methods for non-hierarchical clusters, it can be used to group objects using centroids or the closest average between objects. The steps are to determine the number of groups to be formed (k), then bring each object into the k cluster, after that calculating the centroid value using the formula, $v_i = \frac{\sum_{i=1}^{n} X_i}{n}$ where v_i : the central value of the group; X_i: the object of observation too -i; n : the number of objects that are members of the cluster. Followed by calculating the distance of each object to each centroid of each cluster. Then group by nearest centroid using Euclidean distance using the formula, where $d = \sum_{i=1}^{n}\left(x_i - v_i\right)^2$ x_i: the object of observation too-i; v_i : centroid to -i; n : many objects that are members of the cluster. Mrecalculates the centroid for new or outgoing members and skips the 3rd and 4th steps until there are no change in-group members [12]. After analysis with K-Means continued discriminant analysis

The variables used in the discriminant analysis are dependent variables with category data (nominal and ordinal) and independent

ISSN No. 1978-3787
Open Journal Systems

variables with interval or ratio data. Discriminant analysis is used to classify observations into mutually free and thorough groups based on several explanatory variables [16]. The basic model on discriminant analysis can be presented in (12).

$$D = b_1 + b_1X_1 + b_2X_2 + ... + b_kX_k \quad (12)$$

The form of the general function in discriminant analysis is presented in the equation (13).

$$\hat{Y}_i = \beta_0 + \beta_1X_1 + \beta_2X_2 + ... + \beta_pX_p \quad (13)$$

Where D: Discriminant score; b: discriminant coefficient; X: predictor (independent variable); Y: dummy variables indicating groups and X_j: Free variables. APER or apparent error rate is one of the measures used in the classification and can be interpreted as a fraction of observations in experimental samples grouped by classification functions [12]. The precision of the classification is to be presented in Table 2.

Table 2 Accuracy of Classification

		Predictions	
Current		π_1	π_2
	π_1	n_{1c}	$n_{1M} = n_1 - n_{1c}$
	π_2	$n_{2M} = n_2 - n_{2c}$	n_{2c}

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \quad (14)$$

Where π_1 is group 1; is π_2 group 2; is n_{1c} the number of groups on π_1 which it is correctly classified as a group π_1; n_{1M} is the number of groups on π_1 the misclassified groups π_1; is n_{2c} the number of groups that are classified as true π_2; is the number of n_{2M} groups on π_2 the misclassified as a group π_2; n_1 is the observation of π_1; n_2 is an observation of π_2. Before conducting a discriminant analysis, the data must be carried out an anguish-Covariance Homogeneity.

A homogeneity test is a test to find out whether the variance and covariance of the studied variable are the same or different. Homogeneity testing can use Box's M test statistics [17]. Discriminant analysis requires the condition of a homogeneous variance-covariance matrix so that the Box's M test is performed [12]. The hypothesis is H_0: $\sum_1 = \sum_2 = ... = \sum_g = \sum$ (Homogeneous covariance variance matrix) and H_1: There is at least one $\sum_i \neq \sum_j$ for $i \neq j$ (variance matrix and inhomogeneous covariance). Test statistics are

$$C = (1-\mu)M \quad (15)$$

$$= (1-\mu)\left\{\left[\sum_\ell (n_\ell - 1)\right]\ln\left|\mathbf{S}_{pooled}\right| - \left[\sum_\ell (n_\ell - 1)\ln\left|\mathbf{S}_\ell\right|\right]\right\} \quad (16)$$

dengan,

$$\mathbf{S}_{pooled} = \frac{1}{\sum_l (n_l - 1)}\left\{(n_1 - 1)\mathbf{S}_1 + ... + (n_p - 1)\mathbf{S}_g\right\} \quad (17)$$

$$\mu = \left[\sum_l \frac{1}{(n_l - 1)} - \frac{1}{\sum_l (n_l - 1)}\right]\left[\frac{2p^2 + 3p - 1}{6(p+1)(g-1)}\right] \quad (18)$$

Where p: number of variables; g: number of groups / groups; n_l: sample size of each group; : $\mathbf{S}$ sample of n_l the covariance variance matrix; $\mathbf{S}_{pooled}$: a combination of the sample of the covariance variance matrix; . $\ell = 1, 2, ..., g$ The decision that is expected to proceed to discriminant analysis is to reject H_0 with $C \geq \chi_{\alpha;\frac{(p(p+1)(g-1))}{2}}$ or P-value < $\alpha = 0.05$ so that it can be concluded that matriks variance of covariance is not homogeneous or there are differences between groupsk. It can also be interpreted that the variance condition of the error (or Y) is not identical [18].

RESULTS AND DISCUSSION

This chapter will discuss the results of the analysis using factor analysis and then *K-Means* followed by discriminant analysis. The characteristics of poverty indicators in East Java in 2020 used average values, standard deviations, minimums, and maximums.

Table 3 shows the characteristics of the variables of poverty indicators in East Java in 2020. The average use of contraceptives (KB) in women aged 15-49 years (X_1) is quite good, but there are still areas with fairly low contraceptive users, namely in Sidoarjo Regency. The average delivery of women aged 15-49 years assisted by medical personnel (X_2) is good, but there are still areas where childbirth is helped by medical personnel needs to be improved, namely in Sampang Regency. The average household that has access to proper sanitation (X_3) is above Indonesia's decent sanitation achievements, but there are still areas with access to proper sanitation that are very lacking, namely in Bondowoso Regency. The average child aged 0-23 months according to breastfeeding (X_4) of 96.016% almost touched the 100% figure, so it needs to be improved again, especially in East Java in Jember Regency. The average of the last completed education in East Java is still quite low at the elementary (X_5), junior high school (X_6), and high school (X_7) levels in the population aged 10 years and over, some areas need attention, namely in Madya Mojokerto City, Bondowoso Regency, and Sampang Regency. Judging the average literacy rate aged 10 years and over (X_8) shows that there are still residents aged 10 years and over in East Java who cannot read and write, which especially since it needs to be considered in Sidoarjo regency. The average school participation rate for the age of 13-15 years (X_9) in East Java in 2020 is quite good, and the average length of schooling of residents aged 15 years and over in East Java (X_{10}) has completed compulsory education up to elementary school/equivalent. The labor force participation rate (TPAK) (X_{11}) is still quite low and the open unemployment rate (TPT) (X_{12}) is still quite high, especially in Sidoarjo Regency, meaning that there are still many labor forces in Sidoarjo Regency that is not absorbed. The average poor population (X_{13}) in East Java in 2020 is still quite high, and the highest is in Sampang Regency. The average human development index (X_{14}) in East Java in 2020 is already high, but there are still areas that need to be improved, namely in Sampang Regency. The average household ownership status of their residence (X_{15}) is good, but in the Middle City of Surabaya it is still quite low, so it needs to be improved. The average household using PLN (X_{16}) is already quite good, but there are still areas that are quite low in PLN use, namely in Sumenep Regency. Judging from the 16 variables of poverty indicators in East Java in 2020, there have been efforts to minimize poverty, but this is still not optimal in some indicators, so it is necessary to group intermediate districts/cities to make it easier for the government to reduce poverty.

Table 3 Characteristics of Poverty Indicator Variables

Variable	Mean	StDev	Minimum	Maximum
X_1	60,64 %	6,95 %	47,95 %	73,42 %
X_2	98114%	4025%	81,92 %	100%
X_3	82,07 %	13,1 %	44,07 %	98,71 %
X_4	96016%	2923%	88,64 %	100%
X_5	26658%	5395%	17,31 %	35,87 %
X_6	19259%	2802%	13,01 %	24,29 %
X_7	24115%	5968%	14,44 %	36,35 %
X_8	6799%	4433%	0,88 %	16,86 %
X_9	97516%	1888%	92,37 %	99,53 %
X_{10}	8,360 Year	1,178 Year	6,19 Year	10,65 Year
X_{11}	70620%	3,16 %	65,75 %	80,36 %
X_{12}	5624%	2003%	2,28 %	10,97 %
X_{13}	11021%	4568%	3,89 %	22,78 %
X_{14}	71,871	5,046	62,7	82,23
X_{15}	87,28 %	9,98 %	54,4 %	96,14 %
X_{16}	99560%	1901%	88,31 %	100%

Before grouping, *a multivariate* normal distribution check is carried out, and a correlation test (*bartlett* test) and a data adequacy test (KMO) are carried out.

ISSN No. 1978-3787
Open Journal Systems

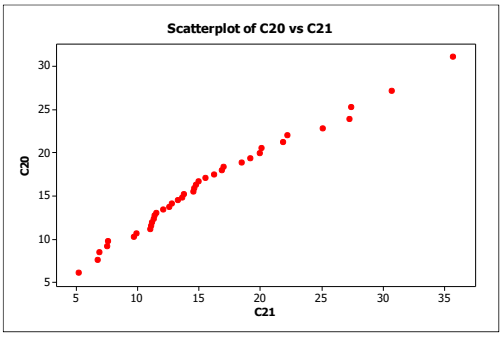

Figure 1 Plot Distribusi Normal Multivariate

Figure 1 shows the results of the *multivariate* normal distribution examination, while visually QQ plot tends to follow a straight line with value *t* the resulting proportion of 0.605 or 60.5% and is still around 50% so that it can be concluded that the poverty indicator data in East Java in 2020 is normally distributed in a *multivariate* manner so that it can be continued to factor analysis. The results of the correlation test on the 2020 poverty indicator data are presented in Table 4.

Table 4 Test Bartlett

Test Bartlett			
χ^2	df	*P-value*	χ^2_{table}
601,604	120	0,00	146,567

Table 4 shows a χ^2 value of 601.604 greater than χ^2_{table} 146.567 or a P-value of 0.00 smaller than $\alpha = 0,05$ so that it can be decided Reject H_0, meaning that there is a correlation between variables (between dependent variables) so that it can proceed to factor analysis.

The results of the KMO test show a KMO value of 0.756 where more than 0.5, it can be decided that it failed to reject H_0 or it can be concluded that the poverty indicator data in East Java in 2020 is sufficient to be factored in.

The method used in factor analysis is the principal component method and because the units between variables are different, it uses a correlation matrix and uses varimax rotation. The results of the factor analysis are in Table 5.

Table 5 Eigenvalue

Component	Initial Eigenvalue		
	Eigenvalue	Variance %	Compulsive Variance %
1	7.831	48.944	48.944
2	2.225	13.909	62.853
3	1.3	8.124	70.977
4	1.171	7.317	78.294

Table 5. shows of the 16 variables in the poverty indicators in East Java in 2020, there were four factors formed, this was seen from the eigenvalue of > 1. The four factors have a variance compulsive value of 78.294 percent, meaning that the four factors can explain the original variable of 78.294 percent. The fact that factors are formed, can also be visualized using a scree plot.

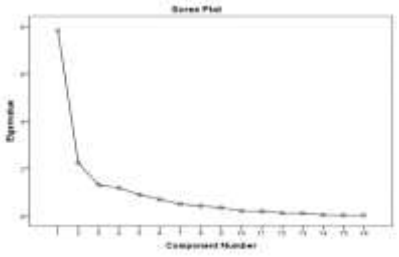

Figure 2 Scree Plot

Figure 2 shows four new factors formed, judging by the considerable slope at factors one, two, three, and four. The scree plot results are also supported by the eigenvalue value in table 5 where there are four eigenvalue values > 1. If you look at which variables are included in the factor it will be presented in Table 6.

Table 6 is the result of the correlation values between variables and factors. The correlation value marked positive has the same meaning between the variable and the factor, but if the value is negative, it means that the relationship between the two is inversely proportional. Factor one is formed by 7 variables, namely the variable Population (10 years and above) of the last education in

elementary and high school, the average length of schooling (15 years and above), the labor force participation rate (TPAK), the open unemployment rate (TPT), the human development index, and households having their residence. Factor two is formed from 6 variables, namely the variables of childbirth assisted by medical personnel, households having access to proper sanitation, population (10 years and above) the last education in junior high school, literacy rate (10 years and above), school participation rate (13-15 years) and the number of poor people. Factor three is formed from 2 variables, namely the Baduta variable (0-23 months) which is breastfed and households using PLN. Meanwhile, factor four is formed from one variable, namely the variable Women (15-49 years) are using contraceptives (KB).

Table 6. Factors Forming Factors for East Java Poverty Indicators in 2020

Variable	Component			
	1	2	3	4
Women (15-49 years old) are on contraceptives (kb)	-0,245	0,004	0,087	**0,872**
Childbirth helped by medical personnel	0,211	**0,797**	0,184	0,392
Households have access to proper sanitation	0,435	**0,605**	0,213	-0,351
Breastfed baduta (0-23 months)	0,019	-0,026	**-0,654**	0,061
Residents (10 years and older) last primary education	**-0,796**	-0,229	0,279	0,108
Residents (10 years and older) last junior high school education	-0,19	**0,731**	0,4	-0,333
Residents (10 years and older) last high school education	**0,849**	0,453	-0,01	-0,105
Literacy rate (10 years and above)	-0,569	**-0,602**	-0,36	0,266
School participation rate (13-15 years)	0,12	**0,785**	-0,146	0,032
The Average length of schooling (15 years and above)	**0,778**	0,538	0,114	-0,228
Labor force participation rate (TPAK)	**-0,709**	0,137	-0,261	-0,03
Open unemployment rate (TPT)	**0,863**	0,018	0,211	-0,108
Percentage of the poor population	-0,58	**-0,637**	-0,184	-0,113
Human development index	**0,777**	0,566	0,021	-0,189
Households have their place of residence	**-0,894**	-0,143	0,076	0,05
Households using PLN	0,176	0,197	**0,732**	0,223

Factor one is named the educational welfare factor, factor two is named economic

ISSN No. 1978-3787
Open Journal Systems

welfare, factor three is named PLN user and baduta breastfeeding, and factor four is named contraceptive (KB) user. From the four factors, it can be known that Sumenep Regency has all negative factor scores, this will affect the results of grouping. The grouping of Madya Regencies/Cities based on poverty indicators in East Java in 2020 uses a factor score which is the result of factor analysis.

The data used in the *K-Means* analysis are data from factor analysis, namely educational welfare factors, economic welfare factors, pln user factors and Baduta breastfeeding, and contraceptive user factors (KB). Grouping using *K-Means* uses a k value of 3 because later it will be classified into low, medium, and high poverty. The results of cluster analysis using *K-Means* will be presented in table 7.

Table 7 Cluster Iterations of Poverty Indicators

Iteration	Change in *Cluster* Centers		
	1	2	3
1	1,990	0,000	1,799
2	0,206	0,000	0,276
3	0,077	0,000	0,079
4	0,000	0,000	0,000

Table 7 shows the iteration process that occurs in K-Means, where there are 4 iterations to obtain the results of grouping. The minimum distance between cluster centers from the results of the iteration was 5,200.

Table 8 Average Values and Variances of Each Group

Average			
Factor	Cluster		
	1	2	3
f1	-0.5312	-1.022	0.6851
f2	0.5397	-0.958	-0.5786
f3	0.1752	-4.2908	0.0462
f4	-0.0773	-1.8231	0.1982
Average	0.0266	-2.02348	0.087725
Variance			
Factor	Cluster		
	1	2	3
f1	0.317246	0	1.019032
f2	0.136191	0	1.373598
f3	0.219064	0	0.861008
f4	0.570256	0	1.378356
Variance	0.035421	0	0.067519

Table 8 shows the averages and variances of each factor in each cluster used to classify the poverty rate in each cluster. Cluster one has an average factor score of 0.0266 meaning that cluster one is included in the madya districts/cities which has a moderate poverty rate. Cluster two has an average factor score of -2.0235 meaning that group two is included in the madya districts/cities which has a high poverty rate. Cluster three has an average factor score of 0.0877 meaning that cluster three is included in the madya districts/cities which has a low poverty rate. However, if we look at the variance value, the variance of clusters one and three is not much different, so cluster one remains in the category of moderate poverty, and cluster three remains in the category of low poverty. Cluster two has no variance because the members of cluster two consist of only one district.

The average value of factors in cluster one shows negative values on factors one and four, meaning that the relationship is the opposite, namely the higher the value of the education welfare factor and the factor of contraceptive users of birth control, the lower the level of poverty in the area, or the higher the poverty rate, but in factors two and three the relationship is comparable, meaning that the higher the economic welfare and users of PLN and breastfeeding to Baduta, then the higher the level of poverty or the lower the poverty in the area. The average value of the factors in cluster two shows negative signs, all from factor 1 to factor 4, meaning that the relationship is the opposite, that is, the higher the value, the lower the poverty level in the area, or the higher the poverty level. The average value in cluster three shows positive values in factors one, three, and four, the

relationship is comparable, meaning that the higher the welfare of education, the factor of PLN users and breastfeeding in Baduta and the factor of contraceptive users (KB), the higher the level of poverty or the lower the poverty in the area, but in the second factor shows a negative value, meaning that the relationship is inversely proportional, or the higher the value of the economic welfare factor, the lower the poverty level in the area, or the higher the poverty level.

Table 9 Grouping of Poverty Indicators

Group	Member
1 (moderate poverty)	Regency: Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Banyuwangi, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik,
	Madya City: Batu
2 (high poverty)	Sumenep Regency
3 (Low Poverty)	Regency: Jember, Bondowoso, Situbondo, Probolinggo, Pasuruan, Sidoarjo, Bangkalan, Sampang, Pamekasan
	Madya City: Kediri, Blitar, Malang, Probolinggo, Pasuruan, Mojokerto, Madiun, Surabaya

Table 9 shows the results of grouping districts / Madya Cities based on poverty indicators in East Java in 2020. Group 1 is a group that has a moderate poverty rate occupied by 19 districts and 1 intermediate city. Group 2 is a group that has high poverty occupied by 1 district, namely the sumenep district. Group 3 is a group that has a low poverty rate occupied by 9 districts and 8 Madya Cities.

Figure 7. Poverty Mapping in East Java in 2020

Table 10 Region Description

No	Regency/City	No	Regency/City
1	Pacitan Regency	20	Malang City
2	Ponorogo Regency	21	Probolinggo City
3	Trenggalek Regency	22	Pasuruan City
4	Tulungagung Regency	23	Mojokerto City
5	Lumajang Regency	24	Madiun City
6	Bondowoso Regency	25	Surabaya City
7	Pasuruan Regency	26	Batu City
8	Jombang Regency	27	Blitar Regency
9	Nganjuk Regency	28	Kediri Regency
10	Madiun Regency	29	Mojokerto Regency
11	Magetan Regency	30	Banyuwangi Regency
12	Ngawi Regency	31	Gresik Regency
13	Bojonegoro Regency	32	Jember Regency
14	Tuban Regency	33	Malang Regency
15	Lamongan Regency	34	Probolinggo Regency
16	Bangkalan Regency	35	Sampang Regency

ISSN No. 1978-3787
Open Journal Systems

No	Regency/City	No	Regency/City
17	Pamekasan Regency	36	Sidoarjo Regency
18	Kediri City	37	Situbondo Regency
19	Blitar City	38	Sumenep Regency

Figure 7 shows the results of grouping based on poverty indicators. Red color indicates high poverty in Sumenep Regency, yellow color indicates areas with moderate poverty and green color indicates areas with low poverty rates. The next step is to perform a discriminant analysis to find out the differences between clusters. The data used in the discriminant analysis for independent variables are data from factor analysis, namely Educational welfare factors, economic welfare factors, pln user factors and Baduta breastfeeding, and contraceptive user factors (KB), while dependent variables are the result of the K-Means cluster formed by 3 groups. Before conducting a discriminant analysis, then first check the homogeneity using the Box's M test.

Box's M = 53,105; P value = 0,00

Based on the results of the Box's M test, a P value of 0.00 or less than $\alpha = 0,05$ can be obtained so that it can be decided to reject H_0, meaning that the matrix of variants-covariants between groups on the poverty indicator variable is not homogeneous, or there is at least one different group. Discriminant analysis in this study used the stepwise method. The results of the discriminant analysis will be presented in Table 11.

Table 11 Model Determination by Stepwise Method

Step	Variable	F	df 1	df 2	$F_{(0,05;df1;df2)}$	P value
1	PLN Users and Baduta Breastfeeding (F_3)	18,596	2	35	3,267	0,00
2	Educational Welfare (F_1)	14,801	4	68	2,506	0,00
3	Economic Well-Being (F2)	18,709	6	66	2,239	0,00
4	Contraceptive Users (KB) (F_4)	17,023	8	64	2,086	0,00

Table 11 shows the variables included in the model, where the variables of PLN User and Baduta breastfeeding have an F value of 18,596 where more than $F_{(0.05;2;35)}$ of 3.267 and P value of 0 is less than $\alpha = 0,05$ that of the means that the PLN user variable and Baduta Breastfeeding are included in the discriminant model. The Education welfare variable has an F value of 14.801 where more than $F_{(0.05;4;68)}$ of 2.506 and a P value of 0 is less than $\alpha = 0,05$ the meaning that the Education welfare variable is included in the discriminant model.

The economic welfare variable has an F value of 18.709 where more than $F_{(0.05;6;66)}$ of 2.239 and a P value of 0 is less than $\alpha = 0,05$ that means the economic welfare variable is included in the discriminant model. The contraceptive user variable (KB) has an F value of 17,023 were more than $F_{(0.05;8;64)}$ is 2.086 and a P value of 0 is less than $\alpha = 0,05$ that of the contraceptive user variable (KB) entered into the discriminant model. It can be concluded that the whole variable is included in the

discriminant model. The discriminant function using significant variables entered into the model as described in table 12 states that the four factors are distinguishing variables so that the discriminant function will be obtained which will be presented in table 12.

Table 12 Discriminant Functions

	Function	
	1	2
Educational Welfare (F_1)	1.338	0.073
Economic Well-Being (F_2)	-1.086	0.564
PLN Users and Baduta Breastfeeding (F_3)	0.194	1.457
Contraceptive Users (KB) (F_4)	0.425	0.547
(Constant)	0	0

Table 12 shows that there are two functions formed, of which the one function is the function used to compare groups one and two, while the second function is a function used to compare groups two and three. A discriminant function with four distinguishing variables will be presented below.

Function 1=1,338 (F_1) -1,086 (F_2) + 0,194 (F_3) + 0,425 (F_4)

Function 2= 0,073 (F_1) + 0,564 (F_2) + 1,457 (F_3) + 0,547 (F_4)

The coefficient value in the first function states that every increase in educational welfare will increase the discriminant score by 1.338, every increase in economic welfare will reduce the discriminant score by 1.086, while every increase in PLN users and Baduta breastfeeding and an increase in contraceptive users (KB) will increase the discriminant score by 0.194 and 0.425. The second function states that every increase in education welfare will increase the discriminant score by 0.073, every increase in economic welfare will increase the discriminant score by 0.564, while every increase in PLN users and Baduta breastfeeding and the increase in contraceptive users (KB) will increase the discriminant score by 1.457 and 0.547. The discriminant linear functions of each group will be presented in table 13.

Table 13 Discriminant Linear Functions

	Group		
	1	2	3
Educational Welfare (F_1)	-1,699	-3,165	2,185
Economic Well-Being (F_2)	1,677	-2,341	-1,835
PLN Users and Baduta Breastfeeding (F_3)	0,447	-11,836	0,171
Contraceptive Users (KB) (F_4)	-0,289	-5,121	0,641
(Constant)	-2,053	-33,898	-2,445

Table 13 shows the linear functions of the discriminants in each group. The linear functions of the discriminants in each group will be presented below.

Group 1= -2,053 -1,699 (F_1) +1,677 (F_2) + 0,447 (F_3) – 0,289 (F_4)

Group 2= -33,898 – 3,165 (F_1) – 2,341 (F_2) – 11,836 (F_3) – 5,121 (F_4)

Group 3= -2,445 + 2,185 (F_1) – 1,835 (F_2) + 0,171 (F_3) + 0,641 (F_4)

Based on the results of the equation of the linear fisher function discriminant in the three groups showed that the value of educational welfare decreased the largest in group 2 by -3.165, this shows that group 2 has a poor characteristic in overcoming poverty, this is also shown by the value of economic welfare, PLN users and baduta breastfeeding and contraceptive (KB) users also decreased, by -2,341, -11,836 and – 5,121. Group 3 has good characteristics in poverty reduction, this is shown by the increasing welfare value of education and contraceptive users of birth control by 2,185 and 0,641. Group 1, it has fairly good characteristics in overcoming poverty, this is shown by the increasing value

ISSN No. 1978-3787
Open Journal Systems

of economic welfare and PLN users and breastfeeding, by 1,677 and 0,447. After this is carried out the calculation of the accuracy of the classification of discriminant models is to be presented in Table 14.

Table 14 Accuracy of Model Classification

Group	Predicted Group Membership			Total
	1	2	3	
1	20	0	0	20
2	0	1	0	1
3	0	0	17	17

Table 14 shows that in group 1 there were 20 observations that were correctly classified and there were no observations that were not precisely classified. Group 2 there is 1 observation that is appropriately classified and there is no observation that is not properly classified. Group 3 had 17 observations that were appropriately classified and no observations that were improperly classified. The calculation of the level of accuracy on the accuracy of the classification will be shown below.

$$APER = \frac{0+0+0+0+0+0}{20+1+17} x100\% = 0\%$$

Accuracy of classification

= 100%-*APER*=100%-0%=100%

The calculation above shows the number of observations that are incorrect in the classification or *Apparent Error Rate (APER)* of 0%, and the accuracy rate of the model classification accuracy is 100%.

CONCLUSIONS AND SUGGESTIONS

Conclusion

The conclusion from the results of the analysis of mapping poverty indicators in East Java in 2020 is

Based on 16 variables of poverty indicators in East Java in 2020, there have been efforts to minimize poverty, but this is still not optimal on several variables in the 2020 East Java poverty indicators.

Mapping of Madya Regencies/Cities in East Java in 2020 formed three groups: Group one is a group with a moderate poverty rate consisting of 19 regencies and 1 madya city, group two is a group with a high poverty rate consisting of 1 regency, and group three is a group with a low poverty rate consisting of 9 regencies and 8 madya cities.

The variables included in the discriminant model are education welfare factors, economic welfare factors, pln user factors and Baduta breastfeeding, and contraceptive user factors (KB). The resulting classification accuracy value is high, which is 100%.

Suggestions

A suggestion for further research is to add other clustering methods so that they can be compared to poor regions clusters, so it can be a comparison. Additional addition of more research variable indicators so that results are more complex.

REFERENCES

[1] Farhan, W., Hartono, I. W., & Meganingrum, Y. (2020). *Application of Big Push Theory in Local Economic Development to Overcome Jember Poverty*. MATRAPOLIS: Journal of Urban and Regional Planning, 1(1), 15. https://doi.org/10.19184/matrapolis.v1i1.19219

[2] Muryani, S. (2021). *Understanding The Goals And Targets Of The Sdgs*. LEGUNDI.NGAWIKAB.ID. Retrieved December 19, 2021, from https://legundi.ngawikab.id/2021/03/pengertian-tujuan-dan-target-sdgs/

[3] Perdana, D. (2021). *Poor Population in East Java Increases to 11.46 Percent*. Suarasurabaya.Net. Retrieved December 19, 2021, from https://www.suarasurabaya.net/ekonomibisnis/2021/penduduk-miskin-di-jatim-meningkat-jadi-1146-persen/

[4] Firmansyah, F. (2021). *Unemployment Rate in East Java 2020 Rises, BPS: Dominated by Urban, Male and High School Graduates*. TribunJatim.Com.

Retrieved December 19, 2021, from https://jatim.tribunnews.com/2021/02/22/angka-pengangguran-di-jawa-timur-2020-meninggi-bps-didominasi-perkotaan-laki-laki-dan-lulusan-sma

[5] Pratiwi, R. D. A. (2014). *Grouping of District/City Poverty Indicators in East Java Province in 2002 and 2012 Using the Cluster Analysis Method*. Institut Teknologi Sepuluh Nopember (ITS)

[6] Komariyah, N., & Akbar, M. S. (2011). *Grouping of regencies/cities in East Java province based on poverty indicators using the cluster analysis method*.

[7] Abidin, Z. (2017). *Grouping of Regencies/Cities in East Java Based on Poverty Indicators using Hierarchical Cluster Analysis.* Institut Teknologi Sepuluh Nopember (ITS)

[8] Badan Pusat Statistik. (2021). *East Java Province In Figures 2021*. Publication. Retrieved December 17, 2021, from https://jatim.bps.go.id/publication/2021/02/26/78c43a895e7f8ea378ffafc4/provinsi-jawa-timur-dalam-angka-2021.html

[9] Badan Pusat Statistik. (2020b). *East Java Provincial Health Statistics 2020*. Publication. Retrieved December 17, 2021, from https://jatim.bps.go.id/publication/2021/08/05/a70cbc1ca224552d5e0f5000/statistik-kesehatan-provinsi-jawa-timur-2020.html

[10] Badan Pusat Statistik. (2020c). *Education Statistics of East Java Province 2020*. Publication. Retrieved December 17, 2021, from https://jatim.bps.go.id/publication/2021/12/06/143ff261cca315e5cbfb82b5/statistik-pendidikan-provinsi-jawa-timur-2020.html

[11] Badan Pusat Statistik. (2020a). *Human Development Index*. Retrieved December 17, 2021, from https://jatim.bps.go.id/indicator/26/36/1/ipm.html

[12] Johnson, R.A., & Wichern, D. . (2007). *Applied Multivariate Statistical Analysis* (6th edition). Prentice-Hall Inc.

[13] Djaali dan Muljono, P. (2007). *Measurement in the Field of Education*. Grasindo

[14] Telussa, A. M., Persulessy, E. R., & Leleury, Z. A. (2013). Application of Partial Correlation Analysis to Determine the Relationship between the Implementation of Personnel Management Functions and Employee Work Effectiveness. *BREAKING: Journal of Mathematical and Applied Sciences*, *7*(1), 15–18. https://doi.org/10.30598/barekengvol7iss1pp15-18

[15] Morisson, D. F. (2005). *Multivariate Statistical Methods*. Mc_Graw Hill.

[16] Mattjik, A. A., & Sumertajaya, I. M. (2011). *Double Changer Fingerprints*. IPB PRESS.

[17] Rencer, A. (2002). *Methods of Multivariate Analysis*. John Wiley & Sons, Inc.

[18] Setiawan, & Kusrini, D. E. (2010). *Econometrics*.ANDI.

ISSN No. 1978-3787
Open Journal Systems

HALAMAN INI SENGAJA DIKOSONGKAN