BAREKENG: Journal of Mathematics and Its Applications
June 2025 Volume 19 Issue 2 Page 1009–1022
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol19iss2pp1009-1022

COMPARISON BETWEEN BICLUSTERING AND CLUSTER-BIPLOT RESULTS OF REGENCIES/CITIES IN JAVA BASED ON PEOPLE'S WELFARE INDICATORS

Yekti Widyaningsih[1*], Alfia Choirun Nisa[2]

[1,2]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Indonesia*
Jl. Prof. Dr. Mahar Mardjono Kampus UI, Depok, 16424, Jawa Barat, Indonesia

*Corresponding author's e-mail: *yekti@sci.ui.ac.id*

Article History:

Received: 29th August 2024
Revised: 3rd February 2025
Accepted: 1st March 2025
Published: 1st April 2025

Keywords:

Clustering;
Plaid Model Biclustering;
Plaid Model;
Silhouette Method;
Ward Method.

ABSTRACT

The success of a country's development can be known from the well-being of its people. Improving the welfare of the population is the main goal of the development activities carried out by the government. To ensure that development is effective and targeted, grouping is needed to understand the characteristics of the region. This study discusses the grouping of regencies/cities in Java Island based on the people's welfare indicators in 2022. The measured welfare is material well-being. Variables used in this study are the percentage of the poor population, GDP per capita at current prices, average length of schooling, expected length of schooling, percentage of per capita expenditure on food, open unemployment rate, population, population density, and life expectancy. There are two approaches used in grouping regencies/cities along with their variables. The first approach is to simultaneously group regencies/cities and their variables using Plaid Model biclustering. The second approach is to group regencies/cities using the Ward clustering method followed by the biplot method. This study aims to compare the results of these two approaches, namely the biclustering and cluster-biplot results, on data from 119 regencies/cities in Java Island in 2022 based on people's welfare indicators. Based on the results of this study, the number of groups from each approach is 2, with group 1 being more prosperous than group 2. Judging from the standard deviation values, the Plaid Model biclustering result groups have lower standard deviation values than the cluster-biplot result groups. Therefore, in general the first approach produces better groups as they are more homogeneous than the second approach.

How to cite this article:

Y. Widyaningsih and A. C. Nisa., "COMPARISON BETWEEN BICLUSTERING AND CLUSTER-BIPLOT RESULTS OF REGENCIES/CITIES IN JAVA BASED ON PEOPLE'S WELFARE INDICATORS," *BAREKENG: J. Math. & App.*, vol. 19, iss. 2, pp. 1009-1022, June, 2025.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

Indonesia, the fourth most populous country in the world, has a highly uneven population distribution, with 56.10% of its total population concentrated on Java Island, which comprises only 7% of the country's total area [1]. Generally, this unevenness occurs because Java Island has almost all the facilities needed by the population, making it attractive for people to try their luck there [2]. Population density in one area can lead to various social problems that will affect the welfare of its inhabitants.

One of the elements of a country's development success can be seen in the condition of the welfare of its people. Welfare is defined as the ability of families to meet all the needs to live decently, healthy, and productive [3]. The welfare of the people is highly regarded by every country because it affects the economy and governmental stability. Therefore, improving the welfare of the people becomes the main target in the development activities carried out by the government. To ensure effective and targeted development, steps that can be taken include grouping regions based on certain indicators. Region grouping can help the government understand the characteristics of each group. Therefore, grouping regencies/cities in Java Island based on indicators of people's welfare can be one source of information to support the success of development equalization efforts carried out by the local government.

Based on the issues previously outlined, regencies/cities in Java Island are grouped along with variables related to indicators of people's welfare. This grouping aims to divide regencies/cities in Java Island into several groups based on their level of welfare to determine which regencies/cities need to prioritize improving their people's welfare. The results of this grouping are expected to provide an overview of the welfare conditions of the people in Java Island, enabling the government to make more targeted policies. Thus, the government's goal of development equalization can be achieved, and the issues arising from the dense population in Java Island can be addressed.

A study on the clustering of regions based on indicators of people's welfare has previously been conducted by [4] titled “Analisis Klaster untuk Pengelompokan Kabupaten/Kota di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat.” This study aimed to group 35 districts/cities in Central Java and to understand the characteristics of each group based on the people's welfare indicators from 2010. The variables used included Gross Regional Domestic Product per capita, population density, the number of poor people, the workforce, adjusted real per capita expenditure, life expectancy, and average years of schooling. In this study, hierarchical clustering analysis was performed using the average linkage method with an agglomerative technique and Euclidean distance as the distance measure. The results revealed three groups, each with distinct value trends across the variables.

A similar study [5] titled “Penerapan Fuzzy C-Means Cluster dalam Pengelompokan Provinsi Indonesia Menurut Indikator Kesejahteraan Rakyat.” This research aimed to group provinces in Indonesia using Fuzzy C-Means Clustering. The variables used included population density, the percentage of poor people, population growth rate, life expectancy, school participation rate, labor force participation rate, open unemployment rate, and average expenditure per capita. This study resulted in two groups.

In this study, two approaches are used to group regencies/cities along with their variables related to indicators of people's welfare. The first approach involves simultaneously grouping regencies/cities and their variables using the Plaid Model biclustering. The second approach involves grouping regencies/cities using Ward’s clustering method and mapping the grouping results using the biplot method. The method used in the second approach is called cluster-biplot. Subsequently, the results of both approaches are compared using standard deviation values. The novelty of this study is in comparing the results of the biclustering method and the results of the cluster-biplot method.

2. RESEARCH METHODS

This chapter discusses the data used, standard deviation, Plaid Model biclustering method, and cluster-biplot method.

2.1 Data

The data used in this study are secondary data sourced from publications by the Central Statistics Agency of the Republic of Indonesia [6]. The data consist of 119 observations (regencies/cities) and 9 variables of people's welfare indicators.

Table 1. Variable Description

Variable	Variable Description	Unit
PPM	Percentage of the poor population	Percent
PDRB	GDP per capita at current prices	Million rupiah per year
RLS	The average length of schooling	Year
HLS	Expected length of schooling	Year
PKM	Percentage of per capita expenditure on food consumption	Percent
TPT	Open unemployment rate	Percent
JP	Population	Person
KP	Population density	Person per km^2
AHH	Life expectancy	Year

2.2 Standard Deviation

Standard deviation is a value that indicates the variability of data or how spread out observations are from their mean [7]. This method will be used to compare the results of biclustering with the results of cluster-biplot. The larger the standard deviation value, the more diverse the values or the less accurate towards the mean value. Conversely, the smaller the standard deviation value, the more similar the values or the more accurate towards the mean value. In other words, the smaller standard deviation value indicates that the data is more homogeneous. For sample data with n observations and p variables, the formula for standard deviation is as follows.

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_{ij}-\mu_j\right)^2}{n-1}}, \qquad i = 1,2,\dots,n, \qquad j = 1,2,\dots,p \tag{1}$$

where:

σ_j : standard deviation value of the j-th variable
Y_{ij} : observation value of the i-th on the j-th variable
μ_j : mean value of the j-th variable
n : number of observations

2.3 Plaid Model Biclustering

One of the biclustering analysis methods that is quite flexible is the Plaid Model method [8]. In the Plaid Model, the observation value is denoted as Y_{ij}, where $i = 1,2,\dots,n$ is the index for rows (observations) and $j = 1,2,\dots,p$ is the index for columns (variables). The observation value in the Plaid Model has the following form.

$$Y_{ij} = \Theta_{ij0} + \sum_{k=1}^{K} \Theta_{ijk}\rho_{ik}\kappa_{jk} + \varepsilon_{ij} \tag{2}$$

$$Y_{ij} = \left(\mu_0 + \alpha_{i0} + \beta_{j0}\right) + \sum_{k=1}^{K} \Theta_{ijk}\rho_{ik}\kappa_{jk} + \varepsilon_{ij} \tag{3}$$

where:

Y_{ij} : observation value of the i-th row and j-th column,
Θ_{ij0} : background effect of the i-th row and j-th column,
k : bicluster index, $k = 1,2,\dots,K$,

K : number of bicluster,
Θ_{ijk} : background effect of the i-th row and j-th column in the k-th bicluster,
ρ_{ik} : membership parameter of the i-th row in the k-th bicluster,
κ_{jk} : membership parameter of the j-th column in the k-th bicluster,
μ_0 : mean effect,
α_{i0} : effect of the i-th row,
β_{j0} : effect of the j-th column,
ε_{ij} : error of the i-th observation and j-th variable.

According to [9], the steps in conducting Plaid Model biclustering are as follows.

1. Converting the data into a data matrix ($\boldsymbol{Y}$).
2. Create an initial model and calculate the effect values, namely the grand mean ($\hat{\mu}_0$), row effect ($\hat{\alpha}_{i0}$), and column effect ($\hat{\beta}_{j0}$).

$$\begin{aligned} \hat{\mu}_0 &= \bar{Y}_{..} \\ \hat{\alpha}_{i0} &= \bar{Y}_{i.} - \bar{Y}_{..} \\ \hat{\beta}_{j0} &= \bar{Y}_{.j} - \bar{Y}_{..} \end{aligned} \tag{4}$$

   where:
   $\hat{\mu}_0$: estimated grand mean,
   $\hat{\alpha}_{i0}$: estimated effect of the i-th row,
   $\hat{\beta}_{j0}$: estimated effect of the j-th column,
   $\bar{Y}_{..}$: mean of the data matrix ($\boldsymbol{Y}$),
   $\bar{Y}_{i.}$: mean of the i-th row in the data matrix ($\boldsymbol{Y}$),
   $\bar{Y}_{.j}$: mean of the j-th column in the data matrix ($\boldsymbol{Y}$).

3. Forming the residual matrix ($\boldsymbol{Z}_{n\times p}$). Before any bicluster is found, the elements of the residual matrix are determined using the following equation.

$$Z_{ij} = Y_{ij} - \left(\hat{\mu}_0 + \hat{\alpha}_{i0} + \hat{\beta}_{j0}\right) \tag{5}$$

   where:
   Y_{ij} : observation value of the i-th row in the j-th column
   Z_{ij} : residual value of the i-th row in the j-th column
   $\hat{\mu}_0$: estimated grand mean in the data matrix
   $\hat{\alpha}_{i0}$: estimated effect of the i-th row
   $\hat{\beta}_{j0}$: estimated effect of the j-th column

   After $(l-1)$ biclusters are found, the elements of the residual matrix can be determined using the following equation.

$$Z_{ij} = Y_{ij} - \widehat{\Theta}_{ij0} - \sum_{k=1}^{l-2} \widehat{\Theta}_{ijk}\hat{\rho}_{ik}\hat{\kappa}_{jk} \tag{6}$$

   where:
   Z_{ij} : residual value of the i-th row and the j-th column,
   Y_{ij} : observation value of the i-th row and the j-th column,
   $\widehat{\Theta}_{ij0}$: estimated background effect of the i-th row and the j-th column,
   $\widehat{\Theta}_{ijk}$: estimated background effect of the k-th bicluster at the i-th row and the j-th column,
   $\hat{\rho}_{ik}$: estimated membership parameter of the i-th row in the k-th bicluster,
   $\hat{\kappa}_{jk}$: estimated membership parameter of the j-th column in the k-th bicluster.

4. Determining initial bicluster candidates using the k-means clustering algorithm to group observations and variables separately. Then, select the clusters with fewer members. Subsequently, the observation clusters are paired with variable clusters to form initial biclusters.

5. Estimating bicluster effect parameters ($\hat{\mu}_k$, $\hat{\alpha}_{ik}$, and $\hat{\beta}_{jk}$). Suppose we have obtained the bicluster $\boldsymbol{Z}^*$, which is a submatrix of the residual matrix ($\boldsymbol{Z}$). The estimation of bicluster effect parameters can be found using the following equation.

$$\begin{aligned} \hat{\mu}_k &= \bar{Z}^*_{..k} \\ \hat{\alpha}_{ik} &= \bar{Z}^*_{i.k} - \bar{Z}^*_{..k} \\ \hat{\beta}_{jk} &= \bar{Z}^*_{.jk} - \bar{Z}^*_{..k} \end{aligned} \tag{7}$$

where :
$\hat{\mu}_k$: estimated mean effect of the k-th bicluster,
$\hat{\alpha}_{ik}$: estimated effect of the i-th row in the k-th bicluster,
$\hat{\beta}_{jk}$: estimated effect of the j-th column in the k-th bicluster,
$\bar{Z}^*_{..k}$: mean residual value of k-th bicluster,
$\bar{Z}^*_{i.k}$: mean residual value of the i-th row in the k-th bicluster,
$\bar{Z}^*_{.jk}$: mean residual value of the j-th column in the k-th bicluster.
The estimation of bicluster effect parameters is continuously updated for S iterations.

6. Estimating bicluster membership parameters ($\hat{\rho}_{ik}$ and $\hat{\kappa}_{jk}$) is done by trying possible membership parameter values (0 or 1) in **Equations (8)** and **Equations (9)**. Use the membership parameter value that produces the smallest value in the equation.

$$\sum_{j=1}^{J} \left[\hat{Z}_{ijk} - \hat{\rho}_{ik}\hat{\Theta}_{ijk}\hat{\kappa}_{jk}\right]^2 \tag{8}$$

$$\sum_{i=1}^{I} \left[\hat{Z}_{ijk} - \hat{\kappa}_{jk}\hat{\Theta}_{ijk}\hat{\rho}_{ik}\right]^2 \tag{9}$$

where:
$\hat{Z}_{ijk}$: estimated residual value of the i-th row and j-th column in the k-th bicluster,
$\hat{\Theta}_{ijk}$: estimated background effect of the i-th row and j-th column in the k-th bicluster,
$\hat{\rho}_{ik}$: estimated membership parameter of the k-th bicluster at the i-th row,
$\hat{\kappa}_{jk}$: estimated membership parameter of the k-th bicluster at the j-th column.

This parameter estimation is conducted S times as determined by the researcher.

7. Pruning bicluster based on threshold values (τ_1 and τ_2) determined by the researcher. The threshold values serve as limits when conducting bicluster pruning, where τ_1 as the limit for pruning on observations and τ_2 as the limit for pruning on variables. Here are the bicluster membership parameter values based on the bicluster pruning process.

$$\tilde{\rho}_i = \begin{cases} 1, & if\ \ \hat{\rho}_i = 1\ and\ \sum_{j:\hat{\kappa}_j=1} \left(\hat{Z}_{ij} - \hat{\Theta}_{ij}\right)^2 < (1-\tau_1) \sum_{j:\hat{\kappa}_j=1} \left(\hat{Z}_{ij}\right)^2 \\ 0, & other. \end{cases} \tag{10}$$

$$\tilde{\kappa}_j = \begin{cases} 1, & if\ \hat{\kappa}_j = 1\ and\ \sum_{j:\hat{\rho}_i=1} \left(\hat{Z}_{ij} - \hat{\Theta}_{ij}\right)^2 < (1-\tau_2) \sum_{j:\hat{\rho}_i=1} \left(\hat{Z}_{ij}\right)^2 \\ 0, & other. \end{cases} \tag{11}$$

where:
$\tilde{\rho}_i$: bicluster membership parameters of the i-th row based on the bicluster pruning process,
$\tilde{\kappa}_j$: bicluster membership parameter of the j-th column based on the bicluster pruning process,
τ_1 : threshold value in pruning row membership parameters,
τ_2 : threshold value in pruning column membership parameters,
$\hat{Z}_{ij}$: estimated value of the i-th row and j-th column in the residual matrix,
$\hat{\Theta}_{ij}$: estimated background effect value of the k-th bicluster.

The threshold value is between 0 and 1. A threshold value that is close to 1 will produce bicluster results that are more coherent or have more similar values, and vice versa. However, using a small threshold value can accept biclusters more easily so that it can accept many biclusters. It is suggested to use a threshold value between 0.5 and 0.7.

8. Backfitting or multiple calculations to achieve convergence. According to [10], to obtain better bicluster results, use one to three backfitting calculations as a midpoint between bicluster effect accuracy and computational load.

2.4 Cluster-Biplot

Cluster-biplot consists of two methods: clustering and biplot. Clustering aims to group observations into several clusters so that observations within the same cluster exhibit relatively homogeneous characteristics. After obtaining clusters, the next step involves mapping the cluster members and their variables using the biplot method.

2.4.1 Clustering Assumption Checking

According to [11], two assumptions must be met before conducting Clustering.

1. Sample Represents Population

This assumption is examined using the Kaiser-Meyer-Olkin (KMO) test. The KMO value ranges from 0 to 1. The sample is said to represent the population if the KMO value is ≥ 0.5.

$$KMO = \frac{\sum\sum_{x\neq y} c_{xy}^2}{\sum\sum_{x\neq y} c_{xy}^2 + \sum\sum_{x\neq y} a_{xy}^2} \quad (12)$$

where:

$a_{xy} = -\frac{R_{xy}}{\sqrt{R_{xx}R_{yy}}}$, $R_{xy} = (-1)^{x+y}M_{xy}$,

c_{xy} : sample correlation of the variable x and y,

a_{xy} : partial correlation matrix of the variables x and y,

R_{xy} : cofactor of the c_{xy} element in the correlation matrix $\boldsymbol{C}$,

M_{xy} : determinant of the remaining submatrix after the x-th row and y-th column in the matrix $\boldsymbol{C}$ are removed.

2. No Multicollinearity

This assumption is checked using the Bartlett Test of Sphericity. If so, then each variable is only correlated with itself. According to [12], the steps in conducting the Bartlett Test of Sphericity are as follows.

a. Hypothesis

H_0: $\mathbf{C} = \mathbf{I}$ (The correlation matrix is an identity matrix).
H_1: $\mathbf{C} \neq \mathbf{I}$ (The correlation matrix is not an identity matrix).

b. Test Statistic

$$\chi^2_{calculated} = -\ln|\mathbf{C}|\left[(n-1) - \frac{2p+5}{6}\right] \quad (13)$$

where:

$\chi^2_{calculated}$: chi-square test statistic,
$|\mathbf{C}|$: the determinant value of the correlation matrix,
n : number of observations,
p : number of variables.

c. Critical Area

H_0 will be rejected if $\chi^2_{calculated} \geq \chi^2_{\alpha,\frac{p(p-1)}{2}}$. It can also be observed through the significance value or p-value, where H_0 will be rejected when p-value < 0.05, indicating the presence of multicollinearity.

2.4.2 Principal Component Analysis

The principal component analysis aims to see data patterns by reducing data dimensions (variables) into smaller dimensions and maintaining the information in the data. In principal component analysis, we will look for an equation (principal component) that contains a linear combination of various variables that can explain the diversity of the data maximally. The main components are independent of each other so that principal component analysis can overcome the problem of multicollinearity in the data. Based on [13], the steps in conducting principal component analysis are as follows.

1. Suppose there is a random vector $\boldsymbol{X} = [x_1, x_2, \dots, x_p]$. Calculate the covariance matrix ($\boldsymbol{\Sigma}$) of the vector, which has pairs of eigenvalues (λ) and eigenvectors ($\boldsymbol{v}$), namely $(\lambda_1, \boldsymbol{v}_1)$, $(\lambda_2, \boldsymbol{v}_2)$, … , $(\lambda_p, \boldsymbol{v}_p)$, where eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.
2. Finding the characteristic equation.
$$(\boldsymbol{\Sigma} - \lambda \boldsymbol{I})\boldsymbol{v} = 0 \tag{14}$$
When $|\boldsymbol{\Sigma} - \lambda \boldsymbol{I}| \neq 0$, then $\boldsymbol{v} = 0$ will be the only solution. Thus, setting $|\boldsymbol{\Sigma} - \lambda \boldsymbol{I}| = 0$ which is called the characteristic equation.
3. Calculating eigenvalues using the characteristic equation.
4. Calculating eigenvectors using the equation $(\boldsymbol{\Sigma} - \lambda \boldsymbol{I})\boldsymbol{v} = 0$.
5. Forming $Y_i$or the i-th principal component using the following equation.
$$Y_i = \boldsymbol{v}_i' \boldsymbol{x} = v_{i1}x_1 + v_{i2}x_2 + \cdots + v_{ip}x_p, \qquad i = 1,2, \dots, p \tag{15}$$
All the principal components that have been obtained are ultimately uncorrelated.

According to [13], several methods can be used to determine the number of principal components.

1. Using eigenvalues, which involves retaining principal components with eigenvalues greater than 1.
2. Using a scree plot, by retaining the number of principal components used when the curve starts to decline.
3. Using the total variance explained by the principal components.

2.4.3 Silhouette Method

The silhouette coefficient is a method used to assess the quality of clusters, indicating how well objects are placed within a cluster. According to [14], the steps to calculate the silhouette coefficient are as follows.

1. Calculate the average distance between observations within the same cluster. For example, suppose there is cluster A with observation i belonging to cluster A. The average distance of observation i to other observations within cluster A is denoted by $a(i)$.
$$a(i) = \frac{1}{n_A - 1} \sum_{r=1, r \neq i}^{n_A} d(i, r) \tag{16}$$
where:
n_A : number of members in cluster A,
$d(i, r)$: euclidean distance between observation i and r, where $i, r \in A$.
2. Calculate the average distance of an observation to all observations in a different cluster. Suppose there is cluster C where cluster A ≠ C. The average distance of observation i (in cluster A) to all observations in cluster C is denoted as $a(i, C)$.
$$a(i, C) = \frac{1}{n_C} \sum_{j=1}^{n_C} d(A_i, C_j) \tag{17}$$
where:
n_C : the number of members in cluster C,

$d(A_i, C_j)$: the Euclidean distance between observation i in cluster A and observation j in cluster C.

3. Calculate the value of $b(i)$ using the following equation.

$$b(i) = \min_{A \neq C} a(i, C) \tag{18}$$

4. Calculate the silhouette value denoted by $s(i)$ using the following equation.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{19}$$

5. Calculate the silhouette coefficient (KS) of clustering with n observations using the following equation.

$$KS = \frac{1}{n}\sum_{i=1}^{n} s(i) \tag{20}$$

2.4.4 Ward's Method Clustering

According to [15], the steps for conducting clustering using Ward's method are as follows.

1. Create k clusters, with each cluster containing one observation.
2. Calculate the value of I (increase in the sum of squared errors when merging two clusters) for each possible pair of clusters.

$$I_{AC} = \frac{n_A n_C}{n_A + n_C}(\bar{y}_A - \bar{y}_C)'(\bar{y}_A - \bar{y}_C) \tag{21}$$

where:
I_{AC} : the increase in the sum of squared errors when merging clusters A and C,
y_i : the value of observation i within a cluster,
$\bar{y}_A, \bar{y}_C, \bar{y}_{AC}$: the mean values of observations in clusters A, C, and AC, respectively,
n_A, n_C, n_{AC} : the number of observations in clusters A, C, and AC, respectively.

3. Merge two clusters that result in the smallest value of I.
4. Repeat steps 2 and 3 until we obtain 1 cluster containing all observations.

2.4.5 Biplot Method

Biplot is an exploratory analysis that presents multivariate data in a two-dimensional plane space. The information a biplot provides includes observations and variables simultaneously represented in a two-dimensional plane space. According to [15], the steps in biplot analysis are as follows.

1. Form the data into a matrix $\boldsymbol{X}$ of size $(n \times p)$, where n is the number of observations and p is the number of variables.
2. Calculate the matrix $\boldsymbol{X'X}$.
3. Calculate the eigenvalues (λ) of the matrix $\boldsymbol{X'X}$.
4. Compute the matrices $\boldsymbol{U}$, $\boldsymbol{\Lambda}$, and $\boldsymbol{V'}$. Based on Singular Value Decomposition (SVD), the data matrix ($\boldsymbol{X}$) can be written as follows.

$$\boldsymbol{X} = \boldsymbol{U\Lambda V'} \tag{22}$$

where:
$\boldsymbol{U}$: an orthogonal matrix of size $(n \times r)$,
$\boldsymbol{\Lambda}$: a diagonal matrix of size $(r \times r)$, where $\boldsymbol{\Lambda} = \text{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_r}\right)$,
$\boldsymbol{V}$: an orthogonal matrix of size $(p \times r)$,
r : the rank of matrix **X**.

The diagonal elements of the matrix $\boldsymbol{\Lambda}$ are the square roots of the eigenvalues of the matrix $\boldsymbol{X'X}$, with $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \cdots \geq \sqrt{\lambda_r}$. Matrix $\boldsymbol{V}$ contains the corresponding eigenvectors of the eigenvalues of the

matrix $\boldsymbol{X'X}$, where $\boldsymbol{V} = [\boldsymbol{v_1}, \boldsymbol{v_2}, \dots, \boldsymbol{v_r}]$ with its columns being normalized column vectors. The elements of matrix $\boldsymbol{U}$, denoted as U_i, can be computed using the following equation.

$$U_i = \frac{1}{\sqrt{\lambda_i}} \times \boldsymbol{v_i} \tag{23}$$

where:
U_i : the i-th element of matrix **U**,
λ_i : the i-th eigenvalue of the matrix $\boldsymbol{X'X}$,
$\boldsymbol{v}_i$: the i-th element of matrix $\boldsymbol{V}$.

5. Calculate matrices $\boldsymbol{G} = \boldsymbol{U\Lambda}^{\alpha}$ and $\boldsymbol{H'} = \boldsymbol{\Lambda}^{1-\alpha}\boldsymbol{V'}$ where $0 \leq \alpha \leq 1$. Thus, the equation obtained is as follows.

$$\begin{aligned} \boldsymbol{X} &= \boldsymbol{U\Lambda V'} \\ &= \boldsymbol{U\Lambda}^{\alpha}\boldsymbol{\Lambda}^{1-\alpha}\boldsymbol{V'} \\ &= \boldsymbol{GH'} \end{aligned} \tag{24}$$

The (i, j)-th element of matrix $\boldsymbol{X}$ can be written as follows.

$$x_{ij} = \boldsymbol{g}_i'\boldsymbol{h}_j \; ; \; i = 1,2, \dots, n; j = 1,2, \dots, p \tag{25}$$

Where $\boldsymbol{g}_i'$ and $\boldsymbol{h}_j$ are respectively row vectors from matrix $\boldsymbol{G}$ and column vectors from matrix $\boldsymbol{H}$ with r dimensions. If $\alpha = 0$ is used, it provides conformity to data diversity [16].

6. Find observation coordinates by taking the first two columns from each matrix $\boldsymbol{G}$ and $\boldsymbol{H}$, written as $\boldsymbol{G_2}$ and $\boldsymbol{H_2}$.
7. Create a biplot using matrices $\boldsymbol{G_2}$ and $\boldsymbol{H_2}$. Each row in matrix $\boldsymbol{G_2}$ represents coordinates (x, y) for each observation and each row in matrix $\boldsymbol{H_2}$ represents coordinates (x, y) for each variable.
8. Measure the quality of the biplot mapping using the following equation.

$$\rho^2 = \frac{\lambda_1 + \lambda_2}{\sum_{k=1}^{r} \lambda_k} \tag{26}$$

where:
λ_1: the first largest eigenvalue,
λ_2: the second largest eigenvalue,
λ_k: the k-th largest eigenvalue, $k = 1,2, \dots, r$.

When the value of ρ^2 approaches one, it indicates that the biplot is better at presenting true data information.

3. RESULTS AND DISCUSSION

In this chapter, analysis and discussion will be conducted on the grouping of regencies/cities on Java Island, along with their indicators of public welfare, using the biclustering method and the cluster-biplot method. The results of the two methods will be compared to determine which method is better for the data in this study. The comparison of the results from both methods will be evaluated using standard deviation values.

3.1. Plaid Model Biclustering Result

The background effect model used contains the mean effect, which is $\Theta_{ijk} = \mu_k$. The threshold values chosen by the researcher are $\tau_1 = 0.5$ and $\tau_2 = 0.5$. Then, backfitting is conducted three times. Two groups are obtained, namely Bicluster 1 and Bicluster 2. Bicluster 1 has a size of 26×2, meaning there are 26 regencies/cities and 2 variables (PPM and PKM (percentage of per capita expenditure on food consumption)). It can be detected that these 26 regencies/cities have relatively smaller values for the PPM and PKM variables. Meanwhile, Bicluster 2 has a size of 30×5, meaning there are 30 regencies/cities and 5 variables (RLS, HLS, KP, PDRB, and TPT). It can be detected that these 30 district/city variables have relatively smaller

values for the RLS (average length of schooling), HLS (expected length of schooling), KP (population density), PDRB (GDP per capita at current prices), and TPT (open unemployment rate) variables. Thus, generally, the regencies/cities in Bicluster 2 are less prosperous than Bicluster 1. The distribution map of members of Bicluster 1 and Bicluster 2 is as follows. Members of Bicluster 1 are generally urban areas, while members of Bicluster 2 are generally rural areas.

(a) **(b)**

Figure 1. Mapping, (a) Bicluster 1, (b) Bicluster 2

3.2 Cluster-Biplot Result

3.2.1 Clustering Assumption Checking

1. Sample Represents Population

   Using the Kaiser Meyer Olkin (KMO) test, a KMO value of $0.76 \geq 0.5$ was obtained, indicating that the sample represents the population.

2. No Multicollinearity

   Using the Bartlett test of sphericity, we obtained a p-value $= 2.22 \times 10^{-16}$, then p-value < 0.05. With a confidence level of 95%, it can be concluded that there is multicollinearity in the data, so this assumption is not met. Before proceeding to cluster analysis, a principal component analysis must first be conducted to reduce correlated variables into several components.

3.2.2 Principal Component Analysis

The number of principal components is determined based on eigenvalues > 1, resulting in two principal components from the reduction of nine existing variables. The total data variance explained by these two principal components is 64.3%. The dominant variables in principal component 1 are PPM, RLS, HLS, PKM, KP, and PDRB, while the dominant variables in principal component 2 are AHH, TPT, and JP.

3.2.3 Silhouette Method

From **Figure 2**, it is observed that the highest silhouette coefficient is when the "Number of clusters k" is two or $k = 2$. In general, the higher the "Average silhouette width" is, the better the grouping quality.

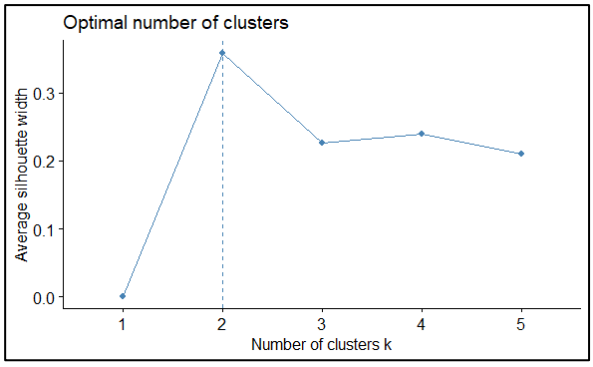

Figure 2. Silhouette Coefficient Value of the Ward's Clustering Method

3.2.4 Ward's Method Clustering

Next, clustering is conducted using the Ward method with two groups formed: Cluster 1 and Cluster 2. Cluster 1 contains 33 regencies/cities, and Cluster 2 contains 86 regencies/cities. Cluster 1 is predominantly comprised of regencies/cities that are generally urban areas. Meanwhile, Cluster 2 is dominated by regencies/cities that are generally rural areas.

3.2.5 Biplot

Subsequently, mapping the two formed groups, observation points, and the variables used is carried out using the biplot method. The percentages on Dim1 and Dim2 represent the amount of data variance explained by each PC1 (principal component 1) and PC2 (principal component 2).

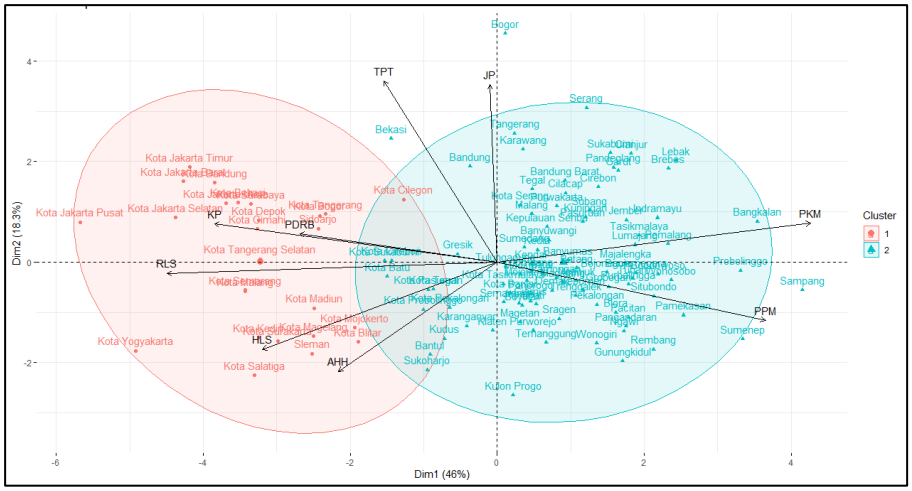

Figure 3. Biplot

Based on **Figure 3**, the proportion of data variance explained by the biplot is 64.3%. This indicates that the biplot method adequately describes the relationship between regencies/cities and the variables representing people's welfare indicators.

Table 2. The Results of the Cluster-Biplot Method

Cluster	Number of Members	Members
1	33	Kota Tangerang, Kota Tangerang Selatan, Kota Jakarta Selatan, Kota Jakarta Timur, Kota Jakarta Pusat, Kota Jakarta Barat, Kota Jakarta Utara, Kota Bogor, Kota Sukabumi, Kota Bandung, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Banjar, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, Kota Tegal, Sleman, Kota Yogyakarta, Sidoarjo, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun, Kota Surabaya, Kota Batu
2	86	Pandeglang, Lebak, Tangerang, Serang, Kota Cilegon, Kota Serang, Kepulauan Seribu, Bogor, Sukabumi, Cianjur, Bandung, Garut, Tasikmalaya, Ciamis, Kuningan, Cirebon, Majalengka, Sumedang, Indramayu, Subang, Purwakarta, Karawang, Bekasi, Bandung Barat, Pangandaran, Kota Tasikmalaya, Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Boyolali, Klaten, Sukoharjo, Wonogiri, Karanganyar, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kulon Progo, Bantul, Gunungkidul, Pacitan, Ponorogo, Trenggalek, Tulungagung, Blitar, Kediri, Malang, Lumajang, Jember, Banyuwangi, Bondowoso, Situbondo, Probolinggo, Pasuruan, Mojokerto, Jombang, Nganjuk, Madiun, Magetan, Ngawi, Bojonegoro, Tuban, Lamongan, Gresik, Bangkalan, Sampang, Pamekasan, Sumenep

3.3 Comparison of Plaid Model Biclustering Results and Cluster-Biplot Results

In general, the biclustering results group has relatively smaller standard deviation values compared to the cluster-biplot results group. Thus, the biclustering method produces better groups because they are more homogeneous compared to the cluster-biplot method.

Table 3. Standard Deviation Value of The Bicluster and Cluster

Variable	Standard Deviation			
	***Bicluster* 1**	***Cluster* 1**	***Bicluster* 2**	***Cluster* 2**
PPM	1.449	1.646	3.149	3.365
PKM	3.762	3.951	2.955	4.633
RLS	0.673	0.452	0.672	1.026
HLS	1.148	1.157	0.411	0.697
AHH	1.544	2.072	2.526	2.514
TPT	1.768	1.695	1.312	2.335
PDRB	145.459	163.750	11.082	39.620
KP	5872.425	5639.483	264.417	1671.190
JP	1034699.329	968068.868	251776.271	839464.094

4. CONCLUSIONS

Based on the results and discussion, both methods produce two groups where Bicluster 2 and Cluster 2 represent less prosperous regions than Bicluster 1 and Cluster 1. Considering the standard deviation values, the biclustering method is superior because it produces more homogeneous groups compared to the cluster-biplot method.

Suggestions that can be conveyed include the idea that further research could use different methods to compare with the results of the Plaid Model biclustering. Further research could also utilize different metrics to evaluate the effectiveness of the grouping performed. Additionally, expanding the observation area to include several islands or using more specific observation objects, such as districts, could enhance the accuracy of the analysis. Suggestions for the local government include prioritizing less prosperous areas and implementing policies tailored to the needs of each region to improve the welfare of the population.

ACKNOWLEDGMENT

The Directorate of Research and Development of Universitas Indonesia (DRPM UI) funded this study as an additional paper through a grant of Hibah Publikasi Terindeks Internasional (PUTI) Q2 2022-2023 No.: NKB-668/UN2.RST/HKP.05.00/2022.

REFERENCES

[1] Badan Pusat Statistik, HASIL SENSUS PENDUDUK 2020, Jakarta: Badan Pusat Statistik, 2021. https://www.bps.go.id/id/pressrelease/2021/01/21/1854/hasil-sensus-penduduk-2020.html

[2] A. F. N. Arif and N. Nurwati, "PENGARUH KONSENTRASI PENDUDUK INDONESIA DI PULAU JAWA TERHADAP KESEJAHTERAAN MASYARAKAT," *Jurnal Ilmu Kesejahteraan Sosial "Humanitas" Fisip Unpas,* vol. 4, no. 1, pp. 54-

70, 2022. https://www.semanticscholar.org/paper/PENGARUH-KONSENTRASI-PENDUDUK-INDONESIA-DI-PULAU-Arif-Nurwati/c4d70585b832f16c08c1e0aa207edb2002c1dec0

[3] R. A. Mulia and N. Saputra, "ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KESEJAHTERAAN MASYARAKAT KOTA PADANG," *Jurnal El-Riyasah*, vol. 11, no. 1, pp. 67-83, 2020. http://dx.doi.org/10.24014/jel.v11i1.10069

[4] S. Yulianto and K. H. Hidayatullah, "ANALISIS KLASTER UNTUK PENGELOMPOKAN KABUPATEN/KOTA DI PROVINSI JAWA TENGAH BERDASARKAN INDIKATOR KESEJAHTERAAN RAKYAT," *Jurnal Statistika Universitas Muhammadiyah Semarang*, vol. 2, no. 1, pp. 56-63, 2014. https://doi.org/10.26714/jsunimus.2.1.2014.%25p

[5] N. Dwitiyanti, N. Selvia and F. R. Andrari, "PENERAPAN FUZZY C-MEANS CLUSTER DALAM PENGELOMPOKAN PROVINSI INDONESIA MENURUT INDIKATOR KESEJAHTERAAN RAKYAT," *Faktor Exacta*, vol. 12, no. 3, pp. 201-209, 2019. http://dx.doi.org/10.30998/faktorexacta.v12i3.4526

[6] Badan Pusat Statistik, STATISTIK INDONESIA 2023, Jakarta: Badan Pusat Statistik, 2023. https://archive.bps.go.id/publication/2023/02/28/18018f9896f09f03580a614b/statistik-indonesia-2023.html

[7] Imasdiani, I. Purnamasari and F. D. T. Amijaya, "PERBANDINGAN HASIL ANALISIS CLUSTER DENGAN MENGGUNAKAN METODE AVERAGE LINKAGE DAN METODE WARD (STUDI KASUS: KEMISKINAN DI PROVINSI KALIMANTAN TIMUR TAHUN 2018," *Jurnal Eksponensial*, vol. 13, no. 1, pp. 9-18, 2022. https://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/875/356

[8] H. A. e. a. Majd, "EVALUATION OF PLAID MODELS IN BICLUSTERING OF GENE EXPRESSION DATA," *Hindawi Publishing Corporation*, 2015. https://pubmed.ncbi.nlm.nih.gov/27051553/

[9] T. Siswantining, A. E. Aminanto, D. Sarwinda and O. Swasti, "BICLUSTERING ANALYSIS USING PLAID MODEL ON GENE EXPRESSION DATA OF COLON CANCER," *Austrian Journal of Statistics*, vol. 50, pp. 101-114, 2021. https://ajs.or.at/index.php/ajs/article/view/1195

[10] H. B. T. &. K. W. Turner, H. Turner, T. Bailey and W. Krzanowski, "IMPROVED BICLUSTERING OF MICROARRAY DATA DEMONSTRATED THROUGH SYSTEMATIC PERFORMANCE TESTS," *Computational Statistics & Data Analysis*, vol. 48, no. 2, pp. 235-254, 2005. https://doi.org/10.1016/j.csda.2004.02.003

[11] J. F. Hair, W. C. Black, B. J. Babin and R. E. Anderson, MULTIVARIATE DATA ANALYSIS, 7th Edition, New York: Pearson Education, 2014.

[12] N. Shrestha, "Factor Analysis as a Tool for Survey Analysis," *AMERICAN JOURNAL OF APPLIED MATHEMATICS AND STATISTICS*, vol. 9, no. 1, pp. 4-11, 2021. https://dx.doi.org/10.12691/ajams-9-1-2

[13] R. A. Johnson and D. W. Wichern, APPLIED MULTIVARIATE STATISTICAL ANALYSIS, vol. 6, UK: Pearson London, 2014.

[14] L. Kaufman and P. J. Rousseeuw, FINDING GROUPS IN DATA: AN INTRODUCTION TO CLUSTER ANALYSIS, vol. 344, John Wiley & Sons, 2009.

[15] A. C. Rencher and W. F. Christensen, METHODS OF MULTIVARIATE ANALYSIS, 3rd Edition, New jersey: John Wiley & Sons, 2012.

[16] I. T. Jolliffe, PRINCIPAL COMPONENT ANALYSIS, 2nd Edition, New York: Springer, 2010.