BAREKENG: Journal of Mathematics and Its Applications
September 2024 Volume 18 Issue 3 Page 1839–1852
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol18iss3pp1839-1852

GROUPING REGENCIES/CITIES IN WEST JAVA PROVINCE BASED ON PEOPLE'S WELFARE INDICATORS USING BIPLOT AND CLUSTERING

Priscilla Ardine Puspitasari[1], Defi Yusti Faidah[2*], Triyani Hendrawati[3]

[1,2,3]*Statistics Department, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran*
Jl. Raya Bandung Sumedang KM 21, Jatinangor, Sumedang, 45363, Indonesia

Corresponding author e-mail: ***** ***defi.yusti@unpad.ac.id***

Article History:

Received: 20th February 2024
Revised: 10th April 2024
Accepted: 8th July 2024
Published: 1st September 2024

Keywords:

Biplot;
People's Welfare;
Ward's Method;
West Java Province.

ABSTRACT

The level of people's welfare in West Java Province still requires improvement in each indicator. People's welfare indicators include poverty, employment, education, housing, consumption patterns, health, and population. The level of people's welfare can be known by reviewing all dimensions based on linear relationships between regencies/cities to produce information on indicators that still need improvement. These efforts can assist the West Java Provincial Government determine regional policies and programs for equitable distribution and improve people's welfare in all regencies/cities. The data used in this study are secondary data derived from the Website of the BPS of West Java Province 2023, West Java Open Data Province 2023, and Diskominfo Statistics Division (Jabar Digital Service). The grouping of regencies/cities was done using Principal Component Analysis based on Singular Value Decomposition biplot analysis, and it continued with Ward's Method Clustering based on Euclidean distance calculation. The analysis results formed four groups with different people's welfare indicators characteristics. The group that needs top priority in improvement is group 2 because it has a low level of people's welfare. Cluster 1 contains regencies/cities with high people's welfare characteristics in the housing and employment indicators. Cluster 3 includes regencies/municipalities with high people's welfare characteristics in the consumption pattern level, poverty, employment, and health indicators. Cluster 4 contains cities with high people's welfare characteristics in education and population indicators.

How to cite this article:

P. A. Puspitasari, D. Y. Faidah and T. Hendrawati., "GROUPING REGENCIES/CITIES IN WEST JAVA PROVINCE BASED ON PEOPLE'S WELFARE INDICATORS USING BIPLOT AND CLUSTERING," *BAREKENG: J. Math. & App.*, vol. 18, iss. 3, pp. 1839-1852, September, 2024.

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

People's welfare indicators show the level of development of the welfare of the Indonesian people between periods and its comparison with other regions [1]. The level of welfare can be described by seven indicators: population, health and nutrition, education, employment, consumption levels and patterns, housing and environment, poverty, and other social factors that affect the quality of human life. The government implements many development programs to improve people's welfare. It is stated in the five strategic issues of the West Java Regional Medium-Term Development Plan that it is a priority, one of which is to improve the socio-economic welfare of the people of West Java [2].

However, one indicator of people's welfare in West Java Province, whose value still needs improvement, is the poverty indicator. Based on BPS West Java data, the percentage of poor people in West Java province in 2022 is 8.06% [3]. That is still below the target poverty percentage for West Java in 2023, which is 7.5% [4].

In addition to the poverty indicator, it is necessary to review several other indicators to determine priority regencies/municipalities that will be given attention because people's welfare has various dimensions of problems such as employment, education, housing, consumption patterns, health, and population [1]. That means that each regency/city will likely have problems different from those of other regencies/municipalities.

The difference in distribution causes differences in the characteristics possessed by regencies/cities. If objects have the same distribution, they will have linear similarity. This linear similarity can cause similar characteristics so that objects that have the same distribution will be close together. Therefore, a method is needed that can provide information simultaneously regarding indicators that still need improvement by looking at groupings based on linear relationships between regencies/cities. The lowest value of the people's welfare indicator in a district/city does not necessarily indicate the need for improvement because, compared with the indicator values of other districts/cities, it is not as bad as imagined. Therefore, it is necessary to simultaneously review each variable by looking at the values in other regencies/municipalities so that the level of people's welfare is more evenly distributed. In addition, this study also aims to provide results of grouping districts/cities based on all aspects of people's welfare as a consideration in determining regional programs and government policies for equalizing people's welfare in West Java Province. In a study conducted by Leleury, a grouping was produced based on the position of the object point in the biplot quadrant [6]. The resulting grouping is ineffective because the mapping results will show the number of groups according to the number of biplot quadrants. Therefore, in this study, clustering will be carried out using cluster analysis to determine a more optimal number of groups.

Over the past five years, the West Java Provincial Government has provided five excellent education, society, and economy programs to improve people's welfare. Later on, data processing of people's welfare indicators is expected to be used to determine policies and evaluate West Java regional programs, both those that have been implemented and those that will be implemented, so that they are right on target. Therefore, information is needed on the grouping of regencies/cities based on indicators of people's welfare, the relationship between regencies/cities and indicators of people's welfare, and the relationship between indicators to determine each group's characteristics.

The method that can be used to map regencies/city trends with their characteristics is the Principal Component Analysis (PCA) Biplot. This method is a visualization and dimension reduction technique through principal components. This method has the advantage of being able to visualize the relationship of objects and variables simultaneously. The research continued using the ward method to obtain existing regencies/cities groupings. Ward's method is formed based on the distance between two groups obtained from the value of the sum of squared deviations on the average group of each observation object (Sum of Square). Ward's method is the best compared to other hierarchical cluster analyses because the grouping is based on the minimum increase in the sum of the squared errors of each cluster [5]. The research that has been carried out previously is biplot analysis on mapping the characteristics of poverty in Maluku province [6]. Apart from that, previous research has also been carried out using the ward's method and comparing it with other hierarchical methods on poverty variables in Indonesia [7]. The difference with this research lies in the objects, variables and methods used. This research is more optimal than existing research because it collaborates biplot with the ward's method clustering as the best grouping method to determine the characteristics of each regency/city along with more accurate grouping results and avoids subjectivity in biplot grouping. This research aligns with the vision of the medium-term development of West Java Province

in 2018-2023, which states that the development of West Java aims to improve and equalize the community's welfare. Therefore, the study's results can support the government's political policies in the future.

2. RESEARCH METHODS

2.1 Research Data

The data used in this study are secondary data derived from the Website of the BPS of West Java Province 2023, West Java Open Data Province 2023, and Diskominfo Statistics Division (Jabar Digital Service). West Java Open Data can be accessed through the Website at https://opendata.jabarprov.go.id/id. The number of observations in the data is 27 regencies/cities in West Java Province. The variables represent several indicators of people's welfare: population indicators, health and nutrition, education, employment, consumption levels and patterns, housing and environment, and poverty. The data consists of 8 variables representing seven indicators of people's welfare. The data matrix used in the study is 27 × 8. The variables and units that will be analyzed for each regency/city in West Java Province in 2022 are as follows:

X_1 : Expenditure per Capita (Thousand Rupiah/Person/Year)
X_2 : Poverty Line (Rupiah/Capita/Month)
X_3 : Education Index (Percent)
X_4 : Average Years of Schooling (Years)
X_5 : Percentage of Households with Access to Adequate Drinking Water (Percent)
X_6 : Percentage of Population with Adequate Sanitation (Percent)
X_7 : Open Unemployment Rate (Percent)
X_8 : Population Density (Person/Area)

2.2 PCA Biplot Analysis

The PCA method converts most of the initial interconnected variables into new variables in the form of principal components that are no longer correlated with each other [8]. Biplot analysis is a statistical technique that aims to present visually and together a group of observations with variables into a flat graph [6]. This method is based on a standardized data matrix's Singular Value Decomposition (SVD).

2.2.1 Data Matrix

The data consisting of 27 regencies/cities with eight research variables are presented in an initial data Y matrix of size $n\ x\ p$ as shown in the following equation:

$$\mathbf{Y}=\begin{bmatrix} y_{1;1} & \cdots & y_{1;8} \\ y_{2;1} & & y_{2;8} \\ \vdots & \ddots & \vdots \\ y_{27;1} & \cdots & y_{27;8} \end{bmatrix} \quad (1)$$

There are different units on the variables in the research data, so it is necessary to standardize the data first to equalize the value in the data because there is a considerable difference in unit size [9]. Data standardization can be done through the following equation:

$$x_{ij} = \frac{y_{ij}-\mu_{.j}}{\sigma_{.j}} ;\ i=1,2,\ldots,n\ ;\ j=1,2,\ldots,p \quad (2)$$

where,
x_{ij} : data after standardization of i-th row and j-th column
y_{ij} : initial data of i-th row and j-th column
$\mu_{.j}$: mean of the initial data of j-th column
$\sigma_{.j}$: standard deviation of the initial data in j-th column

After standardization, the matrix **X** is formed as the standardization result matrix:

$$\mathbf{X}=\begin{bmatrix} x_{1;1} & \cdots & x_{1;8} \\ x_{2;1} & & x_{2;8} \\ \vdots & \ddots & \vdots \\ x_{27;1} & \cdots & x_{27;8} \end{bmatrix} \tag{3}$$

2.2.2 Eigen Value and Eigenvectors

Before calculating the singular values of SVD, it is necessary to know the eigenvalues and eigenvectors first. The eigenvalues and eigenvectors will be obtained from the covariance matrix $\mathbf{X'X}$ with the following formula [15]:

$$|\mathbf{X'X} - \lambda_i \mathbf{I}| = 0 \tag{4}$$

$$(\mathbf{X} - \lambda_i \mathbf{I})\, \boldsymbol{a} = 0 \tag{5}$$

where,
λ_i : i-th eigen value
$\boldsymbol{a}$: eigen vectors
$\mathbf{I}$: identity matrix

2.2.3 *Singular Value Decomposition*

The equation for obtaining the singular decomposition value (SVD) is as follows [10]:

$$\mathbf{X}_{(n\times p)} = \mathbf{U}_{(n\times q)}\mathbf{L}_{(q\times q)}\mathbf{A'}_{(q\times p)} \tag{6}$$

where,

$\mathbf{U}$ and $\mathbf{A}$ are matrices with orthonormal columns ($\mathbf{U'U} = \mathbf{A'A} = \mathbf{I}$)

$\mathbf{L}$ is a diagonal matrix whose diagonal elements are the roots of the eigenvalues of $\mathbf{X'X}$ i.e. $\sqrt{\lambda_1} \geq \sqrt{\lambda_2} \geq \ldots \geq \sqrt{\lambda_q}$

The diagonal elements of matrix $\mathbf{L}$ are called the singular values of matrix $\mathbf{X}$ and the columns of matrix $\mathbf{A}$ are the eigenvectors of $\mathbf{X'X}$. The matrix $\mathbf{U}$ is the column of the multiplication of matrix $\mathbf{A}$, which contains the eigenvectors of matrix $\mathbf{X'X}$, the inverse matrix of singular values, and the initial data matrix with the following equation:

$$\mathbf{XAL^{-1}} = \mathbf{U} \tag{7}$$

The diagonal elements of the matrix $\mathbf{L}$ are the singular values of the matrix $\mathbf{X}$. The elements are defined as $\mathbf{L}^{\alpha}$ with $0 \leq \alpha \leq 1$ which is a diagonal matrix of size $q \times q$ with the diagonal element $\sqrt{\lambda_1}^{\alpha} \geq \sqrt{\lambda_2}^{\alpha} \geq \ldots \geq \sqrt{\lambda_q}^{\alpha}$. This definition also applies to $\mathbf{L}^{1-\alpha}$ with its diagonal elements being $\sqrt{\lambda_1}^{1-\alpha} \geq \sqrt{\lambda_2}^{1-\alpha} \geq \ldots \geq \sqrt{\lambda_q}^{1-\alpha}$. Based on the matrix $\mathbf{X}$, we can suppose that $\mathbf{G} = \mathbf{UL}^{\alpha}$ and $\mathbf{H'} = \mathbf{L}^{1-\alpha}\mathbf{A'}$ with α is from 0 to 1. Therefore, **Equation (6)** can be explained as follows:

$$\mathbf{X} = \mathbf{UL}^{\alpha}\mathbf{L}^{1-\alpha}\mathbf{A'} = \mathbf{GH'} \tag{8}$$

If the elements are *(i,j)*, the matrix $\mathbf{X}$ can be denoted through the row vector g_i and column vector h_j as follows:

$$x_{ij} = g_i' h_j$$

where,
The i-th row is matrix $\mathbf{G}$ with $i = 1,2,\ldots,n$
The j-th column is matrix $\mathbf{H}$ with $j = 1,2,\ldots,p$
Matrix $\mathbf{G}$ can be used as the object coordinates on the biplot, while matrix $\mathbf{H}$ is used as the variable coordinates on the object [11].

2.2.4 Percentage of Data Variance

The next step is to create a biplot according to the number of dimensions based on the proportion of diversity that the biplot can explain. The interpretation of data presented through the PCA Biplot is obtained from the value of the proportion of diversity and the number of dimensions of the biplot. Therefore, it is necessary to know the amount of diversity that can be described through biplots using the following equation:

$$\tau = \frac{\sum_{d=1}^{D} \lambda_d}{\sum_{\ell=1}^{L} \lambda_\ell} \quad (9)$$

where:
τ : proportion of diversity of the biplot
λ_d : *eigen value* d-th with $d = 1,2, \dots D$
D : number of dimensions used
λ_ℓ : *eigen value* ℓ-th of matrix $\mathbf{X}'\mathbf{X}$ with $l = 1,2, \dots, L$
L : number of principal components

In general, biplots are presented in two dimensions or d = 2. If the value of the proportion of cumulative diversity is closer to one (≥ 70%), then the PCA Biplot with d dimensions better explains the information in the data [12]. The percentage of cumulative diversity in the data will produce the number of principal components selected.

2.2.5 Identification of PCA Biplot Mapping Results

The mapping results of PCA Biplot consist of four things that need to be considered, namely the closeness of characteristics between observed objects, variable diversity, correlation between variables, and variable values on objects [12]:

1. Proximity Between Observed Objects
2. Biplot analysis can provide information about objects that have similar characteristics to other objects. Two objects will be grouped based on the same characteristics if indicated by two points with adjacent positions.
3. Variable Diversity
4. Biplot analysis can provide information about variables with almost the same diversity value for each object. In a biplot, variables with small diversity values are depicted as short vectors. In contrast, variables with large diversity values are depicted as long vectors.
5. Correlation Between Variables
6. Biplot analysis can provide information on how a variable can influence or be influenced by other variables. Two variables that have a positive correlation value will be indicated by two vector lines with the same direction or form a narrow-angle (taper). Meanwhile, two variables that have a negative correlation value will be indicated by two vector lines with opposite directions or create wide angle (obtuse). For two variables that are not correlated, it will be indicated by two lines with angles close to 90 degrees (right angle).
7. Variable Value of an Object
8. Biplot analysis can provide information to see the superiority of each object. If the object is located in the direction of the variable vector, then the object has an above-average value. However, suppose the object is located opposite to the variable vector. In that case, the object has a value below the average. Meanwhile, objects almost in the middle indicate that the object has a value close to the average.

2.2.6 Ward's Method Clustering

After biplot mapping, clustering analysis will be carried out based on proximity distance through Ward's method to avoid subjective clustering. Ward's method will produce the best clustering compared to other hierarchical cluster analyses [13]. In addition, Ward's method of clustering will minimize the diversity of each cluster, and the number of members of each cluster is relatively balanced [5].

2.2.7 Measures of Similarity Between Objects

The distance measure between objects to determine the similarity will be obtained by calculating the Euclidean Distance. The calculation for this distance can use the following equation with x,y as the observation object [14]:

$$d(x,y) = \sqrt{\sum_{j=1}^{p}(x_j - y_j)^2} \quad (10)$$

$d(x,y)$: euclidean distance between two objects
x_j : first object distance
y_j : second object distance

2.2.8 Sum Square Error

The purpose of this method is to maximize homogeneity in a cluster; it can be measured by the sum of squares of objects in the cluster, often called the Sum of Square Error (SSE). This SSE value will be used as an objective function. Ward's algorithm groups n objects into n, n-1, n-2, and so on to form one cluster. SSE can be calculated if each cluster has more than one object. Ward's method can be calculated through the following equation [15]:

$$\text{SSE} = \sum_{i=1}^{n}(\boldsymbol{g}_i - \bar{\boldsymbol{g}})'(\boldsymbol{g}_i - \bar{\boldsymbol{g}}) \quad (11)$$

where,
$\boldsymbol{g}_i$: i-th object coordinate with $i = 1,2,3,\ldots,n$
$\bar{\boldsymbol{g}}$: average value of objects in the cluster
n : number of clusters

The first cluster is formed after combining two clusters with the smallest SSE value. Then, Ward's method will determine the number of clusters and their members. Ward's method iteration process is complete if the number of clusters equals 1. If the number of clusters is not equal to 1, repeat the step of determining the SSE value [15].

2.2.9 Determination of the Number of Clusters

In this study, the NBClust library package in Rstudio was used using all existing indices. This package provides 30 indices that determine the optimal number of clusters in a data set and provide the best clustering scheme [16]. This allows researchers to determine the most appropriate number of clusters and clustering scheme. The determination of the optimal number of clusters in NbClust is based on two criteria, namely the maximum and minimum index values or the critical value of each index.

2.3 Flowchart

The following are the stages of data analysis using the biplot and ward's clustering methods presented through a flowchart:

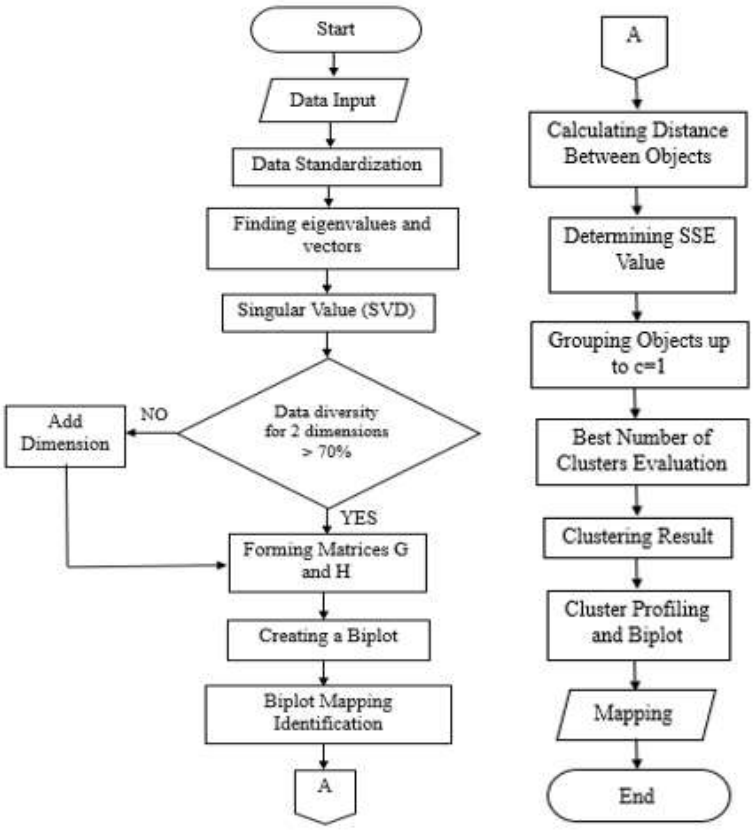
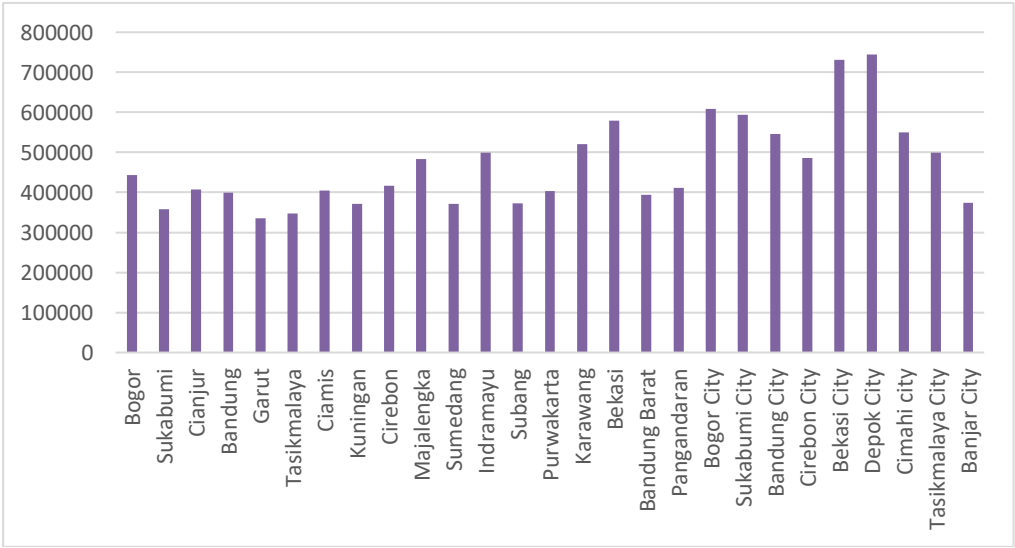


Figure 1. Methodology flowchart

3. RESULTS AND DISCUSSION

3.1 Data Description

BPS uses the basic needs approach to measure the poverty rate based on the poverty line. This poverty line can later be used to review the characteristics of people's welfare, especially poverty indicators. The acquisition of the poverty line in all regencies/cities in West Java province in 2022 is presented in **Figure 1**.

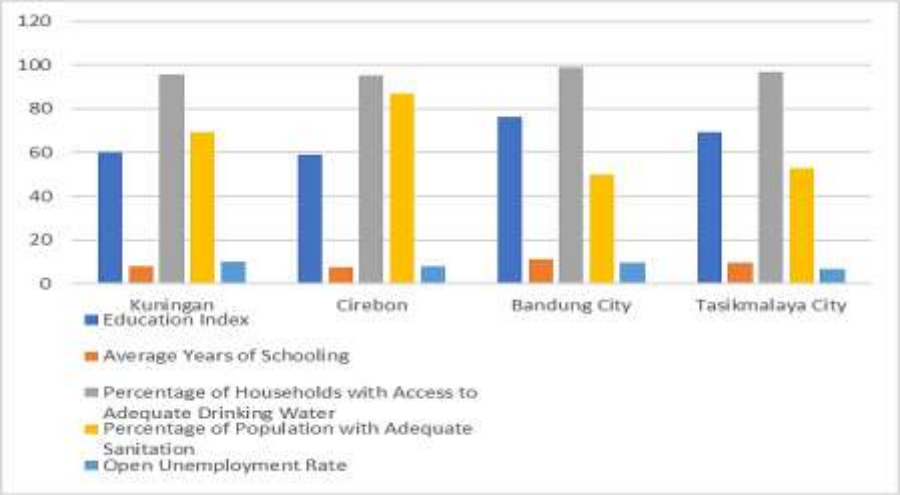


Data source: BPS West Java Province

Figure 2. Poverty lines of regencies/cities in West Java Province 2022

In **Figure 2**, it can be seen that there are differences in poverty lines in each regency/city. The highest poverty line is in Depok City, which is Rp744,771.00, while the lowest is in Garut Regency, which is Rp335,134.00. The poverty line determined by BPS in 2022 was Rp505,469.00 per capita per month **[3]**. It illustrates that in **Figure 2** 19 of 27 regencies/cities, or 70.3% in West Java province, have a poverty line below the specified standard.

Not only can poverty indicators determine the level of people's welfare, but several other indicators can determine regency/city priorities that will receive attention because community welfare has various dimensions of problems such as employment, education, housing, consumption patterns, health, and population. There are possible differences and similarities in problems between regencies/cities based on various dimensions supporting the level of people's welfare, requiring a review regarding the value obtained for each region. Therefore, community welfare indicators can be seen through data distribution to get information about districts/cities with similar characteristics. For example, a sample of data distribution is taken from four districts/cities in West Java province along with supporting variables, namely education index, average years of schooling, percentage of households with access to adequate drinking water, percentage of population with adequate sanitation, and open unemployment rate is presented in **Figure 3**.



Data source: BPS West Java Province

Figure 3. Data distribution in regencies/cities of West Java Province

Figure 3 shows differences in data distribution from each regency/city. Kuningan and Cirebon Regency have the lowest scores on the average years of schooling variable, namely 7.88% and 7.4%. Meanwhile, the cities of Bandung and Tasikmalaya have the lowest scores on the open unemployment rate variable, namely 9.55% and 6.62%. Apart from that, the distribution of Kuningan Regency and Cirebon Regency is similar, with a high percentage of households having access to proper drinking water and a high percentage of the population having access to proper sanitation. However, Bandung City and Tasikmalaya City have many households with access to decent drinking water and a high education index.

3.2 Singular Value Decomposition

The data on people's welfare indicators from each regency/city that have been standardized will be arranged into a matrix **X** as follows:

$$\boldsymbol{X}_{(27x8)} = \begin{bmatrix} -0.121 & \cdots & -0.431 \\ -0.826 & & -0.692 \\ \vdots & \ddots & \vdots \\ -0.075 & \cdots & -0.494 \end{bmatrix} \quad (12)$$

After obtaining the matrix **X**, the matrix $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ and eigenvalues will be calculated, the results of which are presented as follows:

$$\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X}_{(8x8)} = \begin{bmatrix} 26 & \cdots & 21.305 \\ 19.826 & & 19.837 \\ \vdots & \ddots & \vdots \\ 21.305 & \cdots & 26 \end{bmatrix} \quad (13)$$

$$\boldsymbol{\lambda} = \begin{bmatrix} 129.4508 \\ 32.5364 \\ 20.4349 \\ 11.0541 \\ 6.1982 \\ 5.8184 \\ 2.1833 \\ 0.3235 \end{bmatrix} \tag{14}$$

The matrix $\boldsymbol{\lambda}$ is referred to as the eigenvalue matrix whose size (8×1) contains the i-th eigen value of the matrix $\mathbf{X^T X}$ with $i = 1,2,\ldots,8$ where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_8$. The first largest eigenvalue indicates that the first principal component contributes the most to the variation in the data. The second to smallest eigenvalues indicate additional contributions to the variation in the data that were not considered by the previous principal components.

Decomposing the singular values of matrix $\mathbf{X}$ will result in matrices $\mathbf{U}$, $\mathbf{L}$, and $\mathbf{A}'$ which can be written using the following equation:

$$\mathbf{X}_{(27\times 8)} = \mathbf{U}_{(27\times 8)}\mathbf{L}_{(8\times 8)}\mathbf{A}'_{(8\times 8)}$$

Through the help of R software, the matrices $\mathbf{U}$, $\mathbf{L}$, and $\mathbf{A}'$ are obtained as follows:

$$\mathbf{U}_{(27x8)} = \begin{bmatrix} 0.0367 & \cdots & -0.0907 \\ 0.2163 & & -0.1914 \\ \vdots & \ddots & \vdots \\ 0.0341 & \cdots & -0.0562 \end{bmatrix} \tag{15}$$

$$\mathbf{L}_{(8x8)} = \begin{bmatrix} 11.377 & \cdots & 0 \\ 0 & & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0.568 \end{bmatrix} \tag{16}$$

$$\mathbf{A}'_{(8x8)} = \begin{bmatrix} -0.4056 & \cdots & -0.4152 \\ 0.1089 & & -0.2057 \\ \vdots & \ddots & \vdots \\ 0.0028 & \cdots & -0.0259 \end{bmatrix} \tag{17}$$

The matrix $\mathbf{U}$ is a matrix with orthonormal columns where the columns of the matrix are row singular vectors. The diagonal elements in matrix $\mathbf{L}$ are called the singular values of matrix $\mathbf{X}$. Matrix $\mathbf{A}$ is a matrix with orthonormal columns where the columns in the matrix are the eigenvectors of $\mathbf{X^T X}$. While matrix $\mathbf{A}'$ is the transpose of matrix $\mathbf{A}$.

3.3 Percentage of Data Variance

Previously, the value of λ_1 was 129.4508 and λ_2 was 32.5364 based on **Equation (14)**. So, the calculation of the percentage value of diversity is as follows:

$$\tau = \frac{\lambda_1 + \lambda_2}{\sum_{\ell=1}^{8} \lambda_\ell} = \frac{129.4508+32.5364}{207.9996} = 0.7787 \tag{18}$$

The percentage of data diversity in the biplot that the two main components can explain is 77.88%. Therefore, a two-dimensional biplot map can represent information on the grouping of districts/cities in West Java Province based on community welfare indicators quite well if the percentage is above 70% **[15]**.

3.4 Formation of Biplot Coordinates

Based on the results of the previous data diversity percentage, 2 main components were obtained so that the mapping can be mapped in two-dimensional form. Matrix $\boldsymbol{G}^{(2)}$ describes the points of observation objects, namely 27 regencies/cities in West Java Province obtained from the first two columns of the matrix $\mathbf{G}$, while the matrix $\boldsymbol{H}^{(2)}$ describes the points of observation variables consisting of 8 variables supporting

the people's welfare indicators obtained from the first two columns of the matrix **H**. Based on the α used of 0,5 so $\mathbf{G} = \mathbf{UL}$ and $\mathbf{H}' = \mathbf{L}^{0,5}\mathbf{A}'$. Therefore, matrix **G** and matrix **H** can be arranged as follows:

$$\mathbf{G}_{27\times 2} = \begin{bmatrix} 0.124 & -0.148 \\ 0.729 & -0.315 \\ \vdots & \vdots \\ 0.115 & 0.570 \end{bmatrix} \quad (19)$$

$$\mathbf{H}_{8\times 2} = \begin{bmatrix} -1.368 & 0.260 \\ -1.320 & 0.208 \\ \vdots & \vdots \\ -1.400 & -0.491 \end{bmatrix} \quad (20)$$

3.5 PCA Biplot Analysis Results

Based on the results obtained, two dimensions will be used to form biplots. Where the biplot will present points that show the coordinates of regency/city objects and vectors that show the coordinates of the people's welfare indicator variables. Object and variable coordinates are obtained from matrix **G** and matrix **H** in the previous calculation. Based on the coordinates obtained, the biplot map of the people's welfare indicators of all regencies/cities in West Java Province can be presented in **Figure 4**.

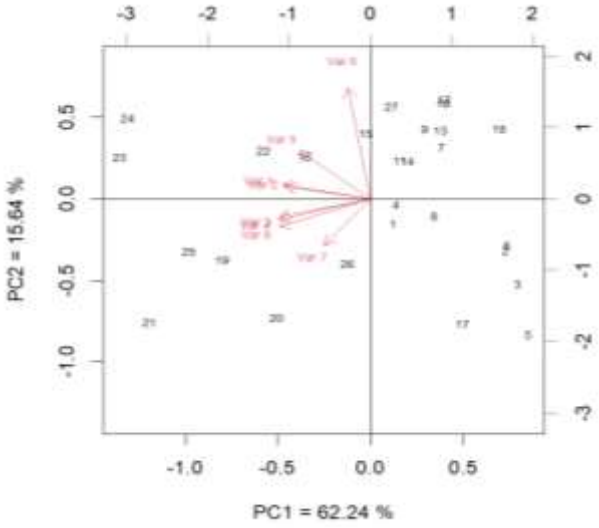


Figure 4. Results of biplot analysis

Based on the direction of the vector in **Figure 4**, it can be seen that information related to the relationship between supporting variables of people's welfare indicators, namely the per capita expenditure variable (Var_1) and the poverty line (Var_2), form an acute angle, which means that they are positively correlated with all existing variables, namely the education index, average years of schooling, percentage of households with access to proper drinking water, percentage of population with proper sanitation, open unemployment rate, and population density. This shows that the increase in per capita expenditure and poverty line in West Java Province will also show an increase in all variables supporting the people's welfare indicators. Meanwhile, the variable percentage of the population with proper sanitation (Var_6) forms an obtuse angle, negatively correlating with the variable open unemployment rate (Var_7).

3.6 Ward's Method Clustering

Calculation of SSE value based on **Equation (11)** with the help of processing through the cluster package in R software. The following are the biplot results with four clusters:

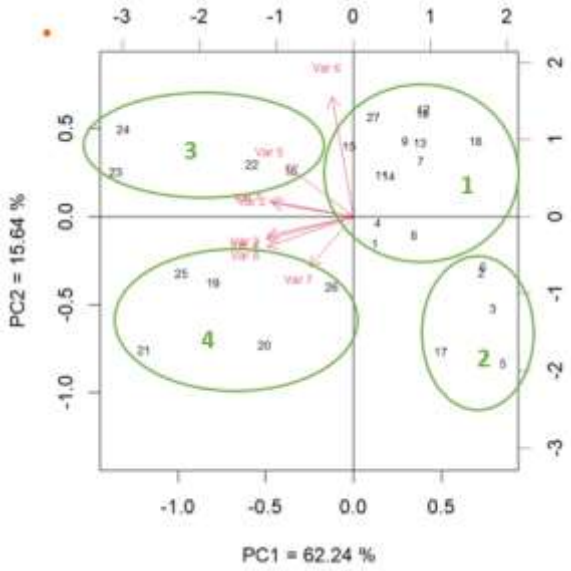


Figure 5. Ward's method clustering

Determining the number of clusters in this study uses a package in R Studio, namely Nbclust. This package provides results in 30 index values that can determine the optimal number of clusters in the ward method. In the end, with the help of R software, the results of the five most indexes were obtained, giving the optimal number of clusters as many as four.

Table 1. Regency/city clusters in West Java Province

Cluster	Regency/City Name
1	Bogor Regency, Bandung Regency, Ciamis Regency, Kuningan Regency, Cirebon Regency, Majalengka Regency, Sumedang Regency, Indramayu Regency, Subang Regency, Purwakarta Regency, Karawang Regency, Pangandaran Regency, and Banjar City
2	Sukabumi Regency, Cianjur Regency, Garut Regency, Tasik Regency, and Bandung Barat Regency
3	Bekasi Regency, Cirebon City, Bekasi City, and Depok City
4	City Bogor, Sukabumi City, Bandung City, Cimahi City, and Tasikmalaya City

Each cluster has different characteristics of people's welfare indicators. In Figure 5, the character of each cluster can be seen from the value and location of the object or coordinate point against the variable vector. Therefore, it can be determined which objects in a cluster have values below or above the average for a variable. The following are the values of objects against variables described based on the characteristics of each cluster:

1. Cluster 1 contains regencies/cities with high people's welfare characteristics in the housing and employment indicators. It can be seen from the direction of the vector of variable 6, namely the percentage of population with proper sanitation, which is in the same direction as all regency/city object points in cluster 1, and the vector of variable 7, namely the open unemployment rate, which is in the opposite direction to all object points in cluster 1. Therefore, the percentage of the population with proper sanitation in cluster 1 is above the average, and the open unemployment rate is below the average of all regencies/cities in West Java Province. Based on this, it can be concluded that the regencies/cities in cluster 1 need improvement in expenditure per capita, poverty line, education index, average years of schooling, percentage of households with access to proper drinking water, and population density.
2. Cluster 2 contains regencies that mostly have the lowest people's welfare characteristics in almost all indicators, starting from the level of consumption patterns, poverty, education, health, housing, and population, except employment. It can be seen from the direction of the eight variable vectors, namely

per capita expenditure, poverty line, education index, average years of schooling, percentage of households with access to proper drinking water, percentage of population with proper sanitation, open unemployment rate, and population density which are in the opposite direction to all regency object points in cluster 2. Therefore, all variables supporting the people's welfare indicators in cluster 2 have values below the average of all regencies/cities in West Java Province, so they need improvement, except for the open unemployment rate because the lower the unemployment value, the better for people's welfare.

3. Cluster 3 contains regencies/municipalities with high people's welfare characteristics in the consumption pattern level, poverty, employment, and health indicators. It can be seen from the vector direction of variables 1, 2, and 5, namely per capita expenditure, poverty line, and percentage of households with access to decent drinking water, which is in the same direction as all regency/city object points in cluster 3. In addition, the open unemployment rate variable is in the opposite direction to the regency/city object points in cluster 3. Therefore, per capita expenditure, poverty line, and percentage of households with access to decent drinking water in cluster 3 are above the average. In contrast, the open unemployment rate is below the average of all regencies/cities in West Java Province. Based on this, it can be concluded that the regencies/cities in cluster 3 need improvement in the education index variable, average years of schooling, percentage of population with access to proper sanitation, and population density.
4. Cluster 4 contains cities with high people's welfare characteristics in education and population indicators. It can be seen from the vector direction of variables 3, 4, 7, and 8: the education index, average years of schooling, unemployment rate, and population density, which are in the same direction as all regency/city object points in cluster 4. Therefore, the education index, average years of schooling, open unemployment rate, and population density in cluster 4 are above the average in all regencies/cities in West Java Province. Based on this, it can be concluded that the regencies/cities in cluster 4 need improvement in the variables of per capita expenditure, poverty line, percentage of households with access to proper drinking water, percentage of population with proper sanitation, and open unemployment rate.

4. CONCLUSIONS

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

1. Based on the clustering results, regencies/cities in West Java Province can be grouped into 4 clusters, with 1 cluster that needs to be prioritized by the West Java Provincial Government to increase and improve the majority of people's welfare indicators. Cluster 1 contains 13 regencies/cities, Cluster 2 contains five regencies, Cluster 3 contains four regencies/cities, and Cluster 4 contains five cities.
2. The information obtained from the biplot shows the relationship between indicators of people's welfare and regencies/cities or can be interpreted as the characteristics of each regency/city based on indicators of people's welfare and the relationship between variables supporting indicators of people's welfare in West Java Province.
3. The results of clustering regencies/cities based on the people's welfare indicators show that cluster 2 has the lowest people's welfare characteristics compared to other clusters. Regencies/cities in Cluster 1 have high people's welfare characteristics in the housing and employment indicators. Regencies/cities in Cluster 3 have high public welfare characteristics regarding consumption patterns, poverty, employment, and health. Regencies/municipalities in Cluster 4 have high people's welfare characteristics in the education and population indicators. Therefore, each regency/city in West Java Province needs attention from the government to improve specific indicators of people's welfare.
4. The variables of per capita expenditure and the poverty line are variables that have a positive correlation with all supporting variables of the people's welfare indicators, so it is concluded that with an increase in the value of per capita expenditure and the poverty line, the supporting variables of other people's welfare indicators will also increase. Therefore, it can be a top priority for the West Java Provincial government to focus on the per capita expenditure and poverty line variables.

REFERENCES

[1] BPS, *Statistik Kesejahteraan Rakyat Provinsi Jawa Barat*. Bandung: BPS, 2022.

[2] BAPPEDA, *Rancangan Akhir RPJMD Tahun 2018-2023*. Bandung: BAPPEDA, 2018.
[3] BPS, *Indikator Kesejahteraan Rakyat Provinsi Jawa Barat Tahun 2022*. Bandung: BPS, 2023.
[4] BAPPEDA, *Rancangan Awal RKPD Tahun 2023*. Bandung: BAPPEDA, 2023.
[5] E. Schubert, "HACAM: Hierarchical Agglomerative Clustering Around Medoids – and its Limitations," *CEUR Workshop Proc*, pp. 191–204, 2021.
[6] Z. A. Leleury and A. E. Wokanubun, "Analisis Biplot pada Pemetaan Karakteristik Kemiskinan di Provinsi Maluku," *Jurnal Ilmu Matematika dan Terapan*, vol. 9, no. 1, pp. 21–31, 2015.
[7] N. Afira and A. W. Wijayanto, "Analisis Cluster Kemiskinan Provinsi di Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarki," *Jurnal Sistem Komputer*, vol. 10, no. 2, pp. 101–109, 2021.
[8] R. Firliana, R. Wulanningrum, and W. Sasongko, "Implementasi Principal Component Analysis (PCA) Untuk Pengenalan Wajah Manusia," *Nusantara of Engineering Journal*, vol. 2, no. 1, 2015.
[9] N. Ulinnuh and R. Veriani, "Analisis Cluster dalam Pengelompokan Provinsi di Indonesia Berdasarkan Variabel Penyakit Menular Menggunakan Metode Complete Linkage, Average Linkage, dan Ward," *Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 5, no. 1, 2020.
[10] I. T. Jolliffe, *Principal Components Analysis*, 2nd ed. New York: Springer, 2012.
[11] I. Ginanjar, U. S. Pasaribu, and S. W. Indratno, "A measure for objects clustering in principal component analysis biplot: A case study in inter-city buses maintenance cost data," *AIP Conf Proc*, pp. 1827, 2017.
[12] A. A. Mattjik and I. M. Sumertajaya, *Sidik Peubah Ganda (First)*. Bogor: IPB Press, 2011.
[13] S. Saracli, N. Dogan, and I. Dogan, "Comparison of Hierarchical Cluster Analysis Methods by Cophenetic Correlation," *J Inequal Appl*, vol. 203, no. 1, 2013.
[14] A. M. Jarman, "Hierarchical Cluster Analysis: Comparison of Single Linkage, Complete Linkage, Average Linkage, and Centroid Linkage Method," *Researchgate Publication*, 2020.
[15] A. C. Rencher, *Methods of Multivariate Analysis*, 3rd ed. New York: John Wiley and Sons, 2012.
[16] M. Charrad, N. Ghazzali, V. Boiteau, and A. Niknafs, "NBClust: An R Package for Determining the Relevant Number of Clusters in a Data Set," *J Stat Softw*, vol. 61, no. 6, 2014.