JURNAL STATISTIKA DAN APLIKASINYA

e-ISSN 2620 – 8369
VOLUME 8 (ISSUE 2) DECEMBER 2024: 193-205
https://doi.org/10.21009/JSA.08206

ANALYSIS OF BICLUSTERING ITERATIVE SIGNATURE ALGORITHM ON POVERTY DATA IN SULAWESI ISLAND IN 2022

Nabila Safitri [1], Yekti Widyaningsih[2*]

[1,2]*Statistics Study Program, Faculty of Mathematics and Natural Science, University of Indonesia, Kampus Baru UI, Depok, West Java, 16424, Indonesia*

Corresponding author's e-mail: [1)]**nabila.safitri@sci.ui.ac.id**, *[2)]**yekti@sci.ui.ac.id**

ABSTRACT

Article History:
Received: 30, June 2024
Revised: 24, December 2024
Accepted: 27, December 2024
Published: 31, December 2024
Available online.

Keywords:
Biclustering, Clustering, Mean Square Residue (MSR).

Poverty in Indonesia is still a problem that must be addressed every year. According to the March 2022 Susenas report, Sulawesi Island ranks third among the six major islands in Indonesia in terms of the percentage of the population living in poverty. This shows that there are still many people living in poverty in Sulawesi. Therefore, the government needs to make the right policies to address this problem. One potential approach is to cluster districts or cities in Sulawesi based on poverty-related variables. The objective of this research is to group the data in two directions: first, by districts or cities and, second, by its variables simultaneously. The formation of these groupings will facilitate the development of the right government policies to address poverty. The appropriate method for these groupings is the biclustering method, which can group observations and characteristics simultaneously so that biclusters formed can be characterized differently. One of the biclustering algorithms is the Iterative Signature Algorithm (ISA), which requires an upper threshold value and a lower threshold value. The threshold value is the value used to determine whether a district or city and variables can be included in a bicluster. The best result is selected based on the average Mean Square Residue (MSR) per volume. Biclustering analysis of poverty data in Sulawesi in 2022 using ISA produced 2 biclusters. Based on these results, the government is expected to make the right policy to overcome poverty problems in Bicluster 1 and Bicluster 2.

How to cite this article:
N. Safitri, Y. Widyaningsih, "ANALYSIS OF BICLUSTERING ITERATIVE SIGNATURE ALGORITHM ON POVERTY DATA IN SULAWESI ISLAND IN 2022", Jurnal Statistika dan Aplikasinya, vol. 8, iss. 2, pp. 193 – 205, December 2024

Journal e-mail: *jsa@unj.ac.id*
Research Article · Open Access

1. INTRODUCTION

Poverty in Indonesia is one of the problems that must be considered every year. In March 2022, the percentage of the national poor population was 9.54%, a decrease compared to September 2021 which was previously 9.71% (BPS, 2022). There are 256 districts/cities that have a percentage of poor people above the national figure. This means that there are still many Indonesians who experience poverty, making it a national problem. Several regions in Indonesia are experiencing serious poverty. In May 2022, the islands of Papua and Maluku had the highest percentage of poor population and poverty rates. Among the 10 provinces with the highest poverty rates, two were located on Sulawesi Island: Gorontalo (15.42%) and Central Sulawesi (12.33%). In the March 2022 Susenas data, Sulawesi Island is the third island with the highest total percentage of poor people after Maluku and Papua Islands; and Bali and Nusa Tenggara Islands. According to the poverty percentage distribution map, Sulawesi Island has an average percentage of poor people of around 10%-15%. This shows that the percentage of poor people in Sulawesi Island is still relatively high. Therefore, research on poverty in Sulawesi Island is interesting to study.

The high level of poverty in a region can occur due to several factors. Mahmud et al. (2020) stated that the factors that influence the poverty rate in Sulawesi Island are economic growth, average years of schooling, and unemployment. Aziz et al. (2016) stated that education and unemployment have a significant effect on poverty in Kutai Kartanegara Regency. Safitri et al. (2020) stated that government subsidies can help poverty alleviation efforts so that they can improve the level of community welfare. The poor are people who have an average per capita expenditure below the Poverty Line. The Poverty Line is used to measure poverty indicators, namely the percentage of poor people (Head Count Index-P_0), the poverty depth index (Poverty Gap Index-P_1), and the poverty severity index (Poverty Severity Index-P_2) (BPS, 2022). According to Adisasmita (2005), poverty indicators used in general are salary/wage, life expectancy, government assistance in the form of subsidies, food fulfillment, clean water, population development, literacy rate, health, education, income per capita, and income distribution (Dewi, 2017). Based on the previously described studies, the factors that affect the poverty rate and poverty indicators will be used as research variables.

The percentage of poor people in Sulawesi Island, which is still relatively high, needs to be explored more deeply by the government. Exploration is carried out so that the government can take the right policy to overcome poverty. One of the efforts that the government can make is by clustering. The clustering method is carried out with the aim of grouping observations that have the same characteristics. The clustering method is only done by grouping rows or observations. To obtain results that are more complete and clearer than the results of clustering rows or columns alone, it is necessary to group row (observation) and column (variable) data together. Thus, the appropriate method to use is biclustering. Biclustering is a method of simultaneously clustering rows and columns, which was first introduced by Cheng & Church in 2000. Initially, biclustering was a development of clustering and applied to gene data. Currently, the development of biclustering is not only in the field of bioinformatics but in other fields, such as social fields, economic fields, and others. Previous research that has discussed biclustering is research conducted by Putri et al. (2021) using the Cheng and Church algorithm on poverty indicators in Central Java.

This research focuses on the Iterative Signature Algorithm (ISA) because it applies a thresholding approach. The application of thresholding is done with the aim of producing more significant bicluster data with a lot of noise. According to Hans (2012), one form of noise is outliers. ISA biclustering aims to define a bicluster as a Transcription Module (TM) (Bergmann et al, 2003). Research on ISA biclustering was first conducted by Bergmann (2003) on gene data. Another study was conducted by Ningsih et al. (2022) by clustering 34 provinces in Indonesia and indicators of economic pattern detection and vulnerability index simultaneously.

Based on the background previously described, an Iterative Signature Algorithm (ISA) biclustering analysis will be conducted on poverty data in Sulawesi Island in 2022. The novelty of this research is that the biclustering analysis area is 81 districts/cities in Sulawesi Island using 16 poverty variables.

2. METHODS

Material and Data

This research covers 81 districts/cities on the island of Sulawesi and 16 poverty variables. The results of the analysis of research variables are shown in Table 1.

Table 1. Results of Descriptive Analysis of Poverty Variables

No.	Variable	Symbol	Mean	Standard Deviation	Median	Minimum	Maximum
1.	Percentage of poor people (%)	PPM	10.84	3.803617	11.65	4.57	18.74
2.	GRDP at current prices (billion rupiah)	PDRBHB	16363.44	27950.013	10111	637.5	208935.8
3.	Open unemployment rate (%)	TPT	3.603	2.078932	2.99	0.58	11.82
4.	Poverty severity index	IKPM	0.4094	0.25747	0.34	0.09	1.37
5.	Poverty depth index	IKDM	1.672	0.792746	1.54	0.56	3.84
6.	Average years of schooling (year)	RLS	8.665	1.129833	8.46	6.75	12.52
7.	Expected years of schooling (year)	HLS	13.19	0.933663	13.03	11.61	16.9
8.	Literacy rate of poor population (%)	AMHPM	99.67	1.513224	100	87.33	100
9.	Percentage of poor households using safe water (%)	PRTMAL	88.12	13.0778	93.4	25.16	100
10	Percentage of poor households using their own/shared latrines (%)	PRTJ	78.11	16.10817	83.54	32.82	100
11.	The school enrollment rate for the poor is 13-15 years (%)	APSPM	87.85	12.8714	91.47	32.54	100
12.	Life expectancy (year)	AHH	69.45	2.364392	70.15	62.16	73.93
13.	Percentage of poor people working in the agricultural sector (%)	PPMBSP	29.05	13.11114	29.21	1.67	63.69
14.	Percentage of poor people who graduated from high school (%)	PPMUTSMA	25.03	8.732871	22.7	10.68	45.77
15.	Per capita food expenditure of the poor (%)	PPKMPM	65.85	5.836168	64.72	52.53	80.15
16.	Percentage of poor households receiving the basic food program (%)	PRTMMPS	32.96	15.7654	32.91	0.55	71.28

Research Method

Iterative Signature Algorithm (ISA) is one of the biclustering algorithms first introduced by Bergmann (2003) on gene data with the aim of finding a set of genes and a set of conditions that

Transcription Module (TM). The advantage of this algorithm can work on a large scale (Csardi, 2009). Ihmels et al. (2004) and Bergmann et al. (2003) conducted research using gene data, namely transcriptomic data. In the process, ISA identifies modules with an iterative procedure. The algorithm starts from a seed input (corresponding to some set of genes or samples), which is refined at each iteration by adding and/or removing genes and/or samples until the process converges to a stable set, referred to as transcription modules. The result of the process is a collection of potentially overlapping modules. Each module contains data on genes that are over- and/or under-expressed, in the samples included in that module (Csardi, 2010).

The steps of bicluster formation with the ISA algorithm were modified by researchers based on the research of Sumertajaya et al. (2023), Ningsih et al. (2022), Zhang et al. (2020), A. Freitas et al. (2011), Ihmels et al. (2002) as follows.

1. Set the row threshold values (t_R) and the column threshold values (t_C), seed values, and number of seeds (n).
2. Create matrix $A_{|U|\times|V|} = (U, V)$. Matrix A is a matrix with a set of rows U consisting of I rows denoted by $|U|$ and a set of columns V consisting of J columns denoted by $|V|$.
3. Create the row normalization matrix A (A_R) and column normalization matrix A (A_C).
4. Randomly select a subset of matrices that are multiple row vectors (row samples)
5. Calculate the row average of each column in the matrix subset A ($a^C_{U'v}$).
6. Calculate the overall average, that $\frac{1}{|V|}\sum_{v\in V} a^C_{U'v}$, where $a^C_{U'v}$ is the row average of each column in the matrix subset, and V is the number of columns.
7. Selecting the column scores that satisfy the conditions shown in Equation 1.

$$a^C_{U'v} - \frac{1}{|V|}\sum_{v\in V} a^C_{U'v} > t_C\sigma_C \quad (1)$$

where $a^C_{U'v}$ is the row average of each column in the matrix subset, V is the number of columns, t_C is the column threshold value, and σ_C is the column standard deviation.

8. Calculating the column average of each row in the column score ($a^R_{uV'}$).
9. Calculate the overall average, that $\frac{1}{|V|}\sum_{v\in V} a^C_{U'v}$, where $a^C_{U'v}$ is the row average of each column in the matrix subset, and V is the number of columns.
10. Selecting the row scores that satisfy the conditions shown in Equation 2.

$$a^R_{uV'} - \frac{1}{|U|}\sum_{u\in U} a^R_{uV'} > t_R\sigma_R \quad (2)$$

where $a^R_{uV'}$ is the column average of each row in the column score, U is the number of rows, t_R is the row threshold value, and σ_R is the row standard deviation.

11. Steps 4 and 10 will repeat for n number of seeds when the convergence condition is not met, that is $\frac{|U'\backslash U"|}{|U'\cup U"|} > \varepsilon$. (U' is the set of rows at the n-1 th iteration, U" is the set of rows at the nth iteration, and ε is the accuracy parameter)
12. When the convergence condition is met, that is $\frac{|U'\backslash U"|}{|U'\cup U"|} < \varepsilon$, rows and columns (bicluster) are obtained. (U' is the set of rows at the n-1 th iteration, U" is the set of rows at the nth iteration, ε is the accuracy parameter)
13. Repeat steps 3 to 13 as many times as the number of possible biclusters formed.

The design of the Iterative Signature Algorithm (ISA) biclustering analysis process, which generates the biclusters in this study, is illustrated in Figure 1.

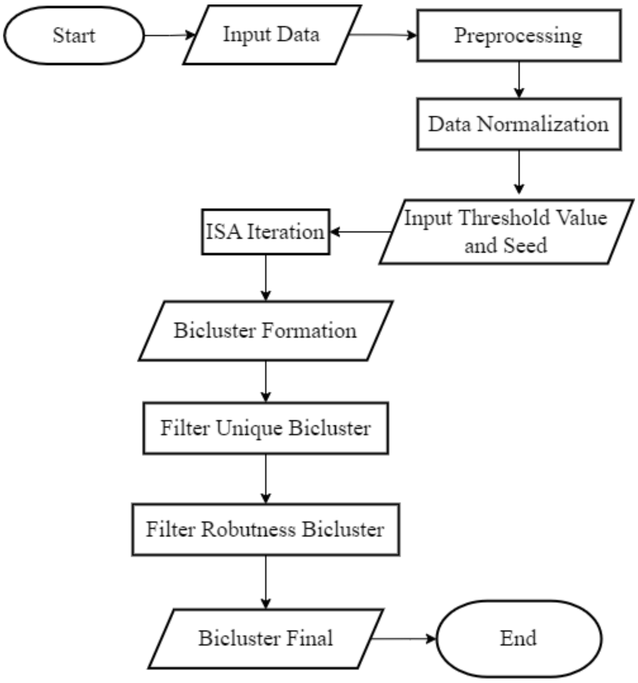


Figure 1. Flowchart of ISA Biclustering Work Steps

The following is an explanation of Figure 1.

1. Data Input

   The data used in this research is numerical data containing variables and research objects.

2. *Preprocessing*

   The preprocessing process is to perform standard normalization on the data. The data normalization method performed in this research is the zero-mean normalization method. This method uses mean and standard deviation. The normalization equation for each matrix element is:

$$z_{ij} = \frac{a_{ij} - \bar{a}_J}{s_J} \tag{3}$$

   for $i = 1,2, \dots, n$ and $j = 1,2, \dots, m$
   with:
   z_{ij} : the result of normalizing the value in the i-th row and j-th column
   a_{ij} : matrix element value of i-th row and j-th column
   $\bar{a}_J$: average of the j-th variable
   s_J : standard deviation of the j-th variable
   n : number of observations
   m : number of variables

3. Row and column normalization in ISA

   In the ISA process, the data matrix can be viewed from two points of view, namely rows and columns. After normalizing the data that equalizes the unit value, normalization of rows and columns is performed again. The results of normalization are normalization to rows (A^r) and normalization to columns (A^c). Both matrices are obtained by centering and scaling to get an average value of 0 and a variance of 1.

4. *Threshold*

   Thresholding is done by calculating the mean and standard deviation of the row vector and column vector. The retained matrix element values are those that are significant or different from the average value. In thresholding, researchers can determine the direction of the bicluster value to be obtained based on the direction parameter, namely "up", "down", and "updown".

5. *Seed*

   In addition to the threshold, the number of seeds is also used in this study. Seeds are randomly selected from a very large search space. Generate seed is used as the initial value for the iteration process.

6. Iteration Process

   The iteration process is carried out by determining the initial value (seed) and the number of seeds (n). Then, it continues by calculating the average of each column in the column normalization matrix subset (the value has been calculated in the data normalization process). If the result is greater than the threshold value then the result becomes the column score. Then, continue to calculate the average of each row on the column score in the row normalization matrix (the value has been calculated in the data normalization process). If the result is greater than the threshold value, then the result becomes the row score. The process is carried out until it converges, that is, the result is always the same as the result of the previous iteration so that it becomes a fixed point. If the result does not converge, calculate the average value again to find the score value until the result converges. After the bicluster is formed, the iteration process is carried out again as many times as the number of seeds (n). produce bicluster (this process is done with the help of R software).

7. Bicluster Formation

   The matrix element values that have converged are then aggregated into a bicluster.

8. Filter Unique Bicluster

   At this stage, duplicate biclusters are removed (Csardi, 2009). This process uses the help of R software in the 'isa2' package using the isa.unique function.

9. Filter Robustness Bicluster

   At this stage filtering is done based on the robustness value of each bicluster, the bicluster with the lowest robustness value is removed (Csardi, 2009). This process uses the help of R software in the 'isa2' package using the isa.filter.robust function.

Evaluation of Biclustering Results

The results of the biclustering algorithm can be evaluated using the Mean Square Residue (MSR) with the following equation.

$$MSR = \frac{\sum_{i \in I} \sum_{j \in J} \left(b_{ij} - b_{iJ} - b_{Ij} + b_{IJ}\right)^2}{|I| \times |J|} \tag{4}$$

with:

MSR : *Mean Square Residue* (MSR) value

b_{ij} : entry value of i is the number of objects and j is an attribute

b_{iJ} : average entry value of object i

b_{Ij} : average value of entries of attribute j

b_{IJ} : average value of the whole bicluster

$|I|$: number of bicluster rows

$|J|$: number of bicluster columns

Bicluster quality can be obtained by calculating the average value of the MSR to volume ratio, which can be calculated using the following equation (Ningsih et al., 2022).

$$\overline{MSR}\ to\ volume\ ratio = \frac{1}{k} \sum_{l=1}^{k} \frac{MSR_l}{Volume_l} \tag{5}$$

with

MSR_l : MSR value pf lth bicluster
$Volume_l$: size (row × column) of the lth bicluster
k : number of biclusters

3. RESULTS

Data Exploration

In this section, the data exploration is illustrated using a heatmap. The results of the scaling matrix data are shown in Figure 2.

Figure 2. Heatmap of Scaling Data Matrix

Fig. 2 is a heatmap of the scaling matrix data, which is the normalized variable values for each region. The heatmap illustrates some of the extreme values of the poverty data variables in a particular region. The more intense the red color in a region, the higher the variable value in that region. In the PPM, IKPM, IKDM, and TPT variables, regions that are colored solid red have a higher poverty rate than regions that are colored pink. On the other hand, for the variables PDRBHB, RLS, HLS, AMHPM, PRTAL, PRTJ, APSPM, AHH, PPMBSP, PPMUTSMA, PPKMPM, and PRTMMPS, regions that are colored solid red have a lower poverty rate than regions that are colored pink. One example of the heatmap depiction in Figure 4.17 is the TPT variable in Makassar City, which has a solid red color. This shows that the number of unemployed people in the region is high, so the poverty rate in the region is also high.

Biclustering Analysis

The biclustering process using the Iterative Signature Algorithm (ISA) starts with standardization based on column matrix and row matrix. After that, determine the initial value, perform data iteration and bicluster formation. The formed bicluster will go through a filtering process to get the best bicluster. The iteration process requires a threshold parameter and uses direction to determine the direction of the desired bicluster value. The biclustering process is done with the help of R software. The parameter values are determined by the researcher, the threshold parameters used are 0.1, 0.2, and 0.01 in each row and column and the direction "up".

Biclustering Evaluation

Evaluation of biclustering results using Mean Square Residue (MSR) against volume. The evaluation results are shown in Table 2.

Table 2. Evaluation of ISA Biclustering Results Using MSR/V

Column Threshold	Row Threshold	Bicluster	Number of Rows	Number of Column	MSRi	MSRi/Vi	MSR/V
0.01	0.01	1	35	8	0.7291595	0.0026041411	0.002548018
		2	36	6	0.5383414	0.0024426754	
0.1	0.1	1	33	7	0.7216307	0.0031239424	0.0028612148
		2	35	6	0.5456823	0.0025984781	
0.2	0.2	1	30	5	0.485214	0.00323476	0.002941897
		1	35	5	0.463581	0.0026490343	

Based on the results of each threshold parameter, the parameters that give the best results are column threshold 0.01 and row threshold 0.01. The best parameter selection is seen from MSR/V. The greater the MSR/V value, the lower the quality of the biclustering results formed. Therefore, to find the best biclustering result, the MSR/V value should be as low as possible to produce good quality.

Biclustering Results and Interpretation

The results of the biclustering process with a column threshold parameter of 0.01 and a row threshold of 0.01 are shown in Table 3.

Table 3. ISA Biclustering Process Results

Bicluster	Variable	District/City
1	PDRBHB, TPT, RLS, HLS, PRTMAL, PRTJ, AHH, PPMUTSMA	Minahasa, Kepulauan Sangihe, Kepulauan Talaud, Mihanasa Selatan, Minahasa Utara, Bolaang Mongondow Utara, Siau Tagulandang Biaro, Bolaang Mongondow Timur, Manado District, Bitung District, Tomohon District, Kotamobago District, Morowali, Toli, Buol, Kota Palu, Maros, Barru, Soppeng, Sindang Rappang, Pinrang, Luwu, Tana Toraja, Luwu Timur, Makassar District, Parepare District, Palopo District, Konawe Kolaka, Kolaka Timur, Kendari District, Baubau District, Gorontalo District, Mamuju, Mamuju Tengah
2	PPM, IKPM, IKDM, PPMBSP, PPKMPM, PRTMMPS	Minahasa Tenggara, Bolaang Mongondow Selatan, Banggai Kepulauan, Poso, Donggala, Toli, Buol, Parigi Muotong, Tojo Una-Una, Sigi, Banggai Laut, Morowali Utara, Kepulauan Selayar, Jeneponto, Pangkajene dan Kepulauan, Enrekang, Luwu Utara, Konawe, Konawe Selatan, Wakatobi, Kolaka Utara, Kolaka Timur, Buton Utara, Konawe Utara, Konawe Kepulauan, Muna Barat, Buton Tengah, Buton Selatan, Baalemo, Gorontalo, Pahuwato, Bone Bolango, Gorontalo Utara, Majene, Polewali Mandar, Mamasa

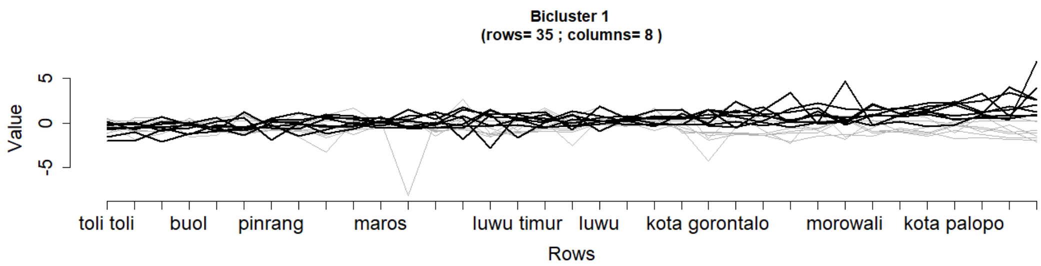
Based on Table 3, there are 2 biclusters from the biclustering process. There are 8 variables in bicluster 1 and 6 variables in bicluster 2. Two other variables are not in bicluster 1 and 2. In addition, there are several districts/cities that fall into bicluster 1 and bicluster 2, namely Toli-Toli District, Buol District, Konawe District, and East Kolaka District. Biclusters that have the same observations or variables mean that there is overlapping between the biclusters, so it can be concluded that there are similarities in the 2 biclusters.

4. DISCUSSIONS

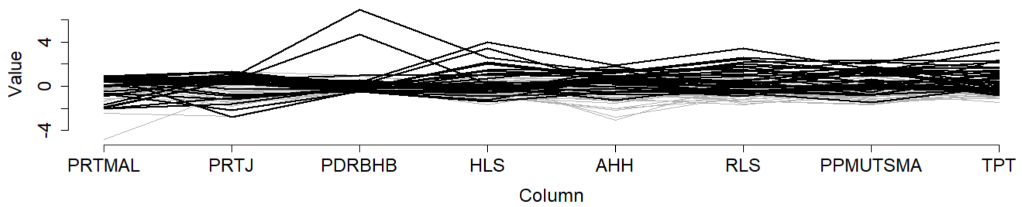
In this section we interpret biclusters based on the heatmap and parallel coordinate Bicluster 1 and Bicluster 2.

The results of the heatmap and parallel coordinates Bicluster 1 are shown in Figure 3.

(a)



(b)

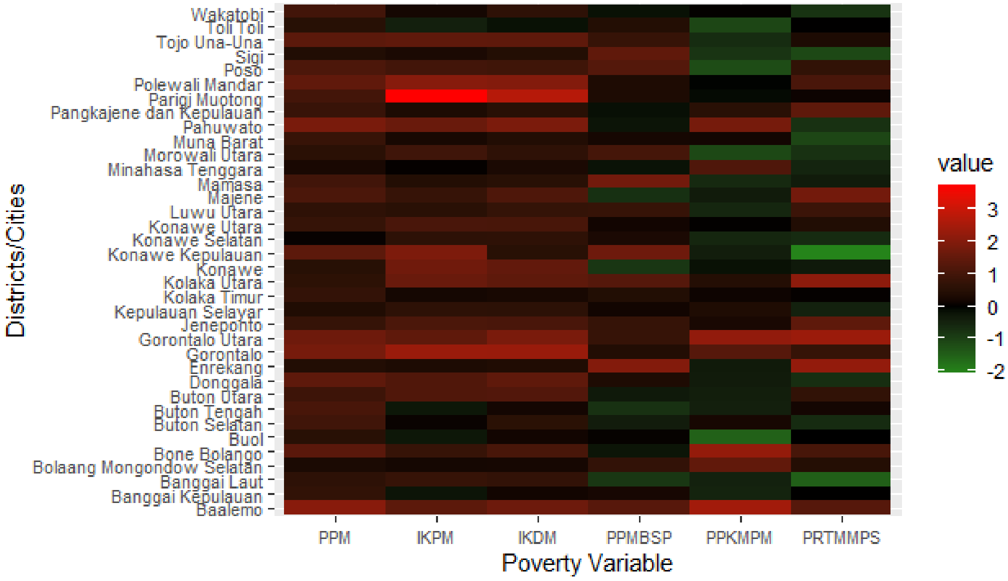


(c)

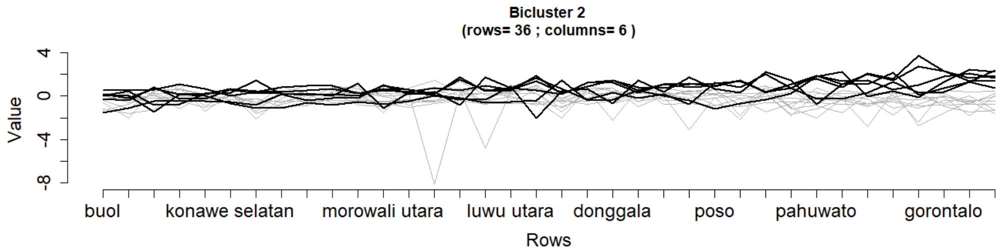
Figure 3. Heatmap and Parallel Coordinate Bicluster 1

In Fig. 3, the 35x8 bicluster heatmap section (35 rows and 8 columns) has the highest and lowest results. The highest values are bright red, and the lowest values are bright green. The Parallel Coordinate section explains that there are extreme points in the rows and columns. Figure 3(c) explains one example, namely the highest value in the 2nd variable, namely the variable GDP at Current Prices in Morowali Regency.

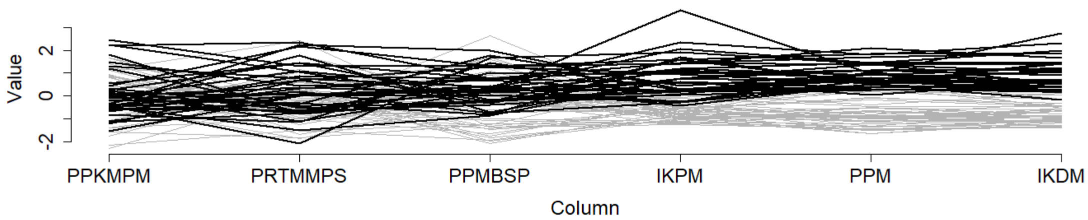
The results of the heatmap and parallel coordinates Bicluster 2 are shown in Figure 4.



(a)



(b)



(c)

Figure 4. Heatmap and Parallel Coordinate Bicluster 2

In Fig. 4, the 36x6 bicluster heatmap section (36 rows and 8 columns) has the highest and lowest results. The highest values are bright red, and the lowest values are green. The Parallel Coordinate section explains that there are extreme points in the rows and columns. Figure 4(c) explains such as the highest value in the 4th variable, the Poverty Severity Index variable in Parigi Muotong District.

The mean and median values of the normalized poverty variables in bicluster 1 and bicluster 2 are shown in Table 4.

Table 4. Mean and Median of Poverty Variables

Variable	Bicluster 1		Bicluster 2		Sulawesi Island	
	Mean	Median	Mean	Median	Mean	Median
PDRBHB	0.2495	-0.1067	-	-	0	-0.3986
TPT	0.6539	0.4220	-	-	0	-0.2947
RLS	0.5567	0.2256	-	-	0	-0.1815
HLS	0.3996	0.0147	-	-	0	-0.1674
PRTMAL	0.0575	0.4473	-	-	0	0.4037
PRTJ	0.2396	0.5147	-	-	0	0.3372
AHH	0.5174	0.6033	-	-	0	0.2946
PPMUTSMA	0.6407	0.6407	-	-	0	-0.2668
PPM	-	-	0.9109	0.8013	0	0.2138
IKPM	-	-	0.8601	0.7792	0	-0.2695
IKDM	-	-	0.9211	0.6727	0	-0.1662
PPMBSP	-	-	0.3896	0.3221	0	0.0121
PPKMPM	-	-	0.0864	-0.1372	0	-0.1937
PRTMMPS	-	-	0.2109	0.0711	0	-0.0035
AMHPM	-	-	-	-	0	0.2149
APSPM	-	-	-	-	0	0.2812

Table 4 shows that the mean and median values of the variables in bicluster 1 and bicluster 2 are greater than the overall mean and median values. Bicluster 1 has a mean value on the variables of GRDP at current prices (PDRBHB), average years of schooling (RLS), expected years of schooling (HLS), percentage of poor households using safe water (PRTMAL), percentage of poor households using their own/shared latrines (PRTJ), life expectancy (AHH), percentage of poor people who graduated from high school (PPMUTSMA) greater than the overall mean value. This means that the population in bicluster 1 has a good fulfillment of life needs, such as education and health. Meanwhile, the average value of the unemployment rate (TPT) is higher than the overall average value. This means that the population in bicluster 1 has a high unemployment rate, which can affect the poverty rate in the area. Bicluster 2 has an average value on the poor working in the agricultural sector (PPMBSP), per capita food expenditure of the poor (PPKMPM), and the percentage of poor households receiving the basic food program (PRTMMPS) higher than the overall average value. This means that the population in the area has good employment and fulfillment of basic needs. Meanwhile, the average value of the percentage of poor people (PPM), the poverty severity index (IKPM), and the poverty depth index (IKDM) is higher than the overall average value. This means that there are still many poor people in bicluster 2 areas.

The resulting biclusters show distinct characteristics. Bicluster 1 includes variables related to education, health, economy, and employment. These variables include average years of schooling, expected years of schooling, the percentage of poor households with high school graduates, health indicators such as the percentage of poor households with access to clean water and sanitation, life expectancy, as well as the Gross Regional Domestic Product (GRDP) at current prices and the open unemployment rate. Bicluster 1 excels in variables such as GRDP at current prices, average years of schooling, expected years of schooling, education, and health indicators, but faces challenges regarding the open unemployment rate, which requires the government's focus on addressing employment issues in the region. On the other hand, Bicluster 2 focuses more on employment, economy, and poverty. The relevant variables include the population working in the agricultural sector, per capita food expenditure, recipients of the food assistance program (sembako), and absolute poverty indicators such as the percentage of the poor population, poverty severity index, and poverty depth index. Bicluster 2 shows strengths in variables like the percentage of the poor working in the agricultural sector, per capita food expenditure for the poor, and the proportion of poor households receiving food assistance. However, Bicluster 2 also faces challenges related to the high percentage of the poor population, as well as the poverty severity and depth indices, which need to be addressed more effectively.

To provide a clearer representation of the bicluster distribution, a combined map of the two biclusters is presented. This map illustrates the spatial arrangement of bicluster 1, bicluster 2, and their overlapping regions.

The results of the combined bicluster distribution map of bicluster 1 and bicluster 2 are shown in Figure 5.

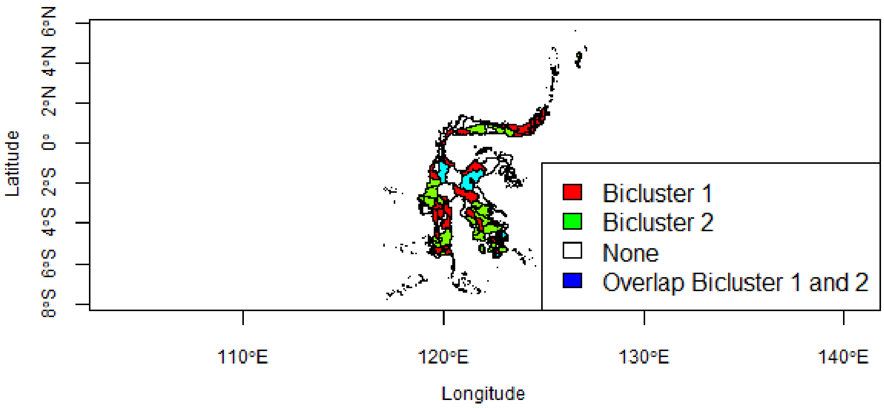


Figure 5. Bicluster Map of Poverty in Sulawesi

In Figure 5, the red color indicates bicluster 1, the green color indicates bicluster 2, the blue color indicates overlapping areas, and the white color indicates no bicluster.

5. CONCLUSION

Based on the results of the research conducted, it can be concluded that clustering districts/municipalities and poverty variables together in Sulawesi Island using the Iterative Signature Algorithm (ISA) produces two biclusters, namely: bicluster 1 that contains 35 districts/cities and 8 variables and bicluster 2 that contains 36 districts/cities and 6 variables. Bicluster 1 has variables on education, economy, and employment. Bicluster 2 has variables on employment, economy, and absolute poverty measurement. In bicluster 1, the government is advised to focus on employment because of the high unemployment rate in bicluster 1. In bicluster 2, the government is advised to focus on the poor because there are still many people whose average per capita expenditure is not below the poverty line in bicluster 2.

6. ACKNOWLEDGMENTS

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.

7. REFERENCES

[1] G. A. Aziz, E. Rochaida, and Wardilan, "Faktor-Faktor yang Mempengaruhi Kemiskinan di Kabupaten Kutai Kartanegara," *Jurnal Ekonomi Keuangan dan Manjemen*, vol. 12, no. 1, pp. 29-48, 2016.
[2] BPS RI, "Data dan Informasi Kemiskinan Kabupaten/Kota di Indonesia," 2022.
[3] BPS RI, "Perhitungan dan Analisis Kemiskinan Makro Indonesia Tahun 2022," 2022.
[4] S. Bergmann, J. Ihmels, and N. Barkai, "Iterative signature algorithm for the analysis of large-scale gene expression data," *Physical Review E*, vol. 67, no. 3, 2003.
[5] G. Csardi, "The Iterative Signature Algorithm," 2009.
[6] G. Csardi, Z. Kutalik, and S. Bergmann, "Modular analysis of gene expression data with R," *Bioinformatics Aplication Note*, vol. 26, no. 10, pp. 1376-1377, 2010.
[7] N. Dewi, "Pengaruh Kemiskinan dan Pertumbuhan Ekonomi Terhadap Indeks Pembangunan Manusia di Provinsi Riau," *JOM Fekon*, vol. 4, no. 1, pp. 870-882, 2017.

[8] A. Freitas, V. Afreixo, M. Pinheiro, J. L. Oliveira, G. Moura, and M. Santos, "Improving the Performance of the Iterative Signature Algorithm for the Identification of Relevant Pattern," *Statistical Analysis and Data Mining*, vol. 4, pp. 71-83, 2011.

[9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann Publisher, USA, 2012.

[10] J. Ihmels, G. Friedlander, S. Bergmann, O. Sarig, O. Ziv, and N. Barkai, "Revealing Modular Organization in the Yeast Transcriptional Network," *Nature Publishing Group*, vol. 31, pp. 370-377, 2002.

[11] S. Kaiser, "Biclustering: Methods, Software and Application," Ph.D. dissertation, 2011.

[12] F. Mahmud, F. Z. Olilingo, and F. H. Y. Akib, "Faktor-Faktor yang Mempengaruhi Kemiskinan di Pulau Sulawesi," *Jurnal Kajian Ekonomi dan Bisnis*, vol. 13, no. 2, pp. 130-147, 2020.

[13] W. A. L. Ningsih, I. M. Sumertajaya, and A. Saefuddin, "Pattern Detection of Economic and Pandemic Vulnerability Index in Indonesia Using Bi-Cluster Analysis," *JUITA: Jurnal Informatika*, vol. 10, no. 2, pp. 273-282, 2022. [Online]. Available: https://doi.org/10.30595/juita.v10i2.14940

[14] C. A. Putri, R. Irfani, and B. Sartono, "Recognizing Poverty Pattern in Central Java Using Biclustering Analysis," *Journal of Physics: Conference Series*, vol. 1863, no. 1, pp. 1-8, 2021. [Online]. Available: https://doi.org/10.1088/1742-6596/1863/1/012068

[15] E. S. Safitri, N. Triwahyuningtyas, and Sugianto, "Analisis Faktor-Faktor yang Mempengaruhi Tingkat Kemiskinan di Provinsi Banten," *Jurnal Ilmiah Bidang Sosial, Ekonomi, Budaya, Teknologi, dan Pendidikan*, vol. 1, no. 4, pp. 259-274, 2022.

[16] I. M. Sumertajaya, W. A. L. Ningsih, A. Saefuddin, and E. Rohaeti, "Biclustering Performance Evaluation of Cheng and Church Algorithm and Iterative Signature Algorithm," *JTAM: Jurnal Teori dan Aplikasi Matematika*, vol. 7, no. 3, pp. 643-661, 2023.

[17] L. Zhang, S. Chen, J. Zhu, J. Meng, and H. Liu, "REW-ISA: Unveiling Local Functional Blocks in Epi-Transcriptome Profiling Data," *Frontiers in Genetics*, vol. 12, no. 5, pp. 1-10, 2020. [Online]. Available: https://doi.org/10.3389/fgene.2021.654820