JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 8, No. 4, October 2024, pp. 1262-1277

Identifying Poverty Vulnerability Patterns in Indonesia using Cheng and Chruch's Algorithm

Irsyifa Mayzela Afnan[1], Hari Wijayanto[1], Aji Hamim Wigena[1]
[1]Department of Statistics/Statistics and Data Science, IPB University, Indonesia
irsyifamayzela@apps.ipb.ac.id

Article History:
Received : 30-07-2024
Revised : 16-10-2024
Accepted : 19-10-2024
Online : 23-10-2024

Keywords:
Biclustering;
Cheng and Chruch;
Poverty;
MSR

ABSTRACT

Poverty remains a significant issue in developing countries, including Indonesia, where in 2022, the number of people living in poverty reached 26.36 million, with a poverty rate of 9.57%. The Central Statistics Agency (BPS) measures poverty using a basic needs approach, defined as the inability to meet essential food and non-food needs through expenditure. Individuals are considered poor if their average monthly per capita expenditure is below the poverty line. Research on poverty has evolved into a more multidimensional understanding, The Multidimensional Poverty Index (MPI), which identifies deprivation across three key dimensions: health, education, and living standards. This study aims to identify patterns of poverty vulnerability by applying the Cheng and Church (CC) algorithm through a biclustering approach using data from BPS. This quantitative method utilizes 13 multidimensional poverty indicators across 34 provinces. The CC algorithm begins by setting a threshold, followed by removing rows and columns with the largest residuals, adding qualifying rows and columns, and substituting elements to prevent overlap. The quality of the bicluster is then evaluated based on the Mean Squared Residue (MSR) value until optimal groups are formed. The results indicate that a threshold of δ = 0.01 generates seven biclusters with the lowest mean squared residual (0.0065), signifying optimal bicluster quality. Further validation using the Liu and Wang index reveals less than 50% similarity with other thresholds, reinforcing the uniqueness of these findings. MSR serves as a measure of homogeneity within the bicluster, similar to how uniform the level of poverty is within a region. If families have similar expenditures and are below the poverty line, they face similar challenges, resulting in a low MSR value. In contrast, the Liu and Wang index compares regions with different poverty alleviation strategies. These findings provide valuable insights for policymakers. For example, in bicluster 7, where specific interventions are needed in Papua and West Kalimantan, which face local challenges such as reliance on agriculture, low education levels, and limited access to sanitation and clean water.

Crossref
https://doi.org/10.31764/jtam.v8i4.25790

A. INTRODUCTION

Proper exploration of data can yield valuable and significant insights. How data is explored and patterns are identified can greatly influence the diversity of the resulting information. One method for finding patterns in data is clustering, also known as classical cluster analysis. By using the similarities between variables (columns), clustering divides a collection of items (rows) into uniform groupings (Castanho et al., 2022). In this method, all variables are considered in clustering objects, and vice versa. However, this method has limitations as it only produces a general (global) model and cannot identify specific (local) models of phenomena occurring at the unit level of observation (Kaban et al., 2019).

To overcome these limitations, there is a need for techniques that can identify subsets of data with similar behavior patterns under specific conditions (Padilha & Campello, 2017). One emerging solution is the biclustering algorithm, which has proven to be more effective in revealing hidden structures in data. Biclustering is a data mining technique that allows for the simultaneous clustering of rows and columns from a matrix (Xie et al., 2019). For illustration, imagine grouping students in a school based on their hobbies and academic grades. In traditional clustering methods, we only consider the students' grades. However, with biclustering, we can group students based on a combination of hobbies and grades, enabling us to identify groups of students who may have similarities in both aspects. Thus, biclustering not only reveals homogeneous clusters of objects and variables within large and complex datasets (Helgeson et al., 2020), but also enhances our understanding of local characteristics and identifies consistent patterns in rows and columns (Acharya et al., 2019). However, it is important to note that biclustering can also produce overlapping biclusters and unclustered samples (Padilha & Campello, 2017), similar to a group of students who may belong to more than one group based on their various hobbies and grades.

Initially, biclustering was used to address issues in bioinformatics, such as analyzing gene expression microarrays (Baruah et al., 2024; Biswal et al., 2022; Maâtouk et al., 2021; Padilha & Campello, 2017; Pontes et al., 2015; Xie et al., 2019). Microarray analysis, a widely used approach, allows researchers to analyze the activities of many genes under many conditions (Charfaoui et al., 2024). However, research in other fields has also been conducted, such as social vulnerability (Kaban et al., 2019), market analysis (Wang et al., 2016), water consumption (Silva et al., 2022), and macroeconomics (Pang, 2022). Several biclustering algorithms have been developed, including Cheng and Church (CC), Bimax, BCBimax, ISA, Plaid Model, and various others. However, there are no specific guidelines regarding the optimal choice of biclustering algorithms for certain data criteria (Wang et al., 2016). Henriques et al. (2015) selected what they believed to be the most advanced biclustering algorithms. Padilha & Campello (2017) based their algorithm selection on criteria such as popularity and implementation availability. Therefore, this study will explore the CC algorithm, considering three criteria: the algorithm's popularity in biclustering, implementation availability, and computational efficiency.

Pontes et al. (2015) state that the CC algorithm seeks to identify maximal biclusters with a high degree of similarity. This algorithm also successfully avoids overlap between biclusters. The goal of this algorithm is to find biclusters from subgroups of rows and columns with a mean squared residue (MSR) value that is less than or not greater than a threshold (δ). This research focuses on applying the CC algorithm to poverty data. Poverty remains a major issue for developing countries, including Indonesia. By September 2022, the number of impoverished individuals in Indonesia had risen to 26.36 million, with a poverty rate of 9.57%. This marked a 0.03 percentage point increase from March 2022 and a 0.14 percentage point decrease from September 2021. According to BPS as of September 2022, the poor population is still concentrated in Java, while only Sumatera shows a decline (BPS, 2022). BPS measures poverty using the basic needs method. Poverty is defined as the inability to meet basic food and non-food needs, as measured by spending. Individuals classified as poor are those whose average monthly per capita spending is less than the poverty line.

Research on poverty has significantly evolved, leading to an understanding of poverty in multiple dimensions (Putri et al., 2021). The significance of this research lies in its potential to provide a deeper understanding of poverty vulnerability in Indonesia, a critical issue for social and economic development. By identifying groups of provinces that face similar poverty challenges, policies can be designed to be more targeted and effective. However, measuring poverty in Indonesia is complex. Many families live in conditions that traditional measurement methods, which often focus solely on income, fail to detect. Factors such as health conditions, access to education, and basic services also play crucial roles in determining poverty status but are frequently overlooked.

Therefore, this study aims to fill the knowledge gap concerning how poverty vulnerability manifests at the provincial level. This research is particularly relevant given the significant shift in the understanding of poverty, which is now defined multidimensionally. The Multidimensional Poverty Index (MPI) is a technique for measuring poverty that identifies deprivations across multiple dimensions and displays the proportion of individuals who are multidimensionally poor. The MPI describes three dimensions of poverty for Indonesian households: health, education, and living standards. The variables used in this study are MPI variables that have been modified by Yuniarto & Kurniawan (2017), including variables for absolute poverty measurement, health dimensions, human resources, and living standards

The application of the CC algorithm will identify subgroups within the data that exhibit similar characteristics of poverty. The results of this analysis can provide important insights for public policy decision-making. For example, by understanding the patterns of poverty vulnerability across various provinces, policymakers can design more targeted intervention programs, such as better resource allocation or the provision of educational and health services that are more aligned with the needs of the region. This analysis can also assist in prioritizing areas that require more attention, allowing for more focused and efficient actions to reduce poverty. For instance, if a province is known to have many families with low education and limited access to healthcare services, interventions aimed at improving education and health can be planned. Thus, the application of biclustering analysis in the context of poverty is expected to enhance understanding of the dynamics of vulnerability and assist in formulating more effective anti-poverty policies at the provincial level.

B. METHODS

1. Data

This research employs a quantitative approach, utilizing secondary data obtained from the National Socioeconomic Survey (SUSENAS) conducted by the Central Statistics Agency (BPS) as of March 2022. The data collection process was carried out by trained BPS field officers through face-to-face interviews with selected households using structured questionnaires. The collected data covers various aspects of social, economic, and demographic conditions, including household consumption and expenditure, health and education status, access to basic services such as clean water, electricity, and sanitation, as well as housing conditions. The analysis focuses on 34 provinces in Indonesia, using these provinces as units of observation. To assess various aspects of poverty, we incorporate 13 multidimensional poverty indicators, which encompass dimensions such as absolute poverty, health, human resources, and living standards. These indicators are derived from the structural poverty model proposed by Yuniarto & Kurniawan (2017), as outlined in Table 1.

Table 1. Research Variables

Dimension	Variable
Poverty	Headcount index (X1)
	Poverty depth index (X2)
	Poverty severity index (X3)
Economic	The percentage of poor individuals aged 15 and older who are unemployed (X4)
	The percentage of poor individuals aged 15 and older working in the agricultural sector(X5)
	The percentage of percapita expenditure on food (X6)
Human Resources	The percentage of poor individuals aged 15 and older who did not complete elementary school(X7)
	The literacy rate of poor individuals aged 15-55 years old^{-1} (X8)
	The average years of schooling^{-1} (X9)
Health	The percentage of women using contraceptives in poor households^{-1} (X10)
	The percentage of poor households using clean water sources for drinking^{-1} (X11)
	The percentage of poor households with own/shared latrine^{-1} (X12)
	Life expectancy^{-1} (X13)

$^{-1}$Invers value

The inverse transformation applied to variables X8, X9, X10, X11, X12, and X13 adjusts the direction of these variables in the context of poverty vulnerability. Increasing values signify higher levels of poverty vulnerability, while decreasing values signify lower levels.

2. Research Stages

a. Data Preparation
First, the direction of variables X8 to X13 is reversed by multiplying them by -1. All variables and objects are then mapped into a dataset matrix. Subsequently, data standardization (scaling) is carried out.

b. Data Exploration
The standardized (scaled) data are explored using heatmaps. A heatmap is a popular method for statistical data visualization that uses different color gradients to represent matrix data. The clustering heatmap is designed to group random samples based on their underlying distributions (Zhang & Shi, 2024). In the heatmap, rows and columns with similar patterns are clustered closely together. (Gu, 2022).

c. Biclustering Analysis
The CC algorithm is used to identify maximal biclusters with high similarity. The objective of this algorithm is to find biclusters from subsets of rows and columns with MSR values less than or not greater than a specified threshold (δ) (Pontes et al. 2015). Kaban et al. (2019) mentioned that a bicluster is considered good if its MSR is below the defined threshold, necessitating the determination of the δ parameter. The submatrix

A_{IJ} is identified as a δ-bicluster if $MSR(I, J) < \delta$ for $\delta \geq 0$, the value of MSR is defined by equation (1) (López-Fernández et al., 2020):

$$MSR(I,J) = \frac{1}{|I||J|}\sum_{i\in I,j\in J}\left(a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}\right)^2 \quad (1)$$

$$a_{iJ} = \frac{1}{|J|}\sum_{j\in J} a_{ij} \quad (2)$$

$$a_{Ij} = \frac{1}{|I|}\sum_{i\in I} a_{ij} \quad (3)$$

$$a_{IJ} = \frac{1}{|I||J|}\sum_{i\in I,j\in J} a_{ij} \quad (4)$$

The notation used includes a_{ij} for the object at row i and column j, I and J as the sets of rows and columns in bicluster A_{IJ}. Meanwhile, $d(i)$ and $d(j)$ represent the average squared residuals for row i and column j, respectively, as calculated by Equations (5) and (6).

$$d(i) = \frac{1}{|J|}\sum_{j=1}^{|J|}\left(a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}\right)^2 \quad (5)$$

$$d(j) = \frac{1}{|I|}\sum_{j=1}^{|I|}\left(a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}\right)^2 \quad (6)$$

Generally, Figure 1 illustrates three phases: deletion (single and multiple node deletion), addition (node addition), and substitution (Pontes et al., 2015).

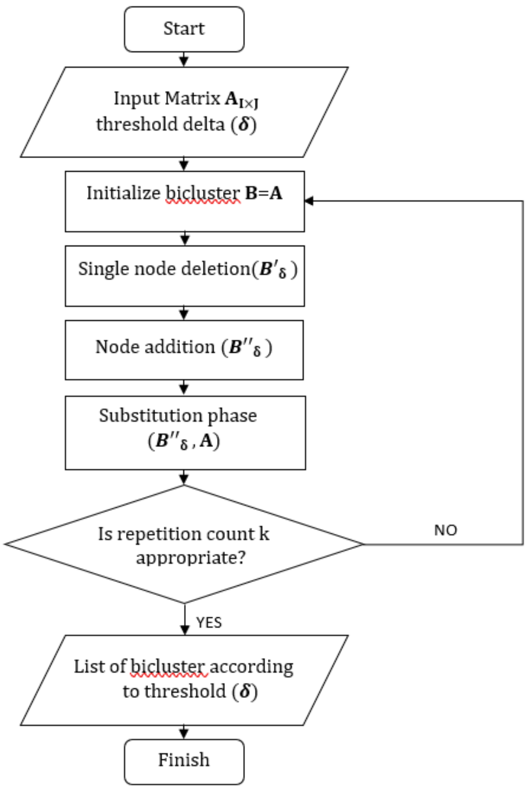

Figure 1. CC Algorithm Flowchart (modified from Sumertajaya et al. (2023)

The following are descriptions of each step in Figure 1 (Pauk & Minta-Bielecka, 2016):

1. Initialization of biclusters; set the threshold delta (δ) based on the MSR obtained from the initial scaled matrix.
2. Single node deletion; If $MS(I, J) > \delta$, the rows or columns with the highest average squared residuals ($d(i)$) or ($d(j)$) will be removed. Since the data utilized in this study has fewer rows or columns than 100 (< 100), the process of multiple node deletion is not carried out.
3. Node addition; this process aims to maximize the bicluster volume by adding rows and columns based on the criteria ($d(i)$) ≤ $MSR(I, J)$ and ($d(j)$) ≤ $MSR(I, J)$, while ensuring the condition $MSR(I', J') \leq MSR(I, J)$.
4. Substitution; replace the elements of the resulting bicluster matrix with random numbers to avoid overlapping among biclusters.
5. Steps 1 through 4 should be repeated until the desired k biclusters are found.

d. Bicluster Evaluation and Selection

The performance of a bicluster can be evaluated by examining its MSR value (Huang et al., 2020). The quality of the bicluster is considered better when the MSR value is smaller. When there are multiple biclusters, performance can be calculated using the average MSR value. The average MSR is an intra-bicluster evaluation function used to measure coherence within a bicluster (Kavitha Sri & Porkodi, 2019). This approach allows a comprehensive assessment of the uniformity and quality of all biclusters produced in a dataset. The average MSR value can be calculated using equation (7) as follows (Pontes et al., 2015):

$$AvMSR = \frac{1}{k}\sum_{i=1}^{k} MSR_i(I,J) \tag{7}$$

In the meantime, two biclusters' similarity can be compared using the Liu and Wang Index, which is used as an inter-bicluster evaluation function. The Liu and Wang Index is defined in equation (8) as follows (Kavitha Sri & Porkodi, 2019):

$$I_{Liu\&Wang}(M_{opt}, M) = \frac{1}{k_{opt}}\sum_{i=1}^{k_{opt}} \max\left(\frac{|R_i \cap R_j| + |C_i \cap C_j|}{|R_i \cup R_j| + |C_i \cup C_j|}\right) \tag{8}$$

where M_{opt} is the set of biclusters with the smallest average value of mean square residue and M is another set of biclusters, k_{opt} is the number of biclusters in M_{opt}, $|R_i \cap R_j|$ is the number of rows (R) in M_{opt} that overlap with rows in M, $|C_i \cap C_j|$ is the number of columns (C) in M_{opt} that overlap with columns in M, $|R_i \cup R_j|$ is the total number of combined rows in M_{opt} and M, $|C_i \cup C_j|$ is the total number of combined columns in M_{opt} and M. The smaller the Liu and Wang Index value, the more different the characteristics of the bicluster members.

C. RESULT AND DISCUSSION

1. Data Exploration

The initial characteristics of each region based on poverty vulnerability indicators from 34 provinces in Indonesia are illustrated through a heatmap of scaled data matrices in Figure 2.

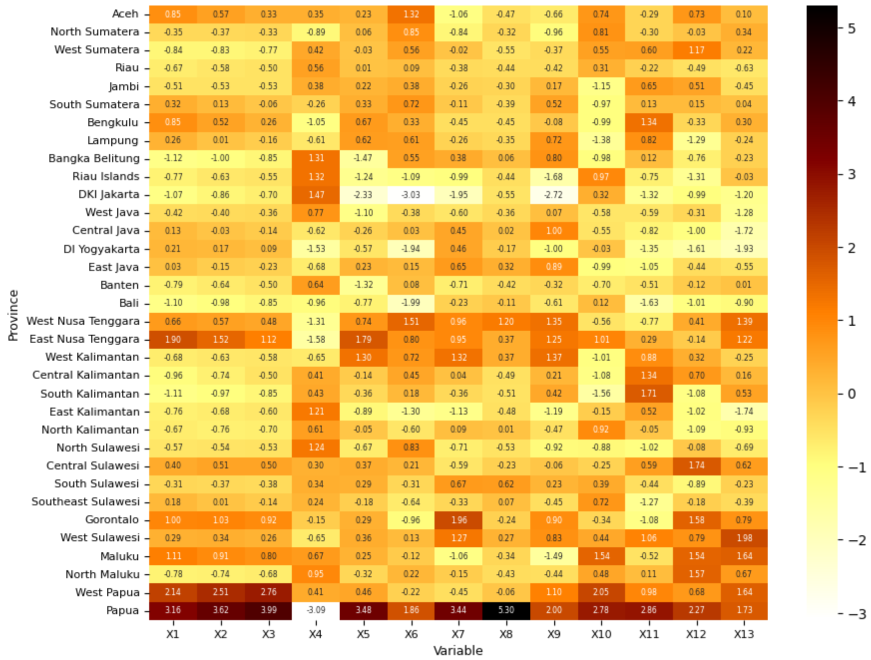

Figure 2. Heatmap of Scaled Data Matrix

In Figure 2, it is evident that some indicators and provinces show extreme values, both very high and very low. Dark colors (black) represent positive extremes, indicating high vulnerability, while light colors (white) represent negative extremes, indicating low vulnerability. The provinces with positive extreme values on certain indicators indicate significant vulnerability in those aspects, whereas negative extreme values indicate relatively better resilience. Through the heatmap, there are indications of outliers in the data matrix, as evidenced by the presence of extreme data points. For example, Papua Province shows high poverty vulnerability on indicator $X8$, which is the literacy rate among poor residents aged 15-55, reflecting serious educational challenges among the poor population. On the other hand, Papua Province shows low vulnerability for indicator $X4$, indicating better employment conditions among the poor population.

2. Optimal Threshold Selection

Setting the threshold δ is an important first step in the CC algorithm. The scaling data matrix's MSR value is used to calculate the threshold δ, and values below the MSR are used to test thresholds. In this study, the MSR value is 0.5787, so the tested thresholds range from 0.01 to 0.56 in increments of 0.01, resulting in a total of 56 tested thresholds to find the best or optimal threshold based on the smallest average MSR value. Figure 3 shows the distribution of Average MSR values obtained from the various-tested δ thresholds.

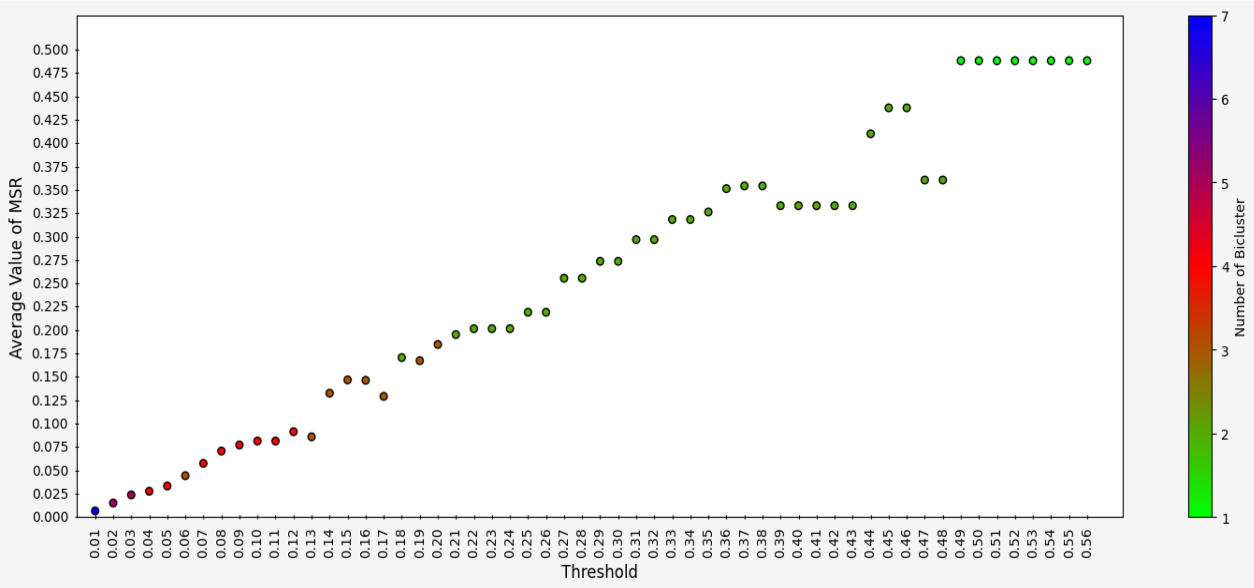

Figure 3. The Average Value of MSR Based on The Tested Delta Thresholds

Figure 3 illustrates the results of testing 56 thresholds. Among these, a threshold of 0.01 yielded the smallest average MSR value of 0.0065, resulting in seven biclusters. Therefore, a threshold of 0.01 is considered as the optimal threshold for biclustering using the CC algorithm. Generally, Figure 3 shows that as the δ threshold increases, the average MSR value tends to become higher and the number of biclusters decreases. A higher average MSR value indicates that the quality of the biclusters is lower, whereas a lower average MSR value indicates higher quality. For tuning parameter settings where δ exceeds 0.48, only one bicluster is formed, which does not provide meaningful results. It is preferable to select a tuning parameter that produces at least two biclusters. The best bicluster is one with the smallest average MSR value and high intra-bicluster similarity. Therefore, a threshold of 0.01 is deemed the most optimal.

Although a threshold of 0.01 is identified as optimal, testing various thresholds validates the model's performance. The Liu-Wang index analysis in biclustering evaluates the quality and similarity of biclusters at different thresholds. Figure 4 shows that a threshold of 0.01 has the maximum Liu and Wang index value of 1 because biclusters formed from the same threshold have perfect similarity. Therefore, the 0.01 threshold should be compared with biclusters from other thresholds for meaningful evaluation. Index values at other thresholds are below 50%, indicating low similarity with the optimal bicluster. This suggests that the quality of biclusters significantly decreases when the threshold differs from 0.01. This analysis is crucial to ensure that the 0.01 threshold is the right choice.

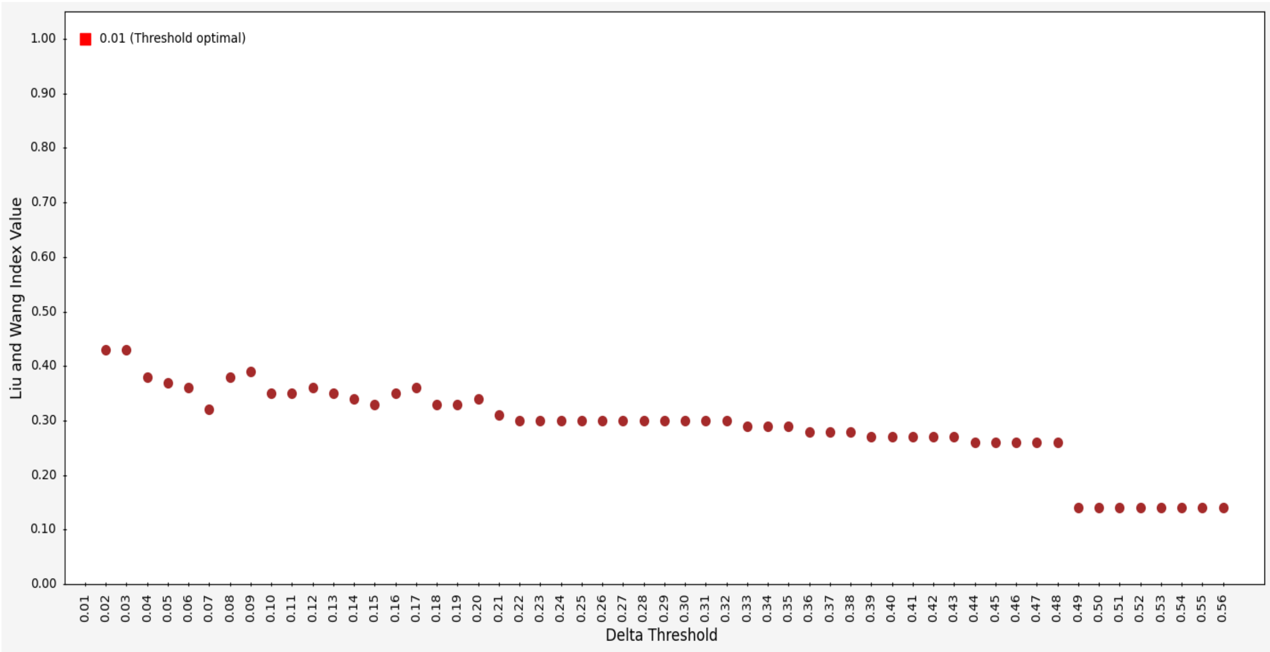

Figure 4. Distribution of Liu and Wang Index Data Based on Delta Thresholds

Based on Table 2, variables X1 (headcount index), X2 (poverty depth index), and X3 (poverty severity index) tend to serve as global descriptors, appearing across nearly all biclusters. This indicates that these variables play a crucial role and have global relevance in explaining poverty patterns across various biclusters. Meanwhile, variables X4 (the percentage of poor individuals aged 15 and older who are unemployed), X6 (the percentage of percapita expenditure on food), and X9 (average years of schooling) do not characterize every bicluster. This suggests that these variables are not significant in differentiating poverty patterns among groups. Additionally, there are non-overlapping variables from the human resources and health dimensions such as X7, X10, X11, and X12. This indicates the uniqueness of these variables for specific biclusters, possibly reflecting the unique characteristics of each specific bicluster in the context of poverty.

Table 2. Membership of Selected CC Biclustering Results

Bicluster	size	Province	Variable
1	7 × 5	North Sumatera, West Sumatera , Riau, Jambi, Central Kalimantan, South Kalimantan, North Maluku	X1, X2, X3, X5, X8
2	9 × 4	Riau Islands, West Java, Central Java, East Java, Banten, East Kalimantan, North Sulawesi, Southeast Sulawesi, West Sulawesi	X1, X2, X3, X8
3	6 × 3	South Sumatera, Bengkulu, Lampung, East Nusa Tenggara, North Kalimantan, South Sulawesi	X2, X3, X13
4	5 × 3	Bangka Belitung Islands, DKI Jakarta, Bali, West Nusa tenggara, Gorontalo	X1, X2, X3
5	3 × 4	Aceh, DI Yogyakarta, Maluku	X1, X2, X3, X5
6	2 × 3	Central Sulawesi, West Papua	X2, X3, X10
7	2 × 5	West Kalimantan, Papua	X5, X7, X11, X12, X13

Furthermore, bicluster profiles were created through visualization plots to further understand the characteristics of each bicluster and identify emerging patterns.

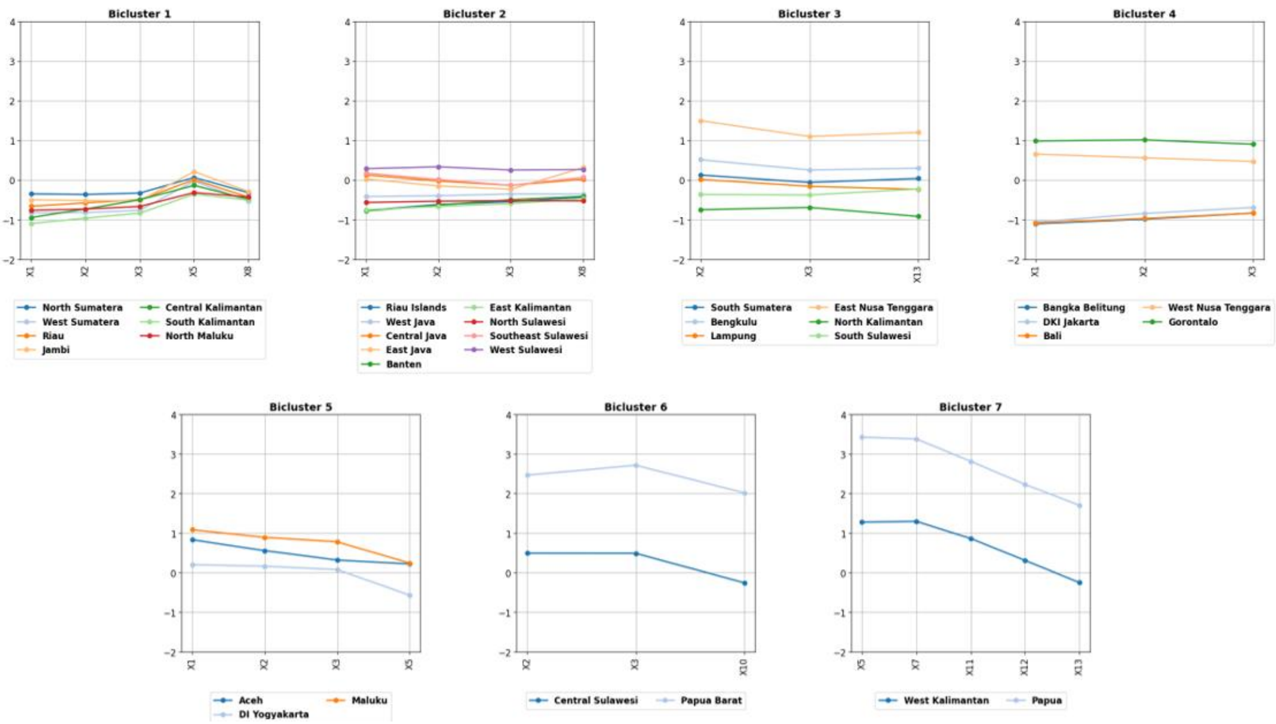

Figure 5. Profiling Plot of Each Bicluster

Line plots are used to observe changes in variable values in biclusters, as illustrated in Figure 5. The profiles in proximity indicate that the bicluster is more homogeneous. Based on Figure 5, the profiles of Bicluster 1 and 2 show more homogeneous patterns, marked by closely aligned lines. Bicluster 1, consisting of North Sumatera, West Sumatera, Riau, Jambi, Central Kalimantan, South Kalimantan, and North Maluku provinces, exhibit relatively high stability in poverty indicators. Variables such as headcount index (X1), poverty depth index (X2), and poverty severity index (X3) tend to be lower compared to other biclusters, reflecting effective poverty management. Although the poor population in these provinces still relies on the agricultural sector (X5), it does not significantly affect the poverty rate. However, this variable remains a challenge that needs attention. Meanwhile, the poor individuals in this age group have high literacy rates, which can help them secure decent formal jobs and higher incomes.

Bicluster 2 shows varying levels of vulnerability among its provinces. The provinces of Riau Islands, West Java, Banten, East Kalimantan, and North Sulawesi show relatively low values across all defining variables, namely X1 (headcount index), X2 (poverty depth index), X3 (poverty severity index), and X8 (the literacy rate of poor individuals aged 15-55 years old). This suggests that these provinces have lower vulnerability compared to others within the same bicluster. The low poverty depth and severity indicators indicate better economic stability. This means that not only are poor communities slightly below the poverty line, but there is also no significant inequality among them. The moderate literacy rate is categorized as low poverty vulnerability or can be interpreted that these provinces have high literacy rates. This suggests widespread and equitable access to basic education. Conversely, Central Java, East Java, Southeast Sulawesi, and West Sulawesi show more moderate levels on these variables, indicating more significant challenges in addressing poverty vulnerability.

Bicluster 3 demonstrates varied levels of vulnerability among its provinces, characterized by variables X2 (poverty depth index), X3 (poverty severity index), and X13 (life expectancy).

East Nusa Tenggara (NTT) stands out with high vulnerability levels in all these variables, indicating significant difficulties in poverty depth, severity, and lower life expectancy. This may be due to limited access to healthcare, education, and inadequate basic infrastructure in the region. Conversely, South Sumatera, Bengkulu, and Lampung show more moderate vulnerability conditions, indicating relative stability and better conditions, although not excellent. Meanwhile, North Kalimantan and South Sulawesi show low vulnerability in all three variables, although South Sulawesi still has a moderate life expectancy. This may be due to factors such as the availability and quality of healthcare services still needing improvement.

Bicluster 4 consists of the Bangka Belitung Islands Province, DKI Jakarta, Bali, West Nusa Tenggara (NTB), and Gorontalo. Bangka Belitung, DKI Jakarta, and Bali have low vulnerability in absolute poverty indicators, namely X1 (headcount index), X2 (poverty depth index), and X3 (poverty severity index). This is because these provinces are areas with relatively high economic growth. They have more diverse economic sectors, including tourism, trade, and services, providing better and stable job opportunities. Strong economic growth helps reduce poverty rates and increase per capita income. Proper investments in education, health, and social infrastructure also help reduce inequality and improve the quality of life. Meanwhile, NTB and Gorontalo still rely on vulnerable traditional sectors, face difficult geographical conditions, and encounter challenges in investment and balanced economic development.

Bicluster 5 consists of Aceh, Yogyakarta Special Region, and Aceh and Maluku provinces. Although facing similar poverty challenges, they have unique characteristics that affect the level and depth of poverty. The graph shows that Maluku is more affected by poverty compared to Aceh, in terms of headcount index (X1), poverty depth index (X2) and severity (X3), and both regions still rely on the agricultural sector (X5). Both regions lack economic diversification. Many other sectors such as manufacturing and services have not developed well, so agriculture remains the primary source of income for many residents. Geographic and climatic conditions also support agricultural activities in these regions. Yogyakarta, while still needing attention, shows relatively better conditions compared to Maluku or Aceh.

Bicluster 6 consists of Central Sulawesi, and West Papua provinces, both showing high vulnerability levels to poverty in several variables. West Papua stands out with very high values in X2, X3, and X10, indicating serious challenges in addressing poverty in terms of depth, severity, and demographic aspects related to family planning policies. On the other hand, Central Sulawesi, although relatively lower in these variables, including X10 indicating a lower percentage of women using family planning tools in households, is still considered a province highly vulnerable to poverty based on a 95% confidence interval. Despite differences in defining variable values, both face similar challenges in reducing significant poverty levels. Many areas in Central Sulawesi and West Papua are remote and difficult to access, causing economic and social isolation. Difficult-to-reach geographic conditions also hinder the distribution of services and assistance, contributing to high poverty depth and severity. Meanwhile, the percentage of women using family planning tools in households can be influenced by education levels, access to possibly ineffective health services, or inadequate reach to remote areas, resulting in low contraceptive use among women in both regions.

Bicluster 7 consists of West Kalimantan, and Papua provinces, which face high vulnerability across several indicators such as X5, X7, X11, X12, and X13. These provinces heavily rely on agriculture, fisheries, and forestry sectors, which provide lower and unstable incomes compared to the industrial and service sectors due to limited economic diversification. Remote areas experience challenges in accessing basic education, hindering children's completion of education. Low education levels impact job opportunities and living standards. Geographic challenges like dense forests complicate the development of clean water infrastructure. Many poor households lack adequate sanitation facilities, contributing to health issues. Access to adequate healthcare remains a major challenge, resulting in lower life expectancy in these regions.

In general, the biclustering results show a similarity in vulnerability patterns within each group, with biclusters 1, 2, and 3 being more homogeneous. However, upon closer examination, there are biclusters that, despite having similar patterns, have significantly different values for each of their distinguishing variables, such as biclusters 6 and 7, which appear heterogeneous. This issue may arise from employing a threshold to reject solutions, which poses a limitation because it is contingent on the specific dataset and must be determined prior to applying the algorithm. This makes it harder to discover an ideal solution and may have an impact on the algorithm's overall performance. Overall, the level of each variable can be obtained by categorizing the mean value of each variable. This level can be identified using a 95% confidence interval to determine the vulnerability level of each variable per bicluster. Table 3 shows the poverty vulnerability level of each bicluster.

Table 3. The Level of Average Value of Poverty Vulnerability Variables for Each Bicluster

Bicluster	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12	X13
1	L	L	L		M			L					
2	M	M	M					M					
3		M	M										M
4	M	M	M										
5	H	H	H		M								
6		H	H							H			
7					H		H				H	H	H

L : Low , M : Moderate, H : High

Based on Table 3, it can be seen how the variables are distributed across each bicluster, allowing for the identification of the most vulnerable groups and determining the key factors contributing to this vulnerability. Groups highly vulnerable to poverty should be the top priority due to facing the greatest risks and severe impacts of poverty, such as those in biclusters 5, 6, and 7. Strategies or actions taken to address poverty vulnerability should be based on the variables that have a significant impact in each bicluster. However, biclusters with moderate and low levels of vulnerability should not be overlooked. Furthermore, observe how the provinces and variables are distributed within each group (bicluster) and quadrant of the biplot. The biplot is presented in Figure 6.

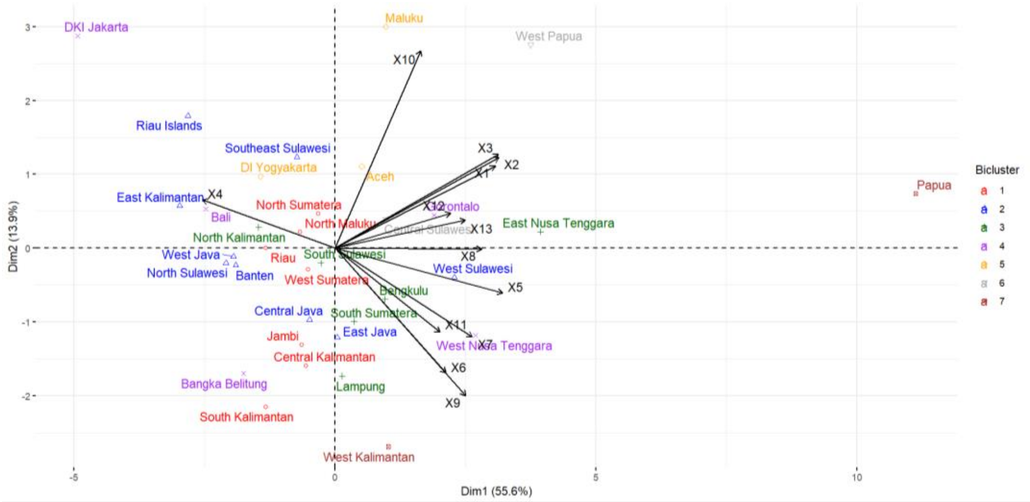

Figure 6. Distribution Bicluster in Biplot

Biclustering provides a specific view of pattern similarities between provinces and a set of certain variables, highlighting groups of provinces with similar characteristics. Biplot PCA offers an overview of the total variation in the data, showing how provinces and variables are distributed in the main two-dimensional space. Figure 6 explains a total variability of 69.5%. The first dimension accounts for 55.6% of the data variability, while the second dimension accounts for only 13.9%. Quadrant 1 includes variables such as X1, X2, X3, X10, X12, and X13, indicating the relevance of these variables for provinces in this quadrant. The same variables, including X2, X3, and X10, also appear in bicluster 6. This suggests that these variables have a significant impact in this quadrant, particularly for the provinces of Central Sulawesi and West Papua. Quadrant 2 of the biplot study suggests that provinces such as DKI Jakarta and Riau Islands are in a position where variable X4 could have a substantial impact. X4 represents the percentage of impoverished people who are not working, which means they are unemployed, looking for employment, or are not part of the labor force (BPS, 2022).

However, in the bicluster analysis, variable X4 does not characterize any bicluster. Several reasons can explain this discrepancy. First, the variable X4 shows high variation among the provinces included in different biclusters. Although X4 is significant for DKI Jakarta (Bicluster 4) and the Riau Islands (Bicluster 2), high variation among other provinces in these biclusters may prevent X4 from emerging as a strong overall characteristic. The pattern shown by X4 is not consistent across all provinces in each bicluster, reducing its influence as a distinguishing variable, and the negative correlation marked by an obtuse angle in the biplot between X4 and other variables might contribute to X4 not significantly characterizing any bicluster.

Additionally, Quadrant 3 includes many provinces with no significant variables, suggesting that the provinces in this quadrant may be influenced by factors not identified within variables X1 to X13. However, in the bicluster analysis, some provinces in Quadrant 3, such as West Java, South Kalimantan, West Sumatera, and Bangka Belitung, are characterized by several low variables such as X1, X2, and X3. From the biplot, this aligns with these provinces being in the opposite direction of vectors X1, X2, and X3. The quadrant 4 highlights variables X5, X6, X7, X8, X9, and X11, which are important for the provinces in this quadrant. The same is shown by bicluster 7, characterized by variables X5, X7, and X11, which include provinces such as West Kalimantan and Papua. Some provinces may appear in different groups between biclustering and biplot PCA due to different clustering methods. For example, the provinces related to many variables in biclustering might be spread across several quadrants in the biplot PCA. The spatial distribution of the biclustering results is illustrated through a map, as shown in Figure 7.

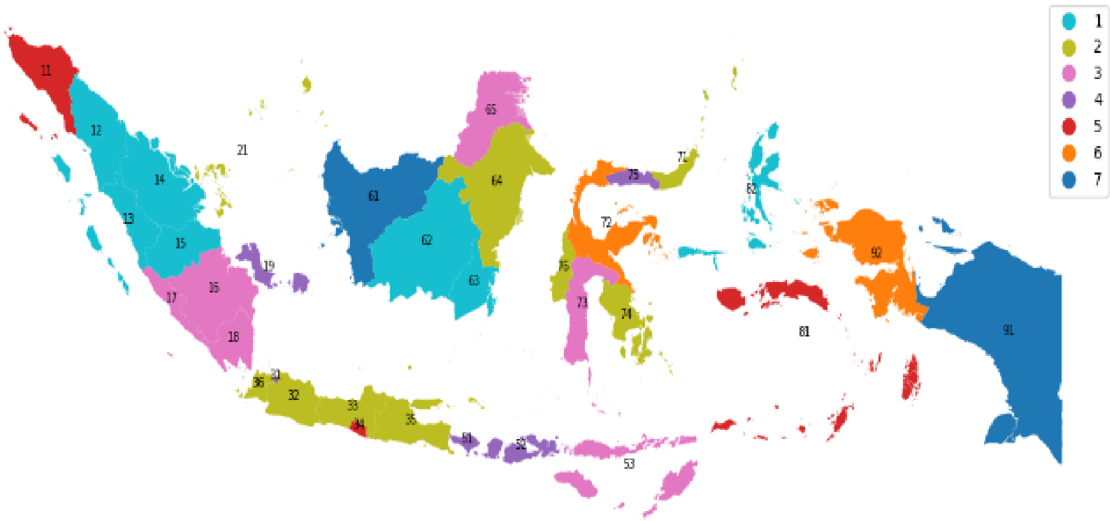

Figure 7. Distribution Map of Optimal Bicluster Results by Province

D. CONCLUSION AND SUGGESTIONS

The use of the CC algorithm in biclustering analysis is explored in this research to identify poverty vulnerability across various provinces in Indonesia. The results show that out of 56 thresholds tested, the threshold (δ) 0.01 produced the lowest average MSR value of 0.0065. This threshold is considered the most optimal as it yields the best bicluster quality with the lowest average MSR value. The study successfully formed seven biclusters with varying indicators and provinces. Variables X1 (headcount index), X2 (poverty depth index), and X3 (poverty severity index) tend to act as global descriptors, appearing in almost all biclusters, indicating their global relevance in explaining poverty patterns. In this context, global descriptors are variables that appear in nearly all formed biclusters, signifying that they play a crucial role in explaining poverty patterns across different provinces. Meanwhile, other variables such as X7 (the percentage of poor people aged 15 and older who did not complete elementary school), X10 (the percentage of women using contraceptives in poor households), X11 (the percentage of poor households using clean water sources for drinking), and X12 (the percentage of poor households with own/shared latrine) show unique characteristics for specific biclusters.

Each bicluster profile reveals distinct vulnerability characteristics, which can inform targeted policy interventions. For instance, in Bicluster 1, which has moderate vulnerability regarding the percentage of poor individuals working in the agricultural sector, policies should encourage economic diversification through the development of non-agricultural sectors and skill training programs. Bicluster 2, which displays moderate vulnerability in poverty indicators and literacy rates, requires a focus on improving access to education and providing education subsidies for poor families. In Bicluster 3, which shows moderate vulnerability in the poverty depth index and life expectancy, it is essential to enhance access to and the quality of health services, as well as implement disease prevention programs. Meanwhile, Bicluster 4, which also shows moderate vulnerability to poverty, requires integrated poverty alleviation programs combined with improved access to basic services. Bicluster 5, which prioritizes poverty indicators, requires additional attention through social protection programs and direct cash assistance. Bicluster 6, with high vulnerability to poverty depth and severity, needs aggressive health interventions and programs focused on women's empowerment. Last, Bicluster 7 must develop basic infrastructure, such as sanitation and access to clean water, along with educational programs to improve job-seeking skills. Thus, policies formulated based on the specific characteristics of each bicluster will be more effective in reducing poverty vulnerability and improving the welfare of communities in each province. While the CC algorithm has been successful in identifying patterns of poverty vulnerability, this study has limitations regarding whether the variables used are sufficiently representative to reflect poverty patterns across all provinces in Indonesia. Therefore, further research is needed to explore other variables that may contribute to a more comprehensive understanding of poverty vulnerability.

REFERENCES

Acharya, S., Saha, S., & Sahoo, P. (2019). Bi-clustering of microarray data using a symmetry-based multi-objective optimization framework. *Soft Comput*, *23*(14), 5693–5714. https://doi.org/10.1007/s00500-018-3227-5

Baruah, B., Dutta, Manash P. Banerjee, S., & Bhattacharyya, D. K. (2024). EnsemBic: An effective ensemble of biclustering to identify potential biomarkers of esophageal squamous cell carcinoma. *Computational Biology and Chemistry*, *110*, 108090. https://doi.org/10.1016/j.compbiolchem.2024.108090

Biswal, B. S., Mohapatra, A., & Vipsita, S. (2022). Ensemble Neighborhood Search (ENS) for biclustering of gene expression microarray data and single cell RNA sequencing data. *Journal of King Saud University - Computer and Information Sciences*, *34*(5), 2244–2251. https://doi.org/10.1016/j.jksuci.2019.11.011

BPS. (2022). *Data dan Informasi Kemiskinan Kabupaten/Kota di Indonesia Tahun 2022*. Badan Pusat Statistik. https://ipb.link/datakemiskinan-periodemaret2022

Castanho, E. N., Aidos, H., & Madeira, S. C. (2022). Biclustering fMRI time series: a comparative study. *BMC Bioinformatics*, *23*(1), 1–30. https://doi.org/10.1186/s12859-022-04733-8

Charfaoui, Y., Houari, A., & Boufera, F. (2024). AMoDeBic: An adaptive Multi-objective Differential Evolution biclustering algorithm of microarray data using a biclustering binary mutation operator. *Expert Systems with Applications*, *238*(Part B). https://doi.org/10.1016/j.eswa.2023.121863

Gu, Z. (2022). Complex heatmap visualization. *IMeta*, *1*(3), 1–15. https://doi.org/10.1002/imt2.43

Helgeson, E. S., Liu, Q., Chen, G., Kosorok, M. R., & Bair, E. (2020). Biclustering via sparse clustering. *Biometrics*, *76*(1), 348–358. https://doi.org/10.1111/biom.13136

Henriques, R., Antunes, C., & Madeira, S. C. (2015). A structured view on pattern mining-based biclustering. *Pattern Recognition*, *48*(12), 3941–3958. https://doi.org/10.1016/j.patcog.2015.06.018

Huang, Q., Chen, Y., Liu, L., Tao, D., & Li, X. (2020). On Combining Biclustering Mining and AdaBoost for Breast Tumor Classification. *IEEE Transactions on Knowledge and Data Engineering*, *32*(4), 728–738. https://doi.org/10.1109/TKDE.2019.2891622

Kaban, P. A., Kurniawan, R., Caraka, R. E., Pardamean, B., Yuniarto, B., & Sukim. (2019). Biclustering method to capture the spatial pattern and to identify the causes of social vulnerability in Indonesia: A new recommendation for disaster mitigation policy. *Procedia Computer Science*, *157*, 31–37. https://doi.org/10.1016/j.procs.2019.08.138

Kavitha Sri, N., & Porkodi, R. (2019). An extensive survey on biclustering approaches and algorithms for gene expression data. *International Journal of Scientific and Technology Research*, *8*(9), 2228–2236. https://ipb.link/paper-kavithaetc

López-Fernández, A., Rodríguez-Baena, D. S., & Gómez-Vela, F. (2020). gMSR: A multi-GPU algorithm to accelerate a massive validation of biclusters. *Electronics (Switzerland)*, *9*(11), 1–15. https://doi.org/10.3390/electronics9111782

Maâtouk, O., Ayadi, W., Bouziri, H., & Duval, B. (2021). Evolutionary Local Search Algorithm for the biclustering of gene expression data based on biological knowledge. *Applied Soft Computing*, *104*, 107177. https://doi.org/10.1016/j.asoc.2021.107177

Padilha, V. A., & Campello, R. J. G. B. (2017). A systematic comparative evaluation of biclustering techniques. *BMC Bioinformatics*, *18*(1), 1–25. https://doi.org/10.1186/s12859-017-1487-1

Padilha, V. A., & Carvalho, A. C. P. de L. F. de. (2019). Experimental correlation analysis of bicluster coherence measures and gene ontology information. *Applied Soft Computing Journal*, *85*(xxxx), 105688. https://doi.org/10.1016/j.asoc.2019.105688

Pang, C. (2022). Construction and Analysis of Macroeconomic Forecasting Model Based on Biclustering Algorithm. *Journal of Mathematics*, *2022*(1), 1-10. https://doi.org/10.1155/2022/7768949

Pauk, J., & Minta-Bielecka, K. (2016). Gait patterns classification based on cluster and bicluster analysis. *Biocybernetics and Biomedical Engineering*, *36*(2), 391–396. https://doi.org/10.1016/j.bbe.2016.03.002

Pontes, B., Giráldez, R., & Aguilar-Ruiz, J. S. (2015). Biclustering on expression data: A review. *Journal of Biomedical Informatics*, *57*, 163–180. https://doi.org/10.1016/j.jbi.2015.06.028

Putri, C. A., Irfani, R., & Sartono, B. (2021). Recognizing poverty pattern in Central Java using Biclustering Analysis. *Journal of Physics: Conference Series*, *1863*(1), 1-8. https://doi.org/10.1088/1742-6596/1863/1/012068

Silva, M. G., Madeira, S. C., & Henriques, R. (2022). Water Consumption Pattern Analysis Using Biclustering: When, Why and How. *Water (Switzerland)*, *14*(12), 1–35. https://doi.org/10.3390/w14121954

Sumertajaya, I. M. S., Ningsih, W. A. L., Saefuddin, A., & Rohaeti, E. (2023). Biclustering Performance Evaluation of Cheng and Church Algorithm and Iterative Signature Algorithm. *JTAM (Jurnal Teori Dan Aplikasi Matematika)*, *7*(3), 643. https://doi.org/10.31764/jtam.v7i3.14778

Wang, B., Miao, Y., Zhao, H., Jin, J., & Chen, Y. (2016). A biclustering-based method for market segmentation using customer pain points. *Engineering Applications of Artificial Intelligence*, *47*, 101–109. https://doi.org/10.1016/j.engappai.2015.06.005

Xie, J., Ma, A., Fennell, A., Ma, Q., & Zhao, J. (2019). It is time to apply biclustering: A comprehensive review of biclustering applications in biological and biomedical data. *Briefings in Bioinformatics*, *20*(4), 1449–1464. https://doi.org/10.1093/bib/bby014

Yuniarto, B., & Kurniawan, R. (2017). Understanding Structure of Poverty Dimensions in East Java: Bicluster Approach. *Signifikan: Jurnal Ilmu Ekonomi*, *6*(2), 289–300. https://doi.org/10.15408/sjie.v6i2.4769

Zhang, J., & Shi, J. (2024). Nonparametric clustering of discrete probability distributions with generalized Shannon's entropy and heatmap. *Statistics & Probability Letters*, *208*. https://doi.org/10.1016/j.spl.2024.110070