Pattern Detection of Economic and Pandemic Vulnerability Index in Indonesia using Bi-Cluster Analysis

by Wiwik Andriyani Lestari Ningsih

Pattern Detection of Economic and Pandemic Vulnerability Index in Indonesia using Bi-Cluster Analysis

Wiwik Andriyani Lestari Ningsih[1], I Made Sumertajaya[2], Asep Saefuddin[3]

[1,2,3]IPB University, Bogor, Indonesia

[1]BPS-Statistics Indonesia, Jakarta, Indonesia

[1]wiwik.andriyani@apps.ipb.ac.id

Abstract - **Bi-clustering is a clustering development that aims to group data simultaneously from two directions. The Iterative Signature Algorithm (ISA) is one of the bi-clustering algorithms that work iteratively to find the most correlated bi-cluster. Detecting economic and pandemic vulnerability using bi-cluster analysis is essential to get spatial patterns and an overview of Indonesia's economic and pandemic vulnerability characteristics. Bi-clustering using ISA requires setting the row and column threshold to form seventy combinations of thresholds. The best is chosen based on the average value of mean square residue to volume ratios. In addition, the similarity of the best bi-cluster with the other is also seen based on the Liu and Wang index values. The -1.0 row and -1.0 column threshold combinations were selected and produced the best bi-cluster with the smallest average value of mean square residue to volume ratios (0.00139). Based on Liu and Wang index values, it has more than 95% similarity with the combination of -1.0 row and -0.9 column thresholds and the -0.9 row and -1.0 column thresholds. These selected threshold combinations produce three bi-clusters with five types of spatial patterns and different characteristics because of the overlap between these three bi-clusters.**

Keywords: Bi-clustering; iterative signature algorithm; Liu and Wang index; mean square residue; pattern detection

I. INTRODUCTION

Bi-clustering is a development of clustering analysis aiming to group data simultaneously from two directions or dimensions. [1] introduced a two-way clustering technique that links a cluster of objects with a cluster of variables together in a data matrix. The bi-cluster analysis was applied by [2] to the gene expression matrix and [3] on microarray data. Bi-clustering was initially applied in biological data analysis to analyze gene expression microarray data. However, in recent years it has become popular to be applied in other fields, such as in the analysis of social data to identify patterns of poverty, food insecurity, and social vulnerability. In addition, bi-cluster analysis of economic data due to the COVID-19 pandemic that occurred in early 2020 is also possible and exciting to apply.

The COVID-19 pandemic in Indonesia has occurred since March 2, 2020, and has significantly impacted various aspects, including the unstable economy. Economic instability is indicated by Indonesia's economic conditions in 2020, experiencing a contraction of 2.07 percent (c-to-c) compared to 2019. In addition, it is also marked by the percentage of the poor population in September 2020, which increased by 0.97 percentage points to September 2019. The occurrence of economic instability has the potential to cause regional economic vulnerability and does not escape the occurrence of vulnerability to the COVID-19 pandemic. This is reflected in the statistics of cases of people who died from COVID-19 in Indonesia, which fluctuated and increased throughout 2020. It was recorded that the number of victims who died from COVID-19 in Indonesia throughout 2020 was 22,138.

The economic and COVID-19 pandemic vulnerability is measured separately by their respective measures. The Department of Economics and Social Affairs, United Nations (UN) made a standard measure to measure economic vulnerability in the form of an index, namely the Economic Vulnerability Index (EVI) [4]. On the other hand, by the National Institute of Environmental Health Sciences (NIEHS), the vulnerability or threat of the COVID-19 pandemic is measured by creating a risk profile in the form of a PVI (Pandemic Vulnerability Index) score or a pandemic vulnerability index for each country in the United States [5]. This measurement focuses more on scoring analysis without considering the region's influence on EVI and PVI constituent indicators. Meanwhile, the bi-cluster analysis (bi-clustering) considers the coherent pattern between the studied area (object) and each of the indicators that make up EVI and PVI (variable). Therefore, bi-clustering can detect spatial patterns of the economic and COVID-19 pandemic vulnerability in an area.

Bi-clustering has various algorithms that can be applied to the case of economic and pandemic vulnerability. However, there are no specific guidelines for selecting the correct bi-clustering algorithm. Therefore, the Iterative Signature Algorithm or ISA was chosen to be applied in this study. According to [6], ISA is classified into a non-metric based linear algebra group. The algorithm excludes evaluation measures in the bi-cluster search process but uses vector spaces and linear mapping between these spaces to describe and find the most correlated sub-matrixes. ISA was proposed by [7] and defined bi-clusters as transcription modules. The size of the transcriptional module depends on the set of two thresholds. Each determines the similarity between the gene and the condition of the module. The use of thresholds in ISA is the most powerful tool that allows structures at different levels to be precise. In addition, ISA works iteratively so that each step will be evaluated repeatedly until convergence occurs.

Based on the discussion above, the purpose of this study is to conduct a bi-cluster analysis using the Iterative Signature Algorithm or ISA in cases of the economic and COVID-19 pandemic vulnerability. In addition, researchers are also intended to obtain an overview of the bi-cluster spatial pattern of the economic and COVID-19 pandemic vulnerability.

II. METHODS

A. Data

This study uses secondary data obtained from the BPS-Statistics Indonesia, the Ministry of Health, the Ministry of Village, Development of Disadvantaged Regions and Transmigration, and the Ministry of Environment and Forestry. The observation unit in this study is the region (province) in Indonesia, with the variables that make up the EVI and PVI indicators based on 2020 data as presented in Table I (except for the X3 and X16 variables).

TABLE I
RESEARCH VARIABLES COMPOSING EVI AND PVI INDICATORS BASED ON SUBCOMPONENTS

Component	Variable (notation)
EVI Indicator	
Size	Population size (X1)
Location	Remoteness and underdeveloped areas (X2)
Environment	Population of the coastal area (X3)
Economic structure	Export concentration (X4)
	Share of agriculture, forestry, and fisheries (Category A) in GRDP (X5)
Trade shocks	Instability in the export of goods and services (X6)
Natural shocks	Instability of agricultural production (X7)
	Natural disaster victims (X8)
PVI Indicator	
Infectious cases	Covid-19 infectious cases (X9)
Spread of disease	Covid-19 death rate (X10)
Population mobility	Daytime population (X11)
	Traffic volume average (X12)
Housing density	Average of total household members (X13)
Testing	Covid-19 testing (X14)
Social distancing	Social distancing score (X15)
Air pollution	Air quality index (X16)
Age distribution	Population aged 65 years and over (X17)
Comorbidities including premature death, smoking, diabetes, and obesity	Morbidity rate (X18)
	Adult residents (20 years and over) smoker (X19)
Health disparities	Residents without insurance (X20)
	Poor population (X21)
	Open unemployment rate (X22)
Hospital beds	Availability of hospital beds (X23)

A brief explanation of some of the variables in Table I is as follows:
1. remoteness and disadvantaged areas using an approach of the inverse value of the Developing Village Index (IDM);

2. the population of the coastal area is estimated using the proportion of the number of villages located by the sea multiplied by the total population;
3. export concentration using the percentage of exports of agricultural, forestry, and fishery product types;
4. instability in the export of goods and services using an approach of the inverse value of the ratio of the contribution of total exports in 2020 divided by the previous year;
5. instability of agricultural production using an approach of the ratio of the Category A GRDP contribution in 2020 divided by the previous year;
6. the daytime population is estimated by multiplying the total population by the proportion of the number of passenger cars and motorcycles;
7. the social distance score using an approach of the ratio of the domestic tourists (number of people traveling other than for work or school) per resident.

The inverse value is used so that the condition of those variables is in line with the concept of vulnerability in the other variables; the higher the value of that variable, the more tend vulnerability will occur. However, non-vulnerable conditions tend to occur along with the lower of those variables.

B. Research Stages

1) Preprocessing: At this stage, three steps will be performed, that are: 1) form a data matrix consisting of 34 provinces as rows and 23 EVI and PVI variables as columns, 2) scale the data matrix using the standard normalization approach, and 3) explore the data to see the initial characteristics of the scaled data matrix using a heatmap.

2) Bi-cluster Analysis

According to [8], bi-clustering is a valuable methodology for finding coherent local patterns hidden in data matrices. In contrast to the traditional clustering procedure, which searches for coherent pattern clusters using the entire feature set (variable), bi-clustering performs pattern classification simultaneously in both directions, the rows and columns of the data matrix. Bi-clustering aims to find a subset whose element objects are the same as each other based on a specific set of variables and as much as possible different from the remaining objects and other variables [9], [10]. According to [6], one of the bi-clustering algorithms is an Iterative Signature Algorithm (ISA), which is classified into the non-metric-based linear algebra group. ISA is a bi-clustering method proposed by [7], with the input in the form of a matrix while the resulting output is a set of bi-clusters. A collection of bi-clusters will be generated as input matrix blocks that are rearranged until they meet several predetermined criteria. ISA works iteratively, so the evaluation will be repeated at each step until convergence occurs. In addition, it takes thresholds as an essential step in running ISA. There are two main parameters: the thresholds of rows and columns in the ISA [11].

At this stage, bi-clustering will be performed using ISA. Bi-cluster analysis using ISA requires determining row and column thresholds. This study will use 70 types of row and column threshold combinations. The row threshold starts from -0.4 to -1.0 with 0.1 sequences, so there are seven thresholds. Meanwhile, the column threshold starts from -0.1 to -1.0 with 0.1 sequences, so there are 10 column thresholds. The determination of the row and column thresholds is based on the condition of the scaling data matrix, which tends to accumulate more at values below zero (negative). The ISA bi-clustering flow chart is presented in Fig. 1.

The description for each stage of the Iterative Signature Algorithm (ISA) flow chart in Fig. 1 is as follows:
1. set the column and row threshold values (t_C, t_G), seed values, and number of seeds (n);
2. create the matrix E^G and E^C;
3. choose several row vectors (row samples) randomly so that the $V'U$ matrix will be formed;
4. calculate the average of each row of the row samples ($e_{uV'}^C$) in the column standardized matrix (E^C);
5. form the U' matrix with the condition that the sum of $e_{uV'}^C > t_C \sigma_C$ and its average becomes a "column score" so that the $V'U'$ matrix will be formed;
6. calculate the average of each column of the U' matrix ($e_{U'v}^G$) in the row standardized matrix (E^G);
7. form the V'' matrix with the condition that the sum of $e_{U'v}^G > t_G \sigma_G$ and its average becomes a "row score" so that the $V''U'$ matrix will be formed;
8. steps 4 to 7 will be repeated for n number of seeds when the convergence condition is not met, that is $\frac{|V' \backslash V''|}{|V' \cup V''|} > \varepsilon$; however

9. when the convergence condition is met, that is $\frac{|V'\backslash V''|}{|V' \cup V''|} < \varepsilon$, then some row and column will be selected (with "score") and it is called transcription modules;
10. steps 3 to 9 will be repeated as many B bi-cluster as may be formed and following a predetermined threshold value
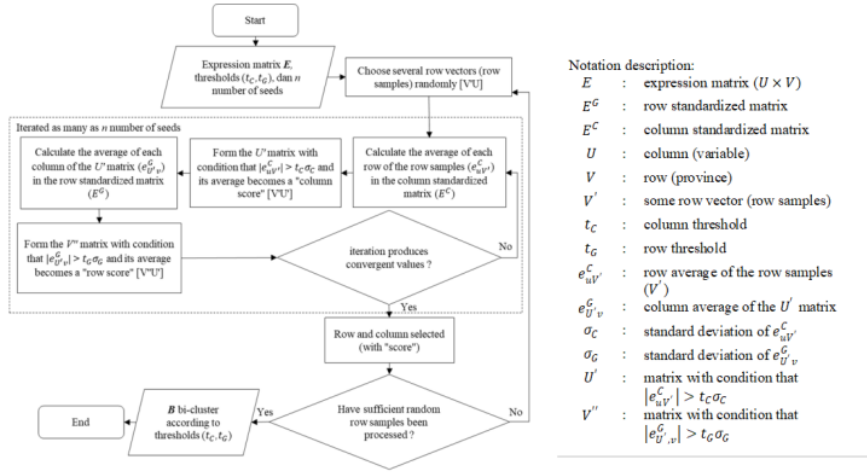

Fig. 1 Iterative Signature Algorithm/ISA Flowchart (modified from [11])

3) *Bi-cluster Evaluation and Selection*

According to [12], the performance of the bi-clustering algorithm can be evaluated using the Mean Squared Residue (MSR) defined by equation 1,

$$MSR_{(I,J)} = \frac{\sum_{i \in I} \sum_{j \in J} (b_{ij} - b_{iJ} - b_{Ij} + b_{IJ})^2}{I \times J} \tag{1}$$

where b_{IJ} is the average across the bi-clusters, b_{Ij} is the average in column j, b_{iJ} is the average in row i, $I \times J$ is the dimension of the bi-cluster, i.e. the size of the bi-cluster row (I) times the size of the bi-cluster column (J). $MSR_{(I,J)}$ represents the variation associated with the interaction between rows and columns in a bi-cluster [12], [13]. According to [14], the quality of a bi-cluster will be better as the residual value (MSR) decreases and/or the volume (rows × column) of the bi-cluster increases and it is obtained by calculating the average value of MSR to volume ratios $\left(\frac{1}{n}\sum_{i=1}^{n} \frac{MSR_i}{Volume_i}\right)$. In addition, to see how well each optimal bi-cluster group (M_{opt}) will have similarities with every other bi-cluster group (M), we can use the size of Liu and Wang index defined by equation 2 [15],

$$I_{Liu\&Wang}(M_{opt}, M) = \frac{1}{K_{opt}} \sum_{i=1}^{K_{opt}} \max\left(\frac{[G_i \cap G_j] + [C_i \cap C_j]}{[G_i \cup G_j] + [C_i \cup C_j]}\right) \tag{2}$$

where M_{opt} is defined as the bi-cluster group that has the smallest average value of mean square residue to volume ratios, K_{opt} is the number of bi-clusters in M_{opt}, $[G_i \cap G_j]$ is the number of rows (G) in M_{opt} that intersect with rows in M, $[C_i \cap C_j]$ is the number of columns (C) in M_{opt} that intersect with columns in M, $[G_i \cup G_j]$ is the number of combined rows (G) of M_{opt} and M, and $[C_i \cup C_j]$ is the number of combined columns (C) of M_{opt} and M.

At this stage, the bi-cluster formed in each type of threshold combination will be evaluated using the average value of MSR to volume ratio. One type of threshold combination is selected for further analysis in the form of spatial patterns and their characteristics. The results of the spatial pattern of economic and COVID-19 pandemic vulnerability using ISA can be described in the form of a map, and those characteristics can be identified. In

addition, it is also seen how well each bi-cluster group in the selected threshold combination has similarities with the bi-cluster group in other threshold combinations using the Liu and Wang index values.

III. RESULTS

A. Data Exploration

After scaling the data, the condition of the used data matrix can be described through the heatmap in Fig. 2. It can be seen from the heatmap that there are several data variables for EVI (X1 to X8) and PVI (X7 to X23) in a particular province whose values tend to be highly positive (dark blue) and extremely negative (broken white).
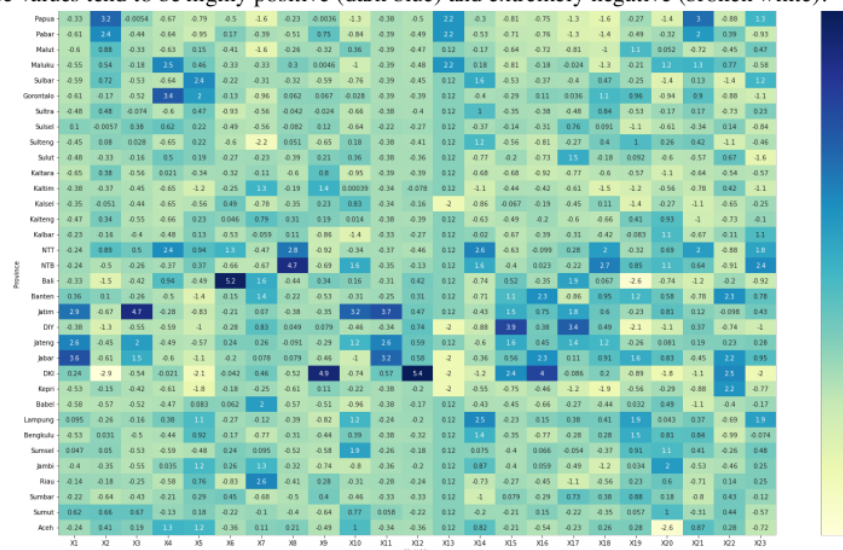
Fig. 2 Heatmap of Scaled Data Matrix

The province with a value that tends to be highly positive indicates that the province in the relevant variable tends to be vulnerable. On the other hand, the province with a value that tends to be highly damaging indicates that the province in the relevant variable tends to be less vulnerable (non-vulnerable). For example, in the PVI X9 variable (covid-19 infectious cases), DKI Jakarta Province has an extreme positive value (dark blue) which indicates that DKI Jakarta Province tends to have a high COVID-19 pandemic vulnerability, especially in the infectious case indicator. Another example in the EVI X2 variable (remoteness and underdeveloped areas), DKI Jakarta Province has a highly negative value (broken white) which indicates that DKI Jakarta Province tends to have a low economic vulnerability, especially on indicators of remoteness and disadvantaged areas.

B. ISA Bi-clustering

Bi-cluster analysis using ISA requires setting the row and column threshold value, seed values, and the number of seeds. In this study, several combinations of row and column thresholds were tried using 0.1 sequences, 070 seed values, and the number of seeds was 100. The row threshold started from -0.4 to -1.0, so there were 7-row thresholds. Meanwhile, the column threshold started from -0.1 to -1.0, so there are 10-column thresholds. Furthermore, from the 70 threshold combinations, the threshold combination is chosen, which produces the smallest average value of MSR to volume ratios. Fig. 3 presents the distribution of the average value of MSR/Volume for each threshold combination and its number of bi-clusters formed.

It can be seen from Fig. 3 that the combination of -1.0 row and -1.0 column threshold has the smallest average value of MSR to volume ratio (0.00139), with the number of bi-clusters formed being 3. Meanwhile, the combination of -0.4 row and - 0.1 column threshold has the highest average value of MSR to volume ratio (0.00435), with the number of bi-clusters formed being 6. In general, it can be seen from the Fig. that the average MSR to volume ratio value tends to be worth higher (towards a dark blue color) when the row and column thresholds are getting closer to

positive values. The larger the average value of MSR to volume ratio, the lower the quality of the bi-cluster formed. Therefore, a combination of -1.0 row and -1.0 column threshold was chosen from the various row and column threshold combinations. Each bi-cluster from the selected row and column threshold combinations will be further analyzed in its characteristics.
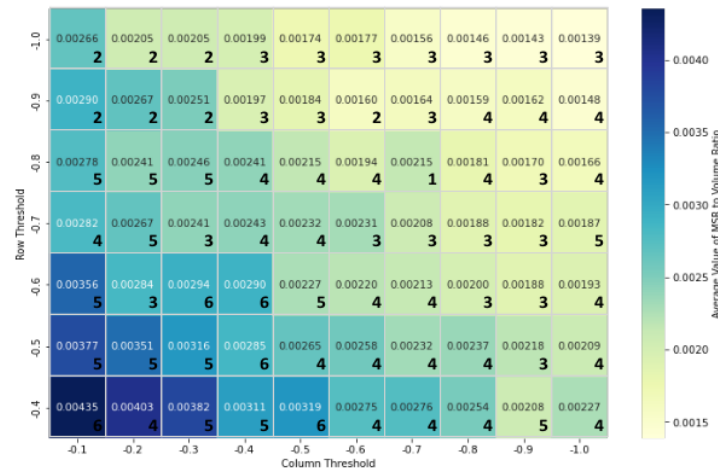

Fig. 3 Heatmap of the Average Value of MSR/Volume and the Number of Bi-clusters according to Combination of Row and Column Threshold

Before further analyzing the characteristics of the selected threshold combination, it is necessary to identify the similarity of the goodness of the bi-cluster in the selected threshold combination with the other bi-clusters. The identification measures the Liu and Wang index values depicted in Fig. 4. The 1.00 value of Liu and Wang index at the -1.0 row and -1.0 column threshold indicates that the bi-cluster group at that threshold is the optimal bi-cluster group (having the smallest average MSR to volume ratio). The Liu and Wang index values, closer to 1, indicate that the bi-cluster group at the relevant threshold has a higher level of similarity with the optimal bi-cluster group. It can be seen from Fig. 4 that the bi-cluster group at the selected threshold has a similarity level above 95% with the bi-cluster group at the -1.0 row and -0.9 column threshold and the -0.9 row and -1.0 column threshold.
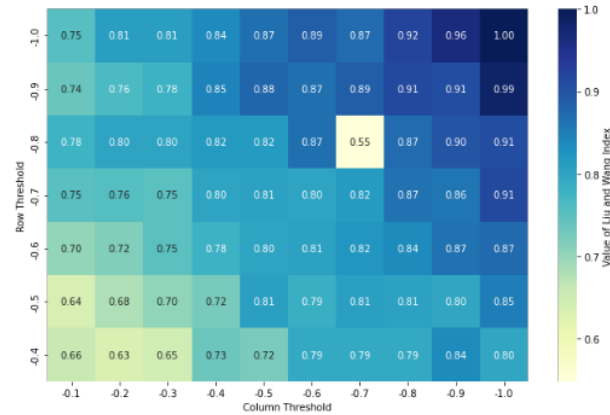

Fig. 4 Heatmap of Liu and Wang Index Values according to Combination of Row and Column Thresholds

The membership of the selected ISA bi-clustering results (-1.0 row and -1.0 column threshold) according to the bi-cluster formed in Table II. It can be seen from the table that the membership of the three bi-clusters formed overlaps each other, both provincial membership and its variables.

TABLE II
MEMBERSHIP OF SELECTED ISA BI-CLUSTERING RESULTS ACCORDING TO BI-CLUSTER

Bi-cluster	Bi-cluster size	Membership	
		Province (code)	Variable
1	26 × 21	11, 12, 13, 14, 15, 16, 17, 18, 19, 21, 36, 52, 53, 61, 62, 65, 71, 72, 73, 74, 75, 76, 81, 82, 91, 94	X1, X2, X3, X4, X5, X6, X8, X9, X10, X11, X12, X13, X14, X15, X16, X17, X18, X19, X20, X21, X23
2	27 × 18	11, 12, 13, 14, 15, 16, 19, 21, 31, 32, 33, 34, 35, 36, 51, 62, 63, 64, 65, 71, 73, 74, 75, 76, 81, 91, 94	X1, X5, X4, X6, X7, X8, X9, X10, X11, X12, X15, X16, X17, X18, X19, X20, X22, X23
3	29 × 17	11, 12, 13, 14, 15, 16, 18, 19, 21, 31, 32, 33, 34, 35, 36, 51, 61, 62, 63, 64, 65, 71, 73, 74, 75, 76, 82, 91, 94	X1, X5, X4, X6, X7, X8, X9, X10, X11, X12, X15, X16, X17, X18, X20, X21, X22

For example, province coded 11 (Aceh), and variable X1 (population size) are included as members of bi-clusters 1, 2, and 3. Province coded 18 (Lampung), and variable X21 (poor population) are included as a member of bi-clusters 1 and 3. Province coded 63 (Kalimantan Selatan) and variable X7 (instability of agricultural production) is included as a member of bi-clusters 2 and 3. Also, province coded 81 (Maluku), and variable X23 (availability of hospital beds) are included as a member of bi-clusters 1 and 2.

The overlap between the three bi-clusters formed is detected to have five different types of spatial patterns and characteristics that could be analyzed further. The spatial pattern is depicted as a map from the result of selected ISA bi-clustering, as shown in Fig. 5.
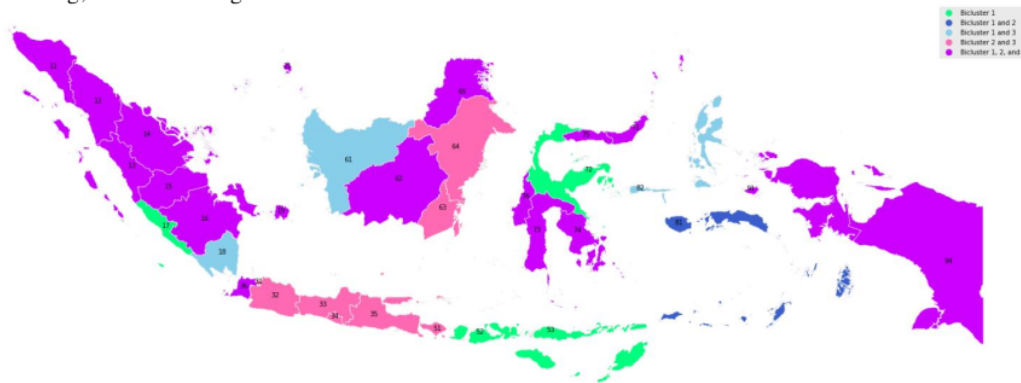

Fig. 5 Map of Selected ISA Bi-clustering Results According to Spatial Pattern Type

Through the Fig.5, it can be seen that there are five types of spatial patterns formed. First is the light green area (bi-cluster 1), which are the provinces coded 17 (Bengkulu), 52 (NTB), 53 (NTT), and 72 (Sulawesi Tengah). The second is the dark blue area (bi-cluster 1 and 2), province-coded 81 (Maluku). The third is the light blue areas (bi-cluster 1 and 3), which are provinces coded 18 (Lampung), 61 (Kalimantan Barat), and 82 (Maluku Utara). Fourth, the colored pink areas (bi-cluster 2 and 3) which are provinces coded 31 (DKI Jakarta), 32 (Jawa Barat), 33 (Jawa Tengah), 34 (DIY), 35 (Jawa Timur), 51 (Bali), 63 (Kalimantan Selatan), and 64 (Kalimantan Timur). The fifth, the colored purple areas (bi-cluster 1, 2, and 3), which are provinces coded 11 (Aceh), 12 (Sumatera Utara), 13 (Sumatera Barat), 14 (Riau), 15 (Jambi), 16 (Sumatera Selatan), 19 (Bangka Belitung), 21 (Kepulauan Riau), 36 (Banten), 62 (Kalimantan Tengah), 65 (Kalimantan Utara), 71 (Sulawesi Utara), 73 (Sulawesi Selatan), 74 (Sulawesi Tenggara), 75 (Gorontalo), 76 (Sulawesi Barat), 91 (Papua Barat), and 94 (Papua).

Besides identifying spatial patterns through maps such as Fig. 5, the variable characteristics for each type of spatial pattern can also be identified and described in Fig.s 6 to 10 in the form of a bar chart. The bar chart illustrates the average value for each variable included in each type of spatial pattern. For example, for the first spatial pattern (bi-cluster 1), all EVI variables except X7 and all PVI variables except X22 are characteristics of the region included in the first spatial pattern. It can be seen from Fig. 6 that four of the seven EVI variables (X2, X4, X5, and X8) in the first spatial pattern tend to be highly vulnerable. It indicates that the Bengkulu, NTB, NTT, and Sulawesi Tengah

regions tend to have a high economic vulnerability on indicators X2 (remoteness and underdeveloped areas), X4 (export concentration), X5 (share of agricultural, forestry, and fishery in GRDP), and X8 (natural disaster victims). In addition, six of the fourteen PVI variables (X13, X14, X19, X20, X21, and X23) also tend to be highly vulnerable. It indicates that the Bengkulu, NTB, NTT, and Sulawesi Tengah regions tend to have a high pandemic vulnerability on indicators X13 (average of total household members), X14 (covid-19 testing), X19 (adult residents smokers), X20 (residents without insurance), X21 (poor population), and X23 (availability of hospital beds). Ten of the twenty-one variables in the first spatial pattern tend to have high vulnerability from an economic and pandemic perspective.
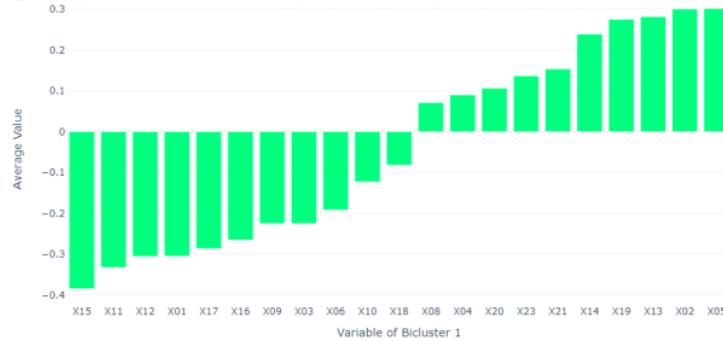

Fig. 6 Characteristics of the First Spatial Pattern Variable (Bi-cluster 1)

The characteristics of variables for the second spatial pattern are depicted in Fig. 7, which shows that not all variables are included to use as the characteristics of areas in this spatial pattern. Seven of the variables in this study did not characterize this second type of spatial pattern: variables X2, X5, X7, X13, X14, X21, and X22. It can be seen that only two of the sixteen variables in the second spatial pattern have high vulnerability from the economic and pandemic perspective, which are X4 (export concentration) and X19 (adult residents smokers). It indicates that the Maluku region tends to have high economic vulnerability only on the export concentration indicator and also tends to have a high pandemic vulnerability only on the indicator of adult residents smokers.
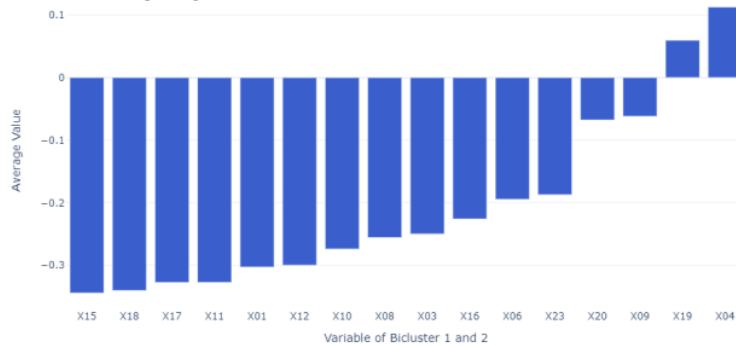

Fig. 7 Characteristics of the Second Spatial Pattern Variable (Bi-cluster 1 and 2)

In contrast to the first and second spatial patterns, the third spatial pattern is depicted in Fig. 8, which shows that all variables that characterize this type of spatial pattern tend to have low vulnerability. Fifteen of the twenty-three variables in this study that characterize this third type of spatial pattern are X1, X3, X4, X6, and X8 (EVI variables), and also X9, X10, X11, X12, X15, X16, X17, X18, X20, and X21 (PVI variables). It indicates that the regions of Lampung, Kalimantan Barat, and Maluku Utara tend to have a low economic vulnerability on indicators X1 (population size), X3 (population of the coastal area), X4 (export concentration), X6 (instability in the export of goods and services), and X8 (natural disaster victims). In addition, those region also tends to have low pandemic vulnerability on indicators X9 (covid-19 infectious cases), X10 (covid-19 death rate), X11 (daytime population), X12 (traffic volume average), X15 (social distancing score), X16 (air quality index), X17 (population aged 65 years and over), X18 (morbidity rate), X20 (residents without insurance), and X21 (poor population).
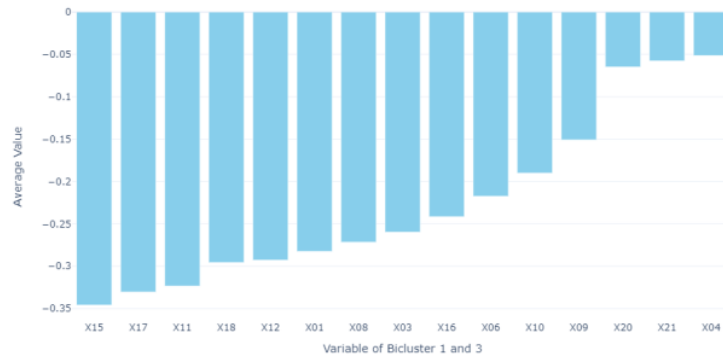
Fig. 8 Characteristics of the Third Spatial Pattern Variable (Bi-cluster 1 and 3)

The variable characteristics of the fourth spatial pattern are depicted in Fig. 9. It can be seen from the figure that seven of the total variables in this study did not characterize this type of spatial pattern, which are X2, X5, X13, X14, X19, X21, and X23 variables. It is also seen from Fig. 9 that eleven of the sixteen variables in this fourth spatial pattern tend to have high vulnerability in terms of the economic side (X1, X3, X6, and X7 variables) and also the COVID-19 pandemic side (X9, X11, X12, X15, X16, X17, and X22 variables). It indicates that the areas of DKI Jakarta, Jawa Barat, Jawa Tengah, DIY, Jawa Timur, Bali, Kalimantan Selatan, and Kalimantan Timur tend to have a high economic vulnerability on indicators X1 (population size), X3 (population of the coastal area), X6 (instability in the export of goods and services), and X7 (instability of agricultural production). In addition, those regions also tend to have a high pandemic vulnerability on indicators X9 (covid-19 infectious cases), X11 (daytime population), X12 (traffic volume average), X15 (social distancing score), X16 (air quality index), X17 (population aged 65 years and over), and X22 (open unemployment rate).

Fig. 9 Characteristics of the Fourth Spatial Pattern Variable (Bi-cluster 2 and 3)

The same as the third spatial pattern, the fifth spatial pattern, which is depicted in Fig. 10, shows that all of the variables that characterize this type of spatial pattern tend to have low vulnerability. Fourteen of the twenty-three variables in this study that characterize the fifth type of spatial pattern are X1, X3, X4, X6, and X8 (EVI variables), as well as X9, X10, X11, X12, X15, X16, X17, X18, and X20 (PVI variables). This indicates that the regions of Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bangka Belitung, Kepulauan Riau, Banten, Kalimantan Tengah, Kalimantan Utara, Sulawesi Utara, Sulawesi Selatan, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Papua Barat, and Papua tends to have a low economic vulnerability on indicators X1 (population size), X3 (population of the coastal area), X4 (export concentration), X6 (instability in the exports of goods and services), and X8 (natural disaster victims). In addition, those regions also tend to have a low pandemic vulnerability on indicators X9 (covid-19 infectious cases), X10 (covid-19 death rate), X11 (daytime population), X12 (traffic volume average), X15 (social distancing score), X16 (air quality index), X17 (population aged 65 years and over), X18 (morbidity rate), and X20 (residents without insurance).
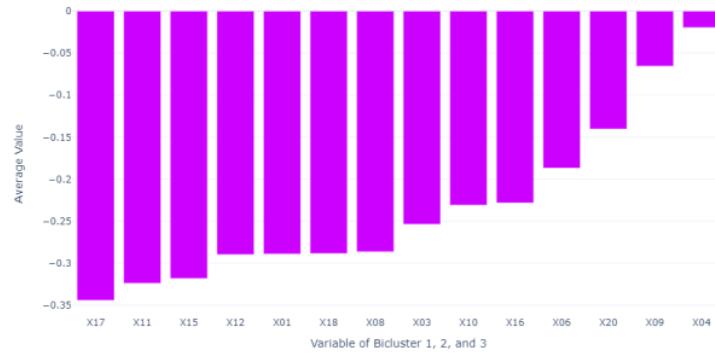
Fig. 10 Characteristics of the Fifth Spatial Pattern Variable (Bi-cluster 1, 2, and 3)

The analysis above shows that most areas in Indonesia are included in the fifth type of spatial pattern with the characteristics of EVI and PVI variables which tend to have low vulnerability. Meanwhile, most areas on Java Island are included in the fourth type of spatial pattern with the characteristics of both EVI and PVI variables; most tend to have high vulnerability. It indicates that most regions in Indonesia tend to have low economic and COVID-19 pandemic vulnerability. However, most areas on Java Island tend to have high economic and COVID-19 pandemic vulnerability.

IV. CONCLUSION

This study has conducted a bi-cluster analysis using the Iterative Signature Algorithm (ISA) in economic and COVID-19 pandemic vulnerability cases. In this case, the results of ISA bi-clustering use the -1.0 row and -1.0 column threshold combination as the optimal threshold with the smallest average value of MSR to volume ratios, which is 0.00139. The optimal threshold has a similarity level above 95% with the -1.0 row and -0.9 column threshold and -0.9 row and -1.0 column threshold. The number of bi-clusters formed with the optimal threshold is three, but it is detected to form five types of spatial patterns and different characteristics due to the overlap between the bi-clusters formed. Most regions in Indonesia tend to have a low economic and COVID-19 pandemic vulnerability pattern, but most areas on Java Island tend to have a high economic and COVID-19 pandemic vulnerability.

ACKNOWLEDGEMENT

REFERENCES

[1]   J. A. Hartigan, "Direct clustering of a data matrix," *J. Am. Stat. Assoc.*, vol. 67, no. 337, pp. 123–129, 1972, doi: 10.1080/01621459.1972.10481214.

[2]   Y. Cheng and G. M. Church, "Biclustering of expression data.," *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, vol. 8, pp. 93–103, 2000.

[3]   Y. Kluger, R. Basri, J. T. Chang, and M. Gerstein, *Spectral biclustering of microarray data: Coclustering genes and conditions*, vol. 13, no. 4. 2003. doi: 10.1101/gr.648603.

[4]   United Nations, "EVI Indicators," 2011. ipb.link/un-evi (accessed Apr. 27, 2021).

[5]   National Institute of Environmental Health Sciences, "Details for PVI Maps," 2020. ipb.link/niehs (accessed Apr. 27, 2021).

[6]   B. Pontes, R. Giráldez, and J. S. Aguilar-Ruiz, "Biclustering on expression data: A review," *J. Biomed. Inform.*, vol. 57, pp. 163–180, 2015, doi: 10.1016/j.jbi.2015.06.028.

[7]   S. Bergmann, J. Ihmels, and N. Barkai, "Iterative signature algorithm for the analysis of large-scale gene expression data," *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, vol. 67, no. 3, p. 18, 2003, doi: 10.1103/PhysRevE.67.031902.

[8]   H. Zhao, A. Wee-Chung Liew, D. Z. Wang, and H. Yan, "Biclustering Analysis for Pattern Discovery: Current Techniques, Comparative Studies and Applications," *Curr. Bioinform.*, vol. 7, no. 1, pp. 43–55, 2012, doi: 10.2174/157489312799304413.

[9]   S. Kaiser, "Biclustering: Methods, Software and Application," *Ph.D Thesis*, p. 178, 2011.

[10]   Nurmawiya and R. Kurniawan, "Pengelompokan Wilayah Indonesia Dalam Menghadapi Revolusi Industri 4.0 Dengan Metode Biclustering," pp. 790–797, 2020.

[11]   A. L. Richards, P. Holmans, M. C. O'Donovan, M. J. Owen, and L. Jones, "A comparison of four clustering methods for brain expression microarray data," *BMC Bioinformatics*, vol. 9, pp. 1–17, 2008, doi: 10.1186/1471-2105-9-490.

[12]   N. Kavitha Sri and R. Porkodi, "An extensive survey on biclustering approaches and algorithms for gene expression data," *Int. J. Sci. Technol. Res.*, vol. 8, no. 9, pp. 2228–2236, 2019.

[13]   H. Cho and I. S. Dhillon, "Coclustering of human cancer microarrays using minimum sum-squared residue coclustering," *IEEE/ACM Trans. Comput. Biol. Bioinforma.*, vol. 5, no. 3, pp. 385–400, 2008, doi: 10.1109/TCBB.2007.70268.

[14]   A. Chakraborty and H. Maka, "Biclustering of gene expression data using genetic algorithm," *Proc. 2005 IEEE Symp. Comput. Intell. Bioinforma. Comput. Biol. CIBCB '05*, vol. 2005, no. 2000, 2005, doi: 10.1109/cibcb.2005.1594893.

[15]   X. Liu and L. Wang, "Computing the maximum similarity bi-clusters of gene expression data," *Bioinformatics*, vol. 23, no. 1, pp. 50–56, 2007, doi: 10.1093/bioinformatics/btl560.

Pattern Detection of Economic and Pandemic Vulnerability Index in Indonesia using Bi-Cluster Analysis

8 conference.iainsalatiga.ac.id
Internet Source
<1%

9 id.scribd.com
Internet Source
<1%

10 www.krakatausteel.com
Internet Source
<1%

11 sport.detik.com
Internet Source
<1%

12 arxiv.org
Internet Source
<1%

13 ijsrset.com
Internet Source
<1%

14 ijcb.in
Internet Source
<1%

15 www.itb.ac.id
Internet Source
<1%

16 J. H. Ihmels. "Challenges and prospects in the analysis of large-scale gene expression data", Briefings in Bioinformatics, 01/01/2004
Publication
<1%

17 homes.esat.kuleuven.be
Internet Source
<1%

18 kuscholarworks.ku.edu
Internet Source
<1%

19	www.mate.polimi.it Internet Source	<1%
20	"Multimodal Analytics for Next-Generation Big Data Technologies and Applications", Springer Science and Business Media LLC, 2019 Publication	<1%
21	Maryam Golchin, Alan Wee-Chung Liew. "Chapter 6 Bi-clustering by Multi-objective Evolutionary Algorithm for Multimodal Analytics and Big Data", Springer Science and Business Media LLC, 2019 Publication	<1%
22	Zhenghong Wei, Hongya Zhao, Lan Zhao, Hong Yan. "Multiscale co-clustering for tensor data based on canonical polyadic decomposition and slice-wise factorization", Information Sciences, 2019 Publication	<1%
23	www.neliti.com Internet Source	<1%