

BCBimax Biclustering Algorithm with Mixed-Type Data

Hanifa Izzati^{1,2}, Indahwati³, Anik Djuraidah⁴

^{1,2,3,4}IPB University, Bogor, Indonesia

²Ministry of Industry, Jakarta, Indonesia

*corr_author: hanifaizzati@kemenperin.go.id

Abstract - The application of biclustering analysis to mixed data is still relatively new. Initially, biclustering analysis was primarily used on gene expression data that has an interval scale. In this research, we will transform ordinal categorical variables into interval scales using the Method of Successive Interval (MSI). The BCBimax algorithm will be applied in this study with several binarization experiments that produce the smallest Mean Square Residual (MSR) at the predetermined column and row thresholds. Next, a row and column threshold test will be carried out to find the optimal bicluster threshold. The existence of different interests in the variables for international market potential and the number of Indonesian export destination countries is the reason for the need for identification regarding the mapping of destination countries based on international trade potential. The study's results with the median threshold of all data found that the optimal MSR is at the threshold of row 7 and column 2. The number of biclusters formed is 9 which covers 74.7% of countries. Most countries in the bicluster come from the European Continent and a few countries from the African Continent are included in the bicluster.

Keywords: BCBimax, Biclustering, MSR, market potential

I. INTRODUCTION

In real data, it is possible to involve various kinds of data which do not only consist of numeric or categorical data, but rather a mixture of both (mixed-type data). Choosing an approach to handling mixed data is still a challenge, especially for biclustering methods because it is very dependent on datasets. There are several approaches to conducting mixed data analysis, one of which is converting all variables into categorical variables and then proceeding with clustering analysis [1]. Another approach is to quantify categorical variables into numerical variables using a dimension reduction technique, followed by a clustering technique [2].

Cluster analysis is a method of grouping objects based on their similar characteristics [3]. Cluster analysis commonly used is one-way clustering. This analysis assumes that objects have similar characteristics in all rows or columns, so that objects in rows are grouped based on similarity in columns or variables in columns are grouped based on similarities in rows. One-way clustering has been carried out separately, namely by grouping objects in rows using a distance measure matrix, then grouping variables in columns using a correlation matrix [4]. Clustering like this still has limitations for two-way data that want to know the relationship of a particular group of objects with a specific group of variables together. Biclustering is a development of cluster analysis and aims to group data simultaneously from two directions or two dimensions. The two-way clustering technique (biclustering) is then applied to the gene expression matrix, namely the matrix data contains real numbers that show the activity of several different genes (rows) and experimental conditions (columns) [5].

Biclustering algorithm classification is divided into five namely greedy iterative search, divide and conquer, exhaustive bicluster enumeration, iterative row and column clustering combination, and distribution parameter identification [6]. In this study, the algorithm to be chosen is the BCBimax algorithm which is a divide-and-conquer category. The BCBimax algorithm is a technique introduced to research market segmentation using customer pain points [7]. This algorithm is a development of the Bimax algorithm which can avoid overlapping and also solve problems in the Bimax algorithm which produces too many small and many biclusters [8]. This method is fast and precise in producing optimal biclusters so that it becomes a reference for other algorithms [9]. The BCBimax algorithm has been widely applied in identifying customer segmentation based on behavior [10] and has

been applied in the health sector to identify subgroups of lymphoma cancer survivors [8].

Biclustering analysis to mixed data is still relatively new because initially this analysis was mostly used on gene expression data that has an interval scale. In addition, so far no algorithm has been developed that can accommodate mixed data at once. So that in this study we will first transform ordinal categorical variables into interval scales. One method that is widely used in transforming ordinal categorical data into an interval scale is the Successive Interval Method (MSI). The MSI transformation is a method of transforming ordinal data into interval data by changing the cumulative proportion of each variable in the category to its standard normal curve value [11].

The open economic system and globalization make international trade even more important because every country uses it to analyze economic development and formulate economic policies [12]. In determining international trade potential, it is necessary to study many factors such as cultural, economic, and demographic factors where these factors consist of mixed data types, namely numerical data and categorical data. In determining market potential, each company also chooses different interests in determining variable priorities. The existence of different interests in the international market potential variables and the many destinations for Indonesia's export countries is the reason for the need for identification regarding the mapping of destination countries based on international trade potential in order to map destination countries for exports abroad appropriately and efficiently. Several studies regarding the grouping of export destination countries have been carried out, such as grouping Turkey's export destination countries based on an assessment of market potential using factor analysis and K-Means [13] and grouping potential markets in developing countries using Markov chains [14].

Based on this discussion, this research will aim to carry out a biclustering analysis using the BCBimax algorithm for potential international trade data, especially in 103 countries in the world. The results of this study are expected to be a reference for government policymaking in increasing exports to potential countries based on the results of the biclusters formed.

II. METHOD

A. Data

The variables used in this study were 15 variables with 10 numerical variables and 5 ordinal categorical

variables. We extract the data of variable X1-X9 and X14-X15 from World Development Integration Data by World Bank, variable X10 from Global Sustainable Competitiveness Index (GSCI) [15], variable X11 using calculation with formulation [16] to [17], variable X12 from Amphori [18], and variable X13 from FM Global [19]. The research data used was obtained from a total of 103 countries as presented in Table I. R-Studio software is used as a data analysis tool.

B. Research Stages

In this research, there are three stages of research: pre-processing, inverse and data transformation, and biclustering analysis. The following is a complete explanation.

1) *Pre-processing*: At this stage, data cleansing will be carried out by deleting data that contains missing values. After obtaining complete data and producing as many as 103 countries, the next step is to standardize the numerical variables. The results of the data that have been standardized will then be carried out with exploratory analysis to see the initial characteristics of the data.

2) *Inverse and data transformation*: The dataset on the variables X2, X5, and X11 will be inverted by -1 to create a meaning that is in line with other variables according to the definition of international trade potential. In addition, the ordinal variables X13, X14, and X15 will be transformed using MSI into interval data. The following are the stages in carrying out the MSI transformation [11]:

1. Calculate the frequency of observations for each category
2. Calculate the proportion in each category
3. From the proportions obtained, the cumulative proportions for each category are calculated.
4. Calculates the Z value (normal distribution) of cumulative proportions
5. Determine the Z limit value (the value of the probability density function in the abscissa of Z) for each category, by (1)

$$\delta(Z) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{Z^2}{2}\right)}, -\infty < Z < +\infty \quad (1)$$

6. Calculating the scale value for each density category lower limit-density upper limit area under the upper limit - area below the lower limit (2).
7. Calculating the score (transformed value) for each category through (3).

$$Scale = \frac{density\ lower\ limit\ value - density\ upper\ limit\ value}{suitable\ cumulative\ proportion - underneath\ cumulative\ proportion} \quad (2)$$

$$score = scale\ Value + |scale\ Value_{min}| + 1 \quad (3)$$

TABLE I
DATA VARIABLE

Indicator	Variable	Units
Export	Number of export (X1)	USD
Tarrifs	Import tariffs (X2) ^{-1*}	% rate
Population	Country Population (X3)	Thousand
Urban population	Percentage of urban population (X4)	% rate
Physical distance	Country distance (X5) ⁻¹	km
GDP growth	GDP growth rate (X6)	% rate
Internet usage	Proportion of individual internet users (X7)	% rate
GDP	GDP per capita (X8)	USD
Final consumption growth	Final consumption growth rate (X9)	% growth rate
Competitiveness	Global competitiveness index (X10)	1-100 score
Cultural distance	Cultural similarity index (X11) ⁻¹	1-100 score
Country risk	Country economic and political risks (X12) [*]	Low, medium, high
Supply chain	Country resilience to supply chains (X13) ^{**}	Very sensitive, sensitive, resilient, very resilient
Income	Country income category (X14)	Low, medium to low, medium to high, high
Economic development	Economic development category (X15)	Developing, transition, developed countries

* 2020 data, ** 2022 data, ⁻¹ Inverse Value

The existence of an inverse value aims to make a meaning that is in line with other variables according to the definition of international trade potential

3) *BCBimax Algorithm*: The idea behind the BCBimax algorithm is to partition the binary data into three submatrices, one of the partitions will only contain 0 elements so they can be discarded. The algorithm is then applied recursively to the remaining two submatrices, namely **U** and **V** which are the results of the step 3 process, the recursion will end if a submatrix containing only 1 elements is formed. As a way to avoid overlapping, the next bicluster search uses an algorithm based on the submatrix that does not include the previous bicluster row. The stages of the algorithm's work are as follows as in Fig. 1.

4) *Optimal bicluster selection*: The evaluation function is an indicator to measure performance on a biclustering algorithm. To choose the most optimal bicluster from several threshold trials, a comparison value is needed. In this study, the Mean Square Residue (MSR) evaluation function will be used as a basis for selecting the optimal bicluster and is defined in (1). Suppose a_{ij} denoted as the average of the row-th i of bicluster (I, J) , a_{Ij} is the average of the column-th j , (I, J) and a_{IJ} is the average of all elements in the bicluster. Based on [5], it can be written as (4).

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

The residues of the elements a_{ij} in the submatrix A_{IJ} are as follows $r_{ij} = a_{ij} - a_{ij} - a_{Ij} + a_{IJ}$. The quality of the bicluster can be evaluated by

calculating the residuals H , which is the sum of all the squared residues of all the elements. The residue H here in after referred to as MSR as follows in (5):

$$MSR(I, J) = a_{IJ} = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (a_{ij} - a_{ij} - a_{Ij} + a_{IJ})^2 \quad (5)$$

where $|I| \times |J|$ is the dimension of the bicluster (volume), ie is the size of $|I|$ the bicluster row and $|J|$ is

the size of the bicluster column. The quality of a bicluster will be better as the residual value decreases

and/or the volume of the bicluster increases. Qualities of the bicluster group the next MSR-based can be measured by calculating the average of MSR divided by the volume or the average MSR per volume and is defined by (6) [20]:

$$\text{average MSR per volume} = \frac{1}{b} \sum_{i=1}^b \frac{MSR_i}{Volume_i} \quad (6)$$

with b is the number of biclusters generated by an algorithm certain.

III. RESULTS AND DISCUSSION

A. Data Exploration

The description of the data related to the initial characteristics of each variable can be seen from the boxplot in Fig. 2(a) Through the boxplot in the figure, it

can be seen that almost all variables have extreme values except for variables X4, X10, and X11 which do not contain outliers. This indicates that no country in the world has an extreme urban population percentage, competitiveness index, and cultural similarity index. Most of the median data tends to be negative (below zero). This indicates that the country in most of the variables concerned tends to be of moderate to low value.

In Fig. 2(b) shows the frequency of the four categorical variables X12 to X15. From the four categorical variables, it can be seen that the proportion of frequency in each value is not significantly different except for X15. At X15, the frequency of countries with code 1, namely developing countries, seems to dominate compared to transitional countries and developed countries.

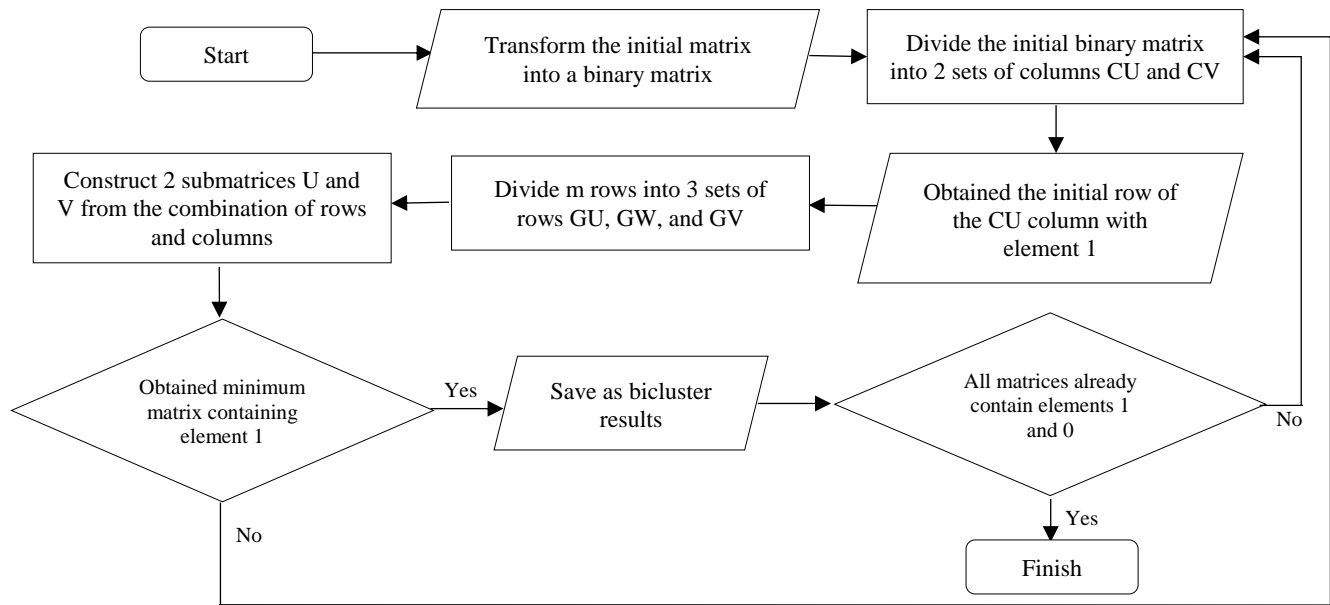


Fig. 1 BCBimax algorithm flowchart diagram

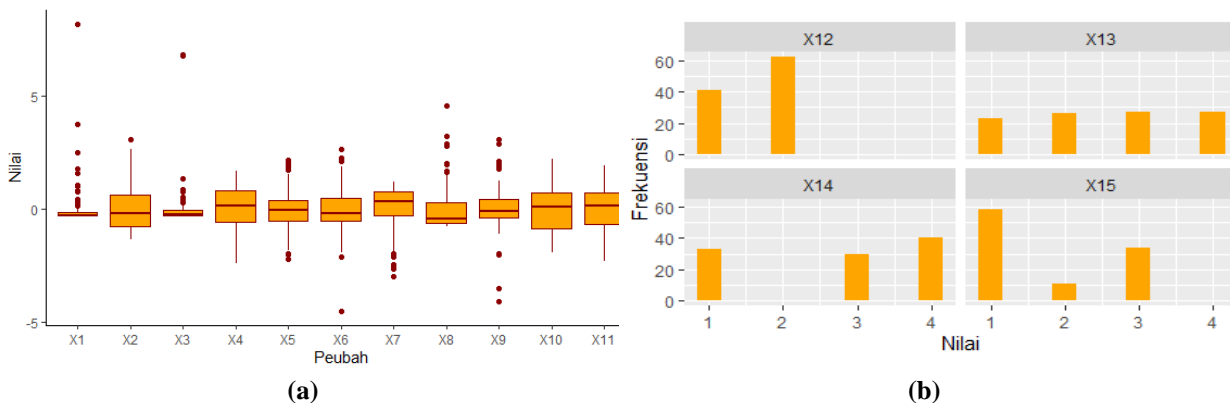


Fig. 1(a) Boxplot matrix of numerical variable scaling data (b) Bar chart of categorical variables

B. Ordinal Data Transformation

Transformation using MSI will change the data scale from ordinal to interval. This is done because the biclustering analysis can only be used for data with an interval scale only or only an ordinal scale.

The results of the MSI transformation are shown in Table II. The variables X13, X14 and X15 are ordinal variables so that the MSI transformation will be carried out, while X12 will not be transformed because it is a binary variable. Variable X13 has 4 categories with almost the same proportion of samples in each category, namely in the range of 0.2, as well as X14, where the proportion of samples in the three categories is not much different. In X15, the proportion is not the same, the majority of countries are developing countries (category 1), then a few countries are transitional countries (category 2). Furthermore, for biclustering analysis, the values of the variables X13, X14, and X15 will be changed with the values from the Interval column which are the resulting values of the transformation.

C. Biclustering Results

In the BCBimax algorithm, a binarization process is required before biclustering analysis is carried out. The binarization process in this study will be carried out in 4 trials, namely: 1) system threshold value, 2) median of all data, 3) median of each variable, and 4) average of all data (which has the same value as the average of each variable). The purpose of the trial was to compare the sample average MSR/volume values of the biclusters formed. This test is applied to the minimum number of rows and columns that are the smallest, namely a minimum row = 2 and a minimum column = 2 as shown in.

Table III shows the results of the comparison of the four threshold test scenarios in the scaling data matrix binarization process. In the first scenario, using the system threshold will produce a binary matrix with a proportion of 0 elements of 0.976. The value of this proportion is very large and unbalanced so it does not form a bicluster output. Furthermore, in scenarios 2 to 4, a fairly balanced proportion value is obtained, which is close to 0.5 and produces a large number of biclusters, so it can be assumed that the binarization of scenarios 2 to 4 can provide sufficient information. In this case, the scenario that has the smallest average MSR/volume value is scenario 4. So that the median value of all data will be used as the basis for binarization as data that will be used for biclustering analysis using the BCBimax algorithm.

The BCBimax algorithm will work by clustering all submatrices whose elements have a value of 1. In this study, 60 minimum threshold combinations of rows and columns will be tested. The minimum column threshold will be taken from 2, 3, to 7, while for the highest threshold, 7 is taken from half of the total number of research variables. The minimum row thresholds generally follow the column minimum thresholds, namely 2, 3, and up to 7. Then for further exploration, a minimum value of rows 10, 13, 15, and 20 will be tried so that there are 60 trial threshold combinations. The highest number of biclusters is 41 resulting from a combination of very small thresholds, namely a minimum row and column of 2. The biclusters that are formed will be bigger when the threshold is smaller, and vice versa, the bigger the threshold value, the biclusters that are formed will be smaller/less.

TABLE II
THE RESULTS OF THE TRANSFORMATION USING THE INTERVAL SUCCESSIVE METHOD

Variable	Category	Proportion	Intervals	Variable	Category	Proportion	Intervals
X13	1	0.223	1.000	X14	1	0.320	1.000
	2	0.252	1.943		2	0.291	2.030
	3	0.262	2.613		3	0.388	3.104
	4	0.262	3.580	X15	1	0.563	1.000
			2		0.107	1.997	
			3		0.330	2.797	

TABLE III
COMPARISON OF BINARIZATION CRITERIA IN THE FORMATION OF BICLUSTERING

No	Binarization scenario	The proportion of elements is "0"	Number of biclusters	Average MSR/volume
1	System threshold (1.81)	0.976	0	-
2	Average of all data (0)	0.546	37	0.013
3	The median of each variable	0.564	36	0.009
4	The median of all data (-0.098)	0.487	41	0.006

The average value of MSR per volume is presented in Fig. 3 which follows the elbow method where this method is commonly used in k-means clustering analysis where the optimal number of clusters will be located at the elbow at a certain point. In this case, the row and column thresholds will produce biclustering information if they have a small average MSR value per volume and form an angle at a point. The results of the calculation of the average MSR per volume that experienced a significant decrease between the 2 threshold points were then followed by a relatively constant value at the minimum threshold in row 7 and column 2 with an average MSR value per volume of 0.0036. This threshold forms a total of 9 biclusters. Thus, the optimal bicluster group from the BCBimax algorithm is at the minimum threshold of row 7 and minimum column 2.

Furthermore, the characteristics of the optimal bicluster results will be presented in Table IV. Based on the table, it can be seen that the BCBimax algorithm in the combination of row 7 and column 2 thresholds will produce clusters of 77 countries or 74.7% in the 9

biclusters formed. There are no overlapping countries in the bicluster or other words the countries in each bicluster are unique. Meanwhile, 26 countries, or 25.3% of the countries in the study that were not included in the bicluster were countries that had a negative average value on the results of standardization or had an average value that was smaller than the mean value of the entire scaling data. This is consistent with the performance of the BCBimax algorithm where this algorithm will only find element "1" which in this study element "1" is the one with a value greater than the median according to the table described in. In addition, another condition for countries that are not included in the bicluster is because these countries do not have the same characteristics of international trade potential as other countries. As for the membership of the variables, it appears that several variables appear in more than one bicluster that is formed. In this case, the overlap in the variables is not a significant problem considering that a country can have several indicators of international trade potential.

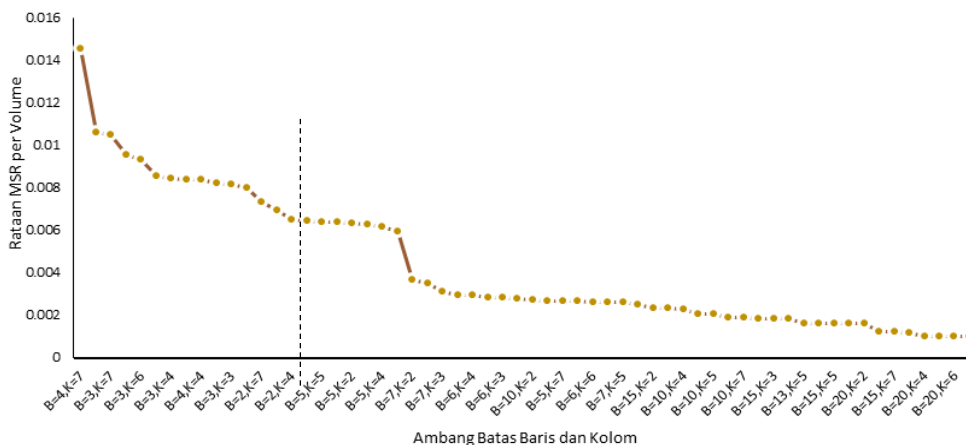


Fig. 2 Average value of MSR per volume

TABLE IV
MEMBERSHIP YIELDS OPTIMAL BICLUSTERING THRESHOLDS

Biclusters	Bicluster size	Membership	
		Country (ISO Alpha-3)	Variable
1	7 × 10	BEL, EST, FRA, LUX, MLT, GBR, USA	X2, X4, X6, X7, X8, X9, X10, X13, X14, X15
2	13 × 9	AUS, DNK, FIN, DEU, GRC, HUN, ITA, JPN, LVA, LTU, NZL, NOR, SWE	X2, X4, X5, X7, X8, X10, X13, X14, X15
3	7 × 8	BGR, HRV, POL, ROU, RUS, SRB, SVN	X2, X4, X5, X7, X8, X10, X13, X14, X15
4	8 × 7	AUT, CAN, IRL, NLD, PRT, SVK, ESP, CHE	X2, X7, X8, X10, X13, X14, X15
5	8 × 6	ARG, BRA, CHL, CRI, ISR, MYS, KOR, URY	X2, X4, X7, X10, X13, X14
6	8 × 6	BIH, CHN, KAZ, MNE, SAU, ZAF, THA, TOUR	X2, X5, X7, X11, X12, X14
7	8 × 5	BGD, ETH, IND, KEN, MDA, PAK, RWA, ZWE	X5, X6, X9, X11, X12
8	9 × 4	ALB, ARM, JOR, LBN, MKD, MNG, OMN, UKR, VNM	X5, X7, X11, X12
9	10 × 3	BEN, BOL, COL, DOM, SLV, GTM, NIC, PAN, PER, SEN	X6, X9, X12

The mapping of international trade potential in Indonesia's export destination countries can be seen in Fig. 4 with a map. In bicluster 1 with orange color, the majority of its members are found on the European continent, namely Belgium, Estonia, France, Luxembourg, Malta, and the United Kingdom, as well as the United States on the American continent. In bicluster 2 the red color is also dominated by countries from Europe, namely Denmark, Finland, Germany, Greece, Hungary, Italy, Latvia, Lithuania, Norway, and Sweden. In addition, there are Japan and Australia as well as New Zealand. This second Bicluster has the most members compared to the other biclusters. Furthermore, in bicluster 3 and 4, it is also dominated by European countries, considering that countries on the European continent also dominate the sample in this study. Most of the Bicluster 5 come from the Americas as well as Bicluster 9. Most countries from the Asian continent are in Bicluster 6 and 8, then Bicluster 7 is dominated by countries from Africa.

Furthermore, to find out the similarity of trading potential, a membership plot will be displayed in Fig. 5.

Each row in the graph shows a variable while the column will represent the bicluster group. The green rectangles depict trade potential variables with the same resemblance so that they are grouped into one bicluster. The brighter the middle color produced, the more countries that have high international trade potential. The light color in this case is interpreted as the color of the light version of the box on the side. The X7 and X14 variables have the brightest color. There are 69.9% of countries with a high proportion of individual internet users and 67.9% of countries with high state income so it can be seen that these two variables appear the most in the biclusters formed, namely in biclusters 1 to 6. Likewise with X2 and X5 where there are 66.9% of countries with low import tariffs and 64.1% of countries with a country that is close to Indonesia. These two variables are also the variables that appear the most in the bicluster. Variables X8 and X15 have a dark color and there are only 34.9% of countries with high per capita GDP and 43.7% of countries with high economic development categories.

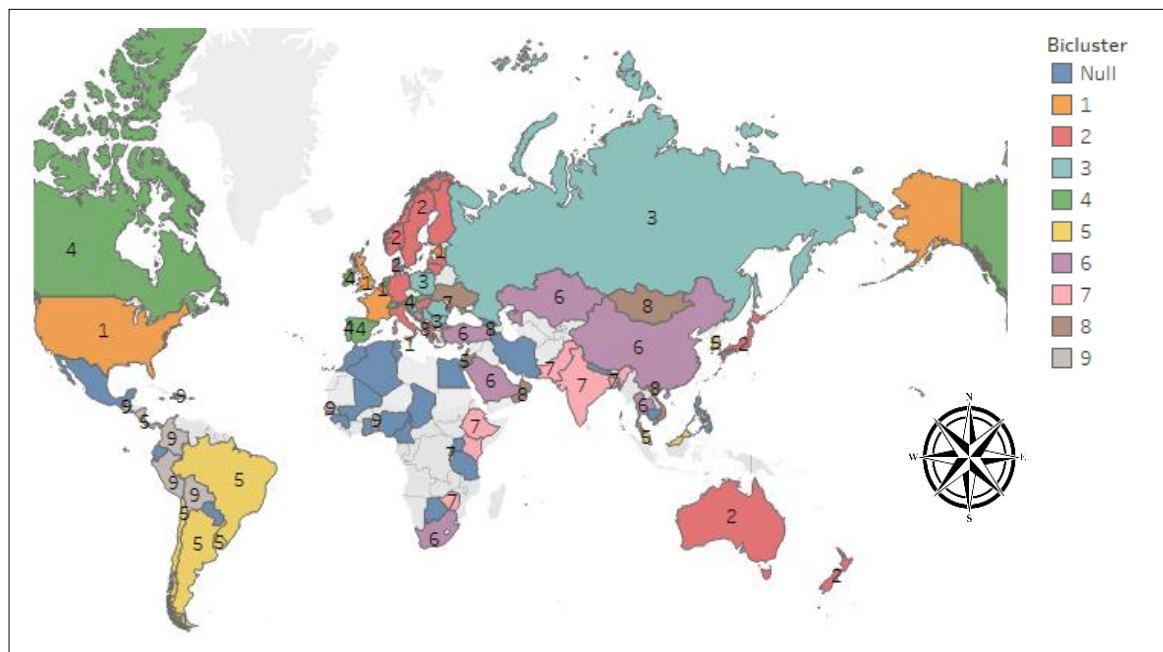


Fig. 3 Distribution map of optimal bicluster result by country without scale

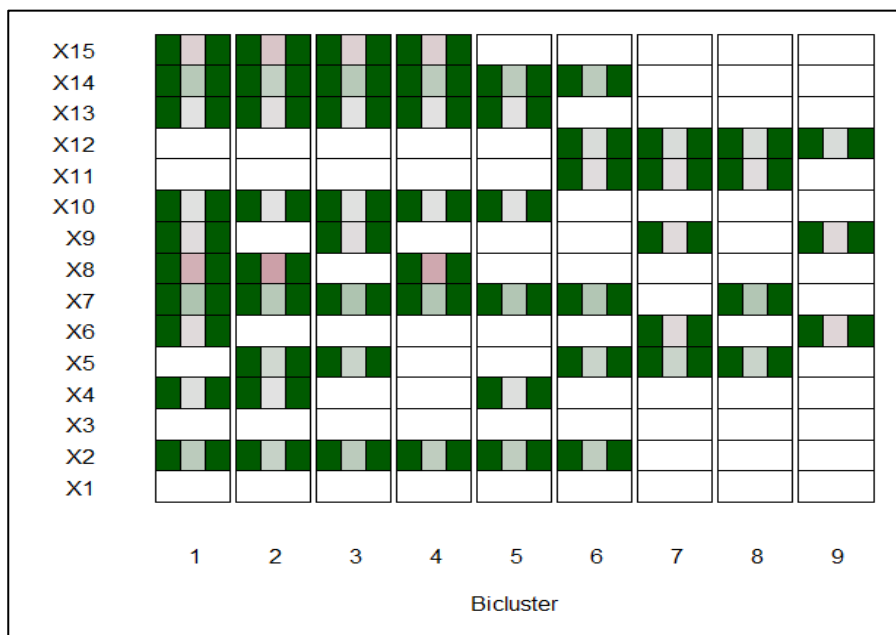


Fig. 4 Biclustering Plot for International Market Potential Variables

IV. CONCLUSION

This study applies the BCBimax algorithm to potential international trade data with 103 countries in the world. Based on an experiment of 60 threshold and row combinations, the BCBimax algorithm produces an optimal bicluster at the combination of rows and columns (7,2) with an average MSR/volume value of 0.0036. The binarization threshold experiment has been carried out and the use of the median value for all data produces the most biclusters and the average MSR/volume which tends to be small compared to the other three experiments. All countries with high trade potential are well clustered and spread out in 9 biclusters. Biclusters are formed have unique country characteristics and the majority of countries on the European continent have good trade potential compared to countries from other continents. Several countries in Africa also have trade potential that is quite good so that it can be used as a consideration for the government to open markets and cooperate. The variables that appear most frequently in the nine biclusters are X2 (import tariffs), X7 (proportion of individual internet users), and X14 (country income category) because the average data on this variable is more from the median, meaning that in this data the majority of countries in the world have import tariffs low, high proportion of individual internet users and state income middle to upper to high.

ACKNOWLEDGMENT

The authors thank the Indonesia Endowment Fund for Education (LPDP) from the Ministry of Finance of the Republic of Indonesia for granting the scholarship and supporting this research.

REFERENCES

- [1] A. H. Foss and M. Markatou, "Kamila: Clustering mixed-type data in R and hadoop," *J. Stat. Softw.*, vol. 83, no. 13, p. 3, 2018, doi: 10.18637/jss.v083.i13.
- [2] L. Hubert and P. Arabie, "Comparing partitions," *J. Classif.*, vol. 2, no. 1, pp. 193–218, 1985, doi: 10.1007/BF01908075.
- [3] A. A. Mattjik and I. M. Sumertajaya, *Sidik Peubah Ganda dengan Menggunakan SAS*. Bogor, West Java, Indonesia, 2011.
- [4] R. C. Tryon and D. E. Bailey, "Cluster Analysis," *Psychol. Med.*, p. 184, 1970, doi: 10.1017/S0033291700000234.
- [5] Y. Cheng and G. M. Church, "Biclustering of expression data," *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, vol. 8, pp. 93–103, 2000.
- [6] 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.
- [7] B. Wang, Y. Miao, H. Zhao, J. Jin, and Y. Chen, "A biclustering-based method for market segmentation using customer pain points," *Eng. Appl. Artif. Intell.*, vol. 47, pp. 101–109, 2016, doi: 10.1016/j.engappai.2016.05.011.

- 10.1016/j.engappai.2015.06.005.
- [8] D. de Smet, "A Biclustering Approach to Symptom Clusters and Subgroup Identification in Non-Hodgkin Lymphoma Survivors," Tilburg University, 2019.
- [9] A. Prelić, Bleuler S, Zimmermann P, Wille A, Bu P, Gruissem W, Hennig L, Thiele L, Zitzler E, "A systematic comparison and evaluation of biclustering methods for gene expression data," *Bioinformatics*, vol. 22, no. 9, pp. 1122–1129, 2006, doi: 10.1093/bioinformatics/btl060.
- [10] A. R. Amna and A. Hermanto, "Implementation of BCBimax algorithm to determine customer segmentation based on customer market and behavior," *Proc. 2017 4th Int. Conf. Comput. Appl. Inf. Process. Technol. CAIPT 2017*, pp. 1–5, 2018, doi: 10.1109/CAIPT.2017.8320694.
- [11] S. Ningsih and H. H. Dukalang, "Penerapan Metode Suksesif Interval pada Analisis Regresi Linier Berganda," *Jambura J. Math.*, vol. 1, no. 1, pp. 43–53, 2019, doi: 10.34312/jjom.v1i1.1742.
- [12] N. Gregory Mankiw, *Pengantar ekonomi makro*. Jakarta: Salemba Empat, 2006.
- [13] A. Kara, "Assessment of Market Potential: a Research on Determining the Potential Markets of Turkish Exporters," *Bus. Manag. Stud. An Int. J.*, vol. 7, no. 5, pp. 2577–2595, 2019, doi: 10.15295/bmij.v7i5.1303.
- [14] D. Huo, "Cluster analysis of market potential in emerging markets: A dynamic research based on Markov chain," *Rom. J. Econ. Forecast.*, vol. 16, no. 4, pp. 218–231, 2013.
- [15] Sustainable Intelligence, "Global Sustainable Competiveness Report," 2022. [Online]. Available: <https://solability.com/the-global-sustainable-competitiveness-index/downloads>.
- [16] G. Hofstede, "Motivation, leadership, and organization: Do American theories apply abroad?," *Organ. Dyn.*, vol. 9, no. 1, pp. 42–63, 1980, doi: [https://doi.org/10.1016/0090-2616\(80\)90013-3](https://doi.org/10.1016/0090-2616(80)90013-3).
- [17] P. Morosini, S. Shane, and H. Singh, "National Cultural Distance and Cross-Border Acquisition Performance," *J. Int. Bus. Stud.*, vol. 29, no. 1, pp. 137–158, 1998, doi: 10.1057/palgrave.jibs.8490029.
- [18] Amfori, "Country Risk Classification," 2022, [Online]. Available: <https://www.amfori.org/resource/country-risk-classification-2022>.
- [19] FM Global, "Resilience Index," 2022. [Online]. Available: <https://www.fmglobal.com/research-and-resources/tools-and-resources/resilienceindex/explore-the-data/?&vd=1>.
- [20] C. A. Putri, R. Irfani, and B. Sartono, "Recognizing poverty pattern in Central Java using Biclustering Analysis," *J. Phys. Conf. Ser.*, vol. 1863, no. 1, 2021, doi: 10.1088/1742-6596/1863/1/012068.

