JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 7, No. 4, October 2023, pp. 1178-1192

Pattern Recognition of Food Security in Indonesia Using Biclustering Plaid Model

Nur Hikmah[1], I Made Sumertajaya[2*], Farit Mochamad Afendi[3]
[1,2,3]Departement of Statistics/Statistics and Data Science, IPB University, Indonesia
[1]Ministry of Agriculture
enhanurnur@apps.ipb.ac.id[1], imsjaya@apps.ipb.ac.id[2*], fmafendi@apps.ipb.ac.id[3]

Article History:
Received : 24-07-2023
Revised : 24-09-2023
Accepted : 02-10-2023
Online : 12-10-2023

Keywords:
Biclustering;
Plaid Model;
MSR;
Layer;
Food Security.

ABSTRACT

Biclustering come in various algorithms, selecting the most suitable biclustering algorithm can be a challenging task. The performance of algorithms can vary significantly depending on the specific data characteristics. The Plaid model is one of popular biclustering algorithms, has gained recognition for its efficiency and versatility across various applications, including food security. Indonesia deals with complex food security challenges. The nation's unique geographic and socioeconomic diversity demands region-specific food security solutions. Identifying province-specific food security patterns is crucial for effective policymaking and resource allocation, ultimately promoting food sufficiency and stability at the regional level. This study assesses the performance of the Plaid model in identifying food security patterns at the provincial level in Indonesia. To optimize biclusters, we explore various parameter tuning scenarios (the choice of model, the number of layers, and the threshold value for row and column releases). The selection criteria are based on the change ratio of the initial matrix's mean square residue to the mean square residue of the Plaid model, the average mean square residue, and the number of biclusters. The constant column model was selected with a mean square residue change ratio of 0.52, an average mean square plaid model residue of 4.81, and it generates 6 overlapping biclusters. The results show each bicluster has unique characteristics. Notably, Bicluster 1 that consist of 2 provinces, exhibits the lowest food security levels, marked by variables X1, X2, X4, and X7. Furthermore, the variables X1, X4, and X7 consistently appear across several biclusters. This highlights the importance of prioritizing these three variables to improve the food security status of the regions.

Crossref
https://doi.org/10.31764/jtam.v7i4.16778

———————— ◆ ————————

A. INTRODUCTION

Clustering is an analytical technique to group similar objects based on their characteristics global (Mandal et al., 2021). It helps to uncover hidden patterns, identify relationships, and gain insights from complex datasets. The primary objective of clustering is to partition the data into homogeneous clusters, where objects within the same cluster are more similar to each other compared to objects in different clusters. Indeed, global clustering may not always be applicable as each object may possess specific subsets of similarity patterns (Padilha & Campello, 2017).

Biclustering analysis, on the other hand, is an extension of clustering that allows for the simultaneous grouping of objects based on both row and column characteristics (Patowary et al., 2020). It recognizes part of the data matrix may exhibit similar patterns or relationships only in specific subsets of objects and subsets of variables namely submatrix (Kléma et al., 2017). Biclustering considers the local characteristics of variables and identifies submatrix that show coherent patterns across both rows and columns (Acharya et al., 2019).

Various studies on biclustering have been performed, but most of them just focus on bioinformatics such as gene expression data (Babichev et al., 2018; Cao et al., 2023; Huang et al., 2020). Recently, biclustering can be implemented in various fields (Silva et al., 2022; Candelise et al., 2021; Kaban et al., 2019; Lin et al., 2019; Novidianto & Irfani, 2020; Wang et al., 2016). Furthermore, biclustering involves a diverse range of algorithms. However, selecting the right algorithm can be a challenge. The performance of biclustering algorithms depends on the specific characteristics of the algorithm and the type of data being used.

One commonly utilized biclustering algorithm is the Plaid model which can yield better results, ease of use and relatively fast (Nicholls & Wallace, 2021). According to Pontes et al. (2015), the Plaid model is an algorithm that works by searching for appropriate model parameters in the grouping process by minimizing the sum of squares of the model residuals. In model-based algorithms, the statistical issue that can be examined is the sensitivity of the algorithm in obtaining the right model parameters to understand the effects that occur in the submatrix. In the previous study, the implementation of the Plaid model only used default parameters while analyzing bicluster. However, this study uses a few parameter scenarios to gain an optimal bicluster.

Plaid model is one of the valuable tools for exploring datasets and extracting meaningful insights for various applications, including food security analysis. Food security is a critical concern in Indonesia, a nation with a diverse archipelago comprising thousands of islands, each with its own unique socio-economic conditions. The country's food security landscape is further influenced by factors such as population growth, poverty rates, and the recent global pandemic COVID -19, which has disrupted supply chains and posed new challenges to access and affordability (Rozaki, 2021). Furthermore, in 2020, Indonesia ranked 65th out of 113 countries in the Global Food Security Index (GFSI) released by The Economic Intelligences Unit (EUI). In 2019, the national prevalence of moderate to severe food insecurity among the population was reported as 5.42% by BPS-Statistics of Indonesia. This figure has since declined to 5.12% in 2020. It's noteworthy that, with a population over 270 million people, at least approximately 12.9 million people in Indonesia are facing food insecurity. These multifaceted dynamics demand a comprehensive approach to understanding and addressing food security issues at various levels, from the national scale down to individual provinces.

One of Indonesia's challenges in maintaining the quality of food security is its geographical condition, thus Indonesia's food security system must be designed to the specific needs and conditions of each region. To address this complexity, there is a need to identify patterns and variations in food security at the province level. Identifying food security patterns in province level of Indonesia holds significant importance because it can describe the characteristics of each specific area. In this context, the application of advanced analytical techniques, such as biclustering using the Plaid model, becomes valuable.

Biclustering using Plaid model can provide insights into the relationship between various food security indicators and geographical regions. It can identify groups of regions that exhibit similar patterns in terms of food availability, food access, and food utilization aspects. It allows for a more precise understanding of the underlying structures within the data, leading to more informed decision-making and targeted interventions in various domains. This aligns with the second Sustainable Development Goals (SDGs) goal, which aims to end hunger, ensure food security, improve nutrition, and promote sustainable agriculture. Therefore, the aim of this study is carrying out research on implementation Plaid model in the case of identifying Indonesian food security in province level.

B. METHODS

This section explains the data used and the steps involved in applying the Plaid model. The main stages of the analysis include data pre-processing, data exploration and biclustering analysis using Plaid model.

1. Data

This research utilized secondary data from the Badan Ketahanan Pangan (BKP)*- Ministry of Agriculture, the Ministry of Health, and SUSENAS (BPS) with variable data based on 2020. The analysis was conducted at the provincial level, with a total of 34 provinces as the observation units. The scope of variables describes three dimensional of food security namely food availability, food access, and food utilization aspect adapt from the Food Security and Vulnerability Atlas (FSVA) analysis (BKP, 2021). Detailed information about these variables is presented in the following Table 1.

Table 1. Research Variable

Aspect	Variable Name (Initials)
Food Availability	The ratio of normative consumption per capita to net production of carbohydrates $(X_1, Ratio)$
Food Access	The percentage of the population who live below line poverty $(X_2, \%)$
	The percentage of households with proportions expenses for food over 65% of the total expenditure $(X_3, \%)$
	Proportion of households without access to electricity $(X_4, \%)$
Food Utilization	Proportion of households without access to clean water $(X_5, \%)$
	Life expectancy $(X_6, Year)^{-1}$
	Ratio of population to healthcare workers against to population density $(X_7, Ratio)$
	Average years of schooling for females above 15 years old $(X_8, Year)^{-1}$
	Proportion of children with stunting criterion $(X_9, \%)$

$^{-1}$Invers value

The utilization of inverse values for variables X_6 and X_8 is intended to align the perception of these variables in the context of food security. A higher value signifies lower food security, or in other words, a higher value will be experiencing food insecurity and vice versa.

2. Research Stages

a. Data Preprocessing

In this stage, we create a data matrix for analysis by scaling the data. Scaling ensures that the variables used have a similar scale, leading to consistent analysis results. Pre-processed matrix scaled data define as $\boldsymbol{A} = (\boldsymbol{N}, \boldsymbol{M})$ where $\boldsymbol{N}$ and $\boldsymbol{M}$ denoting the sets of rows represent the province and set of columns represent the variables.

b. Data Exploration
 In this stage, a descriptive analysis is conducted to gain an initial overview of the data. Exploration is performed by creating a heatmap and PCA biplot. The heatmap provides a visual representation of the data matrix, where each cell's color represents the value of a variable. The PCA biplot displays both the variables and the observations on the same plot, enabling the examination of their relationships.

c. Biclustering Analysis using Plaid Model
 The biclustering process aims to discover subsets of objects in a matrix $\boldsymbol{A} = (\boldsymbol{N}, \boldsymbol{M})$, which is $\boldsymbol{N}$ defined as the set of rows, while $\boldsymbol{M}$ represents the set of columns. The element a_{ij} represents the relationship between the-i^{th} row and the-j^{th} column. Let's assume $\boldsymbol{I} \subseteq \boldsymbol{N}$ while $\boldsymbol{J} \subseteq \boldsymbol{M}$, then $\boldsymbol{B_{ij}}(\boldsymbol{I}, \boldsymbol{J})$ refers to a bicluster or submatrix $\boldsymbol{B}$, with $\boldsymbol{I} = \{i_1, i_2, \dots, i_k\}$ dan $\boldsymbol{J} = \{j_1, j_2, \dots, j_s\}$ (Alberto Magalhães Leite et al., 2016).
 The Plaid is a model-based algorithm that defines the input data matrix as a linear function in layers and its corresponding to a bicluster. The Plaid model is an exceptionally flexible and potent biclustering method based on models, satisfy all essential criteria (Majd et al., 2016). The Plaid model works by partitioning the data into k-layer submatrix formed by subsets of columns and subsets of rows. Each layer represents the effects contributed to the model, resulting in a general model formulation.

$$Y_{ij} = \theta_{ij0} + \sum_{k=1}^{K} \theta_{ijk}\rho_{ik}\kappa_{jk} + \varepsilon_{ij} \tag{1}$$

with Y_{ij} representing the element in the matrix, where, i = 1,2,3,...n; j= 1,2, ...,m; k is the layer/bicluster index starting from 0 for background data up to k biclusters; θ_{ij0} represent the background data effect that consists of sum of overall mean (μ_0), the row-i^{th} effect (α_{i0}) and the column- j^{th} effect (β_{j0}) ; θ_{ijk} represents the sum of the overall mean(μ_k), the row- i^{th} effect(α_{ik}), and the column- j^{th} effect (β_{jk}) in the k-bicluster; ρ_{ik} and κ_{jk} illustrate the membership of rows and columns in the bicluster, and can be overlapping.

$$\sum_i^n \rho_{ik} = \begin{cases} 1 \\ \geq 2 \\ 0 \end{cases} ; \quad \sum_i^m \kappa_{ik} = \begin{cases} 1 \\ \geq 2 \\ 0 \end{cases}$$

ρ_{ik} and κ_{ik} contain 1 while the i^{th} row and j^{th} column in the bicluster. If ≥ 2 occurs when rows or columns are in more than one bicluster. While 0 indicates that the row or column is not in any biclustering. Whereas θ_{ijk} can be μ_k, $\mu_k + \alpha_{ik}$, $\mu_k + \beta_{jk}$ or $\mu_k + \alpha_{ik} + \beta_{jk}$. The algorithm iteratively searches for biclusters until it reaches converged results. At each iteration, the algorithm updates the estimated model parameters. The estimation of model parameters follows the concept of minimizing the model residue using the given formula (2).

$$R = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(\hat{Z}_{ij} - \theta_{ijk}\rho_{ik}\kappa_{jk}\right)^2 \tag{2}$$

where

$$\hat{Z}_{ij} = Z_{ij}^{K-1} = Y_{ij} - \theta_{ij0} - \sum_{k=1}^{K-1} \theta_{ijk}\rho_{ik}\kappa_{jk} \tag{3}$$

$\hat{Z}_{ij}$ represents the initial residue value of matrix $\boldsymbol{A}$. Next, in bicluster pruning, the algorithm will identify the rows and columns that will be part of the bicluster by determining the thresholds row release (τ_1) and column release (τ_2) values within the range of 0 and 1. The application of biclustering with Plaid model requires the specification of several model parameters to obtain optimal results. In this study, the parameters include the type of biclustering model and the number of layers combined with each value of (τ_1) also (τ_2). The selection of model type determines the most fit model for data analysis, while the number of layers affects the complexity and granularity of the biclustering process. Additionally, the row and column release thresholds value for determining which variables and observations should be included in the biclusters in the biclusters (Siswantining et al., 2021).

d. Biclustering Evaluation and Selection

Selecting the optimal biclustering result is crucial. The performance of biclustering is analyzed using the mean square residue (MSR) value (Huang et al., 2020). A lower MSR value indicates a better-formed bicluster, as it reflects higher homogeneity within the bicluster. The MSR is calculated using equation (4)

$$MSR_{\ (I,J)} = \frac{1}{|I||J|} \sum_{i\in I, j\in J} (r_{ij})^2 \tag{4}$$

where

$$r_{ij} = a_{ij} - a_{iJ} - a_{Ij} + a_{IJ} \tag{5}$$

$$a_{iJ} = \frac{1}{J}\sum_{j\in J} a_{ij}\,,\ a_{Ij} = \frac{1}{I}\sum_{i\in I} a_{ij}\ ,a_{IJ} = \frac{1}{|I||J|}\sum_{i\in I, j\in J} a_{ij}$$

a_{iJ} is average value of row i, a_{Ij} is average value of column j and a_{IJ} is average value of all elements in a bicluster. If the biclustering analysis yields multiple biclusters, the average MSR (AvMSR) can be calculated as follows:

$$AvMSR = \frac{1}{n}\sum_{k}^{n} MSR_k\ (I,J) \tag{6}$$

C. RESULT AND DISCUSSION

1. Data Exploration

The analysis begins with exploring the scaled matrix data using a heatmap, as shown in Figure 1.

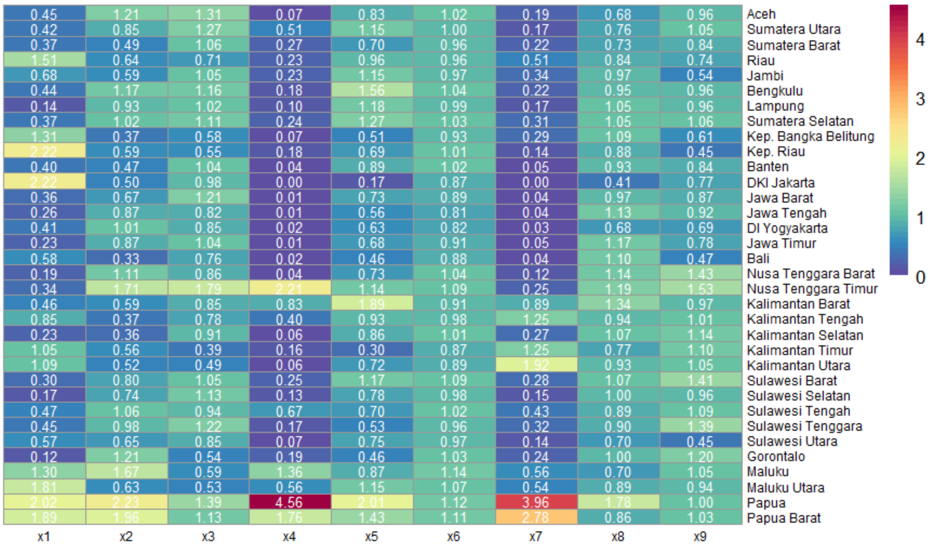

Figure 1. Heatmap Scaled Matrix Data

In this study, the values of each variable have an inline interpretation, where higher values indicate lower food security in the related aspect, or in other words it tends to be more vulnerable. It can be observed that higher values (darker red color) in each aspect of food security variables indicate provinces with lower food security conditions, and vice versa. For example, in the X4 variable (proportion of households without access to electricity) under the food access indicator, Papua Province has a higher value compared to other provinces. This indicates that Papua Province has a higher proportion of households without access to electricity. In other words, it can be inferred that there are households in Papua Province that face limitations in accessing food, as reflected by the high proportion of households without access to electricity. Candelise et al. (2021) stated that access to electricity can directly influence food security in terms of food availability and utilization.

Meanwhile, in Figure 2, a PCA Biplot of the scaled matrix data is presented, dividing provinces into four quadrants. The PCA Biplot illustrates the relative positions of provinces and variables within a two-dimensional framework. Provinces within the same quadrant exhibit similar characteristics in the related variables and differ from provinces in opposing quadrants. For example, in Quadrant I, there are provinces such as North Maluku, Maluku, West Papua, and Papua, which share similar characteristics with high values for variables X1, X7, and X4. On the other hand, provinces in Quadrant III have low values for these three variables. The cumulative total of information diversity explained by the PCA Biplot is 64.6% (PC1=19.6% and PC2=45%). Overall, the diversity depicted is quite good, but it does not fully explain the simultaneous relationship between provinces and variables in grouping the food security conditions of regions. Therefore, the results of the PCA Biplot should only be considered as an initial representation of how provinces are grouped with respect to the variables related to food security aspects.

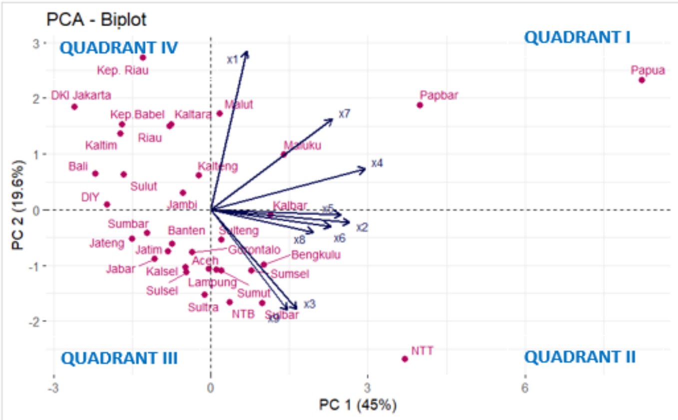

Figure 2. PCA Biplot Result in Scaled Matrix Data

2. Biclustering Result

Biclustering Analysis Using Plaid Model. There are four types of plaid models applied, namely the constant model ($\theta_{ijk} = \mu_k$), the row constant model ($\theta_{ijk} = \mu_k + \alpha_{ik}$), the column constant model ($\theta_{ijk} = \mu_k + \beta_{jk}$) and the two-way model ($\theta_{ijk} = \mu_k + \alpha_{ik} + \beta_{jk}$) . The different thresholds for τ_1 and τ_2 were used ranging from 0.1 to 1 with a step 0.1. This resulted in a total of 10 rows for τ_1 and 10 columns for τ_2. The use of different threshold values is interesting as it helps to obtain optimal biclustering results. While previous studies suggested using thresholds τ_1 and τ_2 between 0.5 and 0.7 (Karim et al., 2019; Siswantining et al., 2021), this study utilized the specified parameter settings. Furthermore, the layer parameter represents the maximum number of layers that can be generated in the process of discovering biclusters. In this study, a total of 9 layers were set, ranging from 2 to 10, to find the optimal biclusters. The combination of model parameters used is described in Table 2, providing a comprehensive overview of the scenarios tested.

Table 2. Scenario Parameter Model

No	Model	Layer	τ_1	τ_2
1	$\theta_{ijk} = \mu_k$	2,3,4,5,6,7,8,9,10	0.1-1 (sequence 0,1)	0.1-1 (sequence 0,1)
2	$\theta_{ijk} = \mu_k + \alpha_{ik}$	2,3,4,5,6,7,8,9,10	0.1-1 (sequence 0,1)	0.1-1 (sequence 0,1)
3	$\theta_{ijk} = \mu_k + \beta_{jk}$	2,3,4,5,6,7,8,9,10	0.1-1 (sequence 0,1)	0.1-1 (sequence 0,1)
4	$\theta_{ijk} = \mu_k + \alpha_{ik} + \beta_{jk}$	2,3,4,5,6,7,8,9,10	0.1-1 (sequence 0,1)	0.1-1 (sequence 0,1)

These models are used to uncover patterns and relationships within the data matrix, allowing for the identification of subsets of objects and variables that exhibit coherent behavior. By appropriately selecting and fine-tuning these model parameters, we aim to obtain meaningful and interpretable biclusters that capture the underlying structure of the data. Based on the given scenarios, each layer of each model generated 100 combinations of biclustering results. The selection of the optimal combination of model parameters in each layer was based on the lowest Average MSR (AvMSR) value. Since there are 9 layers in each model, a total of 9 candidate parameter combinations were obtained for each model. The graph depicting the AvMSR values and the number of biclusters produced for each layer in every model can be seen in Figure 3.

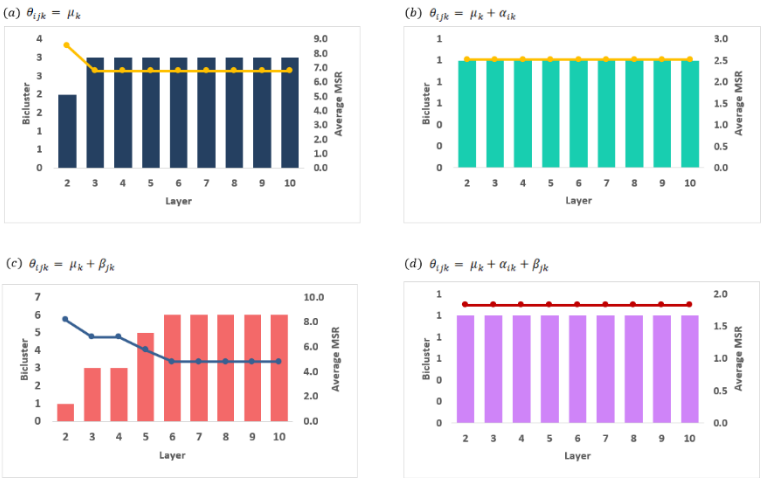

Figure 3. AvMSR in Plaid Model

The two-way model (d) has the lowest AvMSR of 1.83 across all layers but only produces one bicluster. The constant row model (b) also shows constant AvMSR values of 2.52 across all layers and has one bicluster result. The constant column model (c), on the other hand, has a decreasing trend in AvMSR as the number of layers increases. The AvMSR values stabilize at 4.81 from the sixth to the tenth layer. This model yields the highest number of biclusters compared to other models. Meanwhile, the constant model (a) has relatively higher AvMSR values compared to the other models, with an AvMSR of 8.55 in the second layer, decreasing to 6.78 from the third to the tenth layer. In this study, the selection of the appropriate model is first based on the magnitude of the AvMSR change of the Plaid model against the data scaling. If the ratio value is less than 1 $(r < 1)$, it indicates the existence of biclusters/layers in data scaling matrix are better than not performing biclustering, and vice versa. Secondly, it is based on the smallest AvMSR value. Furthermore, another important consideration is how the information obtained from the biclustering results (Wulandari et al., 2023), in other words the model can produce more than one bicluster $(BC > 1)$. Based on Table 3, it can be shown that models 1 and 3 have r value less than 1 $(r < 1)$. Among these two models, model 3 has lower AvMSR in the Plaid model. Therefore, in this study, the constant column model (3) is selected as bicluster optimum with the Plaid model AvMSR of 4.81, $r = 0.52$ and it produces 6 biclusters, as shown in Table 3.

Table 3. AvMSR Ratio in Each Plaid Model

No	Model	AvMSR Data Scaling	AvMSR Plaid Model	Ratio AvMSR (r)	Jumlah BC
1	$\theta_{ijk} = \mu_k$	92.81	6.78	0.07	3
2	$\theta_{ijk} = \mu_k + \alpha_{ik}$	1.84	2.52	1.37	1
3	$\theta_{ijk} = \mu_k + \beta_{jk}$	9.18	4.81	0.52	6
4	$\theta_{ijk} = \mu_k + \alpha_{ik} + \beta_{jk}$	1.25	1.83	1.47	1

Meanwhile, Figure 4 represents the heatmap of the combined values of $\tau_1$and τ_2 for the optimal bicluster. These results were obtained when the layer parameter was set to 6. It can be observed that at τ_1=0.1 and τ_2 =0.4, the lowest AvMSR value was achieved with a total of 6 biclusters. On the other hand, the highest AvMSR value of 12.322 was obtained with several combinations of τ_1 andτ_2. Based on Figure 4, the AvMSR tends to be larger when the values of τ_1 and τ_2 used in the combination are either too small or too large. While, in this study, the AvMSR tends to be smaller when the values of $\tau_1$and τ_2 are set to 0.4 or 0.5, as shown in Figure 4.

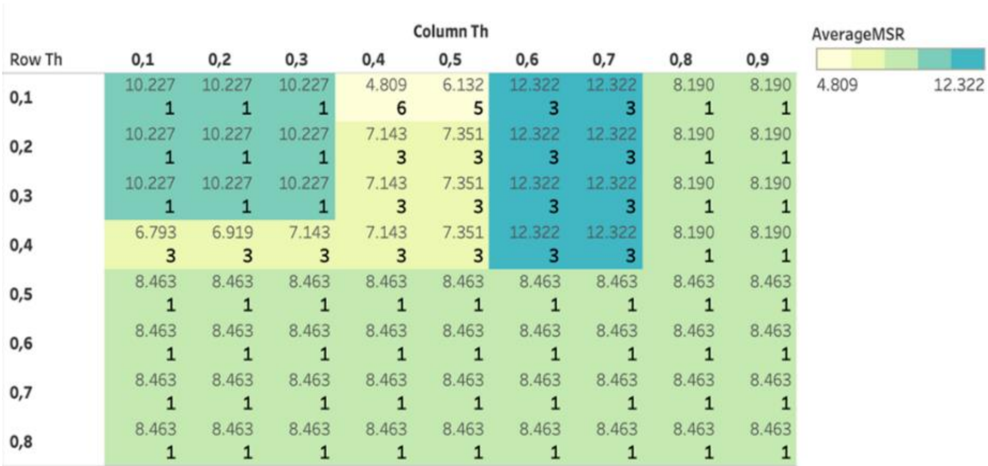

Figure 4. Heatmap of AvMSR and Number of Bicluster in Bicluster Optimum

The results of the optimal bicluster can be observed in Table 4. This model produces 6 biclusters with overlapping membership both in terms of provinces and variables. Bicluster 1, for example, is of size 2x4, indicating the presence of 2 provinces and 4 variables with the same interpretation as the other biclusters, as shown in Table 4.

Table 4. Bicluster Optimal Result of Plaid Constant Column Model

Bicluster	Size	Province	Variable
1	2x4	Papua, West Papua	X1, X2, X4, X7
2	6x1	DKI Jakarta, Bangka Belitung Islands, Riau Islands, Maluku, North Maluku, Riau	X1
3	3x2	Papua, Maluku, East Nusa Tenggara	X2, X4
4	5x2	West Kalimantan, Central Kalimantan, East Kalimantan, North Kalimantan, Papua	X1, X7
5	6x2	Bengkulu, West Kalimantan, East Nusa Tenggara, Papua, West Sulawesi, South Sumatera	X4, X5
6	7x5	Bali, DI Yogyakarta, DKI Jakarta, East Kalimantan, Bangka Belitung Islands, Riau Islands, North Sulawesi	X3, X5, X6, X7, X9

Figure 5 shows the profiling plots for each bicluster. These plots depict the similarity membership within the biclusters. Biclusters become more homogeneous when their profiles are closer and in the same direction. Additionally, Mandal et al. (2021) stated that the relationship profiles among members within a bicluster can be either positively co-expressed (in the same direction) or negatively co-expressed (in opposite directions) with each other. The plots in Figure 5 indicate that the members within each bicluster tend to have positively co-expressed relationships, meaning they exhibit similar trends. For example, in BC 3, the provinces of Papua, East Nusa Tenggara (ENT), and Maluku exhibit a positively co-expressed relationship with variables X2 and X4, as indicated by the decreasing trend of the lines. The profiles also show that the Plaid model successfully creates unique biclusters with distinct characteristic variable.

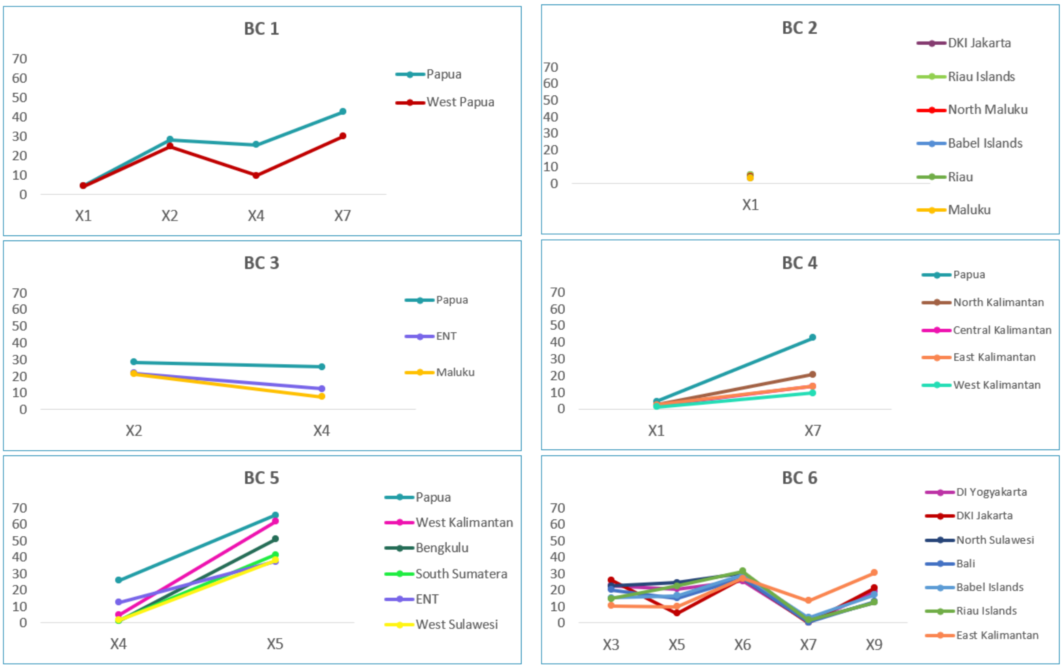

Figure 5. Profiling Plot of Bicluster Variable Result

Furthermore, Table 5 shows the characteristhic of every bicluster. In bicluster 1, the subset of variables (X1, X2, X4, and X7) shows the highest values compared to other biclusters within the same subset of variables. This indicates that Papua and West Papua provinces tend to have higher levels of food insecurity compared to provinces in other biclusters for the related variables. This aligns with the FSVA analysis, which identifies the provinces of Papua and West Papua as highly vulnerable to food insecurity (BKP, 2021). Moreover, studies by Ratnasari et al. (2023) dan Trimanto et al. (2022), have also categorized Papua and West Papua as regions with low food security. This bicluster has a normative food consumption to net production ratio (X1) greater than 1 indicates a food deficit. Furthermore, the variable X7 reveals an average of 1 healthcare worker responsible for an area of 36.19 km^2. In contrast, BC 4 and BC 6, which consist of provinces predominantly located in western and central Indonesia, show lower values for X7. Additionally, variable X2, indicates that 26.50% of the population falls under the poverty line. This finding aligns with the research by Drewnowski (2022) and Ahmad et al. (2021), which highlight poverty as one of the main factors contributing to food insecurity.

BC 2 exclusively includes variable X1 with an average value of 3.90, indicating a normative consumption-to-net production ratio for provinces such as DKI Jakarta, Bangka Belitung Islands, Riau Islands, Maluku, North Maluku, and Riau, which tend to deficits. This result aligns with the findings of Rahayu et al. (2019), who also employed the normative consumption-to-net production ratio variable for mapping food security Indonesia using based clustering model.

In their study, despite DKI Jakarta, Bangka Belitung Islands, Riau Islands, and Riau province being categorized as having high food security, but the X1 variable still low. Furthermore, this bicluster includes island provinces that encounter distinct challenges in terms of food supply and access. These challenges include constraints related to production land availability and food distribution infrastructure (Connell et al., 2020). This suggests that certain areas within these provinces need to ensure food availability by either increasing production through the high food technology or maintaining a stable flow of food distribution between regions. Food distribution between regions can be facilitated by sourcing surplus food from neighboring areas.

In BC 3, the provinces of Papua, Maluku, and East Nusa Tenggara share similar characteristics in terms of food access (X2 and X4). The dominant characteristic is the percentage of the population below the poverty line, which is 23.66%. However, this figure is lower compared to BC 1 but still above the national poverty line percentage in 2020, which is 9.78% (BPS, 2021). Therefore, the provinces in this region can be considered to still have relatively low food security on variable X2.

BC 4 is dominated by provinces in Kalimantan Island with the subset variables X1 and X7. The provinces in this bicluster primarily need attention in terms of food availability (X1) and food utilization (X7). However, when compared to the national average, the average value of variable X7 is much higher than variable X1. Consequently, the provinces within bicluster are more resilient in terms of variable X1 but less resilient in terms of variable X7. Moreover, in Figure 7 (BC 4), it can be observed that Papua province tends to have higher values in variable X7, indicating a lower of healthcare workers than other province. According to Melo et al. (2019), the availability of healthcare indirectly contributes to worsening food insecurity because limited access to healthcare during illness can lead to a decrease in the utilization of quality food. Additionally, it's essential to note that Papua province has overlapping membership in BC 1, with variables X1 and X7 intersecting. This implies that the province is not only vulnerable to variable X7 but also variable X1.

In contrast, BC 5 is characterized by the subset of variables related to food access (X4) and food utilization (X5). The main characteristic variable is X5, suggests that the province within bicluster, on average, 48.96% of households lack access to clean water. The percentage is above the national average, indicating that the provinces in this bicluster tend to have a low food security with respect to variable X5. This aligns with the findings of Dharmawan et al. (2022), which suggest that inadequate access to clean water is a common characteristic among food-insecure households. Ratnasari et al., (2023) similarly, categorized West Kalimantan, East Nusa Tenggara, Papua, and West Sulawesi as having very low-level food security, while Bengkulu and South Sumatra fell into the moderate-level food security group. These categories also exhibited characterized by a high percentage of households without access to clean water.

Meanwhile, in BC 6, there are 7 provinces characterized by variables X3, X4, X5, X6, X7, and X9. This bicluster exhibits the lowest average values compared to other biclusters, all of which are below the national average. This indicates that the regions in this bicluster tend to have a higher food security level in the related subset of characteristic variables. Nevertheless, on average, there are still 18.75% of households with food expenditure more than 65% and 17.9% of children aged under 5 years experiencing stunting. Additionally, the provinces of DKI Jakarta, Bangka Belitung Islands, and Riau overlap with bicluster 2, indicating areas at risk of food insecurity concerning to food availability aspect (X1). This aligns with the Food Security Index (IKP), which categorizes these provinces as moderately (Riau Islands) and highly resilient (DKI Jakarta and Bangka Belitung Islands), yet emphasizes the importance of addressing food availability in these regions. (BKP, 2021), as shown in Table 5 and Figure 6.

Table 5. The Level of Average Values of Food Security Variables with in Each Bicluster

	X1[A]	**X2**[Ac]	**X3** [Ac]	**X4** [Ac]	**X5** [U]	**X6** [U]	**X7** [U]	**X8** [U]	**X9** [U]
BC1	4.42	26.50		17.69			36.19		
BC2	3.90								
BC3		23.66		15.16					
BC4	2.47						19.95		
BC5				7.71	48.96				
BC6			18.75		16.33	28.17	2.91		17.90
National Average	1.70	11.07	24.85	2.58	29.19	30.55	5.74	3.32	26.15

[A] : Food Availability Aspect, [Ac]: Food Access Aspect, [U]: Food Utilization Aspect

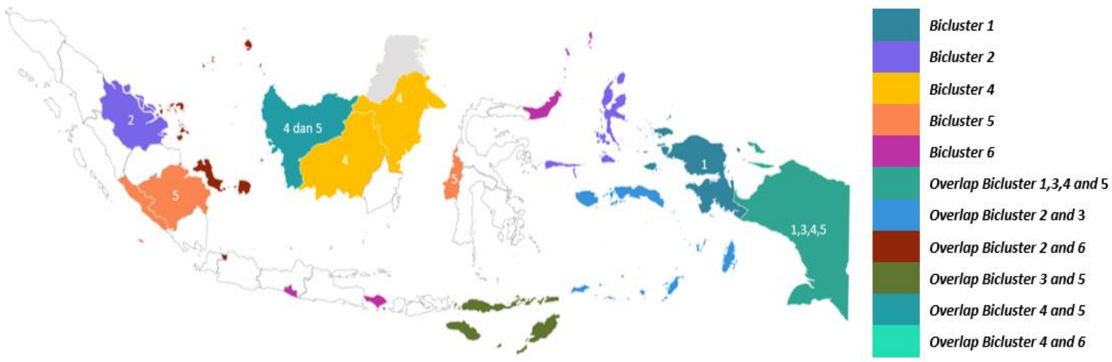

Figure 6. Maps of Bicluster Result in Plaid Constan Column Model

Figure 6 illustrates the mapping of regions based on the results of the Plaid model biclustering. The grouped provinces are predominantly from the eastern regions of Indonesia, Kalimantan Island and parts of Sulawesi also Sumatra Islands. The six overlapping biclusters subsequently produce 11 spatial patterns. Overall, the Plaid model groups the 19 provinces in Indonesia into bi-clusters based on subsets of characteristic variables that exhibit higher values, as shown in the heatmap in Figure 2. In this research, higher values are associated with increased food insecurity risk at the provincial level. As a result, the Plaid model tends to generate biclusters that consist of provinces with higher levels of food insecurity for sepcific variables. The variables X1, X4, and X7 are subsets of variables that frequently appear in several generated biclusters. On the other hand, provinces that have low or moderate values on each variable are not grouped into any bicluster. It can be seen in Figure 6, that fifteen provinces scattered across Java, Sumatra, and Sulawesi are not part of any bicluster.

D. CONCLUSION AND SUGGESTIONS

The application of the Plaid model in biclustering analysis categorizes some provinces in Indonesia into 6 overlapping biclusters based on food security aspects. By identifying characteristic variables within each bicluster, we are able to group 19 provinces that share similar local characteristics. Most provinces are tend to have a low food security and characterized by different variable. Therefore, efforts to maintain food security must be region-specific. Among the six formed biclusters, bicluster 1 stands out as requiring more attention in terms of food security preservation.

Based on this research, the column constant model is considered suitable for several reasons, including the change in AvMSR ratio, the evaluation of AvMSR value, and the information derived from the biclusters. This model was obtained with a configuration of 6 layers, τ_1=0.1 dan τ_2 = 0.4. The selection of the appropriate model type, number of layers, and row and column release thresholds is crucial in optimizing the Plaid model's performance in generating meaningful biclusters. This ensures for resulting coherent and informative bicluster. This study through Plaid model biclustering serves as a valuable tool, in this context, aiding policymakers in making informed decisions to improve food security and enhance overall resilience in the selected provinces by its characteristics within bicluster. This enables us to address food insecurity more effectively by making specific policies and interventions to the unique conditions within each bicluster.

ACKNOWLEDGEMENT

The author would like to express sincere gratitude to the funding organizations, The Agricultural Human Resources Extension and Development Agency (PPSDMP Agency) - Ministry of Agriculture that supported this research.

REFERENCES

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

Ahmad, N., Shahnawaz, S. K., & Alam, Z. (2021). IAR Journal of Humanities and Social Science Food Insecurity: Concept, Causes, Effects and Possible Solutions. *IAR Journal of Humanities and Social Science*, *2*(1), 105–113. https://doi.org/10.47310/jiarjhss.v02i01.016

Alberto Magalhães Leite, C., Luís Fernando Rainho Alves Torgo, O., Associado, P., & Maria Pinto Mora Pinto de Magalhães, C. (2016). *Domain Oriented Biclustering Validation*.

Babichev, S., Osypenko, V., Lytvynenko, V., Voronenko, M., & Korobchynskyi, M. (2018). Comparison Analysis of Biclustering Algorithms with the use of Artificial Data and Gene Expression Profiles. *2018 IEEE 38th International Conference on Electronics and Nanotechnology, ELNANO 2018 - Proceedings*, 298–304. https://doi.org/10.1109/ELNANO.2018.8477439

BKP. (2021). *Peta Ketahanan dan Kerentanan Pangan (Food Security and Vulnerability Atlas) 2021*. Kementerian Pertanian. Jakarta.

BKP. (2021). *Indeks Ketahanan Pangan 2021*. Kementerian Pertanian. Jakarta.

BPS. (2021). *Profil Kemiskinan di Indonesia September 2020*. BRS, 16, 27-42, [Online]. Available: https://ipb.link/kemiskinan20.

Candelise, C., Saccone, D., & Vallino, E. (2021). An empirical assessment of the effects of electricity access on food security. *World Development*, *141*, 105390. https://doi.org/10.1016/j.worlddev.2021.105390

Cao, S., Chang, W., Wan, C., Lu, X., Dang, P., Zhou, X., Zhu, H., Chen, J., Li, B., Zang, Y., Wang, Y., & Zhang, C. (2023). Pipeline for Characterizing Alternative Mechanisms (PCAM) based on bi-clustering to study colorectal cancer heterogeneity. *Computational and Structural Biotechnology Journal*, *21*, 2160–2171. https://doi.org/10.1016/j.csbj.2023.03.028

Connell, J., Lowitt, K., Saint Ville, A., & Hickey, G. M. (2020). Food Security and Sovereignty in Small Island Developing States: Contemporary Crises and Challenges. In *Food Security in Small Island States*, 1–23. Springer Singapore. https://doi.org/10.1007/978-981-13-8256-7_1

Dharmawan, H., Sartono, B., Kurnia, A., Hadi, A. F., & Ramadhani, E. (2022). A Study of Machine Learning Algorithms to Measure The Feature Importance In Class-Imbalance Data Of Food Insecurity Cases In Indonesia. *Communications in Mathematical Biology and Neuroscience*, *2022*, 101. https://doi.org/10.28919/cmbn/7636

Drewnowski, A. (2022). Food insecurity has economic root causes. In *Nature Food* 3(8), 555–556. Springer Nature. https://doi.org/10.1038/s43016-022-00577-w

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

Karim, M. B., Kanaya, S., & Altaf-Ul-Amin, M. (2019). Implementation of BiClusO and its comparison with other biclustering algorithms. *Applied Network Science*, *4*(1), 79. https://doi.org/10.1007/s41109-019-0180-x

Kléma, J., Malinka, F., & železný, F. (2018). Semantic biclustering for finding local, interpretable and predictive expression patterns. *BMC Genomics*, *18*(S7), 752. https://doi.org/10.1186/s12864-017-4132-5

Lin, Q., Zhang, H., Wang, X., Xue, Y., Liu, H., & Gong, C. (2019). A novel parallel biclustering approach and its application to identify and segment highly profitable telecom customers. *IEEE Access*, *7*, 28696–28711. https://doi.org/10.1109/ACCESS.2019.2898644

Majd, H. A., Shahsavari, S., Baghestani, A. R., Tabatabaei, S. M., Bashi, N. K., Tavirani, M. R., & Hamidpour, M. (2016). Evaluation of Plaid Models in Biclustering of Gene Expression Data. *Scientifica*, *2016*, 1–8. https://doi.org/10.1155/2016/3059767

Mandal, K., Sarmah, R., & Bhattacharyya, D. K. (2021). PopBIC: Pathway-based order preserving biclustering algorithm towards the analysis of gene expression data. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *18*(6), 2659–2670. https://doi.org/10.1109/TCBB.2020.2980816

Nicholls, K., & Wallace, C. (2021). Comparison of sparse biclustering algorithms for gene expression datasets. *Briefings in Bioinformatics*, *22*(6), 1–16. https://doi.org/10.1093/bib/bbab140

Novidianto, R., & Irfani, R. (2020). Bicluster CC Algoritm Analysis to Identify Patterns of Food Insecurity in Indonesia. *Jurnal Matematika, Statistika Dan Komputasi*, *17*(2), 325–338. https://doi.org/10.20956/jmsk.v17i2.12057

Padilha, V. A., & Campello, R. J. G. B. (2017). A systematic comparative evaluation of biclustering techniques. *BMC Bioinformatics*, *18*(1), 55. https://doi.org/10.1186/s12859-017-1487-1

Patowary, P., Sarmah, R., & Bhattacharyya, D. K. (2020). Developing an effective biclustering technique using an enhanced proximity measure. In *Network Modeling Analysis in Health Informatics and Bioinformatics,* 9(1), 6. Springer. https://doi.org/10.1007/s13721-019-0211-7

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

Rahayu, R. S., Purwaningsih, Y., & Daerobi, A. (2019). *Mapping Of Provincial Food Security In Indonesia Using Based Clustering Model*. Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi Dan Pembangunan, *20*(1), 69-79. https://doi.org/DOI:10.23917/jep.v20i1.7096

Ratnasari, V., Tri, A., & Dani, R. (2023). Mapping the Provincial Food Security Conditions in Indonesia Using Cluster Ensemble-Based Mixed Data Clustering-Robust Clustering with Links (CEBMDC-ROCK). *International Journal on Advanced Science, Engineering and Information Technology*, *13*(2), 611–617. https://doi.org/DOI:10.18517/ijaseit.13.2.16457

Rozaki, Z. (2021). Food security challenges and opportunities in indonesia post COVID-19. In *Advances in Food Security and Sustainability.* Elsevier Ltd, 6, 119–168.. https://doi.org/10.1016/bs.af2s.2021.07.002

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

Siswantining, T., Eriza Aminanto, A., Sarwinda, D., & Swasti, O. (2021). Biclustering analysis using plaid model on gene expression data of colon cancer. *Austrian Journal of Statistics*, *50*(5), 101–114. https://doi.org/10.17713/ajs.v50i5.1195

Trimanto, R., Yustari, E., Nafisah, Z., Carolina, N., Irsalinda, N., & Setyorini, A. I. (2022). Indonesian provincial clustering using Elbow method for the national food security during pandemic. *Bulletin of Applied Mathematics and Mathematics Education*, *2*(2), 51–58. https://doi.org/10.12928/bamme.v2i2.6166

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

Wulandari, C., Sumertajaya, I. M., & Aidi, M. N. (2023). *Evaluation of Bicluster Analysis Results in Capture Fisheries Using the BCBimax Algorithm*. *11*(1), 57–66. https://doi.org/10.30595/juita.v11i1.15457