Scientific Journal of Informatics
Vol. 12, No. 3, August 2025
p-ISSN 2407-7658 https://journal.unnes.ac.id/journals/sji/index e-ISSN 2460-0040

Biclustering-Based Analysis to Identify Fruit Production Potential in Indonesia Using Plaid Model Algorithm

Nadira Nisa Alwani[1*], I Made Sumertajaya[2], Aji Hamim Wigena[3]
[1,2,3]Department of Statistics, Institut Pertanian Bogor University, Indonesia

Abstract.
Purpose: The application of biclustering using the plaid model aims to simultaneously identify mapping or grouping patterns of provinces and fruit type in Indonesia. The performance evaluation of the plaid model algorithm is used to assess its capability to discover and generate optimal biclusters, thereby representing the relationship between regions and fruit types with similar production characteristics.
Methods: The plaid model algorithm produces optimal biclusters by configuring parameter scenarios such as model selection, managing the number of layers, and determining threshold values for rows and columns. The Average Mean Square Residue (MSR) value and the number of biclusters that can provide the most relevant data are used to determine the optimal parameter selection.
Result: The plaid model algorithm effectively grouped provinces and fruit varieties into multiple biclusters. The row-constant model was choosen based on the average MSR value of 2.0537, which formed five overlapping biclusters across provinces and fruit types. Several provinces, such as Central Java and West Java, demonstrated a high potential for rose apples, breadfruit, and salak. Other provinces showed comparatively moderate levels of production.
Novelty: This study presents a novel way to apply the plaid model biclustering algorithm to data on fruit varieties in various Indonesian provinces. Rarely used in horticulture, this method offers an alternative perspective on structured commodity mapping, especially when identifying specific patterns between fruit varieties and geographic distribution.

Keywords: Biclustering, Average mean square residue, Plaid model
Received June 2025 / **Revised** July 2025 / **Accepted** July 2025

INTRODUCTION

The agricultural sector is important in regional economic growth because it can contribute to the largest Gross Regional Domestic Product (GRDP) [1]. Horticulture is part of the national agricultural sector system that contributes significantly to food availability, increasing economic value, and the welfare of farmers. Indonesia is geographically located on the equator, which provides advantages in horticultural development, especially in tropical fruit commodities [2]. Fruits are part of the food needs of the Indonesian people, along with increasing nutritional awareness, which needs to be supported by adequate fulfilment capabilities [3]. The main challenges include high population growth, urbanization and lifestyle changes, limited arable land, dependence on food imports, and weak infrastructure and distribution [4].

According to the Central Statistics Agency (BPS) in 2022 [5], the high import of fruits such as apples, oranges, and grapes became a national strategic issue, even though these fruits can be cultivated domestically. The lack of accurate and integrated production data triggers this. To overcome this, an efficient data analysis is needed that can find hidden patterns in fruit production data to support distribution, development of superior commodities, and proper production planning. The clustering method is one of the main techniques in data mining used to group data based on the similarity of patterns between objects in one dimension (rows or columns) [6]. This method is often used in bioinformatics, economics, and agriculture to extract information from large and complex data sets.

However, clustering methods have limitations when applied to two-dimensional data, such as the data matrix between provinces and fruit commodity types. One-way clustering techniques alone cannot efficiently organize rows and columns in a data matrix that shows similar patterns [7]. In overcoming these

*Corresponding author.
Email addresses: nadiranisanadira@apps.ipb.ac.id (Alwani)*, imsjaya@apps.ipb.ac.id (Sumertajaya), aji_hw@apps.ipb.ac.id (Wigena)
DOI: 10.15294/sji.v12i3.25054

limitations, biclustering can identify provinces and fruit commodities with similar production patterns because the biclustering method is a more adaptive alternative and can cluster rows and columns simultaneously [8]. Until now, there are no specific guidelines for determining the best bicluster algorithm, so the process of selecting the bicluster algorithm is based on the superior characteristics of each biclustering algorithm used [9].

A biclustering algorithm ideally searches for every statistically significant bicluster in a data set [10]. In biclustering, overlapping patterns are allowed, where an object may belong to more than one group, and some samples may remain unclustered [11]. The data matrix is split based on algorithms to find submatrices [12]. One of the biclustering algorithms that can be used is the plaid model, as this approach allows one province or type of fruit to belong to more than one bicluster pattern. The plaid model is easy to use and has a relatively high computational speed. It is a good choice for various research applications, especially those involving large and complex data sets [13].

Despite its many advantages, the plaid model algorithm has drawbacks, including sensitivity to model type and parameter settings that can affect the biclustering results. Choosing the correct parameters can be challenging when using the plaid model to generate optimal bicluster [7]. The plaid model algorithm uses four model configurations and varies the row, column, and layer parameters to analyze production data of various types of fruits in Indonesia. Based on the model that meets all the requirements, the plaid model is a highly versatile and effective biclustering method [14]. The effectiveness of the algorithm's performance in generating optimal bicluster is evaluated by calculating the average Mean Squared Residue (MSR) and performing bicluster profile analysis.

An earlier study first discovered by Cheng and Church [15] was to apply biclustering to non-agricultural fields such as gene expression data, creating clusters from combinations of rows and columns with the lowest Mean Squared Residue (MSR) values, indicating a high degree of similarity. Since then, the method has evolved and been used in several fields, including assessing the effectiveness of biclustering algorithms in detecting patterns of susceptibility patterns [16], biclustering on biological and biomedical data [17], identifying distribution patterns of Indonesian Export Goods [18], self-learning evolutionary approaches for biomedical data [19], and assessing the potential of capture fisheries [20].

In addition, Ira Audita [21] developed a tropical fruit data mapping system with data mining techniques to predict production distribution. However, the model is based only on single-dimensional clustering and does not consider the relationship between commodity groups and regions. Research by Suyeon Lee [22] said that the lack of comprehensive data is an obstacle in identifying opportunities and challenges and formulating effective policies to encourage the positive impact of agriculture on the environment and sustainable development goals. No study has applied the Plaid biclustering model to national fruit production data, making it a novel approach. This model has the potential to reveal local patterns between provinces and fruit types, support import reduction policies, strengthen data governance, and support distribution optimization in the horticulture sector. Performance evaluation of the Plaid model algorithm is used to assess its ability to find and generate optimal bicluster that can represent the relationship between regions and fruit types with similar production characteristics.

METHODS

The data were obtained from the Central Bureau of Statistics (Badan Pusat Statistik) in 2022 and consist of secondary data on horticultural fruit crops in Indonesia, which can be accesed at https://ipb.link/produksi-buah-bps-2022. The dataset consists of 24 fruits varieties measured in tons, and 34 provinces. Several stages carried out in this analysis are data preprocessing, data exploration, data analysis using biclustering with the plaid model algorithm, and biclustering performance evaluation.

Data preprocessing and exploration

In the preprocessing phase, a size $\boldsymbol{A}_{34\text{x}24}$ matrix was created, with 24 columns representing the 34 rows represent provinces, as listed in Table 1 below. This step is carried out to form the initial structure of the data in a two-dimensional format that is ready to be analyzed using the biclustering technique. The data were then standardized using a standard normal distribution to ensure that all variables used were on the same scale, thereby enabling consistent analytical results. This standardization is important to prevent large-scale variables from dominating the analysis process and to prevent the visualization results from being biased towards extreme values. Principal Component Analysis (PCA) biplot and a heatmap were used to explore the data. A heatmap is a method for visualizing data that displays data distribution in a graphical format to help find pattern and trends, giving a premilinary graphic depiction of the connection between fruit production and geography [23]. Before further modelling, these heatmap help identify extreme values and early indications of the relationship between provinces and fruit types. A statistical method for visually summarizing multivariate data is the PCA biplot, which details variables and samples [24]. Based on the condensed features of the variables, the Principal Component score biplot was utilized to display the regional clustering patterns. PCA was chosen because it can simplify the high-dimensional data structure by summarizing the most significant variance into several main components so that the relationship pattern between regions and fruit types can be visualized more clearly through biplot graphs to interpret biclustering results [25].

Table 1. Data matrix of horticultural production of fruit types

Province (Row)	Research Variables (Column)
34 Provinces	Avocado, Starfruit, Duku, Durian, Guava, Rose Apple, Mandarin Orange, Pomelo, Mango, Mangosteen, Jackfruit, Pineapple, Papaya, Banana, Rambutan, Salak, Sapodilla, Soursop, Breadfruit, Melon, Watermelon, Apple, Grape, and Strawberry.

Biclustering analysis using the plaid model

The Plaid model was chosen as a suitable approach due to its ability to handle overlapping patterns between rows and columns and its flexibility in representing the two-dimensional data structure more thoroughly on provincial and fruit commodity data. In addition, this model organizes bicluster in the form of layers (layered model) [26], so it is more flexible in capturing two-dimensional data structure than other biclustering methods such as Cheng & Church [15] or spectral biclustering [9]. Based on combinations of rows and columns, the model divides the data into multiple k-layer submatrices, each representing its contribution to the overall model:

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

Y_{ij} represents an element in the matrix with indices $i = 1,2,3,\ldots,n$ and $j = 1,2,3,\ldots,m$. The value k indicates the layer index, starting from 0 for the background data to k *biclusters*, θ_{ij0} represents the background effect, which is the sum of $\mu_0, \alpha_{i0}, \beta_{j0}$. Where μ_0 is the overall mean (grand mean). A bicluster's row and column membership are denoted by ρ_{ik} and K_{jk} which may overlap. If the output values of ρ_{ik} and K_{jk} for observation i and variable j are both equal to 1, it means the data point belongs to a single bicluster; if the value is ≥ 2, the variable and object belong to multiple biclusters and if the value is 0, the data point does not belong to any bicluster. This algorithm includes several model types within the background effect (θ_{ij0}) [27]. Apart from background modeling, the plaid model method also influenced by threshold values τ_1 and τ_2 which help obtaining optimal biclustering results. The plaid model assumes that the data consists of overlapping layers, with each layer potentially representing a single bicluster. The layer parameter indicates the maximum number of layers generated to detect biclusters. This study uses threshold values ranging from 0.1 to 1.0 and applies layers from 2 to 10. Table 2 provides an overview of the parameter combinations used in the model:

Table 2. Plaid model parameters

No	Model Types	θ_{ij0}	*Layer*	τ_1	τ_2
1	Constant	$\theta_{ijk} = \mu_k$	2 – 10	0,1 – 1	0,1 – 1
2	Row Constant	$\theta_{ijk} = \mu_k + \alpha_{ik}$	2 – 10	0,1 – 1	0,1 – 1
3	Column Constant	$\theta_{ijk} = \mu_k + \beta_{jk}$	2 – 10	0,1 – 1	0,1 – 1
4	Row and Column Constant	$\theta_{ijk} = \mu_k + \alpha_{ik} + \beta_{jk}$	2 – 10	0,1 – 1	0,1 – 1

Each algorithm iteration updates the model parameters through an estimation process to minimize the residual value [28]. The biclustering method using the plaid model begins by identifying a residual model (Z), which initially consists only of the entire data matrix and does not contain biclusters. Described by the following equation (3), the residual model (Z) is an NxM matrix with residual values Z_{ij}, obtained by subtracting the input data from the effects of the model's rows and columns:

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

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

A pruning stage for biclusters based on user-defined threshold values (τ_1 dan τ_2) is part of the Plaid Model algorithm. This step aims to use threshold parameters to evaluate each row and column in the candidate layer. The values of (τ_1 dan τ_2) range from 0 to 1, and the more closely the threshold value approach 1, the more coherent the resulting bicluster. After passing this pruning stage, candidates are deemed valid and go through a backfitting procedure to find a new bicluster. The algorithm terminates when no further variables or objects remain unpruned [27]:

$$\rho_i = \begin{cases} 1, \ if\ \rho_i = 1\ and\ \sum_{j:\hat{k}_j=1}\left(\hat{Z}_{ij} - \hat{\theta}_{ij}\right)^2 < (1-\tau_1)\sum_{j:\hat{k}_j=1}\left(\hat{Z}_{ij}\right)^2 \\ 0, \ other \end{cases} \quad (4)$$

$$k_j = \begin{cases} 1, \ if\ k_i = 1\ and\ \sum_{i:\hat{\rho}_i=1}\left(\hat{Z}_{ij} - \hat{\theta}_{ij}\right)^2 < (1-\tau_2)\sum_{i:\hat{\rho}_i=1}\left(\hat{Z}_{ij}\right)^2 \\ 0, \ other \end{cases} \quad (5)$$

Bicluster performance evaluation

Data profiling and the Average Square Residue (ASR) are used to assess the algorithm's performance. Combining different threshold values for rows and columns, the plaid model biclustering algorithm's performance is evaluated to identify the best biclustering. Each threshold combination produces different biclusters. Based on the studies of Jiong Yang [29] and Youngrok Lee [30], ASR is used as an evaluation metric to assess the overall quality of the biclusters generated from different threshold combinations. ASR is calculated based on the Mean Square Residue (MSR), using the following formula:

$$MSR_{(I,J)} = \frac{1}{|I|\text{x}|J|}\sum_{i=1}^{|I|}\sum_{j=1}^{|J|}\left(a_{ij} - a_{iJ} - a_{Ij} + a_{IJ}\right)^2 \quad (6)$$

$$ASR = \frac{1}{n}\left(\sum_{i=1}^{n} E_{MSR_i}(I,J)\right) \quad (7)$$

The dimension of a bicluster is expressed as $|I|\text{x}|J|$ representing the product of the number of rows $|I|$ and the number of columns $|J|$ in the bicluster. The smaller the ASR or MSR value ideally approaching zero the better the quality of the resulting bicluster [30].

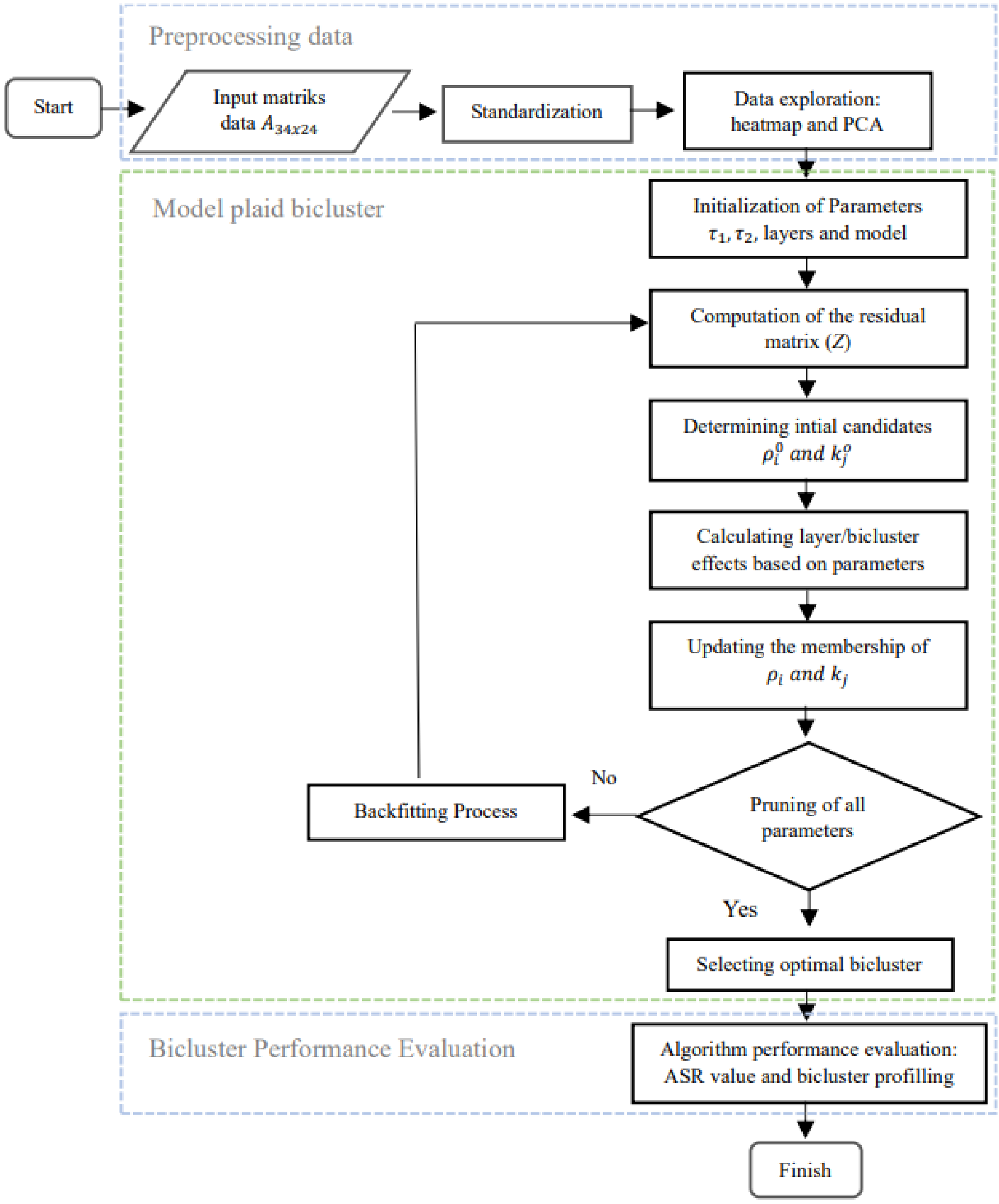

Figure 1. Flowchart

RESULTS AND DISCUSSIONS

Based on previously standardised matrix data, empirical data exploration using a 34×24 matrix was conducted through a heatmap and a PCA biplot. The heatmap provides an overview of the relationships between provinces and types of fruits. Figure 2 shows that several provinces exhibit extreme values across various very high or low variables. Green indicates high positive extremes, representing high vulnerability or potential, while brown indicates negative extremes, representing low values. Yellow shades indicate moderate levels. For example, East Java shows high apple production, whereas Jakarta shows low duku production. Additionally, fruit production across provinces is uneven some regions dominate in producing specific fruits, while others show generally lower output. Therefore, the heatmap is useful for identifying leading provinces for particular fruit types and understanding Indonesia's overall distribution pattern of fruit production.

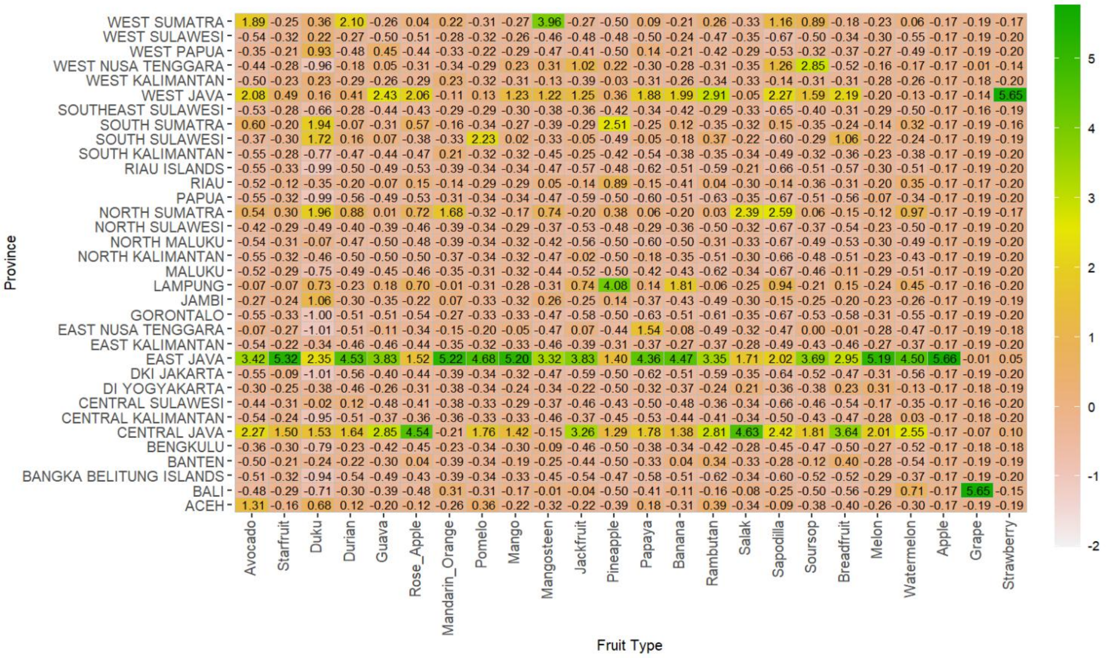

Figure 2. Heatmap of scaled data

Figure 3 presents the PCA biplot analysis of the scaled matrix data, visualizing the relationship between the first dimension (Dim1), which explains 65.8% of the variance, and the second dimension (Dim2), which explains 9.3%. Together, they account for a cumulative total of 75.1% of the variance. Indonesian provinces can be grouped into four quadrants based on the similarity of their characteristics with the variables. For instance, West Java is located in Quadrant IV, showing strong associations with fruits such as rose apple and strawberries, as indicated by the vector directions. This suggests that provinces in this quadrant tend to specialize in specific fruit commodities. East Java, located far in Quadrant III, indicates very different characteristics and likely lower scores on the principal components. Meanwhile, provinces clustered in Quadrants I and II exhibit similar traits and tendencies based on their proximity in the plot.

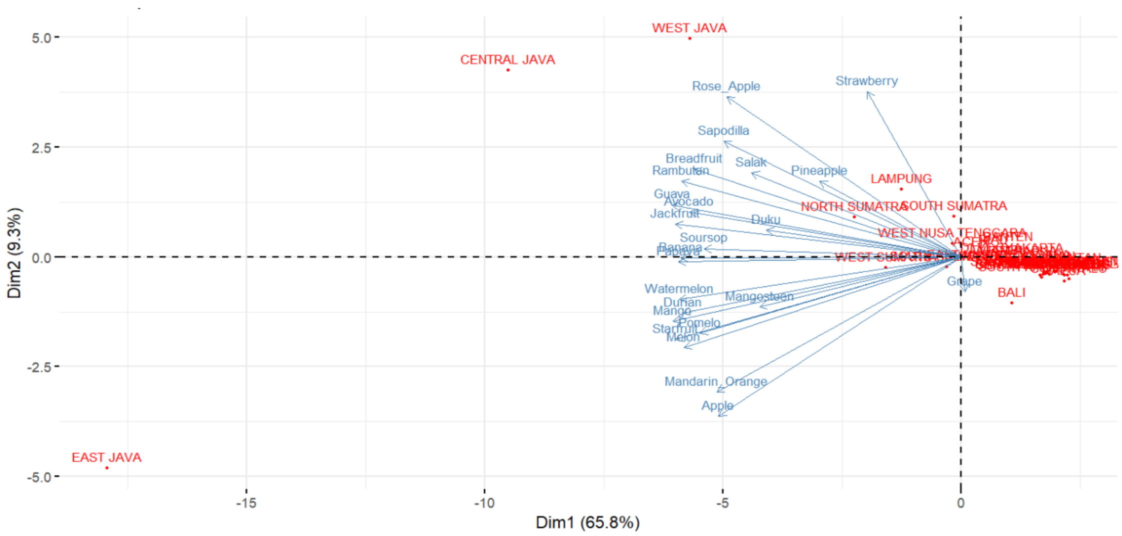

Figure 3. PCA biplot

Performance evaluation using biclustering

To identify patterns and relationships within the data matrix, models such as those illustrated in Table 2 were applied, enabling the mapping of subsets of objects and variables with similar characteristics. The primary objective in selecting and optimizing the model parameters is to generate meaningful and interpretable biclusters that accurately reflect the underlying data structure. The selection of optimal model parameters is guided by identifying the lowest average MSR value across each layer. Figure 4 illustrates the number of biclusters generated and their respective average MSR values.

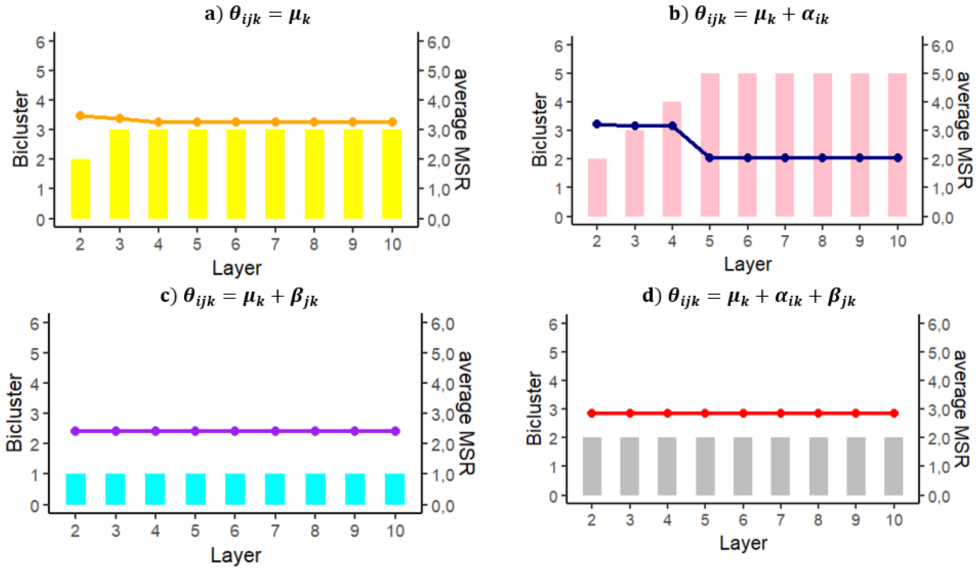

Figure 4. Average MSR for each model

The Constant Model (a) resulted in a relatively high average MSR of 3.2592, stabilizing from the fourth to the tenth layer and producing three biclusters. The Row-Constant Model (b) achieved the lowest average MSR and showed a decreasing MSR trend as the number of layers increased. The MSR stabilized from the fifth to the tenth layer at 2.054, forming five biclusters the highest number among the models. The Column-Constant Model (c) produced a stable average MSR of 2.41 across all layers but formed only one bicluster. The Row-and-Column Constant Model (d) resulted in a constant average MSR of 2.856 across all layers, forming two biclusters. The selection of the most appropriate model is based on the lowest average MSR value from the data scaling, indicating the most optimal biclustering model [28]. In addition to the smallest MSR value, the interpretability and usefulness of the resulting biclusters also play a role in model selection [20].

Table 3. Evaluation of average MSR in the plaid model

Plaid Model	τ_1	τ_2	Average MSR Plaid Model	Number of Bicluster
$\theta_{ijk} = \mu_k$	0,2	0,2	3,2592	3
$\theta_{ijk} = \mu_k + \alpha_{ik}$	0,2	0,1	2,0537	5
$\theta_{ijk} = \mu_k + \beta_{jk}$	0,6	0,7	2,4107	1
$\theta_{ijk} = \mu_k + \alpha_{ik} + \beta_{jk}$	0,1	0,8	2,8555	2

Based on Table 3, the row-constant model yielded the lowest average MSR value among the plaid models. Therefore, this model was selected as the optimal bicluster model, with an MSR of 2.0537, forming five biclusters at the threshold combination of $\tau_1 = 0.2$ and $\tau_2 = 0.1$. The characteristics of the optimal bicluster memberships are shown in Table 4. The membership characteristics of the optimal bicluster are shown in Table 4. These results show that each bicluster represents a particular fruit production pattern unique to the member provinces. Bicluster 1, consisting of Central Java and East Java, has a high potential for pineapple production, while fruits such as grapes and strawberries are only at a moderate level. Bicluster 2 shows a high concentration of production of five fruit commodities (duku, guava, pineapple, salak, and sawo) in three provinces, namely North Sumatra, South Sumatra, and Central Java, indicating a cluster with high and equitable horticultural production capacity. Meanwhile, Bicluster 3 and 4 reflect regions with specific potential, such as guava and rambutan in West Java and Central Java (Bicluster 3), as well as avocado, durian, mango, and sapodilla in western Indonesia, such as North Sumatra and West Sumatra (Bicluster 4). Bicluster 5 consists of eight provinces with similar production patterns for tropical fruits such as salak, breadfruit, and water guava. The results of biclustering have important practical implications for area-based agricultural development planning. The resulting patterns can be used to formulate distribution and development policies for leading commodities so that government interventions are more targeted.

Table 4. Characteristics and membership of optimal biclusters

Biclusters	Size	Province	Fruit Production Potential		
			Low	Medium	High
1	2x3	Central Java, East Java	-	Grapes, Strawberries	Pineapples
2	3x5	North Sumatra, South Sumatra, Central Java	-	-	Duku, Rose Apple, Pineapples, Salak, Sapodilla
3	2x5	West Java, Central Java	-	-	Guava, Rose Apple, Rambutan, Breadfruit, Strawberries
4	3x4	North Sumatra, West Sumatra, West Java	-	-	Avocado, Durian, Mango, Sapodilla
5	8x3	Aceh, West Sumatra, South Sumatra, West Java, Central Java, East Java, West Nusa Tenggara, West Papua	-	-	Rose Apple, Salak, Breadfruit

For example, Bicluster 2 and 4 provinces can be centres of high-value tropical fruits with export potential, such as durian, mango, and salak. In addition, this mapping also helps the efficient allocation of subsidies, technical assistance, and agribusiness infrastructure development in areas with similar production characteristics. This clustering also opens up opportunities for interprovincial collaboration within a cluster for research, investment and market development. These findings support reducing fruit imports by optimizing domestic production based on regional potential while strengthening data-based agricultural policies and supporting the national strategic agenda in the horticulture sector.

Profiling evaluation

The membership profiling of provinces in each bicluster and their associated variables are shown in Figure 5. Bicluster is considered more homogeneous when the profile lines are parallel and close to each other. The profile lines in Biclusters 1 and 5 tend to overlap, indicating high similarity. Bicluster 2 has profile lines that tend to coincide but are not parallel, indicating differences in fruit crop production values in some provinces. South Sumatra province has the highest production potential in pineapple, while Central Java province has the highest production potential in salak compared to other provinces. Bicluster 3 has profile lines that tend not to coincide and are not parallel, in this bicluster West Java province has the highest fruit production potential in rambutan fruit types. Bicluster 4 shows that the profile lines tend not to coincide and are not parallel, and West Sumatra province has high fruit production potential compared to other provinces in this bicluster. The biclustering approach can reveal hidden data patterns from all the generated patterns. This information is important in developing strategies for fruit distribution, promoting superior commodities, and providing a clear picture of area-based policy planning.

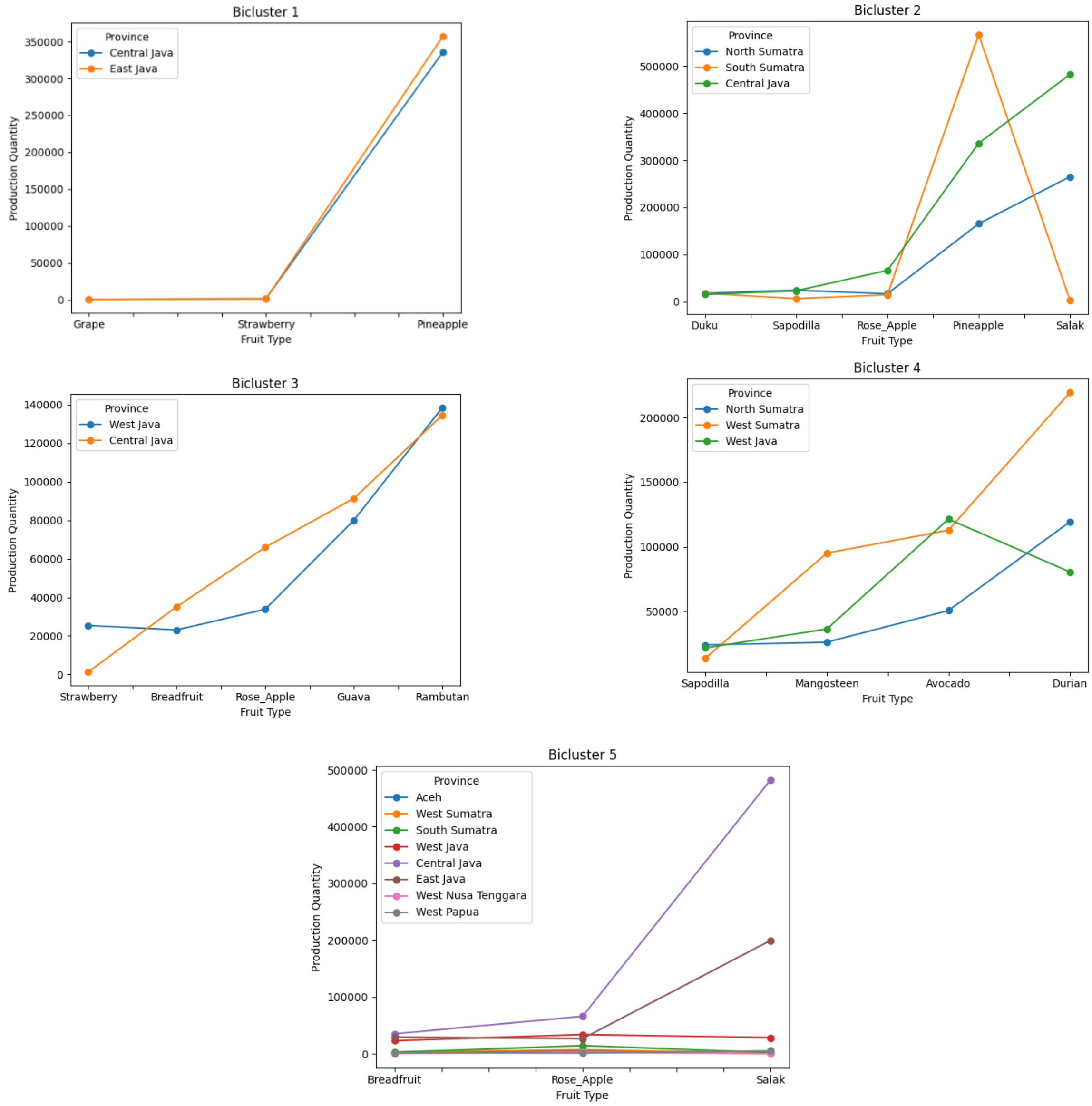

Figure 5. Bicluster profiling

Figure 6 shows a map of regions based on the optimal biclustering result from the plaid model. Most clustered provinces are in Sumatra, Java, West Nusa Tenggara, and West Papua. This map helps visualize how provinces grouped into biclusters. Biclusters 1 and 5 overlap in East Java. Biclusters 2 and 4 include North Sumatra, with high production potential for sapodilla. Biclusters 2 and 5 group South Sumatra, which excels in salak production. Biclusters 4 and 5 include West Sumatra. Bicluster 5 covers Aceh, West Nusa Tenggara, and West Papua, which have high production potential for rose apple, salak, and breadfruit. Biclusters 1, 2, 3, and 5 overlap in Central Java, which shows high production of strawberries, rose apples, breadfruit, and pineapples. Biclusters 3, 4, and 5 include West Java with high breadfruit production. Figure 6 shows that several provinces in Kalimantan, Maluku, and Sulawesi not included in any bicluster. Some provinces and fruit types not clustered in any bicluster are common in biclustering analysis, especially when using the Plaid model. This non-clustering is not solely due to low production values but can be caused by various other factors. One of the leading causes is that the production characteristics are too unique or different compared to the general pattern, so the model does not find simultaneous linkages between rows (provinces) and columns (fruit types) that are strong enough to be included in the bicluster. The Plaid model conceptually allows some data to remain unclustered if it does not show similar patterns. This is done to maintain the quality and homogeneity of each bicluster formed. Therefore, not clustering a province or commodity in a bicluster does not necessarily indicate that the data is unimportant but instead may indicate a very distinctive, extreme, or different pattern that needs to be further explicitly analysed.

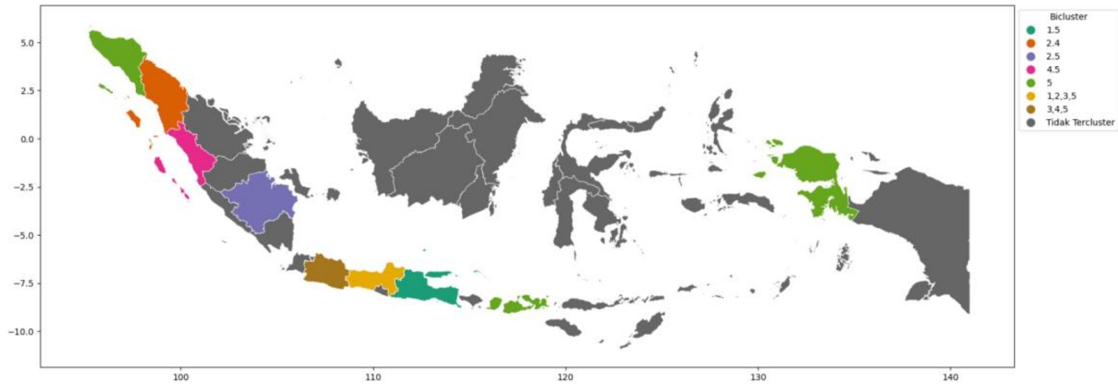

Figure 6. Indonesian map based on plaid model biclusters

CONCLUSION

Applying the plaid model algorithm in the biclustering analysis resulted in five biclusters that reflect local patterns, with overlapping bicluster structures in fruit production across provinces in Indonesia, indicating the complexity and diversity of national fruit production. The main finding of this study is that each bicluster represents distinctive fruit production characteristics, such as the concentration of tropical fruits such as durian and salak in Sumatra and West Java regions. Appropriate model selection, number of layers, and threshold values for rows and columns proved influential in optimizing the performance of the plaid model to produce relevant biclusters. This research's methodological contribution lies in applying the plaid model, which has not been commonly used in agriculture, especially national fruit production, and its ability to reveal two-dimensional patterns that are not detected by ordinary cluster methods. The constant row model was considered the most suitable based on comparing the average MSR value of 2.0537, with layer configuration = 5, τ_1 = 0.2 and τ_2 = 0.1 and revealed very high production potential for various types of fruit.

The results of this biclustering analysis can be a strong basis for the government to design policies to increase fruit production in various types of fruit to be more targeted. The government can strengthen areas with low production through extension strategies and technology and encourage collaboration between regions to distribute and utilise production potential. Biclustering results can strengthen food security, reduce imports and encourage exports of superior commodities. In addition to producing an efficient clustering model, this research provides an analytical framework for sustainable agricultural decision-making. Future research can explore other biclustering models, such as Bayesian, to handle more complex data. Testing the Plaid model on horticultural production data other than fruit, such as vegetables or biopharmaceuticals, is also worth considering as an extension of the model's application in integrated agricultural systems. Thus, the room for development of the results of this research is still wide open, both methodologically and applicatively.

REFERENCES

[1] L. Sundusia, H. Hendrarini, and P. D. Wijayati, “Agribusiness Development Based on Leading Commodities of Fruit Horticulture Subsector in Lamongan Regency,” *Agrisep*, vol. 24, pp. 12–20, 2023.

[2] G. Widhiyoga, H. Wijayati, and R. Alma’unah, “Export Performance Of Indonesia’s Leading Tropical Fruit Commodities To Main Destination Countries,” *J. Ilm. Ekon. Kita*, vol. 12, no. 1, pp. 128–148, 2023.

[3] S. R. Ningsih and W. Hatmi, “Factors Affecting Demand for Imported Grapes at Transmart Carrefour in Palu City,” *AGROTEKBIS; J. Ilmu Pertan.*, vol. 12, pp. 299–310, 2024.

[4] A. Gomez-zavaglia, J. C. Mejuto, and J. Simal-gandara, “Mitigation of emerging implications of climate change on food production systems,” *Food Res. Int.*, 2020, doi: https://doi.org/10.1016/j.foodres.2020.109256.

[5] “Production of Fruits by Crop Type by Province,” Badan Pusat Statistik.

[6] V. Hutagaol, D. Anggraeni, and A. Nata, “Application of Data Mining for Clustering

Underprivileged Communities in Rawang Pasar Vi Village with K-Means Algorithm," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 1, pp. 304–312, 2025.
[7] M. G. Silva, S. C. Madeira, and R. Henriques, "Water Consumption Pattern Analysis Using Biclustering: When, Why and How," *Water (Switzerland)*, vol. 14, no. 12, pp. 1–35, 2022, doi: 10.3390/w14121954.
[8] E. N. Castanho, H. Aidos, and S. C. Madeira, "Biclustering fMRI time series: a comparative study," *BMC Bioinformatics*, vol. 23, no. 1, pp. 1–30, 2022, doi: 10.1186/s12859-022-04733-8.
[9] M. L. Pasaribu, I. M. Sumertajaya, and Erfiani, "Evaluation of Spectral Biclustering Performance in Identifying Potential Production of Horticultural Commodities in Indonesia," *J. Math. Its Appl.*, vol. 21, no. 3, pp. 365–382, 2024.
[10] M. D. M. Noronha, R. Henriques, S. C. Madeira, and L. E. Zárate, "Impact of metrics on biclustering solution and quality: A review," *Pattern Recognit.*, vol. 127, p. 108612, Jul. 2022, doi: 10.1016/j.patcog.2022.108612.
[11] V. A. Padilha and A. C. P. de L. F. de Carvalho, "Experimental correlation analysis of bicluster coherence measures and gene ontology information," *Appl. Soft Comput. J.*, vol. 85, no. xxxx, 2019, doi: 10.1016/j.asoc.2019.105688.
[12] M. N. Aidi *et al.*, "Province clustering based on the percentage of communicable disease using the BCBimax biclustering algorithm," *Geospat. Health*, vol. 18, no. 2, 2023, doi: 10.4081/gh.2023.1202.
[13] 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.
[14] H. A. Majd *et al.*, "Evaluation of Plaid Models in Biclustering of Gene Expression Data," *Scientifica (Cairo).*, vol. 2016, 2016, doi: 10.1155/2016/3059767.
[15] Y. Cheng and G. M. Church, "Biclustering of Expression Data," in *Proceedings of the 8th International Conference on Intelligent Systems for Molecular Biology, ISMB 2000*, 2000, pp. 93–103.
[16] I. M. Afnan, H. Wijayanto, and A. H. Wigena, "Identifying Poverty Vulnerability Patterns in Indonesia using Cheng and Chruch ' s Algorithm," *J. Teor. dan Apl. Mat.*, vol. 8, no. 4, pp. 1262–1277, 2024.
[17] J. Xie, A. Ma, A. Fennell, Q. Ma, and J. Zhao, "It is time to apply biclustering : a comprehensive review of biclustering applications in biological and biomedical data," vol. 20, no. January 2018, pp. 1449–1464, 2019, doi: 10.1093/bib/bby014.
[18] S. Baehera, U. D. Syafitri, and A. M. Soleh, "Comparative Evaluation of the Performance of Cheng and Church Biclustering Algorithm Against the Classic K-Means Clustering Algorithm to Identify the Distribution Pattern of Indonesian Export Goods," *J. Stat. dan Apl.*, vol. 7, no. 2, pp. 149–161, 2023.
[19] A. Segura-ortiz, A. José-García, L. Jourdan, and J. García-Nieto, "Computer Methods and Programs in Biomedicine Exhaustive biclustering driven by self-learning evolutionary approach for biomedical data," *Comput. Methods Programs Biomed.*, vol. 269, no. January, p. 108846, 2025, doi: 10.1016/j.cmpb.2025.108846.
[20] C. Wulandari, I. M. Sumertajaya, and M. N. Aidi, "Evaluation of Bicluster Analysis Results in Capture Fisheries Using the BCBimax Algorithm," *JUITA J. Inform.*, vol. 11, no. 1, pp. 57–66, 2023.
[21] I. Audita, I. Sudahri Damanik, and E. Irawan, "Mapping Fruit Production Results with K-Medoids Data Mining Technique," *J. Tek. Mesin, Ind. Elektro Dan Inform.*, vol. 1, no. 3, pp. 39–53, 2022, doi: 10.55606/jtmei.v1i3.535.
[22] S. Lee, "In the Era of Climate Change: Moving Beyond Conventional Agriculture in Thailand," *Asian J. Agric. Dev.*, vol. 18, no. 1, pp. 1–14, 2021.
[23] B. P. A. Permana and A. H. Limantoro, "Analyze the influence of social media on mental well-being using regression and correlation," *J. IT-Explore*, vol. 04, pp. 44–52, 2025.
[24] S. Wulandary, "Application of PCA Biplot and K-Medoids Cluster on Regional Segmentation Based on Oil Palm Potential," *Media Edukasi Data Ilm. dan Anal.*, vol. 6, Nomor 2, pp. 24–38, 2023.
[25] Y. Widyaningsih and A. C. Nisa, "Comparison Between Biclustering And Cluster- Biplot Results Of Regencies / Cities In Java Based On People ' S Welfare Indicators," *J. Ilmu Mat. dan Terap.*, vol. 19, no. 2, pp. 1009–1022, 2025.
[26] L. Lazzeroni and A. Owen, "Plaid models for gene expression data," *Stat. Sin.*, vol. 12, no. 1, pp. 61–86, 2002.

[27] T. Siswantining, A. Eriza Aminanto, D. Sarwinda, and O. Swasti, “Biclustering analysis using plaid model on gene expression data of colon cancer,” *Austrian J. Stat.*, vol. 50, no. 5, pp. 101–114, 2021, doi: 10.17713/ajs.v50i5.1195.
[28] N. Hikmah, I. M. Sumertajaya, and F. M. Afendi, “Pattern Recognition of Food Security in Indonesia Using Biclustering Plaid Model,” *JTAM (Jurnal Teor. dan Apl. Mat.*, vol. 7, no. 4, p. 1178, 2023, doi: 10.31764/jtam.v7i4.16778.
[29] J. Yang, W. Wang, H. Wang, and P. Yu, “delta-Clusters : Capturing Subspace Correlation in a Large Data Set -Clusters : Capturing Subspace Correlation in a Large Data Set,” no. February, pp. 517–528, 2002, doi: 10.1109/ICDE.2002.994771.
[30] Y. Lee, J. Lee, and C.-H. Jun, “Validation Measures of Bicluster Solutions,” vol. 8, pp. 101–108, 2009.