e-ISSN: 2686-2573
DOI: 10.21512/emacsjournal.v8i2.16534

Comparison of Normalization Methods in Spectral Biclustering on Low Birth Weight Data in Gorontalo Province

Shania Rahmadani P. Panyili[1*], Fahrezal Zubedi[2]
[1,2]Department of Statistics, Faculty of Mathematics and Natural Sciences,
Universitas Negeri Gorontalo,
Gorontalo, Indonesia 96128
shania_s1statistika@mahasiswa.ung.ac.id; fahrezal@ung.ac.id

*Correspondence: shania_s1statistika@mahasiswa.ung.ac.id

Abstract – *Low Birth Weight (LBW) remains a major maternal and child health challenge associated with elevated risks of neonatal mortality and stunting. Conventional one-way clustering algorithms partition health centers or risk indicators separately, failing to identify local interactions across specific subsets of variables simultaneously. This study conducts a comparative evaluation of three normalization schemes within the Spectral Biclustering algorithm—Independent Rescaling of Rows and Columns (IRRC), Bistochastization, and Log-Interactions—to identify optimal bicluster structures on 2024 LBW data across 95 public health centers in Gorontalo Province. Preprocessing procedures addressed missing observations using mean imputation prior to variable standardization via Z-scores. The parameter tuning framework systematically evaluated 105 parameter combinations across variations of singular vectors (E = 2 to 8) and intra-bicluster variance thresholds (withinvar = 1 to 5), yielding 17 candidate models that satisfied stability criteria. Model validation using the Average Squared Residue (ASR) demonstrates that Bistochastization achieves optimal bicluster homogeneity, generating eight biclusters with the lowest ASR value of 0.1755535 at configuration E = 6 and withinvar = 3. In comparison, Log-Interactions and IRRC achieved higher ASR values of 0.3552414 (eight biclusters) and 0.3623521 (six biclusters), respectively. Structural bicluster simplification reveals distinct spatial patterns, grouping maternal health service disparities and maternal physiological risk profiles across specific public health centers. These empirical findings provide an analytical foundation for targeted maternal health interventions across regional health centers.*

Keywords: *Average Squared Residue; Low Birth Weight; Matrix Normalization; Public Health; Spectral Biclustering*

I. INTRODUCTION

Pattern discovery in complex observational data represents an essential analytical step for extracting meaningful insights across multifaceted domains. A conventional analytical approach for examining underlying multivariate structures is data clustering. Centroid-based clustering algorithms, such as K-Means, alongside hierarchical clustering schemes, are frequently employed as standard benchmarks in unsupervised segmentation studies (Jain et al., 1999; Wani, 2024). Nevertheless, these classical methodologies operate under the rigid assumption that inter-object proximity is uniformly shared across the entire feature space simultaneously. In real-world phenomena, observational units frequently exhibit coherent similarities across only an isolated subset of variables, rendering such localized sub-matrix patterns unidentifiable through traditional one-dimensional clustering algorithms (Castanho et al., 2024).

To overcome the structural limitations of classical partitioning methods that evaluate rows and columns in isolation, biclustering algorithms were conceived (Lestari et al., 2024). Biclustering provides a simultaneous, two-way dimension-reduction framework that concurrently groups rows (objects) and columns (attributes), thereby uncovering localized submatrices through diverse algorithmic paradigms, such as Cheng and Church (CC), Plaid Models, SAMBA, and Spectral Biclustering (Castanho et al., 2024). Among various algorithmic paradigms, Spectral Biclustering demonstrates marked advantages due to its global linear algebraic formulation via Singular Value Decomposition, allowing it to effectively capture overarching checkerboard structures within complex matrices (Kluger et al., 2003). Spectral biclustering exhibits enhanced empirical robustness against severe collinearity and data overlap compared to CC algorithms (Hafsah et al., 2025). The practical utility of spectral approaches has been demonstrated in agricultural zoning (Lestari et al., 2024) and genomic subtyping, proving resilient in isolating

Copyright © 2026

localized sub-structures within heterogeneous and noisy environments (Liu et al., 2024).

The integrated data normalization stage is paramount, as it directly governs the geometric coherence of the resulting bicluster space. Kluger et al. (2003) proposed three fundamental normalization formulations: Independent Rescaling of Rows and Columns, Bistochastization, and Log-Interactions, where selecting an appropriate transformation preserves underlying structural contrasts and prevents marginal effects from dominating the decomposition. However, prior evaluations of Spectral Biclustering have predominantly relied on complete observational matrices or strictly controlled synthetic datasets with fixed parameters. While recent studies have examined algorithmic stability against collinearity, cluster overlap, and Gaussian noise (Hafsah et al., 2025; Lestari et al., 2024; Liu et al., 2024), they consistently omitted the missing value imputation procedure from the preliminary data preparation pipeline. Neglecting missing data handling introduces structural distortion into singular vector spaces, potentially leading to valuable information loss, biased outcomes, and degraded bicluster coherence (Han & Kamber, 2006; Pramudita et al., 2025). This research addresses this methodological gap by embedding mean imputation as an essential preparatory stage prior to two-way matrix normalization, ensuring that structural completeness is preserved without compromising sample representation.

The methodological reliability of this integrated pipeline is applied to address a strategic public health priority: Low Birth Weight (LBW). LBW represents a severe global health concern directly associated with neonatal mortality and long-term stunting risks (Dewi et al., 2020; Utami et al., 2024). Previous regional LBW profiling has been confined to conventional one-way approaches, such as K-Means (Mutawalli et al., 2025), K-Medoids (Anggraini & Dzikrullah, 2024), and hierarchical agglomerative clustering (Azzahra & Wijayanto, 2022). According to the 2024 Indonesian Health Profile, Gorontalo Province ranked among the top ten highest LBW prevalence areas nationwide, reaching 7.8% (Kemenkes, 2025). Such elevated prevalence highlights complex local determinant variations across primary healthcare jurisdictions. Therefore, this study conducts an empirical comparative analysis of IRRC, Bistochastization, and Log-Interactions normalization schemes within Spectral Biclustering on 2024 Gorontalo LBW surveillance data, aimed at establishing optimal checkerboard submatrix configurations to support precision maternal health policy interventions.

II. METHODS

2.1 Data

This study utilized secondary Maternal and Child Health (MCH) surveillance data regarding LBW for the year 2024, obtained from the Gorontalo Provincial Health Office. Low Birth Weight is operationally defined as a condition in which an infant is born weighing less than 2,500 grams, recorded within the first hour of life regardless of gestational age (Kemenkes, 2023). A total sampling (saturated sampling) technique was employed, incorporating all 95 public health centers across Gorontalo Province as the study sample. The dataset comprises nine ratio-scale indicator variables, measured as the number of individuals (pregnant women) aggregated at the public health center level, capturing maternal healthcare service coverage, maternal health and nutritional status, and maternal risk profiles. The inclusion and categorization of these variables were established based on prior empirical literature (Afrida et al., 2025; Martinus et al., 2023; Setyawati, 2024). According to Dinas Kesehatan Provinsi Gorontalo (2024), the nine research variables and their supporting literature are presented in Table 1.

Table 1. Research Variables

Notation	Variable
X_1	Number of Pregnant Women with Incomplete Antenatal Care Visits (Fauziah et al., 2023)
X_2	Number of Pregnant Women with Pregnancy Spacing < 2 Years (Jayanti & Ernawati, 2021)
X_3	Number of Pregnant Women with Parity > 4 (Arfiyani et al., 2026)
X_4	Number of Pregnant Women Aged < 20 Years and > 35 Years (Ratnaningsih & Nira, 2021)
X_5	Number of Pregnant Women with Chronic Energy Deficiency (Azzahra et al., 2025)
X_6	Number of Pregnant Women with Anemia (Sulung et al., 2022)
X_7	Number of Pregnant Women Not Consuming Iron Supplement Tablets (Khairunnisa et al., 2023)
X_8	Number of Pregnant Women with Incomplete Tetanus Toxoid Vaccinations (Musfirah et al., 2021)
X_9	Number of High-Risk Pregnant Women (Yusrawati et al., 2023)

All variables were subsequently structured into an input data matrix of size 95 x 9 for the data preprocessing stage and Spectral Biclustering analysis.

2.2 Data Preprocessing

The data preprocessing pipeline was implemented to ensure structural completeness and scale uniformity across variables prior to two-way matrix normalization. Missing observations were first addressed using the mean imputation method (Han & Kamber, 2006) via the zoo package, which successfully resolved a single missing entry at Wanggarasi Public Health Center for variable X_8 (Number of Pregnant Women with Incomplete Tetanus Toxoid Vaccinations), thereby preserving the complete matrix structure of 95 public health centers without sample exclusion. Following imputation, data standardization was performed using the Z-score transformation method to eliminate scale discrepancies among heterogeneous health indicators, ensuring that all variables share a mean of zero and unit variance (Johnson & Wichern, 2007).

To gain initial empirical insights into object distribution patterns, inter-variable relationships, and preliminary clustering tendencies prior to biclustering modeling, an exploratory data analysis was conducted using Principal Component Analysis (PCA) biplot visualization on the standardized data matrix (Lestari et al., 2024). All data preprocessing, standardization, and exploratory procedures were executed using RStudio within the R programming environment utilizing the stats, zoo, and factoextra packages. The preprocessed matrix subsequently served as the direct input for the Spectral Biclustering normalization schemes.

2.3 Spectral Biclustering

Spectral Biclustering is a simultaneous two-way clustering methodology designed to uncover checkerboard-patterned submatrices within complex observational data (Kluger et al., 2003). To eliminate baseline scale disparities and suppress the dominance of specific rows or columns prior to decomposition, three distinct matrix normalization formulations were evaluated: IRRC, Bistochastization, and Log-Interactions (Kluger et al., 2003; Lestari et al., 2024).

2.3.1 Independent Rescaling of Rows and Columns

The IRRC normalization scheme performs separate row and column scaling to balance marginal variations across dimensions and mitigate the dominance of large baseline values. The normalized matrix $\hat{A}$ is obtained using the following formulation:

$$\hat{A} = R^{-1/2} A C^{-1/2} \quad (1)$$

where A denotes the standardized data matrix, R = diag(r_{11}, ..., r_{ii}) represents the diagonal matrix of row sums, and C = diag(c_{11}, ..., c_{jj}) represents the diagonal matrix of column sums.

2.3.2 Bistochastization

The Bistochastization scheme applies an iterative doubly stochastic normalization process via Sinkhorn-Knopp diagonal scaling until a balanced matrix across both row and column dimensions is achieved. The normalized matrix B is expressed as:

$$B = D_1 A D_2 \quad (2)$$

where D_1 = diag(d_{11}, ..., d_{ii}) and D_2 = diag(d_{11}, ..., d_{jj}) are diagonal balancing matrices derived iteratively until convergence.

2.3.3 Log-Interactions

The Log-Interactions scheme involves a two-step transformation. First, a logarithmic transformation is applied to each matrix element to stabilize data variance, incorporating a unit offset (+1) to accommodate zero values:

$$L_{ij} = \ln(A_{ij} + 1) \quad (3)$$

Second, double-centering ANOVA residual extraction is applied to isolate pure interaction effects by removing row and column main effects:

$$K_{ij} = L_{ij} - \bar{L}_{i.} - \bar{L}_{.j} + \bar{L}_{..} \quad (4)$$

where L_{ij} represents the log-transformed observation, $\bar{L}_{i.}$ is the mean of the i-th row, $\bar{L}_{.j}$ is the mean of the j-th column, and $\bar{L}_{..}$ is the grand mean of matrix L.

2.3.4 Singular Value Decomposition

Following matrix normalization, Singular Value Decomposition (SVD) is executed on the normalized matrix M (representing $\hat{A}$, B, or K) to extract primary spectral components:

$$M = U \Lambda V^T \quad (5)$$

where U represents the left singular vectors (public health center partitioning space), Λ is the diagonal matrix of singular values, and V^T represents the transposed right singular vectors (LBW indicator partitioning space).

Bicluster construction was implemented in RStudio using the BCSpectral() method within the biclust package, with a fixed random seed set.seed(123) applied to the K-Means vector partitioning stage to ensure strict computational reproducibility. For the IRRC and Bistochastization schemes, the trivial first singular vector was removed prior to subspace projection because it reflects a constant background effect without group-separating information (Kluger et al., 2003). Systematic hyperparameter tuning evaluated 105 candidate combinations derived by varying the number of singular vectors (E = 2, 3, ..., 8) and within-variance thresholds (*withinvar* = 1, 2, ..., 5). The selection of $E \geq 2$ ensures sufficient partitioning dimensions, while the upper bound $E \leq$

8 corresponds to the n - 1 non-trivial singular vectors for nine indicator variables (Kluger et al., 2003; Lestari et al., 2024). All candidate models were subject to structural constraints requiring a minimum of two rows (*minr* = 2) and two columns (*minc* = 2) per bicluster (Lestari et al., 2024). Model stability was operationally defined as the capacity of a parameter configuration to yield non-empty, non-trivial, and structurally consistent partitions across feature subspaces. Out of the 105 tested combinations, 17 parameter configurations satisfied these stability criteria, as lower threshold values (such as *withinvar* = 1) frequently resulted in empty or collapsed bicluster structures, whereas stable bicluster partitions consistently emerged at *withinvar* ≥ 3.

2.4 Performance Evaluation

Bicluster quality and interaction homogeneity were evaluated using the Mean Squared Residue (MSR) for each k-th bicluster (Lee et al., 2009):

$$MSR_k = \frac{1}{|I||J|}\sum_{i\in I}\sum_{j\in J}\left(A_{ij} - \bar{A}_{iJ} - \bar{A}_{Ij} + \bar{A}_{IJ}\right)^2 \quad (6)$$

where A_{ij} is the cell value within the bicluster, $\bar{A}_{iJ}$ is the row mean within the bicluster, $\bar{A}_{Ij}$ is the column mean within the bicluster, $\bar{A}_{IJ}$ is the biluster grand mean, and $|I||J|$ represents the bicluster volume.

Overall model performance across all K generated biclusters was summarized using the Average Squared Residue (ASR):

$$ASR = \frac{1}{K}\sum_{k=1}^{K} MSR_k \quad (7)$$

The 17 candidate models passing the stability filter were ranked based on ascending ASR values, where lower ASR values indicate higher bicluster homogeneity. The optimal parameter configuration for each normalization scheme was selected at the minimum ASR. Finally, to facilitate public health interpretation, post-processing categorized bicluster interaction levels into High and Low relative to the global normalized dataset mean ($\bar{B} = 0.01052632$), where bicluster interaction means exceeding or equal to $\bar{B}$ denote strong indicator-health center associations (Cesario et al., 2022; Johnson & Wichern, 2007) .

III. RESULTS AND DISCUSSIONS

3.1 Data Exploration

Data exploration was conducted to establish an initial empirical overview of the distributional characteristics and spatial variations of LBW indicators across Gorontalo Province prior to executing Spectral Biclustering. Initial data screening identified a single missing observation at the Wanggarasi Public Health Center for indicator X_8 (Number of Pregnant Women with Incomplete Tetanus Toxoid Vaccinations). Applying the mean imputation procedure yielded a completed value of 106 for this entry, successfully preserving matrix integrity without listwise deletion. Subsequently, Z-score standardization was applied to transform all nine ratio-scale indicators into a uniform zero-mean and unit-variance scale, preventing variables with larger absolute magnitudes from dominating the subsequent matrix decomposition.

The multivariate structural relationships and object proximity of the standardized dataset were examined using a Principal Component Analysis (PCA) biplot, as illustrated in Figure 1.

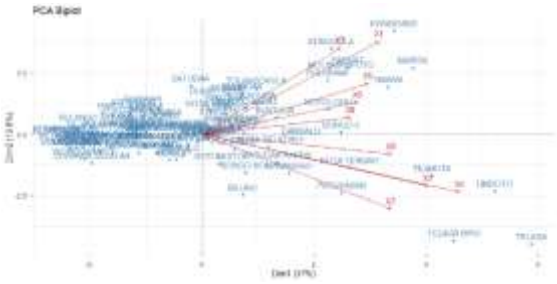


Figure 1. PCA Biplot

As shown in Figure 1, the first two principal components accounted for a cumulative variance of 56.8% (Dim1 = 37.0% and Dim2 = 19.8%), providing a representative low-dimensional projection of the multivariate space. The 95 public health centers in Gorontalo Province were mapped across four distinct quadrants based on the profile similarities of their LBW indicators. The relatively uniform vector lengths across all variables (X_1 to X_9) indicate balanced contributions toward the underlying principal axes. Public health centers situated in close spatial proximity within the same quadrant exhibit similar indicator profiles, whereas facilities positioned opposite an indicator vector display values below the provincial average.

Quadrant I encompasses primary healthcare facilities such as Kwandang, Marisa, Tibawa, Atinggola, and Pulubala, characterized by relatively higher normalized values for indicators X_1, X_5, X_6, X_7, and X_8. Quadrant II includes facilities such as Limboto, Telaga, Telaga Biru, and Tilamuta, which are predominantly distinguished by elevated levels for indicators X_2, X_3, X_4, and X_9 Conversely, public health centers clustered in Quadrants III and IV exhibit generally lower indicator values across both healthcare service coverage and maternal risk profiles. While the PCA biplot effectively highlights global clustering trends, it operates on a one-dimensional projection that cannot isolate specific,

localized sub-matrix structures—thereby necessitating the two-way Spectral Biclustering approach.

3.2 Spectral Biclustering

Bicluster construction was conducted using the Spectral Biclustering algorithm evaluated across the three matrix normalization formulations: IRRC, Bistochastization, and Log-Interactions. Systematic hyperparameter optimization was executed over 105 total parameter combinations derived from varying the number of singular vectors (E = 2, 3, ..., 8) and the within-variance threshold (*withinvar* = 1, 2, ..., 5) across all three normalization schemes ($3 \times 7 \times 5 = 105$ configurations).

Empirical grid-search evaluation revealed that parameter choice substantially governs bicluster formation and structural stability. Configurations with low intra-bicluster variance thresholds (specifically *withinvar* = 1) consistently failed to generate meaningful structures, yielding zero or empty biclusters across all normalization schemes. This occurred because overly restrictive variance limits prevent the partitioning algorithm from grouping observational units into stable clusters. Structural biclusters began to consistently emerge when the variance threshold was relaxed to *withinvar* $\geq$ 2, with optimal stability achieved at *withinvar* = 3. Out of the 105 evaluated combinations, exactly 17 parameter configurations satisfied the operational stability filter by producing non-empty, non-trivial, and structurally consistent checkerboard biclusters. The candidate models passing this initial stability screening across the three normalization schemes are summarized in Table 2.

Table 2. Candidate Parameter Configurations for Each Normalization Scheme

Method	Parameter	Number of Biclusters
IRRC	E = 3, *withinvar* = 3	6
	E = 4, *withinvar* = 3	6
	E = 6, *withinvar* = 3	8
	E = 7, *withinvar* = 3	8
Bistochastization	E = 2, *withinvar* = 3	6
	E = 3, *withinvar* = 3	3
	E = 4, *withinvar* = 3	6
	E = 5, *withinvar* = 3	3
	E = 6, *withinvar* = 3	8
	E = 7, *withinvar* = 3	8
	E = 8, *withinvar* = 3	8
Log-Interactions	E = 2, *withinvar* = 3	6
	E = 3, *withinvar* = 3	6
	E = 5, *withinvar* = 3	6
	E = 6, *withinvar* = 3	8
	E = 7, *withinvar* = 3	8
	E = 8, *withinvar* = 3	8

As shown in Table 2, the candidate models exhibited varying structural characteristics depending on the normalization scheme and singular vector dimensions. The IRRC normalization scheme yielded candidate models generating between 6 and 8 biclusters at *withinvar* = 3. The Bistochastization scheme generated stable structures ranging from 3 to 8 biclusters, whereas the Log-Interactions scheme produced stable configurations containing 6 to 8 biclusters. These 17 stable candidate models were subsequently subjected to quantitative performance evaluation using MSR and ASR metrics to identify the optimal model for each normalization strategy.

3.3 Performance Evaluation

Quantitative model performance and submatrix homogeneity across the 17 stable candidate configurations were evaluated using MSR and ASR metrics. ASR measures the mean residual variance within generated biclusters, where smaller values approaching zero signify superior bicluster coherence and greater internal interaction consistency. The 17 candidate configurations were ranked in ascending order of ASR values to identify the optimal parameter model for each normalization scheme, as presented in Table 3.

Table 3. Parameter Combination Ranking Based on ASR Values

Rank	Method (E, *withinvar*)	Number of Biclusters	ASR
1	Bistochastization (6,3)	8	0.1755535
2	Bistochastization (4,3)	6	0.1899280
3	Bistochastization (7,3)	8	0.2158923
4	Bistochastization (8,3)	8	0.2246200
5	Bistochastization (2,3)	6	0.2288361
6	Bistochastization (5,3)	3	0.2446322
7	Bistochastization (3,3)	3	0.2508451
8	Log-interactions (6,3)	8	0.3552414
9	IRRC (3,3)	6	0.3623521
10	Log-interactions (7,3)	8	0.3673629
11	Log-interactions (8,3)	8	0.3724477
12	Log-interactions (2,3)	6	0.3736087
13	Log-interactions (3,3)	6	0.3736789
14	IRRC (4,3)	6	0.4061327
15	Log-interactions (5,3)	6	0.4176999
16	IRRC (7,3)	8	0.5225343
17	IRRC (6,3)	8	0.5324544

As shown in Table 3, parameter choices directly influenced bicluster homogeneity across all normalization approaches. To select the definitive model for each normalization scheme, the configuration achieving the absolute minimum ASR value was identified. The performance metrics and structural properties of the best-performing model for each normalization scheme are summarized in Table 4.

Table 4. Comparison of MSR and ASR Values of the Best Model for Each Normalization Scheme

IRRC		Bistochastization		Log-interactions	
Bicluster	MSR	Bicluster	MSR	Bicluster	MSR
1	0.5086393	1	0.0614283	1	0.2591425

2	0.4479974	2	0.0229041	2	0.2105486
3	0.4878060	3	0.6269541	3	0.3162434
4	0.1883820	4	0.4992836	4	0.3539887
5	0.4386113	5	0.0547632	5	0.6559213
6	0.1026764	6	0.0168826	6	0.2084900
-	-	7	0.0862874	7	0.3762010
-	-	8	0.0359242	8	0.4613954
ASR	**0.3623521**	**ASR**	**0.1755535**	**ASR**	**0.3552414**

As detailed in Table 4, empirical evaluation demonstrated that the Bistochastization normalization scheme achieved superior performance, yielding the lowest global ASR value of 0.1755535 at configuration $E = 6$ and *withinvar* = 3, generating eight distinct checkerboard biclusters. In comparison, the Log-Interactions scheme achieved an optimal ASR of 0.3552414 (eight biclusters at $E = 6$, *withinvar* = 3), while the IRRC scheme yielded an ASR of 0.3623521 (six biclusters at $E = 3$, *withinvar* = 3).

The methodological superiority of Bistochastization over IRRC and Log-Interactions stems from its underlying mathematical formulation. Bistochastization utilizes an iterative Sinkhorn-Knopp doubly stochastic normalization that balances row and column marginal variations simultaneously. This double-stochastic balancing suppresses extreme baseline disparities across both public health centers and indicator variables without distorting localized interaction patterns. Consequently, when SVD is executed on a bistochastized matrix, the resulting singular vectors preserve localized submatrix contrasts, allowing K-Means projection to recover highly homogeneous bicluster partitions. In contrast, IRRC scales rows and columns independently, which can leave residual marginal imbalances that inflate intra-bicluster variance. Similarly, while Log-Interactions stabilizes variance through logarithmic scaling, double-centering ANOVA residual extraction can attenuate subtle submatrix signals in highly sparse public health surveillance data.

The use of ASR as the primary model selection criterion is methodologically justified because it directly quantifies the average deviation of individual data cells from their respective bicluster row, column, and grand means. Minimizing ASR ensures that the identified public health center groupings reflect genuine, localized co-clustering patterns rather than global scale artifacts.

3.4 Bicluster Interpretation

To convert the complex checkerboard partitions generated by the optimal Bistochastization model ($E = 6$, *withinvar* = 3) into actionable public health insights, a structural bicluster simplification procedure was performed. The original eight biclusters were aggregated across primary dimensions, yielding four distinct public health center groups (P_1, P_2, P_3, P_4) and two coherent indicator clusters (*Cluster* 1 and *Cluster* 2), as summarized in Table 5.

Table 5. Bicluster Simplification Results

Public Health Center	Cluster 1	Cluster 2
TELAGA JAYA, TOTO UTARA, BULANGO SELATAN, BULANGO ULU, KABILA BONE	$X_2, X_3, X_4, X_5, X_6, X_9$	X_1, X_7, X_8
LIMBOTO, LIMBOTO BARAT, TELAGA, PILOHAYANGA, TELAGA BIRU, . . ., KOTA UTARA, KOTA TENGAH, SIPATANA	$X_2, X_3, X_4, X_5, X_6, X_9$	X_1, X_7, X_8
SUWAWA TENGAH, SUWAWA SELATAN, DUMBAYABULAN, WONGGARASI II, WONGGARASI I, LIMBATO	$X_2, X_3, X_4, X_5, X_6, X_9$	X_1, X_7, X_8
BILATO, TALAGA PUNCAK, LAMAHU, PINOGU, PANCA KARSA I, PILOLODAA	$X_2, X_3, X_4, X_5, X_6, X_9$	X_1, X_7, X_8

As shown in Table 5, Cluster 1 encompasses six indicators capturing maternal health and nutritional status as well as maternal risk profiles (X_2, X_3, X_4, X_5, X_6, X_9), whereas Cluster 2 comprises three indicators capturing maternal healthcare service coverage (X_1, X_7, X_8). Across the row dimension, public health centers were categorized into P_1 (5 facilities), P_2 (78 facilities), P_3 (6 facilities), and P_4 (6 facilities), fully accounting for all 95 public health centers across Gorontalo Province.

To determine the relative intensity of indicator-facility relationships, the mean interaction value for each submatrix block was evaluated against the global normalized dataset mean ($\bar{B} = 0.01052632$). Bicluster interaction means greater than or equal to $\bar{B}$ were categorized as High, indicating strong co-clustering associations, whereas values below $\bar{B}$ were classified as Low. The resulting interaction levels and mean values across all bicluster blocks are presented in Table 6.

Table 6. Bicluster Mean Interaction Values

Public Health Center	Cluster 1	Cluster 2
P1	0.00994	0.01170
P2	0.01048	0.01063
P3	0.01107	0.00944
P4	0.01112	0.00935

The spatial distribution and interaction patterns of the simplified bicluster framework are visually mapped in the heatmap in Figure 2.

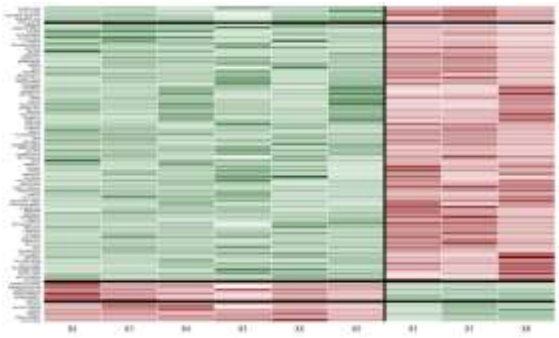


Figure 2. Bicluster Heatmap

As detailed in Table 6 and illustrated in Figure 2, the bicluster profiling uncovers clear regional groupings across Gorontalo Province:

1. Public Health Center Groups P_1 and P_2 (Service Coverage Focus): Facilities in P_1 (5 health centers) and P_2 (78 health centers) exhibit High interaction levels with *Cluster* 2 indicators (X_1, X_7, X_8), recording mean values of 0.01170 and 0.01063, respectively, which are relatively higher than the global threshold $\bar{B}$. Conversely, their interactions with *Cluster* 1 remain Low (0.00994 and 0.01048). This pattern identifies a subset of health centers characterized by relatively elevated normalized rates of pregnant women receiving healthcare services, iron-folic acid supplementation, and tetanus immunizations.
2. Public Health Center Groups P_3 and P_4 (Maternal Risk Profile Focus): Facilities in P_3 (6 health centers) and P_4 (6 health centers) display High interaction levels with *Cluster* 1 indicators (X_2, X_3, X_4, X_5, X_6, X_9), recording mean values of 0.01107 and 0.01112, respectively, which are also relatively higher than the global mean. Their interactions with *Cluster* 2 fall below the global mean (0.00944 and 0.00935). This structure isolates primary care facilities where maternal risk indicators—such as chronic energy deficiency, anemia, high-risk age pregnancies, and obstetric complications—co-occur at relatively higher concentrations.

These empirical groupings demonstrate that LBW determinants in Gorontalo Province do not manifest uniformly across all primary healthcare jurisdictions. By simultaneously grouping row and column dimensions, Spectral Biclustering effectively isolates specific localized submatrix patterns, providing local health authorities with a precise analytical framework to allocate targeted maternal health resources and tailor intervention strategies based on specific facility profile characteristics.

IV. CONCLUSION

A comparative evaluation of three normalization schemes within the Spectral Biclustering algorithm demonstrated that the Bistochastization scheme achieves superior performance for profiling LBW data in Gorontalo Province. Bistochastization generated eight valid checkerboard biclusters with the lowest global ASR value of 0.1755535 at configuration E = 6 and *withinvar* = 3, markedly outperforming Log-Interactions (ASR = 0.3552414; eight biclusters) and IRRC; ASR = 0.3623521; six biclusters). Structural simplification identified two distinct pattern groupings: Public Health Center Groups P_1 and P_2 exhibited higher interaction levels with maternal healthcare service coverage indicators (X_1, X_7, X_8), whereas Groups P_3 and P_4 displayed stronger associations with maternal physiological risk profiles (X_2, X_3, X_4, X_5, X_6, X_9). These empirical findings confirm that Spectral Biclustering effectively isolates localized submatrix structures, supporting precision maternal health policy interventions across primary healthcare jurisdictions.

Nevertheless, this study carries inherent methodological limitations. The empirical analysis relied on secondary aggregated health surveillance data at the public health center level, which may mask individual-level clinical variations. Furthermore, model evaluation relied primarily on the ASR metric to assess bicluster homogeneity without explicitly quantifying feature variance coverage. To address these constraints, future research is recommended to explore alternative biclustering paradigms, such as Cheng and Church or Plaid Models, and incorporate primary individual-level health records from Maternal and Child Health (KIA) registries to enable more granular epidemiological assessments.

AUTHOR'S CONTRIBUTION

Conceptualization, Data curation, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing: Authors.

AVAILABILITY OF DATA AND MATERIALS

The dataset analyzed during the current study is not publicly available due to institutional data privacy policies, but it is available from the corresponding author upon reasonable request and with official permission from the Gorontalo Provincial Health Office.

REFERENCES

Afrida, B. R., Aryani, N. P., Idyawati, S., Jannati, S. H., & Jannah, R. (2025). Analisis Deskriptif Kuantitatif Terhadap Distribusi dan Faktor Penghambat Pemberian Imunisasi Tetanus Toxoid (TT) Pada Ibu Hamil. *JOMIS (Journal of Midwifery Science)*, *9*(2), 253–260. https://doi.org/10.36341/jomis.v9i2.5783

Anggraini, B. S. K., & Dzikrullah, A. A. (2024). Implementasi Analisis Clustering K-Medoids dalam Pengelompokan Bayi Lahir, Gizi Buruk, dan BBLR Berdasarkan Kecamatan Di Kabupaten Sleman Tahun 2020. *Emerging Statistics and Data Science Journal*, *2*(1), 30–40. https://doi.org/10.20885/esds.vol2.iss.1.art4

Arfiyani, R. S., Fibrila, F., & Martini. (2026). Hubungan Usia, Paritas Dan Anemia Terhadap Kekurangan Energi Kronik (KEK) Pada Ibu Hamil Di Puskesmas Yosomulyo Kota Metro. *Jurnal Ners*, *10*(3), 8042–8048. https://doi.org/10.31004/jn.v10i3.57925

Azzahra, A., & Wijayanto, A. W. (2022). Comparison of Agglomerative Hierarchical and K-Means in Grouping Provinces Based on Maternal Health Services. *SISTEMASI*, *11*(2), 481–495. https://doi.org/10.32520/stmsi.v11i2.1829

Azzahra, Hamdiana, H., & Amir, S. (2025). Hubungan Tingkat Pengetahuan Gizi dan Asupan Makanan dengan Kejadian Kurang Energi Kronis (KEK) pada Wanita Usia Subur di Wilayah Kerja UPT Puskesmas Paccing. *Syntax Idea*, *7*(10), 2548–1398. https://doi.org/10.46799/syntaxidea.v7i10.13581

Castanho, E. N., Aidos, H., & Madeira, S. C. (2024). Biclustering data analysis: a comprehensive survey. *Briefings in Bioinformatics*, *25*(4), bbae342. https://doi.org/10.1093/bib/bbae342

Cesario, Rd. M. R., Syafmen, W., & Mardhotillah, B. (2022). Analisis Cluster Metode Average Linkage Berdasarkan Indikator Pendidikan Tiap Provinsi Di Indonesia. *Media Edukasi Data Ilmiah Dan Analisis*, *5*(1), 30–39. https://doi.org/10.64679/median.v5i01.55

Dewi, R., Evrianasari, N., & Yuviska, I. A. (2020). Kadar Hb, LILA, dan Berat Badan Ibu Saat Hamil Beresiko Terhadap Kejadian Stunting Pada Anak Usia 1-3 Tahun. *Jurnal Kebidanan Malahayati*, *6*(1), 57–64. https://doi.org/10.33024/jkm.v6i1.1769

Dinas Kesehatan Provinsi Gorontalo. (2024). *Data KIA BBLR Menurut Puskesmas di Provinsi Gorontalo 2024* .

Fauziah, A., Hilmi, I. L., & Salman. (2023). Factors Related to Pregnant Women's Compliance in Conducting Antenatal Care Visits: Literature Review. *Journal of Pharmaceutical and Sciences*, *6*(1), 127–131. https://doi.org/10.36490/journal-jps.com.v6i1.22

Hafsah, S., Indahwati, I., & Wijayanto, H. (2025). Performance Evaluation of Cheng & Church (CC) and Spectral Biclustering Algorithms under Collinearity and Overlap Conditions. *Scientific Journal of Informatics*, *12*(2), 371–386. https://doi.org/10.15294/sji.v12i2.26413

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). Morgan Kaufmann.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Jayanti, R., & Ernawati, R. (2021). Faktor Jarak Kehamilan yang Berhubungan dengan Kejadian Stunting di Puskesmas Harapan Baru Samarinda Seberang. *Borneo Student Research*, *2*(3), 1705–1710. https://journals.umkt.ac.id/index.php/bsr/article/view/1868

Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall.

Kemenkes. (2023). *Survei Kesehatan Indonesia*. https://www.badankebijakan.kemkes.go.id/ski-2023-dalam-angka/

Kemenkes. (2025). *Profil Kesehatan Indonesia 2024*. https://kemkes.go.id/id/profil-kesehatan-indonesia-2024

Khairunnisa, N., Rahmiwati, A., & Sitorus, R. J. (2023). Pola Makan Dan Kepatuhan Tablet Tambah Darah Pada Anemia Ibu Hamil: Literature Review. *Health Information: Jurnal Penelitian*, *15*(2), e1169. https://myjurnal.poltekkes-kdi.ac.id/index.php/hijp/article/view/1169

Kluger, Y., Basri, R., Chang, J. T., & Gerstein, M. (2003). Spectral biclustering of microarray data: Coclustering genes and conditions. *Genome Research*, *13*(4), 703–716. https://doi.org/10.1101/gr.648603

Lee, Y., Lee, J.-H., & Jun, C.-H. (2009). Validation Measures of Bicluster Solutions. *IEMS*, *8*(2), 101–108.

Lestari, P. M., Sumertajaya, I. M., & Erfiani, E. (2024). Evaluasi Kinerja Spectral Biclustering dalam Identifikasi Potensi

Produksi Komoditas Hortikultura di Indonesia. *Limits: Journal of Mathematics and Its Applications*, *21*(3), 365–382. https://doi.org/10.12962/limits.v21i3.21718

Liu, F., Yang, Y., Xu, X. S., & Yuan, M. (2024). MESBC: A novel mutually exclusive spectral biclustering method for cancer subtyping. *Computational Biology and Chemistry*, *109*, 108009. https://doi.org/10.1016/j.compbiolchem.2023.108009

Martinus, F. D., Bintang, M. R., & Sari, R. N. (2023). Hubungan Usia Ibu, Paritas, dan Jarak Kehamilan Pada Ibu Hamil dengan Kejadian Bayi Berat Lahir Rendah (BBLR) di Rumah Sakit Charis Medika Batam Tahun 2021. *Zona Kedokteran: Program Studi Pendidikan Dokter Universitas Batam*, *13*(1), 295–305. https://doi.org/10.37776/zked.v13i1.1149

Musfirah, M., Rifai, M., & Kilian, A. K. (2021). Faktor yang Memengaruhi Kepatuhan Imunisasi Tetanus Toksoid Ibu Hamil. *Jurnal Ilmiah Kesehatan Sandi Husada*, *10*(2), 347–355. https://doi.org/10.35816/jiskh.v10i2.619

Mutawalli, L., Zaen, M. T. A., Tantoni, A., & Suhriani, I. F. (2025). Segmentasi Risiko Kesehatan Bayi dan Balita Menggunakan Algoritma K-Means. *Bulletin of Computer Science Research*, *5*(4), 382–391. https://doi.org/10.47065/bulletincsr.v5i4.504

Pramudita, C., Wicaksono, N. A. A., Hartiyah, A. K., Maisaroh, A., & Kharisudin, I. (2025). Menangani Missing Value dengan Metode Imputasi Menggunakan Package MICE di R. *Prosiding Seminar Nasional Matematika*, *8*, 399–406. https://proceeding.unnes.ac.id/prisma/article/view/4341

Ratnaningsih, D., & Nira, M. F. A. (2021). Tingkat Pengetahuan Ibu Hamil Primigravida Tentang Tanda Bahaya Pada Kehamilan Trimester III di Klinik Pratama Widuri Sleman. *Jurnal Permata Indonesia*, *12*(1), 35–43. https://doi.org/10.59737/jpi.v12i1.23

Setyawati, S. A. (2024). *Hubungan Pengetahuan, Kepatuhan Konsumsi Tablet Fe, dan Status Gizi (KEK) dengan Kejadian Anemia pada Ibu Hamil di Wilayah Kerja Puskesmas Ngarip Kecamatan Ulubelu Kabupaten Tanggamus Tahun 2024* [Poltekkes Tanjungkarang]. http://repository.poltekkes-tjk.ac.id/id/eprint/6314

Sulung, N., Najmah, Flora, R., Nurlaili, & Slamet, S. (2022). Faktor-Faktor yang Berhubungan dengan Kejadian Anemia pada Ibu Hamil. *Journal of Telenursing (JOTING)*, *4*(1), 28–35. https://doi.org/10.31539/joting.v4i1.3253

Utami, S. R., Benvenuto, A. F., Wanadiatri, H., & Prajitno, S. (2024). Prevalensi Kematian Neonatal dengan Berat Badan Lahir Rendah (BBLR) di RSUD Praya Lombok Tengah. *MAHESA: Malahayati Health Student Journal*, *4*(6), 2374–2382. https://doi.org/10.33024/mahesa.v4i6.14511

Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, *10*, e2286. https://doi.org/10.7717/peerj-cs.2286

Yusrawati, Desmawati, Amir, A., Serudji, J., Basyir, V., Defrin, Karmia, H. R., Insani, A. A., Safaringga, M., Evareny, L., Agus, M., Driza, N. A., Arnita, M., Kornelia, K., Wizia, L., Angraini, V., Amilia, F., Sintia, E., Mardhiyah, R., … Azkiyaunnisa. (2023). Deteksi Dini Kehamilan Risiko Tinggi di Wilayah Kerja Puskesmas Pegambiran Kota Padang Tahun 2022. *Jurnal Pengabdian Inovasi Dan Teknologi Kepada Masyarakat*, *3*(1), 6–12. https://doi.org/10.58794/jdt.v3i1.429