Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.4, August 2026, pp. 3448~3454
e-ISSN: 2548-6861

Divorce Determinants Clustering in Indonesia Using K-Means and Agglomerative Hierarchical Methods (AHC)

Danang Hilal Kurniawan [1]*, Nurul Alfajar Gumel [2]*, David Boby C. Nainggolan [3]*, M. Syamsuddin Wisnubroto [4]*, Fajri Farid [5]*
*Data Science Study Program, Faculty of Science, Institut Teknologi Sumatera
danang.122450085@student.itera.ac.id[1], nurul.122450127@student.itera.ac.id[2], david.122450048@student.itera.ac.id[3], syamsuddin.wisnubroto@sd.itera.ac.id[4], fajri.farid@sd.itera.ac.id[5]

Article Info

Article history:

Received 2026-03-10
Revised 2026-07-25
Accepted 2026-08-08

Keyword:

Agglomerative Hierarchical Clustering,
K-Means Clustering,
Divorce Determinants,
Spatio Temporal Analysis.

ABSTRACT

This study evaluates the spatial typologies of marital dissolution across 34 Indonesian provinces through a dual algorithm computational framework. By applying a proportional ratio transformation, demographic bias generated by absolute population disparities is mathematically eliminated, enabling an objective spatial analysis of 13 specific divorce determinants. *Agglomerative Hierarchical Clustering* (AHC) and *K-Means* algorithms are integrated to assess structural validity. The optimal spatial partitioning is established at $k = 4$, empirically supported by the inflection point on the *Within Cluster Sum of Squares* (WCSS) curve at 9.7238 and a stable *Silhouette Score* of 0.2417. Structural consistency between the hierarchical and partitional models is validated by an *Adjusted Rand Index* (ARI) of 0.7931. Statistical profiling based on centroid matrices stratifies the regions into four distinct typologies, Cluster 0 (Moderate Multi Factor) exhibiting moderately distributed determinants, Cluster 1 (High Risk Deviant Behavior) characterized by average centroid scores exceeding 0.70 for variables including drug abuse, gambling, and forced marriage, Cluster 2 (Economic Driven) demonstrating an absolute dominance in economic factors with a *centroid score* of 0.9152, and Cluster 3 (Conflict Pure) fundamentally driven by continuous disputes with a centroid score of 0.7787. These structural configurations mathematically verify that marital dissolution in Indonesia is geographically stratified by highly specific socioeconomic, behavioral, and relational variables.

I. INTRODUCTION

Divorce constitutes a severe social pathology that fundamentally disrupts the structural stability of the family unit, characterized by a critical statistical escalation in recent years. In 2025, the national divorce rate in Indonesia surged to 438,168 cases, representing a nearly 10% escalation from the 399,921 cases documented in the preceding year [10]. Demographically, this phenomenon is overwhelmingly driven by *cerai gugat* (divorces filed by wives) with 346,516 cases, overshadowing the 91,652 cases of *cerai talak* (divorces filed by husbands) [9]. Spatially, marital dissolution is heavily concentrated within densely populated territories, specifically peaking in West Java (98,903 cases), East Java (83,208 cases), and Central Java (67,500 cases). The structural collapse of these households is empirically triggered by highly specific determinants, predominantly continuous disputes and quarrels (282,326 cases), severe economic problems (105,727 cases), spousal abandonment (31,029 cases), domestic violence (7,138 cases), and gambling (4,623 cases) [11].

Deciphering the exact nature of these marital breakdowns across 34 provinces is critically imperative to safeguard individual rights and maintain socio demographic equilibrium. The massive volume of institutionalized separations signifies that divorce is no longer an isolated domestic anomaly, but a multidimensional crisis driven by overlapping socio economic and relational stressors.

Therefore, mathematically mapping the spatial distribution and concentration of these distinct determinants is an absolute urgency to halt further social degradation and construct highly precise intervention frameworks.

To computationally address this national urgency, contemporary methodologies have increasingly relied on *unsupervised machine learning* to partition and analyze regional vulnerabilities. The *K-Means* clustering algorithm, specifically, has been widely deployed as the principal analytical engine to group divorce catalysts based on shared spatial characteristics [3]. Such algorithmic stratifications offer a robust quantitative mechanism, enabling policymakers to efficiently pinpoint high risk jurisdictions affected by severe domestic conflicts [8].

Despite its broad utility, the current body of literature exhibits fundamental methodological deficiencies. The majority of existing research is rigidly confined to local spatial boundaries, thereby failing to capture the macro level socio behavioral complexity of the divorce phenomenon on a national scale [4]. Furthermore, previous analytical models predominantly operate on absolute case volumes, a flawed convention that inherently introduces demographic bias by erroneously penalizing highly populated regions such as West Java and East Java as high risk clusters purely due to their massive demographic size [6]. Additionally, the pervasive reliance on a singular *K-Means* algorithm, without any structural cross validation protocols, significantly degrades the mathematical stability and reliability of the final analytical outputs [12].

To directly eliminate these established limitations, this study engineers a national scale hybrid computational framework explicitly driven by a proportional ratio transformation mechanism. By systematically integrating *K-Means* and *Agglomerative Hierarchical Clustering* (AHC), the proposed methodology mathematically neutralizes population bias to objectively isolate the genuine core determinants within each province. This dual algorithm approach strictly guarantees structural consistency and delivers an unbiased regional profiling matrix, providing a highly robust empirical foundation for optimizing targeted public policy formulations and localized social rehabilitation strategies.

II. METHOD

A. Dataset and Data Preprocessing

The dataset utilized in this study is secondary panel data covering the period from 2021 to 2025. The spatial observation unit is focused on the provincial level in Indonesia. To maintain the consistency of the panel data matrix and the validity of inter year comparisons, four newly autonomous provinces in the Papua region (Papua Pegunungan, Papua Barat Daya, Papua Selatan, and Papua Tengah) were excluded from the modeling, resulting in a final total of 34 provinces as spatial entities. The data matrix involves 13 divorce determinant variables analyzed simultaneously. The structure of each numerical feature is outlined in the Table 1

TABLE 1
DATASET

No	Variable Name
1	Adultery (*Zina*)
2	Alcoholism (*Mabuk*)
3	Drug Abuse (*Madat*)
4	Gambling (*Judi*)
5	Abandonment (*Meninggalkan Salah Satu Pihak*)
6	Imprisonment (*Dihukum Penjara*)
7	Polygamy (*Poligami*)
8	Domestic Violence (*KDRT*)
9	Physical Disability (*Cacat Badan*)
10	Continuous Disputes and Quarrels (*Perselisihan dan Pertengkaran*)
11	Forced Marriage (*Kawin Paksa*)
12	Apostasy (*Murtad*)
13	Economic Factors (*Faktor Ekonomi*)

The data preprocessing pipeline was strictly executed through the following stages:

1. Aggregate Elimination: Removing national scale recording rows ("Indonesia") to avoid sectoral data duplication.
2. Standardization of Missing Data Indicators: Anomalous characters such as "–", "-", "...", blank spaces, and "0" entries were uniformly converted into Missing Value representations or *NaN* (Not a Number).
3. Spatial Median Imputation: *NaN* values were imputed using the median values of each respective determinant column in the corresponding year to preserve the original data distribution characteristics without being distorted by outliers.
4. Proportional Ratio Transformation: To eliminate demographic bias (where provinces with large populations automatically possess a high number of absolute cases), a mathematical transformation was performed from absolute values to proportional ratio values. The value of each determinant was divided by the total number of divorce cases in that province for the current year through the following equation:

$$\text{Ratio}_{ij} = \frac{\text{Cases}_{ij}}{\sum_{k=1}^{13} \text{Cases}_{ik}} \quad (1)$$

   Where Ratio_{ij} is the proportion of the $j-th$ feature in the $i-th$ province [15].
5. Temporal Aggregation: The panel data, having been transformed into ratios over a 5-year period, were then averaged (spatial mean aggregation) to produce a single national scale static profile matrix.
6. Scale Normalization: The aggregated ratio matrix was transformed into a homogeneous scale of 0 to 1 using the *Min Max Scaling* method to equalize the

variance weights during geometric distance calculations.

B. Agglomerative Hierarchical Clustering (AHC)

The second stage operationalizes a bottom-up hierarchical approach via the *Agglomerative Hierarchical Clustering* (AHC) algorithm. This method initially considers each province as a single cluster, then iteratively merges the closest clusters until a complete taxonomic tree (dendrogram) is formed [13]. The distance matrix between provinces was calculated using the Euclidean Distance criterion through the formula:

$$d(x,y) = \sqrt{\sum_{m=1}^{13} (x_m - y_m)^2} \qquad (2)$$

The cluster branch merging process was executed using *Ward's Linkage Method*. The selection of this criterion is based on the mathematical principle of minimizing the increase in the total internal variance (*Within Cluster Sum of Squares*) at each stage of spatial entity merging [14].

C. Optimizing the Number of Clusters (K-Optimum)

The determination of the optimal number of cluster groups K was objectively conducted by combining two internal validation metrics, the *Elbow Method* and *Silhouette* Analysis. The iteration testing was computationally executed from a value of k=2 to k =10. The Elbow Method evaluates the decrease in the *Within Cluster Sum of Squares* (WCSS) value as the number of clusters increases. A sharp inflection point on the curve indicates the efficiency limit of adding clusters [5]. Simultaneously, Silhouette Analysis is used to measure how closely a province is matched to its own cluster compared to its nearest neighboring cluster through the coefficient calculation:

$$s(i) = \frac{b(i) - a(i)}{\max\big(a(i), b(i)\big)} \qquad (3)$$

Where *a(i)* is the mean intra cluster distance and *b(i)* is the mean nearest inter cluster distance. The equilibrium combination between the sharp decline in WCSS and the peak *Silhouette* score definitively establishes the operational number of clusters [2].

D. K-Means Pure Clustering and Stability Validation

After the optimal number of clusters was established from Phase 3, the *K-Means* Pure Clustering algorithm was executed as the primary modeling approach to definitively allocate spatial partitions [1]. The cluster center initialization process was configured using the *k-means* method to optimize the randomized, controlled placement of initial *centroids*, thereby accelerating global convergence and avoiding local minimum traps [7].

To test the stability and consistency of the data partitions from the resulting *K-Means* model, cross validation was performed against the AHC hierarchical structure built in Phase 2. The operationalized measurement metric is the *Adjusted Rand Index* (ARI), which calculates the probability of clustering decision similarity between the two algorithms through chance adjustment [15]. An ARI value approaching 1 indicates a highly stable convergence of the clustering structure [1].

E. Centroid Profiling Framework

The final stage focused on the extraction of quantitative profiles from the formed clusters [5]. The interpretation process was not conducted subjectively but was based on the average mathematical values (*centroid* matrix) of the 13 proportional ratio variables within each cluster group. This quantitative profile matrix was then orthogonally transposed to map the determinant variables as rows and the cluster labels as columns. The evaluation of these *centroid scores* generates the stratification of regional characteristics indicating which clusters are dominated by relational conflicts, pure economic pressures, or high risk social categories [6], [15].

III. RESULT AND DISCUSSION

A. Cluster Evaluation and Optimal Partitioning

Determining the optimal number of clusters K is essential to ensure that regional partitioning accurately captures the patterns of divorce determinants in Indonesia. This study employed a dual validation approach *Elbow Method* (WCSS) and *Silhouette* Analysis to balance cluster complexity and homogeneity. Prior to selecting K, *Agglomerative Hierarchical Clustering* (AHC) was applied to explore the natural grouping structure of provinces. The resulting dendrogram, based on Euclidean distance and *Ward's linkage*, is shown in Figure 1.

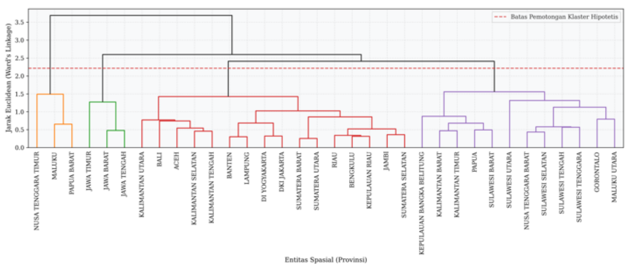

Figure 1. Spatial Dendrogram (AHC)

As shown in Figure 1, the provinces exhibit distinct grouping tendencies, indicating the existence of natural strata within the data. Following this hierarchical exploration, the quantitative determination of the optimal cluster count was performed. The evaluation process, which tested values from

$k = 2$ *to* $k = 10$, is summarized in Table 2 and visualized in Figure 2.

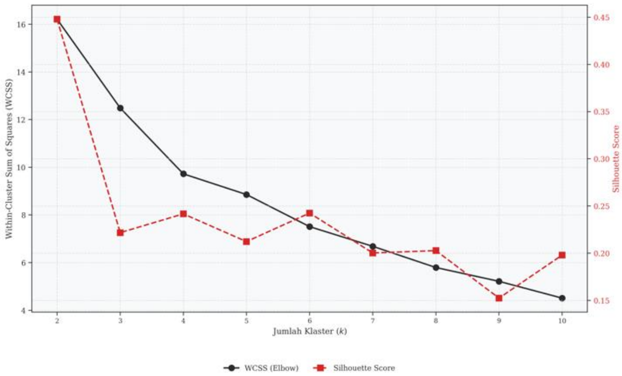

Figure 2. Evaluation of K Optimum using the *Elbow Method* and *Silhouette*

TABLE 2
PERFORMANCE METRICS FOR K-MEANS PARTITIONING

Number of Clusters K	WCSS	*Silhouette Score*
2	16.2131	0.4479
3	12.4859	0.2218
4	9.7238	0.2417
5	8.8567	0.2123
6	7.5097	0.2425
7	6.6839	0.2002
8	5.7948	0.2027
9	5.2152	0.1524
10	4.5108	0.1980

The evaluation of these metrics, visualized in Figure 2, demonstrates the rationale for selecting $k = 4$. While $k = 2$ provides the highest *Silhouette Score* (0.4479), it creates an overly coarse segmentation that fails to distinguish between the nuanced social profiles identified in the data. Conversely, $k = 4$ represents a significant inflection point in the WCSS curve (9.7238), indicating a meaningful reduction in internal variance while maintaining a stable and statistically valid *Silhouette Score* (0.2417).

Furthermore, the robustness of this partition is supported by the *Agglomerative Hierarchical Clustering* dendrogram Figure 2, which validates the natural grouping of provinces. The stability of our model is confirmed by an *Adjusted Rand Index* of 0.7931, measuring the structural similarity between our AHC derived partitions and our *K-Means* results. This high ARI value confirms that the four-cluster configuration is not a stochastic artifact but is structurally consistent across both hierarchical and nonhierarchical algorithms.

B. Cluster Characteristics and Centroid Profiling

This section dissects the identity of the four formed clusters based on the *centroid* values of the 13 divorce determinant variables transformed into proportional ratios. This profiling is conducted to identify the dominant patterns characterizing social behavior within each cluster, enabling an objective stratification of regional characteristics. The average structural characteristics of the four clusters are presented in the *centroid* Table 3.

TABLE 3
CENTROID VALUES

Variable	Cluster 0	Cluster 1	Cluster 2
Zina	0.2232	0.7218	0.3254
Mabuk	0.3892	0.4324	0.0682
Madat	0.1799	0.8022	0.0058
Judi	0.4324	0.7706	0.3012
Meninggalkan Pasangan	0.5429	0.6799	0.1925
Dihukum Penjara	0.2479	0.6274	0.0107
Poligami	0.1398	0.5709	0.0316
KDRT	0.6135	0.8125	0.1373
Cacat Badan	0.2665	0.8319	0.0829
Perselisihan	0.6285	0.4641	0.1174
Kawin Paksa	0.2444	0.8009	0.1964
Murtad	0.2017	0.2138	0.0295
Faktor Ekonomi	0.0864	0.0612	0.9152

To visualize the comparative intensity of these determinants across clusters presented in Figure 3.

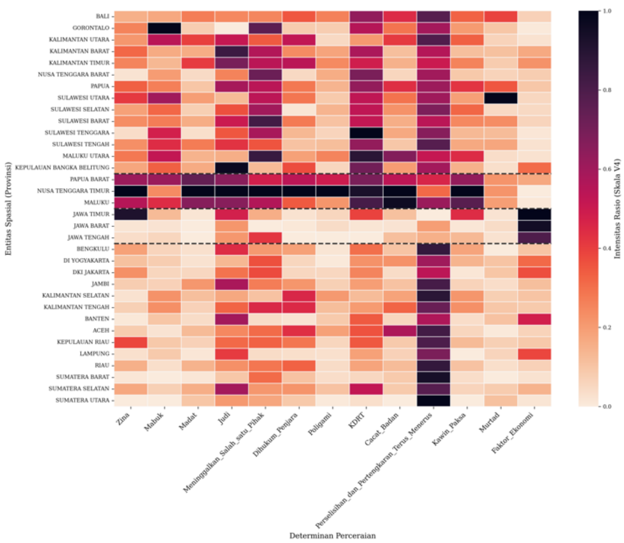

Figure 3. *Heatmap Profiling*

Based on the *centroid* matrix and the heatmap, the characteristics of each cluster are defined as follows:

1. Cluster 0 (Moderate Multi Factor): Displays moderate values for "Domestic Violence" (0.6135) and "Continuous Disputes" (0.6285). This cluster exhibits a pattern where divorce drivers are dispersed, lacking extreme domination by a single factor compared to other clusters.
2. Cluster 1 (High Risk/Deviant Behavior Cluster): This cluster shows extremely high intensity across

nearly all deviant behavior variables, such as "Physical Disability" (0.8319), "Drug Abuse" (0.8022), "Forced Marriage" (0.8009), "Gambling" (0.7706), and "Adultery" (0.7218). This cluster represents the highest social vulnerability.

3. Cluster 2 (Economic Driven Cluster): Characterized by absolute dominance in "Economic Factors" (0.9152). Provinces within this cluster display a unique profile where financial pressure is the fundamental root of marital dissolution.
4. Cluster 3 (Conflict Pure Cluster): Dominated by "Continuous Disputes and Quarrels" (0.7787), while maintaining very low values in behavioral deviation variables compared to Cluster 1. This suggests a pattern of divorce rooted purely in relational disharmony rather than external social pathologies.

C. Spatial Distribution and Regional Patterns

The visualization of the spatial distribution of provinces within the cluster space is presented in Figure 4.

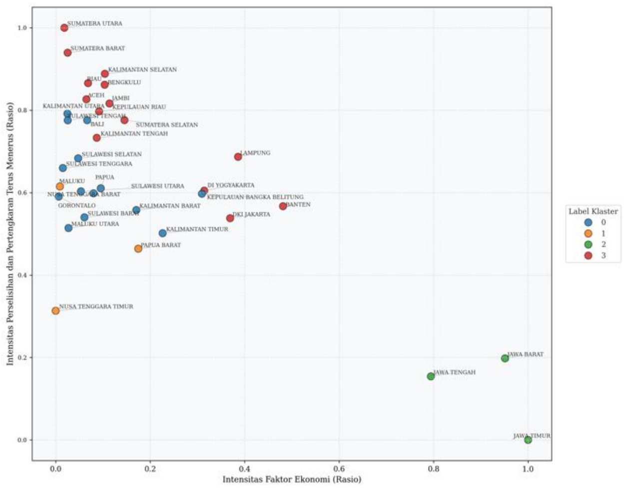

Figure 4. Spatial distribution of provinces based on Economic Factors and Continuous Disputes.

This section analyzes the spatial distribution of the 34 provinces within the established cluster coordinate space. A two-dimensional scatter plot is utilized to map provinces based on the two determinants contributing most significantly to data variance, Economic Factors and Continuous Disputes and Quarrels. This mapping aims to observe whether provinces with similar characteristics tend to cluster geographically or are distributed randomly across the archipelago. Based on the mapping in Figure 4, the observed clustering patterns are as follows:

1. Cluster 2 (Economic Cluster): Provinces categorized into this cluster are clearly concentrated along the high x-axis area (Economic Factors). Spatially, this cluster demonstrates a similarity in patterns where economic issues emerge as the dominant determinant, regardless of differences in geographical location.
2. Cluster 3 (Conflict Pure Cluster): This group is concentrated in the high y-axis area (Continuous Disputes). Provinces within this cluster indicate that relational tension is the primary narrative in their divorce data, without significant intervention from deviant behavior or economic variables.
3. Cluster 1 (High Risk Cluster): Provinces in this cluster are scattered in areas showing high intensity across deviant behavior variables. This group occupies coordinate spaces distinct from Cluster 2 and Cluster 3, indicating that social issues in these regions are multidimensional and significantly more complex than in other areas.
4. Cluster 0 (Moderate Cluster): Situated at the center of the coordinate space, provinces in this cluster exhibit moderate profiles, lacking extreme variable dominance.

The success of this methodology in mitigating population bias compared to previous research is evidenced by the ability to map both small and large provinces on an equivalent proportional scale. By utilizing ratio transformation, the clustering is confirmed to be based on social behavior and divorce causality patterns, rather than merely reflecting absolute population sizes. This pattern reaffirms that sociological characteristics in Indonesia exhibit similarities at the provincial level that transcend physical proximity or mere geographical location.

D. Comparative Analysis of Clustering Methodologies

This study employs a dual methodological approach, integrating *Agglomerative Hierarchical Clustering* (AHC) and *K-Means* Clustering to ensure robust spatial partitioning. Table 4 outlines the comparative technical characteristics of these algorithms.

TABLE 4
COMPARATIVE ANALYSIS OF AHC AND K-MEANS

Characteristic	AHC	*K-Means*
Approach	Hierarchical (Bottom-up)	Partitional (Iterative)
Clustering Logic	Taxonomic tree (dendrogram)	Predefined *centroid* optimization
Primary Utility	Structural exploration	Efficient spatial segmentation
Result Stability	Deterministic structure	High stability via ARI validation

The integration of these models ensures *convergent validity*. AHC serves as an exploratory tool to identify natural provincial groupings without forced constraints, while *K-Means* provides efficient, precise partitioning. By comparing these methods through the *Adjusted Rand Index* (ARI), the study achieves a structural consistency of 0.7931. This high

ARI confirms that the provincial clusters are not stochastic artifacts but represent stable, mathematically validated patterns inherent in the social data.

E. Implications for Public Policy and Social Intervention

Policy interventions must be specifically designed based on the primary determinant profile of each cluster to ensure program efficacy. Table 5 presents the strategic priority framework developed according to the characteristics of each cluster.

TABLE 5
POLICY INVENTION FRAMEWORK BY CLUSTER PROFILE

Cluster	Primary Driver	Policy Focus
Cluster 0	Moderate Factors	Family literacy and community based support services.
Cluster 1	Deviant Behavior	Comprehensive social rehabilitation and legal enforcement.
Cluster 2	Economic Factors	Economic empowerment and poverty alleviation programs.
Cluster 3	Conflict Pure	Conflict mediation and formal marriage counseling.

Regions classified under Cluster 1 require cross agency coordination between social services and law enforcement to address deviant behaviors such as gambling, substance abuse, and forced marriage. For Cluster 2, policies must shift from mere mediation toward economic development, where the primary focus is creating sustainable livelihoods to mitigate the financial pressures triggering marital dissolution.

Meanwhile, for Cluster 3, which is dominated by pure relational conflict, interventions must prioritize psychological support, such as revitalizing marriage counseling and family mediation services as the frontline for conflict resolution. As for Cluster 0, policy focus is directed toward maintaining existing social stability through awareness campaigns on healthy conflict management and providing accessible family services to the community. This targeted approach ensures that government resources are efficiently allocated to address the most relevant root causes for each regional typology.

IV. CONCLUSION

This study successfully modeled the spatial typologies of divorce determinants across 34 Indonesian provinces through a highly robust, dual algorithm computational pipeline. By implementing a proportional ratio transformation, demographic bias inherently caused by absolute population disparities was mathematically eliminated, ensuring that the spatial partitioning strictly reflects underlying socio behavioral patterns rather than mere absolute data scales. The structural validity of the resulting models was rigorously validated by an *Adjusted Rand Index* (ARI) of 0.7931, confirming an objective, highly stable convergence between *Agglomerative Hierarchical Clustering* (AHC) and *K-Means* methodologies.

Quantitative evaluations established the optimal regional segmentation at $k = 4$, explicitly justified by the inflection point of the *Within Cluster Sum of Squares* (WCSS) curve at 9.7238 and a stable *Silhouette Score* of 0.2417. This structural configuration identified four definitive typologies, Cluster 0 (14 provinces, including Lampung and Central Java) exhibiting moderately distributed determinants, Cluster 1 (3 provinces, including Papua and West Papua) representing severe social vulnerabilities with average centroid scores exceeding 0.70 for deviant behaviors such as drug abuse, gambling, and forced marriage, Cluster 2 (3 provinces, including East Nusa Tenggara and Southeast Sulawesi) demonstrating absolute dominance in economic factors with a *centroid score* of 0.9152, and Cluster 3 (14 provinces, including DKI Jakarta and West Java) fundamentally driven by continuous disputes with a *centroid score* of 0.7787.

Despite providing a comprehensive macro level mapping, specific methodological limitations are structurally acknowledged. First, the categorization of relational and economic factors as primary catalysts relies exclusively on unsupervised *centroid* evaluations, therefore, future analytical models must incorporate inferential statistical testing or probabilistic modeling to establish absolute causality. Second, the current computational matrix operates strictly within a closed system of 13 specific divorce determinants, intentionally excluding external macro socioeconomic covariates such as educational indices and urbanization rates, which mathematically influence regional dynamics. Furthermore, the reliance on spatial mean aggregation over the panel data compresses temporal dynamics, obscuring potential inter year pattern shifts across the provinces.

Consequently, subsequent research frameworks are highly encouraged to integrate *Geographically Weighted Regression* (GWR) alongside expanded demographic datasets to capture spatial influences and temporal dynamics more comprehensively. Ultimately, this empirical mapping dictates that national divorce management policies can no longer rely on a generic, homogeneous approach. Interventions must be structurally localized: highly vulnerable regions (Cluster 1) require stringent legal enforcement and social rehabilitation, economically driven regions (Cluster 2) demand targeted poverty alleviation programs, and conflict dominated regions (Cluster 3) necessitate the revitalization of psychological mediation services. This precision based policy framework ensures that public resources are optimally allocated to address the highly specific root causes of marital dissolution across the archipelago.

REFERENCES

[1] S. A. Wardani, N. B. Al Varuq, dan H. T. Santoso, "Implementasi Data Minning Clustering Dalam Mengelompokan Kasus Perceraian di Provinsi Jawa Timur Menggunakan Algoritma K-Means," *JAMI: Jurnal Ahli Muda Indonesia*, vol. 6, no. 1, pp. 68-83, Jun. 2025.

[2] Y. Sopyan, A. D. Lesmana, dan C. Juliane, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *BITS: Building of Informatics, Technology and Science*, vol. 4, no. 3, pp. 1464-1470, Des. 2022.

[3] A. Mulyani dan E. N. Padilah, "Penerapan Algoritma K-Means Untuk Pengelompokan Faktor Perceraian Di Kabupaten Garut," *Jurnal Algoritma*, vol. 22, no. 2, pp. 1470-1480, Nov. 2025.

[4] Yanti, "Analisis Algoritma K-Means Dalam Pengelompokan Perkara Perceraian Berdasarkan Kelurahan Di Kota Jambi," *PROCESSOR: Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer*, vol. 16, no. 1, pp. 9-19, Apr. 2021.

[5] E. Purwaningsih dan E. Nurelasari, "Implementasi Metode K-Means Clustering Dengan Davies Bouldin Index Pada Analisis Faktor Penyebab Perceraian," *Information Management for Educators and Professionals*, vol. 7, no. 2, pp. 134-143, Jun. 2023.

[6] S. Anggaraini dan A. H. Hasugian, "Identifying Dominant Factors of Divorce in Marbau Selatan Village Using K-Means Clustering," *Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE)*, vol. 6, no. 2, pp. 1004-1018, Sep. 2025.

[7] S. Muntari, Sasmita, dan W. Pebrianti, "Implementasi Algoritma K-Means Untuk Mengetahui Faktor Penyebab Perceraian," *Jurnal Ilmiah BETRIK*, vol. 16, no. 2, pp. 182-191, Agu. 2025.

[8] A. Annurfariz, A. I. Purnamasari, dan I. Ali, "Implementasi Algoritma K-Means Pada Kasus Kekerasan Dalam Rumah Tangga Di Jawa Barat," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1904-1910, Apr. 2024.

[9] S. Hariati, "Analisis Hukum Penyebab Terjadinya Perceraian Ditinjau Dari Undang-Undang Nomor 1 Tahun 1974 Tentang Perkawinan Dan Kompilasi Hukum Islam (Studi Di Pengadilan Agama Giri Menang, Lombok Barat)," *Jurnal Kompilasi Hukum*, vol. 8, no. 1, pp. 2-23, Jun. 2023.

[10] A. H. Rani, F. Fernando, M. Bachrudin, dan Y. N. Herwin, "Analisis Faktor-Faktor Penyebab Kasus Cerai Gugat Di Indonesia," *Sriwijaya Journal of Private Law*, vol. 2, no. 2.

[11] M. Hardiansyah, "Penyebab Perceraian dan Akibat Hukumnya Dalam Pemenuhan Hak-Hak Hidup Keluarga," *Misykat Al-Anwar: Jurnal Kajian Islam Dan Masyarakat*, vol. 8, no. 2, pp. 394-432, 2025.

[12] R. D. Y. Yahya, A. P. K. Nisfah, dan D. L. Fitri, "Segmentasi Pengadilan Berdasarkan Jumlah Putusan Dan Lokasi Menggunakan Algoritma Hierarchical Clustering," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 7, no. 3, pp. 388-393, Jul. 2025.

[13] R. O. Pratikto dan N. Damastuti, "Klasterisasi Menggunakan Agglomerative Hierarchical Clustering Untuk Memodelkan Wilayah Banjir," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 6, no. 1, pp. 13-20, 2021.

[14] Irwan, A. Talib, dan A. Lestari, "Penggunaan Hierarchical Agglomerative Clustering dalam Pengelompokan Kabupaten/Kota Berdasarkan Tingkat Kesejahteraan di Sulawesi Selatan," *Proximal: Jurnal Penelitian Matematika dan Pendidikan Matematika*, vol. 9, no. 1, pp. 197-205, 2026.

[15] P. A. E. Ginting, R. I. Situmorang, M. R. Lubis, R. A. H. Sihombing, dan A. Piliang, "Penerapan Metode Agglomerative Clustering Untuk Segmentasi Data Dalam Lingkungan Big Data," *JISKA: Jurnal Sistem Informasi Sunan Kalijaga*. Vol 4 No 1, pp. 70-78, Jan. 2026.