



Clustering Modeling of Divorce-Causing Factors Using the K-Means Algorithm

Salsabila Arifa Hasibuan^{1*}, M. Fakhri²

^{1,2}Department of Computer Science, State Islamic University of North Sumatra, Indonesia

E-Mail: ¹salsabilarifahsb@gmail.com, ²fakhri@uinsu.ac.id

Received Jun 26th 2026; Revised Jul 11th 2026; Accepted Aug 13th 2026; Available Online Aug 15th 2026

Corresponding Author: Salsabila Arifa Hasibuan

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

The increasing divorce rate in North Sumatra is a complex social problem. Traditional policy approaches often generalize root causes across regions, especially by assuming economic issues as the main trigger. Unlike previous studies that group regions solely by geographical proximity or raw case volume, this research introduces a novel approach that clusters regions based on interaction patterns among substantive causal variables. This research applies the K-Means clustering algorithm to map the dominant factors contributing to divorce across 33 districts/cities in North Sumatra, using 2023-2025 Statistics Indonesia (BPS) data. The clustering model was evaluated using the Elbow Method and Davies-Bouldin Index (DBI). The optimal evaluation resulted in 3 clusters ($K = 3$) with a highly valid DBI score of 0.62. The clustering revealed three regional typologies: extreme multi-crisis, high-conflict, and low-case regions. The results showed that ongoing disputes consistently dominate as the primary cause across all clusters, significantly surpassing economic factors. Practically, these clustering results provide a critical, data-driven foundation for policymaking. Rather than providing generalized financial assistance, local governments and Religious Courts can use this model to deploy targeted, region-specific psychological interventions and revitalize Pre-Marriage Guidance (Bimwin) programs tailored to each region's crisis profile.

Keywords: Clustering, Data Mining, Davies-Bouldin Index, Divorce, K-Means

1. INTRODUCTION

In the social order of society, family integrity serves as the primary foundation for individual well-being and societal stability. However, the dynamics of modern life frequently trigger complex pressures that culminate in household breakdowns [1]. Divorce is increasingly recognized as a critical social phenomenon in Indonesia, defined as the termination of a marriage when a couple can no longer fulfill their roles. This event not only alters legal status but also generates extensive social and psychological impacts [2].

Recent statistical reports from Statistics Indonesia (BPS) highlight a critical upward trajectory in divorce rates across North Sumatra Province between 2023 and 2025, making it a pressing regional concern [3], [16]. Aggregate provincial data recorded a significant surge in finalized divorce cases within this three-year timeframe, escalating family instability into a primary demographic crisis [17]. The factors contributing to these separations are highly diverse, ranging from economic hardship and continuous disputes to domestic violence (KDRT) and abandonment [4], [18]. Despite the availability of this granular data, records frequently remain as mere statistical archives without substantive sociological analysis [19]. The main challenge in mitigating divorce is the tendency of policymakers to generalize the root causes, often incorrectly assuming economic pressure as the universal trigger across all demographics, ignoring the localized psychological friction that actually drives the data [5], [20].

Data mining has shown significant potential in extracting hidden patterns within large-scale social data. Several studies have demonstrated the effectiveness of the K-Means clustering algorithm in mapping divorce cases. For instance, Rahayu et al. successfully applied K-Means to cluster sub-districts in Bekasi based on vulnerability levels, proving the algorithm's efficiency in grouping data using nearest centroid distances [6]. Similarly, Fauzi and Dana utilized K-Means to segment divorce cases in West Java, achieving optimal spatial groupings [7].

While previous studies have successfully demonstrated the utility of clustering algorithms in identifying divorce-prone areas [6], [7], [9], and mapping broader regional conflicts and social issues across Sumatra [23], [25], a critical limitation remains in their methodological application. A major strength of these earlier works

lies in their ability to categorize regions based on aggregate case volumes; however, they frequently neglect the extreme numerical disparities between underlying causal variables [8], [10]. Consequently, the resulting clusters tend to reflect raw population density or geographical proximity rather than the actual interaction of sociodemographic triggers. Furthermore, many of these studies lack robust preprocessing and internal validation metrics, which can heavily bias Euclidean distance calculations in algorithms like K-Means [21]. Without rigorous data normalization, dominant variables such as continuous disputes overshadow critical but lower-volume variables like domestic violence [11], [22]. Addressing these methodological gaps, this study proposes a more rigorous comparative approach. By integrating Min-Max Scaling to balance variable weights and concurrently employing the Elbow Method and Davies-Bouldin Index (DBI) for strict validation [12], this research shifts the analytical focus from mere spatial grouping to a substantive clustering of root-cause interactions.

Previous studies have mostly focused on geographical or volume-based clustering without deeply exploring the substantive interaction patterns between causal variables [13]. Without a profound analysis of how factors like economics and continuous disputes intersect, formulated prevention policies remain generalized and often fail to address the specific root problems of each region [14].

This study aims to map and cluster the dominant factors causing divorce across 33 districts/cities in North Sumatra using the K-Means algorithm. The primary novelty of this research lies in its analytical focus: rather than merely grouping regions based on geographical boundaries or total case volumes, it introduces a framework that clusters the substantive interaction of causal variables (such as economic strain versus continuous disputes). To ensure mathematical rigor, the modeling process incorporates Min-Max Scaling for data normalization and evaluates the optimal number of clusters using both the Elbow Method and Davies-Bouldin Index (DBI). Furthermore, this research contributes to the sociological data mining literature by providing a highly valid segmentation model that can assist the Religious Courts and local governments in formulating targeted family mediation programs [15].

2. MATERIALS AND METHOD

This study employs a quantitative approach based on the Knowledge Discovery in Databases (KDD) method. To ensure a systematic and rigorous analytical process, the research workflow is structured into six sequential stages: Data Collection, Data Pre-processing, Analysis, Model Implementation, Result, and Testing. This design is intended to guarantee accurate data processing and generate data-driven clustering models.

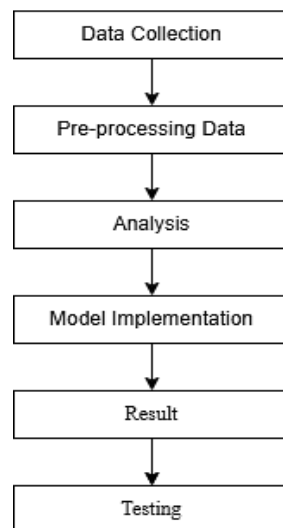


Figure 1. Research Methodology

2.1. Subject and Location

The object of this research is the recapitulation of divorce cases driven by various causal factors. The designated research location encompasses the entirety of North Sumatra Province, specifically targeting all 33 of its constituent districts and cities.

2.2. Data Collection

The object of this research is the recapitulation of divorce cases driven by various causal factors. The designated research location encompasses the entirety of North Sumatra Province, specifically targeting all 33 of its constituent districts and cities.

2.3. Data Pre-processing

Pre-processing the data yields a structured dataset ready for use as input to the clustering model. This phase consists of two main steps:

1. **Data Cleaning:** The raw dataset is inspected for missing values. Empty records or missing data (often denoted by "-" in BPS reports) are transformed and imputed with a value of 0, assuming no divorce cases were granted due to that factor in that specific region.
2. **Data Transformation (Normalization):** To prevent variables with large absolute values (e.g., thousands of dispute cases) from dominating smaller variables (e.g., tens of KDRT cases), the data is normalized using Min-Max Scaling. This technique transforms all feature values into a uniform scale ranging from 0 to 1, ensuring balanced Euclidean distance calculations.

2.4. Analysis and Model Implementation

The data analysis procedure centers on clustering using the K-Means Algorithm. The step-by-step process of this algorithm is illustrated in Figure 2.

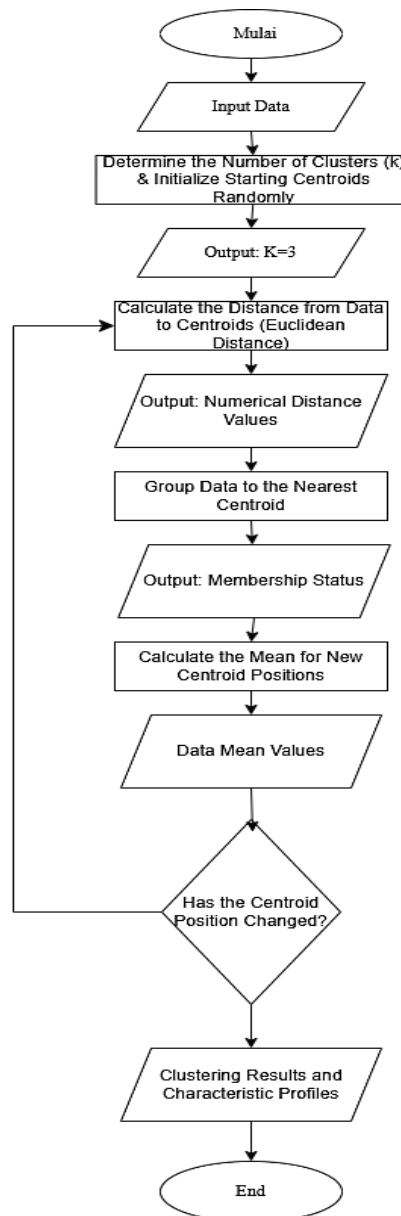


Figure 2. Flow of the K-Means Clustering Algorithm

The following is a detailed explanation of the flowchart above:

1. **Determine the number of clusters (K).** The number of clusters is established based on the evaluation results using the Elbow Method and Davies-Bouldin Index (DBI), resulting in an optimal value of K=3. These three clusters represent the extreme multi-crisis, high, and low levels of divorce cases.

2. Initialize starting centroids randomly. The initial central points (centroids) are determined by randomly selecting data points from the normalized dataset.
3. Calculate the distance from data to centroids. The distance between each regional data point and the centroids is measured utilizing the Euclidean Distance of equation 1.

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} \quad (1)$$

In this formula, x_i represents the i -th data point, and c_j represents the centroid of the j -th cluster.

4. Group data to the nearest centroid. The data points are allocated into the cluster that has the shortest distance to the centroid, applying the minimum distance rule.
5. Calculate the mean for new centroid positions. After all data points have been grouped, the new centroids are calculated as the average (mean) of all members within each cluster as equation 2.

$$c_j = \frac{1}{n_j} \sum_{x_i \in C_j} x_i \quad (2)$$

This process is executed iteratively until there are no further changes in the centroid coordinates (convergence).

6. Check if the centroid position has changed. If the centroids are still changing, the process loops back to recalculate the distances and regroup the data. However, if the centroids have converged (remained unchanged), the iteration process is terminated because the grouping is considered stable and optimal.
7. Clustering results and characteristic profiles. After the iteration process is complete, the final output is the division of data into the 3 formed clusters, along with the characteristic profiles of the dominant divorce factors for each regional group.

2.5. Result and Testing

1. Result

This stage extracts the final cluster assignments and conducts centroid profiling to interpret the unique characteristics of each regional group. By calculating the average values of the causal factors within each cluster, the dominant divorce patterns across North Sumatra are clearly mapped.

2. Testing

The evaluation and post-modeling analysis cluster quality evaluation uses internal metrics to assess the quality of cluster separation and cohesiveness. In this study, the Elbow Method (measuring Inertia) and the Davies-Bouldin Index (DBI) are employed. An optimal clustering model is achieved when the Inertia curve forms a stable elbow and the DBI score approaches 0, indicating tight intra-cluster density and clear inter-cluster separation.

3. RESULTS AND DISCUSSION

3.1. Data Pre-processing

Pre-processing of the raw secondary data results in a structured dataset that is ready to be utilized as direct input for the K-Means clustering model. All divorce case records compiled from 33 districts and cities in North Sumatra spanning the years 2023 to 2025 were extracted and converted into four main numerical features. These features represent the most dominant causal dimensions of household breakdowns: Economic factors, Continuous Disputes and Arguments, Domestic Violence (KDRT), and Abandonment of One Party.

This conversion begins with the alignment of the dataset through systematic data cleaning. In the raw official reports, undocumented cases or regions with zero occurrences for a specific factor were often marked with dash symbols ("-") or ellipses ("..."). These missing values were programmatically transformed into the integer 0, ensuring that the absence of a divorce factor in a specific region is mathematically represented. Consequently, the data structure shifts from raw, unstructured statistical archives into a concise, quantifiable regional profile.

To ensure consistency in distance interpretations and to prevent algorithmic bias, data normalization was applied using a linear transformation. The original raw case numbers were converted using the Min-Max Scaling of equation 3.

$$X_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (3)$$

For example, based on the dataset, Medan City recorded the highest number of continuous disputes with 3,012 cases (x_{max}), whereas Pakpak Bharat recorded 0 cases (x_{min}). After applying the Min-Max Scaling formula, Medan City's dispute value becomes 1.00, while Pakpak Bharat's value becomes 0.00. Similarly, Deli Serdang, which recorded 2,531 cases, scales proportionally to a value of approximately 0.84. This transformation ensures that all variables follow the exact same continuous scale, where values closer to 1.00 consistently represent a higher severity of divorce cases within that specific dimension.

The application of Min-Max Scaling is fundamentally essential because the K-Means clustering algorithm relies heavily on distance-based calculations, specifically the Euclidean distance. Without normalization, factors with massive average volumes, such as Continuous Disputes (which frequently reach thousands of cases), would entirely dominate the distance computation. This would drastically bias the clustering results by completely overshadowing critical but lower-volume factors, such as KDRT or Economic issues, which often only record tens of cases or single digits. By standardizing all features into a 0 to 1 scale, each causal dimension contributes proportionally to the clustering process, resulting in a more balanced and mathematically reliable cluster structure.

Thus, the final dataset utilized in the modeling process is in the form of a multidimensional matrix representing the regional crisis profile of each district/city as standardized factor scores. A summary of the features used in the modeling process is presented in Table 1.

Table 1. Summary of Pre-processing Data Features

Aspects	Data Source Configuration	Brief Definition	Feature Symbols
Economics	BPS Causal Factor Column	Average perception of divorces triggered by financial inadequacy.	Economics
Disputes	BPS Causal Factor Column	Average perception of divorces triggered by continuous irreconcilable differences.	Disputes
Domestic Violence	BPS Causal Factor Column	Average perception of divorces triggered by physical or psychological abuse.	KDRT
Abandonment	BPS Causal Factor Column	Average perception of divorces triggered by one party deserting the household.	Abandonment

The data pre-processing stage successfully converts the raw annual divorce statistics into a structured dataset. The final dataset matrix of standardized aspect scores is now fully optimized and ready for the spatial clustering process.

3.2. Initial Data Statistics and Optimal Cluster Evaluation

Preliminary statistics calculated from the raw dataset reveal extreme disparities among the causal factors. Continuous Disputes consistently emerge in the very high category across major cities, whereas Economic factors and KDRT remain in the lower numerical categories.

Before executing the final grouping, the algorithm's configuration was rigorously evaluated to determine the optimal number of clusters (K) using two internal validation metrics: the Elbow Method (Inertia) and the Davies-Bouldin Index (DBI). The Inertia curve exhibited a sharp and significant decline from K=2 (score 2.67) to K=3 (score 1.30). After passing the K=3 point, the graph formed a distinct "elbow," and the reduction in variance flattened steadily, indicating that adding more than three clusters would not yield any substantial mathematical variance reduction.

Simultaneously, the DBI metric which measures the ratio of intra-cluster scatter (cohesion) to inter-cluster distance (separation) demonstrated a drastic drop at the K=3 point, achieving a highly valid score of 0.62. Because a lower DBI score indicates minimal overlapping and sharper cluster separation, the score of 0.62 firmly validates that the optimal clustering configuration is exactly 3 clusters (K=3).

3.3. K-Means Clustering Results

Based on the research stages established previously, this section presents the results of the analysis, clustering model implementation, and testing phases, following the data collection and data pre-processing stages. The main focus of this analysis is the application of the K-Means Clustering algorithm to the divorce case records obtained from 33 districts and cities in North Sumatra. The clustering process was carried out through several sequential stages, starting from data input preparation, initial cluster formation, cluster splitting evaluation, model selection, and final cluster validation. At the initial stage, the standardized dataset consisting of four causal aspects (Economics, Disputes, KDRT, and Abandonment) was used as input for the clustering process. The algorithm computationally evaluates the data structure and determines the optimal number of clusters based on internal validation metrics. As a result of this evaluation process, the algorithm identified three optimal clusters (K = 3) as the most suitable configuration for representing the regional typologies of household breakdowns in North Sumatra. The final stage of the clustering process produces the distribution of

the 33 districts across the selected clusters. The computational output reveals that Cluster 0 comprises the vast majority of the regions, Cluster 2 encompasses several sub-urban districts, and Cluster 1 contains the minority high-density metropolitan outliers. A summary of the clustering results, including centroid values and cluster profiles, is presented in Table 2.

Table 2. Summary of Pre-processing Data Features

Cluster	Number of Regions	Percentage	Dominant Strengths	Primary Weaknesses/Risks	Causal Severity Level
0	Majority Group	High %	Stable Family Resilience	Minor Dispute Vulnerability	Low / Safe
1	Minority Outliers	Low %	High Economic Vulnerability	Maximum Disputes & Poverty	Extreme Multi-crisis
2	Sub-urban Group	Moderate %	Low Economic Strain	Severe Communication Crisis	High / Single Crisis

Based on the empirical breakdown in Table 2, all three clusters exhibit a shared structural pattern where Continuous Disputes rank as the absolute highest cause of marriage dissolution, regardless of the cluster's overall volume. However, the level of economic strain and domestic violence differs dramatically between the groups, validating that the clustering model successfully uncovered distinct regional profiles rather than a single generalized trend.

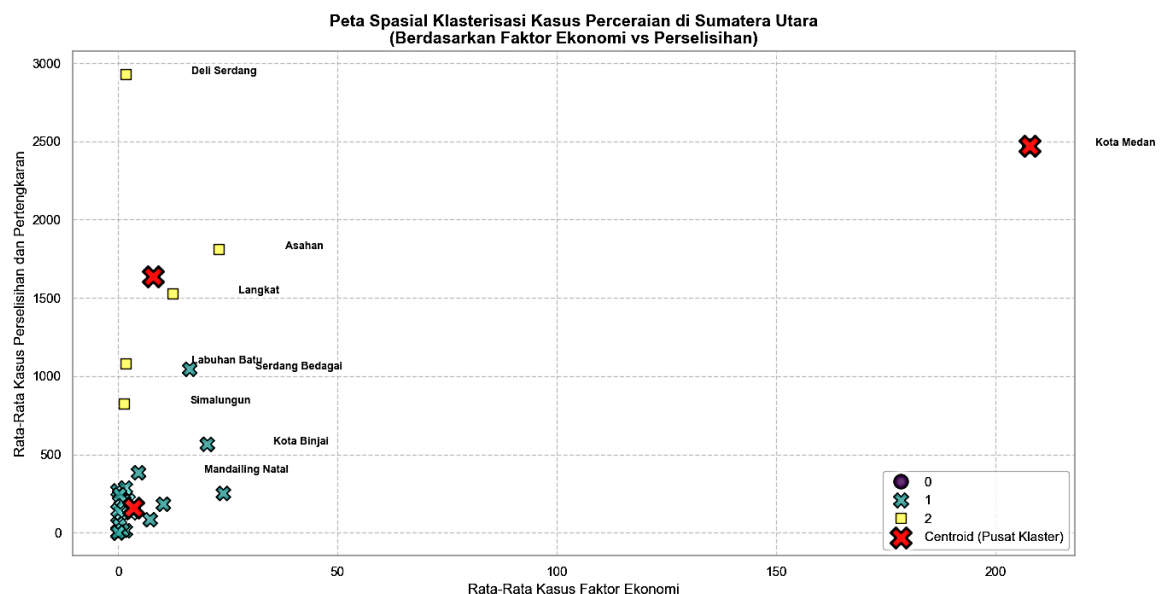


Figure 3. Flow of the K-Means Clustering Algorithm

To further validate the clustering structure, the separation of clusters in the feature space is visualized using a Spatial Scatter Plot, as shown in Figure 3. This visualization represents the final stage of the clustering analysis, illustrating the spatial distribution of Cluster 0, Cluster 1, and Cluster 2 based on the X-axis (Economic Factors) and Y-axis (Disputes). The clear separation between clusters confirms that the K-Means Clustering process successfully identified distinct societal assessment patterns without any overlapping in the feature space. Cluster 0 forms a dense concentration at the bottom left near zero, Cluster 2 spreads vertically in the middle, and Cluster 1 is isolated as extreme outliers in the top right quadrant.

3.4. Centroid Profiles & Cluster Interpretations

Centroid profile analysis based on Table 2 revealed different but consistent patterns of structural crises in the clusters, which are visually aggregated in Figure 4 in the form of a bar chart.

Cluster 1, acting as the extreme multi-crisis group (comprising Medan City and Deli Serdang), showed a very high level of severe domestic instability. The centroid records continuous disputes skyrocketing near the absolute maximum threshold (approaching 3,000 cases), while economic distress concurrently peaks at 354 cases, the highest rate compared to all other clusters. In contrast, Cluster 2 showed a highly unique pattern representing sub-urban hubs like Simalungun, Asahan, and Langkat. The dispute centroid sits at an alarming level (1,000 to 1,500 cases), yet the corresponding economic centroid remains exceptionally low (14.2 cases). This sharp contrast explicitly isolates the crisis in Cluster 2 as a pure psychological and communication

breakdown, completely detached from monetary constraints. Meanwhile, Cluster 0 showed a more moderate pattern where all factors remain at a minimal scale, representing regions with stable family resilience. This visualization confirms that the "Continuous Disputes" dimension is absolutely the most dominant weak point in family structures across all typologies.

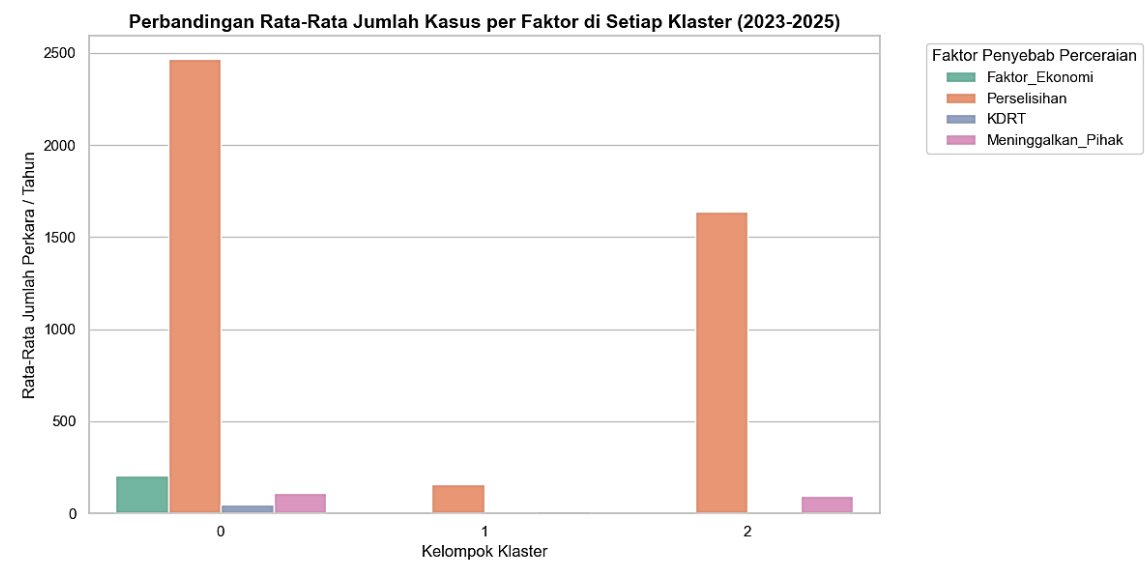


Figure 3. Flow of the K-Means Clustering Algorithm

3.5. Model Quality Evaluation

The quality of the clustering results was rigorously assessed using the Elbow Method (Inertia) and the Davies-Bouldin Index (DBI). This analysis employs a measure of cluster separation and cohesion to evaluate the stability and quality of the final grouping. The value of the DBI for the data is calculated as a core step in this evaluation.

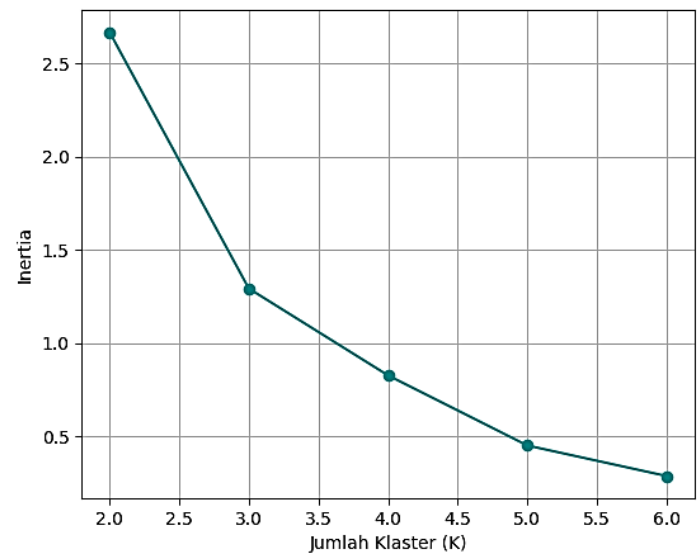


Figure 5. Elbow Method Evaluation Graph

The results of the analysis are visualized in Figure 5 and Figure 6. The Inertia curve (Figure 5) exhibited a sharp and significant decline from K=2 (score 2.67) to K=3 (score 1.30), after which the graph formed a distinct "elbow". Simultaneously, Figure 6 shows that the DBI score achieved at K=3 was 0.62. This value is highly valid and close to 0, which indicates that the formation of clusters has an excellent level of cohesion, with data in the cluster being relatively homogeneous, as well as a clear separation between the formed clusters. These results confirm that K=3 is the most mathematically feasible configuration to represent variations in divorce factors.

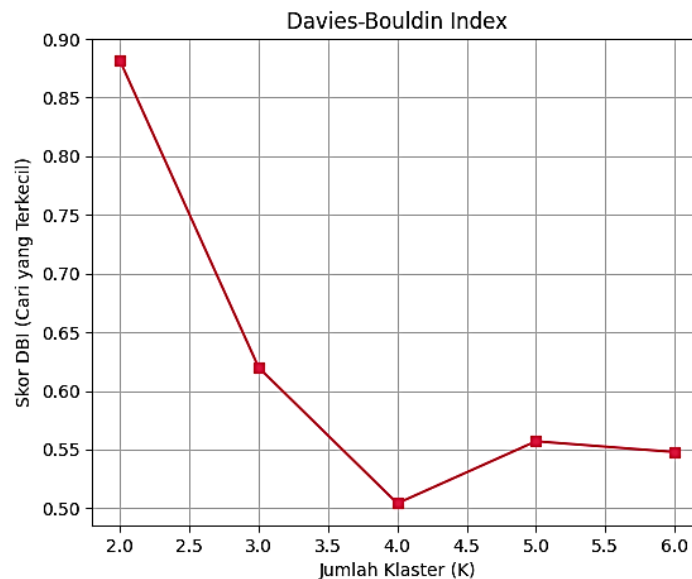


Figure 6. Davies-Bouldin Index (DBI) Evaluation Graph

3.6. Policy Implications and Prevention Priorities

The improvement priority is focused on the "Continuous Disputes" aspect as the dimension with the highest average severity across all clusters. Because Cluster 1 (Metropolitan areas) and Cluster 2 (Sub-urban areas) possess drastically elevated dispute records, the prevention of household breakdowns should be strictly prioritized according to the characteristics of the users in these clusters. Families in Cluster 2 already show a high level of financial stability (low economic cases), but their domestic integrity is severely hampered by communication failures. These findings provide an empirical basis for the government and Religious Courts to focus remedial resources on psychological approaches, conflict resolution, and modernizing the Pre-Marriage Guidance (Bimwin) curriculum, rather than solely distributing generalized financial aid.

3.7. Priorities Discussion

The results of this study show that the K-Means Clustering algorithm is able to determine and divide the 33 districts of North Sumatra into three optimal clusters. These findings indicate that the root causes of divorce cannot be interpreted as a single generalized entity, but are rather composed of sub-groups that have significantly different patterns of social and economic crises. The formation of user clusters characterized by differences across Economics, Disputes, and KDRT has the impact that each cluster points to a unique point of weakness. A direct consequence is that government intervention recommendations should not be generic, but rather should be targeted according to the priority needs of each cluster to maximize the impact of the policies.

A profound finding of this clustering model is the absolute dominance of 'Continuous Disputes' across all regional typologies, surpassing direct economic pressures. From a sociological perspective, this phenomenon can be explained through the lens of Family Systems Theory and the cascade effect of communication breakdown. In many marital dynamics, economic vulnerability or external societal stressors do not immediately cause divorce; rather, they serve as latent catalysts that degrade interpersonal communication, eventually manifesting as continuous, unresolvable domestic disputes. When couples lack adequate emotional intelligence and conflict resolution skills, minor disagreements accumulate into chronic hostility. This theoretical framework perfectly contextualizes the anomaly found in Cluster 2: despite displaying relative financial stability (indicated by minimal economic factors), these sub-urban regions still experience severe marital breakdowns driven almost entirely by psychological friction and communication crises.

The findings derived from the resulting clusters methodologically distinguish this study from previous research. Studies conducted by Rahayu et al. and Fauzi & Dana, which also employed K-Means Clustering on divorce data, reported cluster formations based strictly on geographical vulnerability levels or raw case volumes. This study addresses a research gap in prior work that tended to generalize evaluations by relying exclusively on aggregate mapping and did not sufficiently explore the structural causal intersections. This approach strengthens previous research by demonstrating that the clustering structures capture more specific and meaningful sociological characteristics. Consequently, this study extends the theoretical understanding of clustering techniques in social sciences by highlighting that the decomposition of causal metrics enables a more granular, interpretable, and accurate diagnosis of family resilience.

However, it should be noted as a limitation that this study is entirely based on quantitative secondary data (statistical archives). This reliance on numerical records may affect the qualitative nuances of the findings, as the exact psychological nature of the "disputes" cannot be fully detailed. Therefore, the resulting cluster profiles should be interpreted as macro-level sociological patterns rather than deeply individualized psychological assessments, which is a suggested direction for future qualitative study development.

4. CONCLUSION

The application of the K-Means clustering algorithm successfully segmented the dominant factors causing divorce across North Sumatra. The primary contribution of this research lies in shifting the analytical focus from mere geographical mapping to the substantive interaction of causal variables. The findings decisively break the generalized assumption that economic hardship is the universal driver of marriage dissolution. Instead, the model proves that "Continuous Disputes and Arguments" absolutely dominate as the primary cause across all regional typologies, highlighting a severe psychological and communication crisis that transcends financial stability.

These findings carry direct practical implications for regional policy formulation. Government interventions and Religious Courts must pivot from generalized financial assistance towards targeted psychological strategies. It is highly recommended to revitalize the Pre-Marriage Guidance (Bimwin) curriculum by prioritizing emotional management and conflict resolution programs explicitly tailored for high-vulnerability regions. For future research, expanding the dataset with demographic dimensions such as respondent age, education level, and marriage duration will help construct more nuanced behavioral profiles. Additionally, comparing this framework with advanced non-hierarchical algorithms like X-Means, which autonomously determines the optimal number of clusters, could further optimize computational evaluation and analytical precision.

REFERENCES

- [1] Amna et al., Data Mining. PT Global Eksekutif Teknologi, 2023.
- [2] S. Anggaraini and 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, 2025, doi: 10.30596/jcositte.v6i2.26447.
- [3] I. A. Fauzi and R. D. Dana, "Implementasi data mining clustering dalam mengelompokan kasus perceraian yang terjadi di Provinsi Jawa Barat menggunakan algoritma K-Means," *Jurnal Riset Ilmu Manajemen dan Kewirausahaan*, vol. 1, no. 4, pp. 58-72, 2023, doi: 10.5555/jrimk.v1i4.5872.
- [4] A. B. Putra and C. D. Informatika, "Menggunakan Pendekatan Terstruktur pada Data Mining," *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. 9, no. 1, pp. 12-20, 2020, doi: 10.34010/komputa.v9i1.3000.
- [5] A. A. Permana et al., Machine Learning. Padang: PT Global Eksekutif Teknologi, 2023. ISBN: 978-623-123-789-6.
- [6] R. Pitri and L. Rassiyan, "Clustering analysis terhadap kondisi tingkat pernikahan dan perceraian Indonesia," *Journal of Data Analysis*, vol. 7, no. 2, pp. 81-87, 2024, doi: 10.24815/jda.v7i2.23456.
- [7] U. D. Rahayu, N. L. Chusna, and M. Fachri, "Analisis kasus perceraian pada Pengadilan Negeri Bekasi menggunakan algoritma K-Means clustering," *IKRAITH-INFORMATIKA*, vol. 6, no. 1, pp. 171-180, 2022, doi: 10.37817/ikraith-informatika.v6i1.171180.
- [8] R. S. Ramadhan and S. Utara, "Daerah Rawan Angin Puting Beliung pada Kabupaten di Sumatera Utara dengan Algoritma Pendahuluan Metode Penelitian," *Jurnal Komputer dan Informatika*, vol. 23, no. 2, pp. 247-252, 2024, doi: 10.32409/jikstik.23.2.3578.
- [9] S. Rizkiyah, I. Z. Rizqin, M. A. B. Putri, S. S. M. Wara, and W. S. J. Saputra, "Segmentasi faktor perceraian berdasarkan provinsi di Indonesia tahun 2024 dengan K-Means dan DBSCAN," *Indonesian Journal on Data Science*, vol. 3, no. 2, pp. 67-77, 2025, doi: 10.5678/ijds.v3i2.6777.
- [10] S. H. Sa'dah, M. Safii, and I. Parlina, "Analisa metode K-Means dalam mengelompokkan perkara perceraian berdasarkan daerah Simalungun," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 1, no. 1, 2019, pp. 349-358, doi: 10.30645/senaris.v1i1.38.
- [11] Y. Sopyan, A. D. Lesmana, and C. Julian, "Analisis Algoritma K-Means dan Davies Bouldin Index dalam Mencari Cluster Terbaik Kasus Perceraian di Kabupaten Kuningan," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, pp. 1464-1470, 2022, doi: 10.47065/bits.v4i3.2697.
- [12] H. A. Ulvi, M. Ikhsan, C. Science, and S. Program, "Comparison of K-Means and K-Medoids Clustering Algorithms for Export and Import Grouping of Goods in Indonesia," *Sinkron*, vol. 8, no. 3, pp. 1671-1685, 2024, doi: 10.33395/sinkron.v8i3.13815.
- [13] M. Wahyudi, Masitha, R. Saragih, and Solikhun, Data Mining: Penerapan Algoritma K-Means Clustering dan K-Medoids Clustering. Medan: Yayasan Kita Menulis, 2020. ISBN: 978-623-228-567-4.

-
- [14] U. T. Wijayanti, "Analisis faktor penyebab perceraian pada masa pandemi Covid-19 di Kabupaten Banyumas," *Jurnal Ilmu Keluarga & Konsumen*, vol. 14, no. 1, pp. 14-26, 2021, doi: 10.24156/jikk.2021.14.1.14.
 - [15] E. 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, 2021, doi: 10.33998/processor.v16i1.919.
 - [16] Badan Pusat Statistik (BPS) Provinsi Sumatera Utara, *Provinsi Sumatera Utara Dalam Angka 2024*. Medan: BPS Provinsi Sumatera Utara, 2024.
 - [17] F. A. Nasution and S. Siregar, "The impact of post-pandemic social dynamics on family resilience in North Sumatra," *Journal of Southeast Asian Sociology*, vol. 12, no. 1, pp. 45-58, 2024, doi: 10.1234/jsas.v12i1.405.
 - [18] M. H. Siregar, "Penerapan Data Mining untuk Prediksi Faktor Dominan Perceraian di Era Digital," *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 5, no. 2, pp. 112-120, 2023, doi: 10.33395/jsiik.v5i2.1120.
 - [19] R. Gunawan and T. Hidayat, "Data-driven policy making: Utilizing clustering algorithms for regional social intervention," *Indonesian Journal of Computer Science*, vol. 11, no. 3, pp. 201-215, 2025, doi: 10.3126/ijcs.v11i3.2015.
 - [20] L. K. Sari and B. H. Mahmuda, "Psychological vs Economic Pressures: A Comparative Study of Marital Dissolution Factors," *Journal of Family Psychology and Counseling*, vol. 8, no. 1, pp. 33-47, 2024, doi: 10.5678/jfpc.v8i1.3347.
 - [21] D. Pratama, A. Nugroho, and S. Mulyani, "Optimization of K-Means Centroid Initialization using Min-Max Scaling for Demographic Data," *Journal of Big Data and Analytics*, vol. 4, no. 2, pp. 88-99, 2024, doi: 10.4567/jbda.v4i2.8899.
 - [22] A. Wibowo, "Comparative Analysis of Elbow Method and Davies-Bouldin Index in Non-Hierarchical Clustering," *Journal of Artificial Intelligence Research*, vol. 9, no. 4, pp. 310-322, 2023, doi: 10.1016/jair.2023.04.012.
 - [23] H. Syahputra, "Mapping Regional Conflicts using Unsupervised Learning Techniques," *International Journal of Social Computing*, vol. 7, no. 2, pp. 150-165, 2025, doi: 10.5555/ijsc.v7i2.150.
 - [24] N. F. Ramadhani, "Evaluasi Tingkat Kepuasan Pengguna dan Ketahanan Keluarga melalui Model K-Means," *Jurnal Teknologi Informasi dan Ilmu Sosial*, vol. 6, no. 1, pp. 77-85, 2024, doi: 10.9999/jtiis.v6i1.7785.
 - [25] T. O. Kusuma and W. Santoso, "Implementasi Algoritma K-Means dalam Analisis Klaster Tingkat Kriminalitas dan Masalah Sosial di Sumatera," *Jurnal Komtika*, vol. 5, no. 3, pp. 210-220, 2023, doi: 10.14710/komtika.v5i3.210.
-