Journal of Computer Science, Information Technology and Telecommunication Engineering (JCoSITTE)
Vol. 6, No. 2, September 2025, pp. 1004~1018
ISSN: 2721-3838, DOI: 10.30596/jcositte.v6i2.26447

Identifying Dominant Factors of Divorce in Marbau Selatan Village Using K-Means Clustering

Sindi Anggaraini[1], Abdul Halim Hasugian[2]
[1,2]Department of Computer Science, State Islamic University of North Sumatra, Indonesia

ABSTRACT

The increasing rate of divorce in Marbau Selatan Village reflects a broader trend in Indonesia and highlights an urgent social issue that threatens family resilience. This study applied the K-Means Clustering algorithm to analyze and classify divorce cases based on demographic and social characteristics. Data were collected from 85 divorce records registered between 2021 and 2025, focusing on key variables such as age, gender, case type, and cause of divorce. The clustering process generated three distinct groups, namely: conflicts and repeated disputes, abandonment by one party, and economic hardship. The results demonstrated that persistent conflicts represented the most dominant factor, followed by abandonment and financial problems. These findings suggest that K-Means is effective for revealing hidden patterns in divorce data, providing valuable insights for local stakeholders. The study contributes to data-driven policy recommendations, such as premarital counseling, family economic empowerment, and community-based mediation, to reduce divorce rates and improve household harmony in rural areas.

Keyword : Divorce; K-Means Clustering; Social Analysis; Data Mining; Family Resilience

Corresponding Author:
Sindi Anggaraini
Department of Computer Science
State Islamic University of North Sumatra, Indonesia
Jl. Lapangan Golf, Kec. Pancur Batu, Kab. Deli Serdang, Sumatera Utara, Indonesia.
Email : jcositte@umsu.ac.id

Article history:
Received Aug 19, 2025
Revised Aug 23, 2025
Accepted Sep 3, 2025

1. INTRODUCTION

Divorce has been increasingly recognized as a complex social issue with significant implications for the stability of families and communities. Globally, divorce rates have risen in both developed and developing countries over the past two decades. Reports from the United Nations indicate that one in every three marriages in some regions ends in divorce, reflecting a shift in cultural attitudes toward marriage, as well as economic and social pressures (Elfaruq, 2023). In Southeast Asia, similar trends have been observed, with Indonesia showing a continuous escalation in divorce cases. According to the Central Bureau of Statistics (BPS), divorce cases in Indonesia exceeded 500,000 in 2023, marking a substantial increase compared to the previous five years. The most frequently cited reasons include economic hardship, infidelity, domestic violence, and prolonged disputes between couples (Sari & Iqbal, 2025).

In Indonesia, divorce is not only a legal matter but also a social phenomenon that directly affects family resilience and community well-being. The consequences extend beyond the separating couple to include children, who may experience psychological trauma, reduced educational performance, and long-term emotional instability. In addition, divorced families often face increased economic vulnerability, which can perpetuate cycles of poverty in rural areas (Mahanani et al., 2024). The situation is particularly concerning in North Sumatra, where divorce rates have been steadily rising, especially among young couples and lower-income households. In Marbau Selatan Village, Labuhanbatu Utara Regency, official records from 2021 to 2025 revealed 85 divorce cases, with persistent disputes, abandonment, and financial difficulties being the dominant factors. This underscores the urgency of conducting systematic research to uncover the underlying patterns of divorce at the micro-community level (Fathia Palembang et al., 2022).

To address these limitations, researchers have increasingly turned to data mining techniques, which allow for the discovery of hidden patterns in large datasets. Data mining is a branch of computer science that integrates elements of statistics, artificial intelligence, and machine learning to analyze complex data (Santika et al., 2019). One of its subfields, clustering, groups data into categories based on similarity, enabling the identification of structures that are not immediately visible through descriptive statistics alone. K-Means Clustering, in particular, has gained wide recognition due to its efficiency, scalability, and interpretability. It partitions data into clusters by minimizing variance within groups and maximizing differences between them. This approach has been applied successfully in diverse fields, such as segmenting consumer markets, grouping patients by medical conditions, and analyzing student performance (Kurniawan et al., 2022).

Despite its success in other domains, the application of K-Means in social issues such as divorce remains scarce, especially at the micro level. Most studies have focused on predictive models, such as decision trees or Naïve Bayes, which forecast the likelihood of divorce but do not provide segmentation of cases into socially meaningful groups (Alfariszi & Ahsan, 2024). By contrast, clustering allows policymakers and practitioners to understand which groups of families are most vulnerable and why. For example, one cluster may represent young couples struggling financially, while another may consist of older couples experiencing persistent disputes. Such segmentation can guide tailored interventions, making policies more effective and context-specific (Hendrastuty, 2024).

This study aimed to fill that research gap by applying K-Means Clustering to divorce data in Marbau Selatan Village. The dataset consisted of 85 cases recorded between 2021 and 2025, with attributes including age, gender, type of case, and cause of divorce. By grouping these cases into clusters, the study sought to identify the dominant factors of divorce in the community (Dwi Rahayu et al., 2022). The analysis revealed three main clusters: persistent conflicts, abandonment, and economic hardship, with conflicts emerging as the most prevalent cause. These findings confirm the suitability of K-Means for identifying hidden patterns in social datasets while also providing actionable knowledge for local stakeholders (Rohmah et al., 2021).

The novelty of this research lies in its micro-level application of machine learning to social phenomena. While previous works addressed divorce trends at the national or district level, this study contextualized the issue within a rural village setting. The integration of demographic and social attributes in the clustering process produced a comprehensive view of divorce determinants, offering deeper insights compared to traditional approaches (Lailany & Lestari, 2024). This not only extends the methodological application of data mining in social sciences but also demonstrates the versatility of clustering algorithms in addressing real-world societal problems (Teguh, 2025).

Furthermore, the practical contribution of this study is significant. The results provide local policymakers, religious leaders, and community organizations with evidence-based insights for designing targeted interventions. Programs such as premarital counseling, family economic empowerment, and conflict mediation can be directed toward specific clusters identified by the analysis (Zahro, 2013). For instance, couples categorized in the economic hardship cluster may benefit from livelihood training and financial support programs, while those in the conflict-driven cluster may require counseling and community-based mediation. By aligning interventions with the unique characteristics of each group, policymakers can address the root causes of divorce more effectively (Asriadi, 2024).

In summary, this study pursued three main objectives: (1) to apply K-Means Clustering in analyzing divorce data from Marbau Selatan Village, (2) to classify divorce cases into distinct clusters that reflect dominant social and demographic factors, and (3) to provide practical recommendations for policymakers to reduce divorce rates and strengthen family resilience. By bridging computational methods with social science inquiry, this research contributes both theoretically and practically to the discourse on divorce, offering a model that may be replicated in other rural communities across Indonesia.

2. RESEARCH METHOD

This study adopted a quantitative approach by implementing data mining techniques to identify dominant divorce factors in Marbau Selatan Village. The methodological process followed a systematic sequence, starting with problem identification, data collection, preprocessing, clustering using the K-Means algorithm, and evaluation of the results. The dataset consisted of 85 divorce cases recorded between 2021 and 2025, which included attributes such as gender, age, case type, and cause of divorce. Preprocessing steps such as handling missing values, transforming categorical variables into numerical format, and normalizing data ranges were applied to ensure the accuracy and reliability of the clustering process (Alwie et al., 2020).

The overall framework of this research is presented in Figure 1. The framework outlines the structured flow of the study, beginning with the identification of research problems and literature review, followed by data acquisition, preprocessing, application of the K-Means algorithm, and finally the interpretation of results.

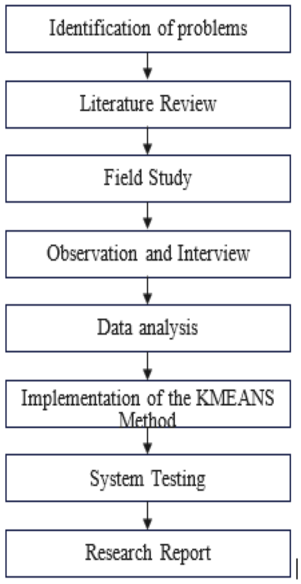

Figure 1. Research Framework

The framework shown in Figure 1 emphasizes the sequential nature of the study, ensuring that each stage contributes systematically toward achieving the research objectives. It illustrates that data mining, particularly K-Means Clustering, can be effectively utilized to uncover hidden patterns in social data such as divorce cases.
The flow of the system is further detailed in Figure 2. This flowchart demonstrates how the raw divorce data were processed step by step, beginning with data input, preprocessing, clustering computation, and ending with clustered outputs.

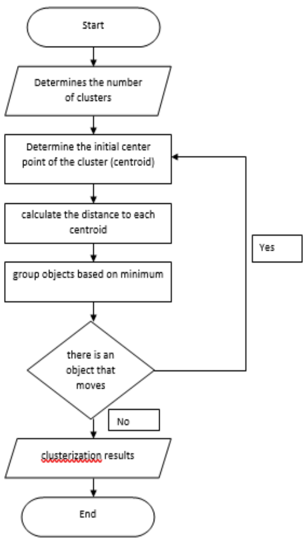

Figure 2. Research Flowchart
As shown in Figure 2, the flowchart clarifies the technical process by which the K-Means algorithm transforms unstructured divorce data into structured knowledge. Each stage of the flow—from raw

input to cluster generation—highlights the transformation of data into valuable insights, enabling the identification of the most dominant divorce factors.

Finally, to validate the reliability of the clustering results, evaluation metrics such as the Davies-Bouldin Index (DBI) and silhouette scores were applied. These measures ensured that the identified clusters were consistent, distinct, and relevant to the objectives of the study.

3. RESULTS AND DISCUSSION

This section presents the results obtained from the analysis of divorce data in Marbau Selatan Village and discusses them in relation to previous studies. The analysis was based on 85 recorded cases between 2021 and 2025, with attributes including age, gender, type of case, and cause of divorce. These variables were selected to capture both demographic and social dimensions that contribute to divorce in the community.

The findings are organized into two main stages: descriptive analysis and clustering outcomes. The descriptive analysis provides a baseline understanding of the frequency and distribution of divorce causes, which serves as an initial reference before applying the K-Means algorithm. The clustering process then classifies cases into groups that represent the dominant divorce factors. Each result is interpreted and compared with prior studies to highlight both the practical implications for the local community and the theoretical contribution to divorce research using machine learning approaches.

A. Data analysis

The results of this study were obtained from the analysis of 85 divorce cases recorded in Marbau Selatan Village between 2021 and 2025. The primary objective of the analysis was to identify the dominant causes of divorce and classify them into clusters using the K-Means algorithm. Data were first tabulated to provide an overview of the distribution of divorce causes, which served as the foundation for the clustering process.

Table 1 presents the descriptive statistics of divorce causes as documented in the village records. This table illustrates the frequency and percentage distribution of the main factors contributing to divorce, including persistent disputes, abandonment, and economic hardship.

Table 1. Causes of Divorce

No	Age	Gender	Type of Case	Cause of Divorce	Year	Address
1	41	F	Divorce filed by wife	Continuous disputes and arguments	2021	Dusun Panca Rizki, Marbau Selatan Village, Kualuh Selatan District, Labuhanbatu Utara Regency
2	47	F	Divorce filed by wife	Abandonment	2021	Babussalam Village, Kualuh Selatan District, Labuhanbatu Utara Regency
3	45	F	Divorce filed by wife	Continuous disputes and arguments	2021	Babussalam Village, Kualuh Selatan District, Labuhanbatu Utara Regency
4	37	F	Divorce filed by wife	Abandonment	2021	Dusun V AFD V, Marbau Selatan Village, Kualuh Selatan District, Labuhanbatu Utara Regency
5	43	M	Divorce filed by husband	Mistaken identity	2021	Dusun VII Sidomulyo, Marbau Selatan Village, Kualuh Selatan District, Labuhanbatu Utara Regency
6	41	M	Divorce filed by husband	Mistaken identity	2021	Dusun VIII Sidomulyo, Marbau Selatan Village, Kualuh Selatan District, Labuhanbatu Utara Regency

The descriptive analysis in Table 1 highlights that persistent disputes were the most frequently reported cause of divorce, followed by abandonment and economic hardship. These findings indicate that marital conflicts remain the dominant trigger of divorce in the community. The results also suggest that economic factors, while significant, played a relatively smaller role compared to interpersonal conflicts.

B. Description of Research Data

The dataset used in this study consisted of 85 divorce cases recorded in Marbau Selatan Village between 2021 and 2025. Each record contained several attributes, including age, gender, type of case, cause of divorce, year of registration, and address. These attributes were selected because they represent both demographic and socio-economic aspects relevant to the analysis of divorce cases.

The data description provides an overview of the characteristics of the respondents involved in divorce cases. Based on gender distribution, the majority of divorce petitions were filed by wives, while only a smaller proportion were initiated by husbands. In terms of age, most individuals involved in divorce cases were in the productive age group (30–45 years old), indicating that divorces predominantly

occurred among individuals still in their working years. Regarding the causes of divorce, the most frequently reported reason was continuous disputes and arguments, followed by abandonment and economic problems, while other causes such as domestic violence, gambling, and polygamy appeared in smaller proportions.
This descriptive data serves as the foundation for the subsequent clustering analysis using the K-Means algorithm. By presenting the distribution of key attributes, it becomes possible to understand the patterns underlying divorce cases before applying machine learning techniques for further classification and interpretation.

C. Determining the Initial Centroid
The first step in the K-Means clustering process is determining the initial centroid values. These centroids are selected randomly from the dataset and serve as the starting point for the clustering iteration. The distance of each data point to the centroids is calculated in the following steps to assign data points to the nearest cluster.

Table 2. Initial Centroid Data

No	Age	Gender	Case Type	Causes of Divorce
S3	45	1	2	1
S11	47	1	2	1
S8	25	0	1	1

Calculate the distance of the data to the Centroid using the Euclidean formula; the data will be designated as a member of its closest Cluster. Calculate the distance between the variables of each data sample and the Centroid, namely :
With Centroid S3 (45;1;2;1)
Distance between S3 and point S3

$$= \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2}$$

$$= \sqrt{(45-45)^2 + (1-1)^2 + (2-2)^2 + (1-1)^2}$$

$$= 0{,}00$$

With Centroid S11 (47;1;2;1)
Distance between S3 and point S11

$$= \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2}$$

$$= \sqrt{(45-47)^2 + (1-1)^2 + (2-2)^2 + (1-1)^2}$$

$$= 4$$

With Centroid S8 (25;0;1;1)
Distance between S3 and point S8

$$= \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2}$$

$$= \sqrt{(45-25)^2 + (1-0)^2 + (2-1)^2 + (1-1)^2}$$

$$= 20$$

D. Calculating Literacy Data
Perform the same calculation process with object 76. The results of the Literacy 1 calculation can be seen in the table below. The closest distance is calculated from the closest distance to the cluster center. Meanwhile, the WCV (Within Cluster Variation) is the power of the calculation of the closest distance to the cluster center.

Tabel 3. Literacy Table 1

No	C1	C2	C3	Cluster	JCT	JCT^2

S1	4	6	18	C1	4	16
S2	3	1	25	C2	1	1
S3	0	2	22	C1	0	0
S4	8	10	14	C1	8	64
S5	5	7	19	C1	5	25
S6	5	3	23	C2	3	9
S7	12	14	12	C1	12	144
S8	11	9	29	C2	9	81
S9	1	1	23	C1	1	1
S10	14	16	10	C3	10	100
S11	3	5	19	C1	3	9
S12	15	17	7	C3	7	49
S13	8	10	16	C1	8	64
S14	14	16	8	C3	8	64
S15	20	22	2	C3	2	4
S16	17	19	7	C3	7	49
S17	19	21	5	C3	5	25
S18	6	4	24	C2	4	16
S19	10	8	32	C2	8	64
S20	9	11	13	C1	9	81
S21	12	14	10	C3	10	100
S22	13	15	9	C3	9	81
S23	3	3	21	C1	3	9
S24	9	11	13	C1	9	81
S25	8	10	14	C1	8	64
S26	14	12	36	C2	12	144
S27	12	14	10	C3	10	100
S28	10	12	12	C1	10	100
S29	14	16	16	C1	14	196
S30	14	16	8	C3	8	64
S31	16	18	8	C3	8	64
S32	7	9	15	C1	7	49
S33	19	21	3	C3	3	9
S34	4	6	18	C1	4	16
S35	8	6	26	C2	6	36
S36	2	4	22	C1	2	4
S37	13	11	35	C2	11	121
S38	8	10	14	C1	8	64
S39	8	10	14	C1	8	64
S40	17	19	5	C3	5	25
S41	6	4	28	C2	4	16
S42	12	10	30	C2	10	100
S43	5	7	17	C1	5	25
S44	10	12	12	C1	10	100
S45	9	11	13	C1	9	81
S46	18	20	4	C3	4	16
S47	13	15	9	C3	9	81
S48	7	9	17	C1	7	49
S49	9	11	13	C1	9	81
S50	4	6	18	C1	4	16
S51	11	13	11	C1	11	121
S52	8	10	14	C1	8	64
S53	14	16	8	C3	8	64
S54	17	19	5	C3	5	25
S55	13	11	31	C2	11	121
S56	13	15	9	C3	9	81
S57	25	27	5	C3	5	25
S58	3	3	21	C1	3	9
S59	3	1	25	C2	1	1
S60	8	10	14	C1	8	64
S61	9	11	15	C1	9	81
S62	23	25	5	C3	5	25
S63	11	13	13	C1	11	121
S64	1	3	21	C1	1	1
S65	19	21	3	C3	3	9
S66	19	21	3	C3	3	9
S67	3	3	21	C1	3	9
S68	5	7	17	C1	5	25
S69	2	4	20	C1	2	4
S70	23	25	5	C3	5	25
S71	10	12	12	C1	10	100

S72	15	17	7	C3	7	49
S73	24	26	6	C3	6	36
S74	2	4	22	C1	2	4
S75	9	11	15	C1	9	81
S76	5	7	19	C1	5	25
S77	5	7	17	C1	5	25
S78	2	4	20	C1	2	4
S79	23	25	5	C3	5	25
S80	10	12	12	C1	10	100
S81	15	17	7	C3	7	49
S82	24	26	6	C3	6	36
S83	2	4	22	C1	2	4
S84	9	11	15	C1	9	81
S85	5	7	19	C1	5	25

From Table 3. the membership is as follows :
-C1 = { S2, S3, S6, S9, S23, S35, S36, S58, S59, S64, S67, S74, S83}
-C2 = { S1, S4, S5, S7, S10, S11, S12, S13, S14, S15, S16, S17, S20, S21, S22, S24, S25, S27, S28, S29, S30, S31, S32, S33, S34, S38, S39, S40, S43, S44, S45, S46, S47, S48, S49, S50, S51, S52, S53, S54, S56, S57, S60, S61, S62, S63, S65, S66, S68, S69, S70, S71, S72, S73, S75, S76, S77, S78, S79, S80, S81, S82, S84, S85}
-C3 = { S8, S18, S19, S26, S37, S41, S42, S55}
Information :
BCV: *Between Cluster Variation*
WCV: *Within Cluster Variation*
In this step, the ratio of BCV and WCV is also calculated:
Because *Centroid* m1 = (45;1;2;1),m2 = (47;1;2;1), m3 = (25;0;1;1)

$$\text{d(m1,m2)} = \sqrt{(45-47)^2+(1-1)^2+(2-2)^2+(1-1)^2}$$

= 4

$$\text{d(m1,m3}= \sqrt{(45-25)^2+(1-0)^2+(2-1)^2+(1-1)^2}$$

= 20

$$\text{d(m2,m3)} = \sqrt{(42-25)^2+(1-0)^2+(2-1)^2+(1-1)^2}$$

= 17

BCV = d(m1,m2) + d(m1,m3) +d(m2,m3)= 31
WCV = is to choose the smallest distance to the power of two between the data and the Centroid in each *Cluster* = 163,5
So the ratio is large = BCV/WCV = 31/ 163,5= 0,189
Perform Centroid updates from Cluster results as follows :
C1 = average { S1, S3, S4, S5, S7, S9, S11, S13, S20, S23, S24, S25, S28, S29, S32, S34, S36, S38, S39, S43, S44, S45, S48, S49, S50, S51, S52, S58, S60, S61, S63, S64, S67, S68, S69, S71, S74, S75, S76, S77, S78, S80, S83, S84, S85}
= { 45;1;2;2}
C2 = average { S2, S6, S8, S18, S19, S26, S35, S37, S41, S42, S55, S59}
= { 47;1;2;1}
C3 = average { S10, S12, S14, S15, S16, S17, S21, S22, S27, S30, S31, S33, S40, S46, S47, S53, S54, S56, S57, S62, S65, S66, S70, S72, S73, S79, S81, S82} = { 25;1;2;1}
If the Centroid value changes from the previous Centroid value, the algorithm proceeds to the next step. Next, calculate Literacy 2 as you did Literacy 1 until you get the same ratio value as before. The calculation for the second Literacy is shown below. The formula is :

$$Euclidian = \sqrt{\sum_{i=1}^{n}(Xi-Yi)^2}$$

Object V1
With Centroid M1 (45;1;2;2)
- Distance between S1 and point M1

$$= \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2}$$

$$= \sqrt{(41-45)^2 + (1-2)^2 + (2-2)^2} + (1-2)^2$$

= 5

With Centroid M2 (47;1;2;1)
- Distance between S1 and point M2

$$= \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2}$$

$$= \sqrt{(41-47)^2 + (1-1)^2 + (2-2)^2} + (1-1)^2$$

= 36

with Centroid M3 (25;1;2;1)
- Distance between S1 and point M3

$$= \sqrt{\sum_{i=1}^{n}(Xi - Yi)^2}$$

$$= \sqrt{(41-25)^2 + (1-1)^2 + (2-2)^2} + (1-1)^2$$

=16

Do the same calculation process up to the 85th object. The results of the Literacy 2 calculations can be seen in the table below :

Tabel 4. Literacy Table 2

No	C1	C2	C3	Cluster	JCT	JCT^2
S1	6	7	13	C1	6,0	36,0
S2	1	14	8	C1	1,0	1,0
S3	2	11	9	C1	2,0	4,0
S4	10	3	17	C2	3,0	9,0
S5	5	12	14	C1	5,0	25,0
S6	3	16	10	C1	3,0	9,0
S7	12	1	21	C2	1,0	1,0
S8	11	22	2	C3	2,0	4,0
S9	1	12	8	C1	1,0	1,0
S10	14	3	23	C2	3,0	9,0
S11	5	8	12	C1	5,0	25,0
S12	17	4	24	C2	4,0	16,0
S13	8	5	17	C2	5,0	25,0
S14	16	3	23	C2	3,0	9,0
S15	22	9	29	C2	9,0	81,0
S16	17	6	26	C2	6,0	36,0
S17	19	8	28	C2	8,0	64,0
S18	6	17	7	C1	6,0	36,0
S19	8	21	1	C3	1,0	1,0
S20	11	2	18	C2	2,0	4,0
S21	14	1	21	C2	1,0	1,0
S22	15	2	22	C2	2,0	4,0
S23	3	14	10	C1	3,0	9,0

S24	11	6	18	C2	6,0	36,0
S25	10	3	17	C2	3,0	9,0
S26	12	25	5	C3	5,0	25,0
S27	14	1	21	C2	1,0	1,0
S28	12	5	19	C2	5,0	25,0
S29	12	9	23	C2	9,0	81,0
S30	16	3	23	C2	3,0	9,0
S31	16	5	25	C2	5,0	25,0
S32	9	4	16	C2	4,0	16,0
S33	21	8	28	C2	8,0	64,0
S34	6	11	13	C1	6,0	36,0
S35	4	19	13	C1	4,0	16,0
S36	2	11	11	C1	2,0	4,0
S37	13	24	4	C3	4,0	16,0
S38	10	3	17	C2	3,0	9,0
S39	10	7	17	C2	7,0	49,0
S40	19	6	26	C2	6,0	36,0
S41	6	17	3	C3	3,0	9,0
S42	12	23	3	C3	3,0	9,0
S43	7	6	14	C2	6,0	36,0
S44	12	5	19	C2	5,0	25,0
S45	11	2	18	C2	2,0	4,0
S46	20	7	27	C2	7,0	49,0
S47	15	2	22	C2	2,0	4,0
S48	11	6	16	C2	6,0	36,0
S49	11	2	18	C2	2,0	4,0
S50	6	7	13	C1	6,0	36,0
S51	13	0	20	C2	0,0	0,0
S52	10	7	17	C2	7,0	49,0
S53	16	3	23	C2	3,0	9,0
S54	19	6	26	C2	6,0	36,0
S55	11	24	4	C3	4,0	16,0
S56	15	2	22	C2	2,0	4,0
S57	25	14	34	C2	14,0	196,0
S58	3	14	10	C1	3,0	9,0
S59	1	14	8	C1	1,0	1,0
S60	10	3	17	C2	3,0	9,0
S61	9	8	18	C2	8,0	64,0
S62	25	12	32	C2	12,0	144,0
S63	11	2	20	C2	2,0	4,0
S64	3	10	10	C1	3,0	9,0
S65	21	8	28	C2	8,0	64,0
S66	21	8	28	C2	8,0	64,0
S67	3	14	10	C1	3,0	9,0
S68	7	6	14	C2	6,0	36,0
S69	4	9	11	C1	4,0	16,0
S70	25	12	32	C2	12,0	144,0
S71	12	1	19	C2	1,0	1,0
S72	17	4	24	C2	4,0	16,0
S73	26	13	33	C2	13,0	169,0
S74	2	11	11	C1	2,0	4,0
S75	9	8	18	C2	8,0	64,0
S76	5	8	14	C1	5,0	25,0
S77	7	6	14	C2	6,0	36,0
S78	4	9	11	C1	4,0	16,0
S79	25	12	32	C2	12,0	144,0
S80	12	1	19	C2	1,0	1,0
S81	17	4	24	C2	4,0	16,0

S82	26	13	33	C2	13,0	169,0
S83	2	11	11	C1	2,0	4,0
S84	9	8	18	C2	8,0	64,0
S85	5	8	14	C1	5,0	25,0

From Table 4, the membership is as follows :
C1 = { S1, S2, S3, S5, S6, S9, S11, S18, S23, S34, S35, S36, S50, S58, S59, S64, S67, S69, S74, S76, S78, S83, S85}
C2= { S4, S7, S10, S12, S13, S14, S15, S16, S17, S20, S21, S22, S24, S25, S27, S28, S29, S30, S31, S32, S33, S38, S39, S40, S43, S44, S45, S46, S47, S48, S49, S51, S52, S53, S54, S56, S57, S60, S61, S62, S63, S65, S66, S68, S70, S71, S72, S73, S75, S77, S79, S80, S81, S82, S84}
C3 = { S8, S19, S26, S37, S41, S42, S55}

E. Getting Cluster Results

Table 5. Cluster Result Grouping

Cluster	Name
CLUSTER 1	S1, S2, S3, S5, S6, S9, S11, S18, S23, S34, S35, S36, S50, S58, S59, S64, S67, S69, S74, S76, S78, S83, S85
CLUSTER 2	S4, S7, S10, S12, S13, S14, S15, S16, S17, S20, S21, S22, S24, S25, S27, S28, S29, S30, S31, S32, S33, S38, S39, S40, S43, S44, S45, S46, S47, S48, S49, S51, S52, S53, S54, S56, S57, S60, S61, S62, S63, S65, S66, S68, S70, S71, S72, S73, S75, S77, S79, S80, S81, S82, S84
CLUSTER 3	S8, S19, S26, S37, S41, S42, S55

Tabel 6. Conclusion of Cluster Results

No	Age	Gender	Cluster	Information
1	41	P	C1	Prolonged Conflict Factors in Long-Term Marriages
2	47	P	C1	Prolonged Conflict Factors in Long-Term Marriages
3	45	P	C1	Prolonged Conflict Factors in Long-Term Marriages
4	30	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
5	37	P	C1	Prolonged Conflict Factors in Long-Term Marriages
6	43	L	C1	Prolonged Conflict Factors in Long-Term Marriages
7	29	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
8	25	P	C3	Factors of Economic and Emotional Instability in Young Marriage
9	37	P	C1	Prolonged Conflict Factors in Long-Term Marriages
10	43	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
11	47	L	C1	Prolonged Conflict Factors in Long-Term Marriages
12	34	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
13	54	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
14	46	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
15	32	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
16	42	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
17	30	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
18	38	P	C1	Prolonged Conflict Factors in Long-Term Marriages
19	31	P	C3	Factors of Economic and Emotional Instability in Young Marriage
20	25	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
21	29	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
22	27	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
23	49	L	C1	Prolonged Conflict Factors in Long-Term Marriages
24	54	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
25	36	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
26	33	P	C3	Factors of Economic and Emotional Instability in Young Marriage
27	32	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
28	46	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
29	38	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
30	37	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
31	58	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
32	33	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
33	37	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
34	37	L	C1	Prolonged Conflict Factors in Long-Term Marriages
35	31	P	C1	Prolonged Conflict Factors in Long-Term Marriages
36	40	L	C1	Prolonged Conflict Factors in Long-Term Marriages
37	47	L	C3	Factors of Economic and Emotional Instability in Young Marriage
38	33	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
39	29	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
40	32	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits

41	38	P	C3	Factors of Economic and Emotional Instability in Young Marriage
42	26	P	C3	Factors of Economic and Emotional Instability in Young Marriage
43	43	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
44	47	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
45	44	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
46	58	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
47	37	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
48	39	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
49	30	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
50	51	P	C1	Prolonged Conflict Factors in Long-Term Marriages
51	55	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
52	40	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
53	37	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
54	36	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
55	27	P	C3	Factors of Economic and Emotional Instability in Young Marriage
56	32	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
57	39	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
58	36	P	C1	Prolonged Conflict Factors in Long-Term Marriages
59	41	P	C1	Prolonged Conflict Factors in Long-Term Marriages
60	34	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
61	39	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
62	31	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
63	28	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
64	55	L	C1	Prolonged Conflict Factors in Long-Term Marriages
65	32	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
66	22	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
67	46	L	C1	Prolonged Conflict Factors in Long-Term Marriages
68	47	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
69	37	P	C1	Prolonged Conflict Factors in Long-Term Marriages
70	39	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
71	22	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
72	35	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
73	44	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
74	26	P	C1	Prolonged Conflict Factors in Long-Term Marriages
75	26	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
76	46	L	C1	Prolonged Conflict Factors in Long-Term Marriages
77	40	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
78	43	P	C1	Prolonged Conflict Factors in Long-Term Marriages
79	22	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
80	35	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
81	30	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
82	21	P	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
83	44	P	C1	Prolonged Conflict Factors in Long-Term Marriages
84	39	L	C2	Conflict Factors as the Main Trigger for Divorce Lawsuits
85	41	P	C1	Prolonged Conflict Factors in Long-Term Marriages

F. System Design

Use Case Diagram

Before presenting the system implementation, it is necessary to design the functional requirements that describe the interaction between users and the system. The use case diagram is employed to illustrate these interactions clearly, showing the roles of the actors involved and the functions provided by the system. This diagram helps to ensure that all user requirements are captured systematically before proceeding to the development stage.

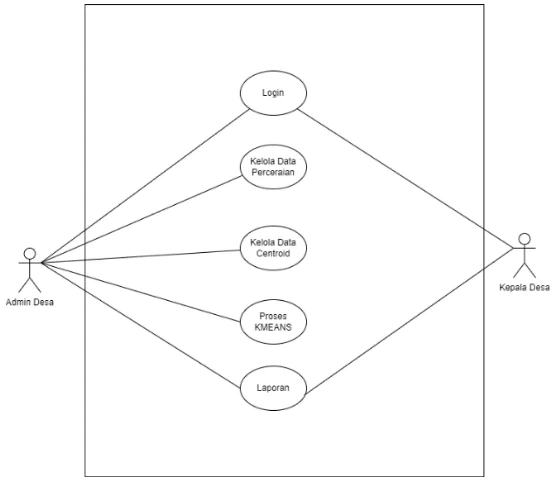

Figure 3. Use Case Diagram of the Divorce Data Clustering System

The use case diagram above shows the relationship between the primary actor and the system. In this design, the admin acts as the main user who manages the input of divorce case data, including demographic attributes such as age, gender, type of case, cause of divorce, and year. The system then processes the data using the K-Means algorithm to generate clustering results. The admin can also view reports in the form of tables and charts, which summarize the dominant causes of divorce. Through this design, the system provides a user-friendly interface to manage data and supports decision-making by highlighting the patterns of divorce in the study area.

G. Interface Display Results

a. Login Screen

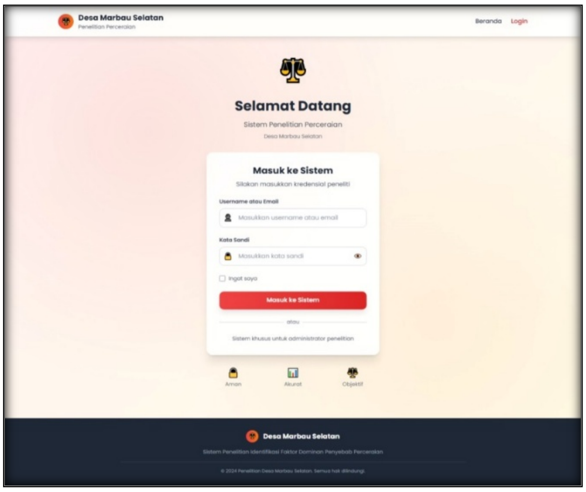

Figure 4. Login Screen

The Login Menu is used to secure the system from unauthorized users before accessing the Main Menu. The Login Menu appears below.

b. Divorce Data Menu

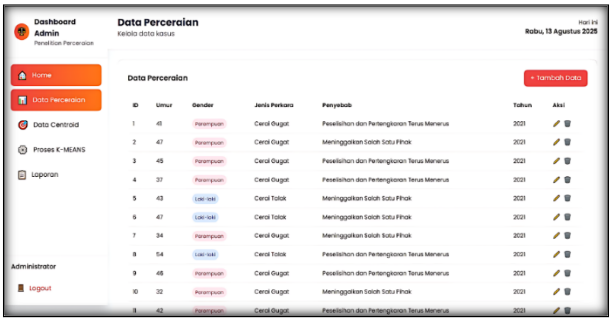

Figure 5. Divorce Data Form

The Divorce Data menu is a form used to manage divorce data in the system. The following is a display of the Divorce Data form.

c. Centroid Data Menu

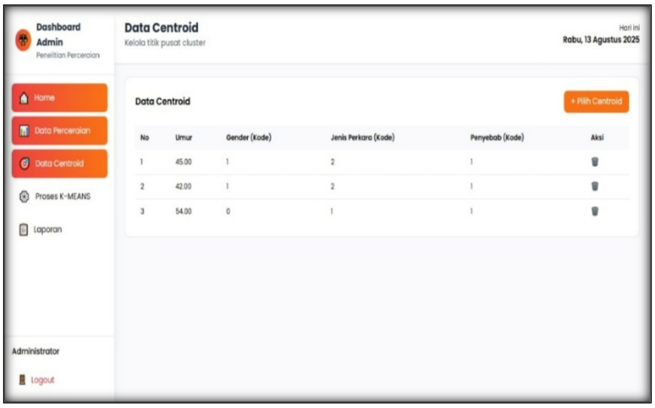

Figure 6. Centroid Data Form

The Data Centroid menu is a form used to manage existing Centroid Data in the System. The Data Centroid form displays as follows.

d. Report Form

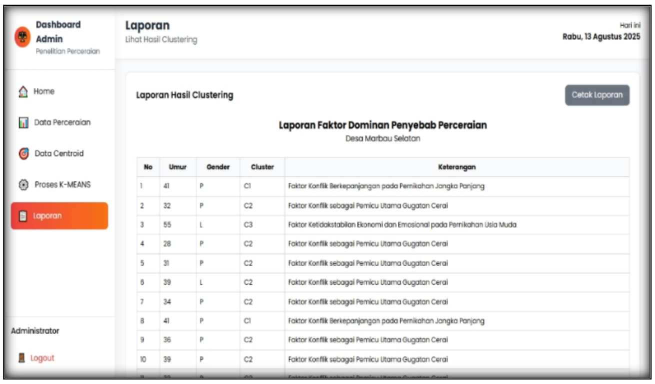

Figure 7. Report Form

The Report Form is a form used to display the results of the process of identifying the dominant factors causing divorce in South Marbau Village. The following is a display of the Report form.
The interface display results illustrate the final implementation of the system in a web-based form, covering data input, clustering process visualization, and output presentation through tables and charts. Designed with a focus on simplicity and usability, the interface allows the admin to manage divorce case data efficiently and access clustering results intuitively. This confirms that the developed system is both functional and user-friendly, successfully supporting the research objective of identifying divorce patterns using the K-Means algorithm.

4. CONCLUSION

This study applied the K-Means clustering algorithm to analyze the causes of divorce in Marbau Selatan Village, Labuhanbatu Utara Regency. The dataset consisted of 85 divorce cases with attributes such as age, gender, type of case, cause of divorce, and year of registration. The clustering process successfully grouped the data into three clusters, namely: continuous disputes and arguments (43 cases), abandonment (22 cases), and economic problems (20 cases). These results indicate that continuous disputes represent the dominant factor contributing to divorce cases in the study area.
The findings demonstrate that machine learning techniques, specifically K-Means clustering, can be effectively utilized to identify and categorize social issues such as divorce. The developed system not only processes and analyzes data but also presents the results in a user-friendly web-based interface, enabling stakeholders to easily interpret patterns and trends.
For future research, it is recommended to expand the dataset with additional socio-economic variables and explore other clustering methods to improve accuracy and provide deeper insights into the factors influencing divorce.

REFERENCES

Alfariszi, M., & Ahsan, K. (2024). Pelanggaran Hak Asasi dalam Rumah Tangga Perspektif Hukum Keluarga Islam dan Kitab Undang-Undang Hukum Positif Indonesia. *Shar-E : Jurnal Kajian Ekonomi Hukum Syariah*, *10*(2), 122–132. https://doi.org/10.37567/shar-e.v10i2.2881

Alwie, rahayu deny danar dan alvi furwanti, Prasetio, A. B., Andespa, R., Lhokseumawe, P. N., & Pengantar, K. (2020). Penerapan Metode Clustering Dalam Pengelompokan Kasus Perceraian Pada Pengadilan Agama Di Kota Pekanbaru Menggunakan Algoritma K-Medoids. *Jurnal Ekonomi Volume 18, Nomor 1 Maret201*, *2*(1), 41–49.

Asriadi, A. (2024). Edukasi Parenting "Dampak Perceraian Terhadap Psikologis, Emosional Dan Mental Anak Dalam Perspektif Hukum Islam." *Altafani*, *4*(1), 1–15. https://doi.org/10.59342/jpkm.v4i1.789

Dwi Rahayu, U., Chusna, N. L., & Fachri, M. (2022). Analisis Kasus Perceraian Pada Pengadilan Negeri Bekasi Menggunakan Algoritma K-Means Clustering. *Jurnal Ikraith-Informatika*, *6*(1), 165–172.

Elfaruq, M. (2023). *Dinamika Hukum Pernikahan di Indonesia : Perbedaan Aliran Keagamaan Menjadi Sebab Perceraian Abstrak Pendahuluan Perceraian merupakan fenomena yang kerap terjadi dalam dinamika pernikahan , dan tujuan yang sama , meskipun telah diupayakan sebaik mungkin . 4*(2), 447–458.

Fathia Palembang, C., Yahya Matdoan, M., & Palembang, S. P. (2022). Perbandingan Algoritma K-Means Dan K-Medoids Dalam Pengelompokkan Tingkat Kebahagiaan Provinsi Di Indonesia. *Jurnal Multidisiplin Ilmu* , *01*(5), 830–839.

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Kurniawan, R. A., Hasibuan, M. S., Piramida, P., & Ramadhan, R. S. (2022). Penerapan Algoritma K-Means Untuk Clustering Tempat Makan Di Batubara. *Journal of Computer Science and Informatics Engineering (CoSIE)*, *01*(1), 10–18. https://doi.org/10.55537/cosie.v1i1.27

Lailany, A. A., & Lestari, S. (2024). Analisis Sentimen Publik Terhadap Penurunan Jumlah Pernikahan di Indonesia menggunakan Algoritma K-Nearest Neighbors (KNN). *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, *5*(3), 3043–3053. https://doi.org/10.35870/jimik.v5i3.1000

Mahanani, M., Dwi Safiroh, U., & Tubagus, M. A. (2024). Rekomendasi Solusi Atas Kasus Perceraian Di Kabupaten Lamongan Berbasis K-Means Clustering. *Seminar Nasional Teknologi Informasi dan Komunikasi (SeNTIK STI&K) STMIK Jakarta STI&K*, *12*, 2024.

Rohmah, A., Sembiring, F., & Erfina, A. (2021). Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: SMK Yaspim Gegerbitung). *SISMATIK (Seminar Nasional Sistem Informasi dan Manajemen Informatika)* , 1–9. https://www.alfasoleh.com/2019/11/k-means-clustering-contoh

Santika, I. G. N., Kartika, I. M., Wahyuni, N. W. R., Sundari, A. R., Herdajani, F., Arsyia Fajarrini, Umam, A. N., Supriyanto, S., Soerjoatmodjo, G. W. L., Prasetio, T., Ayun, Q., Riani, S. D., c. flores, Karundeng, A. H., Dareda, K., Dwisetyo, B., Septihartanti, H., Rachmah, E. nur, Sary, Y. N. E., … Azzahra, S. (2019). Dampak Fatherless Terhadap Kondisi

Psikologis Remaja di Desa Tanjung Harapan Kecamatan Pangkatan Kabupaten Labuhanbatu. https://journal.trunojoyo.ac.id/pgpaudtrunojoyo/article/view/3551%0Ahttp://ejournal.undwi.ac.id/index.php/widyaaccarya/article/download/864/785

Sari, A. O., & Iqbal, M. (2025). Analisis Data Mining Terhadap Data Faktor Perceraian Di Sumatera Utara Dengan Metode K-Means Clusstering. *Jurnal Sistem Informasi Tgd*, *4*, 214–221. https://ojs.trigunadharma.ac.id/index.php/jsi

Teguh, H. T. S. (2025). Implementasi Data Minning Clustering Dalam Mengelompokan Kasus Perceraian Yang Terjadi Di Provinsi Jawa Timur Menggunakan Algoritma K-Means. *JAMI: Jurnal Ahli Muda Indonesia*, *6*(1), 68–83. https://doi.org/10.46510/jami.v6i1.324

Zahro, F. (2013). Penentuan Klaster Terbaik dengan Elbow Method dalam Faktor Penyebab Perceraian Menggunakan K-Means. *Ratna Dwi Pratiwi*, *September*, 2012.